



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Web scraping y análisis de sentimientos para captar la valoración del
cliente hacia los servicios de una empresa de turismo

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Apaza Ramos, Gilberto Romer (orcid.org/0000-0003-2955-2602)

Ñamo Alayo, Edinson Manuel (orcid.org/0000-0002-8689-1718)

ASESOR:

Dr.Chumpe Agosto, Juan Brues Lee (orcid.org/0000-0001-7466-9872)

LÍNEA DE INVESTIGACIÓN:

Sistema de información y comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2022

Dedicatoria

A nuestros Padres por proporcionarnos siempre su protección incondicional, sus consejos, y su dura motivación que nos empujó en torno a adelante.

Agradecimiento

A Dios por estar siempre con nosotros, ofreciéndonos vigor e instrucción para pro avanzar y percibir las tareas que existen en la vida, porque sin él no llegaremos a nuestros sueños.

Afluir a la Universidad Privada César Vallejo, la prestigiosa cimentación de los investigadores, para la educación especializada durante su transición por las aulas del centro y docentes de nuestra ilustre escuela de Ingeniería de Sistemas.

Al final, todo saldrá bien. Si no está bien, no es el final" John Lennon.

Índice de contenido

| | | |
|------|---|--------------------------------------|
| I. | INTRODUCCIÓN | 10 |
| II. | MARCO TEÓRICO | 7 |
| III. | METODOLOGÍA | 23 |
| | 3.1 Tipo y diseño de investigación | 24 |
| | 3.2 Variables y operacionalización | 24 |
| | 3.3 Población, muestra y muestreo | 25 |
| | 3.4 Técnicas e instrumentos de recolección de datos | 26 |
| | 3.5 Procedimientos | 28 |
| | 3.6 Método de análisis de datos (Hipótesis nula y alterna) | 29 |
| | 3.7 Aspectos éticos | 30 |
| IV. | RESULTADOS | ¡Error! Marcador no definido. |
| V. | DISCUSIÓN | |
| VI. | CONCLUSIONES | |
| VII. | RECOMENDACIONES | |
| | REFERENCIAS | 48 |
| | ANEXOS | 62 |

Índice de tablas

| | |
|--|----|
| Tabla 1: Validación de lenguaje de programación | 15 |
| Tabla 2: Validación para el marco de trabajo realizada por expertos. | 19 |
| Tabla 3: Técnica e instrumento a la variable dependiente | 27 |
| Tabla 4: Prueba de normalidad | 30 |
| Tabla 5: Incremento en la tasa de crecimiento de ventas | 33 |
| Tabla 6: Prueba de normalidad de la tasa de crecimiento de ventas | 33 |
| Tabla 7: Prueba de Wilcoxon-Ventas | 34 |
| Tabla 8: Estadístico de prueba de Wilcoxon-Incremento de ventas | 35 |
| Tabla 9: Incremento en la tasa de crecimiento de clientes | 36 |
| Tabla 10: Prueba de normalidad de la tasa de crecimiento de Clientes | 36 |
| Tabla 11: Prueba de Wilcoxon-Clientes | 37 |
| Tabla 12: Estadístico de prueba de Wilcoxon-Incremento de Clientes | 38 |
| Tabla 13: Tasa de crecimiento de ventas | |
| Tabla 14: Tasa de crecimiento de clientes | |
| Tabla 15: Matriz de consistencia | |
| Tabla 16: Matriz de operacionalización de variables | |
| Tabla 17: Validación del instrumento | |
| Tabla 18: Validación de instrumento de cliente | |
| Tabla 19: Evaluación de Marco de trabajo | |
| Tabla 20: Análisis estadístico de SPSS | |
| Tabla 21: Prueba de Normalidad | |
| Tabla 22: Correlación de Pearson Pre-Test | |
| Tabla 23: Pre-test de Clientes | |
| Tabla 24: Pre-test de Ventas | |
| Tabla 25: Matriz de verificación de originalidad | |
| Tabla 26: Presupuesto de proyecto | |
| Tabla 27: Detalle de financiamiento | |
| Tabla 28: Juicio de expertos para lenguaje de programación | |
| Tabla 29: Role Scrum | |
| Tabla 30: Sprint Backlog | |
| Tabla 31: Análisis estadístico de SPSS Post-Test | |
| Tabla 32: Prueba de normalidad Post-Test | |

Tabla 33: Correlación de Pearson Post-Test

Tabla 34: Post-Test clientes

Tabla 35: Post-Test ventas

Índice de figuras

| | |
|---|----|
| Figura 1 Metodología de 10 etapas Data Science | 17 |
| Figura 2 Estructura CRISP-DM | 18 |
| Figura 3 Flujo de Scrum para un Sprint | 19 |
| Figura 4 Fórmula de la tasa de crecimiento | 21 |
| Figura 5 Distribución T-Student | 30 |
| Figura 6 Cronograma de actividades | |
| Figura 7 Tablero Jira | |
| Figura 8 Fases de metodología CRISP-DM | |
| Figura 9 Página de Atipax | |
| Figura 10 Facebook de Atipax | |
| Figura 11 Captación de clientes desde Facebook | |
| Figura 12: Entendimiento de la data | |
| Figura 13: Ventas de Abril | |
| Figura 14: Ventas de Mayo | |
| Figura 15: Filtro de Abril | |
| Figura 16: Filtro de Mayo | |
| Figura 17: Primera parte de prototipo | |
| Figura 18: Segunda parte del prototipo | |
| Figura 19: Web Scraping de Twitter y Google Place | |
| Figura 20: Recuperación de la data | |
| Figura 21: programación de routers | |
| Figura 22: programación de routers 1 | |
| Figura 23: Servicio de Autenticación | |
| Figura 24: Obtención del usuario | |
| Figura 25: Validaciones del usuario | |
| Figura 26: Validación de variables | |
| Figura 27: Validaciones de ejecución | |
| Figura 28: HTML del front | |
| Figura 29: Base de Datos | |
| Figura 30: Rama master del proyecto | |
| Figura 31: Login del sistema | |

Figura 32: Lista de Búsquedas

Figura 33: Búsqueda por Twitter

Figura 34: Búsqueda por Google

Figura 35: Evidencia 1

Figura 36: Evidencia 2

Figura 37: Modelado de análisis de sentimiento

Figura 38: Obtención de datos y limpieza

Figura 39: Servicio del análisis y creación de gráfico

Figura 40: Visualización de análisis

Índice de anexos

- Anexo 1: Ficha de registro de ventas
- Anexo 2: Ficha de registro de clientes
- Anexo 3: Matriz de consistencia
- Anexo 4: Matriz de operacionalización de variables
- Anexo 5: Validación de Instrumentos
- Anexo 6: Validación del marco de trabajo de desarrollo
- Anexo 7: Análisis estadístico de SPSS POST-MEDIA
- Anexo 8: Prueba de normalidad Pre-Test
- Anexo 9: Arquitectura de desarrollo
- Anexo 10: Correlación de Pearson Abril-Mayo
- Anexo 11: Pre-test de Clientes
- Anexo 12: Pre-test de Ventas
- Anexo 13: Prototipo
- Anexo 14: Declaratoria de autenticidad de los autores
- Anexo 15: Compromiso de Confidencialidad
- Anexo 16: Solicitud para realizar la Tesis en la empresa
- Anexo 17: Matriz de verificación de originalidad
- Anexo 18: Consentimiento informado
- Anexo 19: Aspectos Administrativos
- Anexo 20: Cronograma de ejecución
- Anexo 21: Metodología Híbrida Scrum- CRISP-DM
- Anexo 22: Análisis estadístico de SPSS Post-Test
- Anexo 23: Prueba de normalidad Post-Test
- Anexo 24: Correlación de Pearson Post-Test
- Anexo 25: Instrumento Crecimiento de cliente Post-Test
- Anexo 26: Instrumento Crecimiento de ventas Post-Test
- Anexo 27: Validación de lenguaje de programación
- Anexo 28: Flujo del sistema

Índice de abreviaturas

| Sigla | Significado | Pág. |
|----------|---|------|
| MYPE | Micro, pequeñas y medianas empresas (Alberto, H.S.C., <i>et al</i> , 2021 pp. 370) | 354 |
| SPIn | Siena Project Incubator (Kassel, 2019. Pp.44) | 4 |
| KNIME | platform Konstanz Information Miner (KRALJ, JUKIČ y BREN, 2022, pp. 5727) | 10 |
| BDA | big data analytics (DAHIYA, JOHN y WATSON, 2022. pp pp. 193) | 175 |
| ERPB | Enterprise resource planning and business intelligence system (Zheng y Khalid, 2022. pp180) | 40 |
| IBM | International Business Machines (Porkodi, Srihari, y Vijayakumar ,2022,pp582) | 563 |
| PNL | Procesamiento de Lenguaje Natural (Saura, Reyes y Palos, 2019, p58) | 27 |
| CRISP-DM | Cross-Industry Standard Process for Data Mining (Bosch, Ngongoni y Grobbelaar, 2022. pp95) | 78 |
| SCRUM | Marco de trabajo para desarrollo de software de forma ágil (AFSHARI y GANDOMANI, 2022, pp. 2921) | 3 |
| KDD | Knowledge Discovery in Databases (ADHIKARI y ADHIKARI, 2015, p.) | 2 |
| PBI | Producto Bruto Interno (Labroneé, 2018, pp450) | 265 |
| SPSS | Statistical Package for Social Sciences (Habes y Pasha, 2021, pp46) | 34 |
| ISO | International Standards Organization (Slakey, <i>et. al</i> , 2021, pp.601) | 3 |
| HTML | Lenguajes de marcas de HiperTexto (Beynon-Davies, 2018. Pp.73) | 688 |
| VADER | Valence Aware Dictionary for sEntiment Reasoner (Skeen, Jones, Cruse y Horvath., 2022.) | 17 |
| BERT | Representación de codificadores bidireccionales de transformadores. (Mingyu. 2022) | 20 |
| BilsTM | Memoria bidireccional a corto plazo. (Mingyu. 2022) | 21 |

Resumen

Esta investigación describe la implementación, desarrollo e influencia de un sistema de web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo, que se ha visto afectada por pandemia, y diversos problemas ocurridos en las actividades del área de ventas; para lo cual se planteó el objetivo de determinar la tasa de crecimiento de ventas y clientes; para ello se utilizó una metodología híbrida de CRISP-DM y el marco de trabajo Scrum para el desarrollo y lenguaje de programación Python con HTML y FastAPI, implementando técnicas de análisis de sentimientos y web scraping.

La investigación es aplicada con enfoque cuantitativo - explicativo y de diseño pre experimental. La población fue de 30 fichas de registros, en un periodo de un mes. Los resultados muestran una mejoría en la tasa de incremento de ventas de paquetes, de 65.6% y una mejoría de la tasa de incremento de captación de clientes, de un 40.1%. Por tal motivo, se concluye que el sistema gestión de datos con web scraping y análisis de sentimientos influyo en el proceso de ventas y clientes, en base a los datos obtenidos del pre y post test.

Palabras clave: Análisis de sentimientos, Tasa de Crecimiento, Web Scraping, CRISP-DM, Valoración del cliente

Abstract

This research describes the implementation, development and influence of a web scraping and sentiment analysis system to capture customer appreciation towards the services of a tourism company, which has been affected by pandemic, and various problems occurred in the activities of the sales area; for which the objective was to determine the growth rate of sales and customers; for this a hybrid methodology of CRISP-DM and the Scrum framework for development and Python programming language with HTML and FastAPI was used, implementing sentiment analysis techniques and web scraping.

The research is applied with quantitative - explanatory approach and pre-experimental design. The population was 30 record cards, in a period of one month. The results show an improvement in the rate of increase in package sales of 65.6% and an improvement in the rate of increase in customer acquisition of 40.1%. Therefore, it is concluded that the data management system with web scraping and sentiment analysis influenced the sales and customer process, based on the data obtained from the pre and post test.

Keywords: Sentiment analysis, Growth Rate, Web Scraping, CRISP-DM, Customer rating.

I. INTRODUCCIÓN

En la presente investigación se usaron métodos de Data Science conocidos como Web Scraping y análisis de sentimientos para evaluar el impacto de las promociones de las empresas de turismo en las redes sociales, esto les permitirá poder tener un mejor posicionamiento de los paquetes por lanzar.

Julca (2020) nos menciona que el sector turístico arroja pérdidas cuantiosas debido a la clausuras de fronteras y vuelos cancelados. A consecuencia de la rápida propagación del virus COVID-19, se calcula que el turismo mundial tuvo una caída entre 1% y 3% en comparación con el 2019.

La cobertura a nivel nacional de la vacunación y el retiro progresivo de las restricciones a los viajes ayuda a recuperar la confianza de los consumidores, y esto ayuda a la recuperación del turismo en 2022. Sin embargo, son niveles que se encuentran entre el 50% y un 63% inferiores a la pandemia de covid-19 (Promperú, 2022)

Perú aun esta algunos pasos atrás de otros países latinoamericanos en cuanto al conocimiento completo de los métodos de data Science. (Cabrera y Villarejo, 2019)

La empresa turística Atipax Group, también se ha visto afectada, a tal estado que en comparación con el año 2019 antes de que se den las disposiciones dadas por el gobierno para neutralizar la pandemia COVID-19 y el años 2022 se ha tenido una caída de 28% en el monto total de ventas y una disminución del 18% en el número de clientes por ende se genera una necesidad según la problemática en su incremento de ventas y clientes.

La mayoría de las agencias turísticas han implementado el uso de plataformas digitales y herramientas para promoción de sus productos. Esto se ve reflejado en que el 37% de las PYME y MYPE usan las social networks para la difusión de sus productos turísticos. De la misma forma el 84% de las empresas han optado por la implementación de nuevas tecnologías (Promperú, 2022). “El análisis de sentimientos es un campo interdisciplinario, cuyo propósito es convertir la información en valores para reportar, dictaminar, conjeturar y también sugerir

respuestas o modernizar a productos, prestaciones y / o fases". (Batistic y Van-Der 2019).

Egger, Kroner y Stockl (2022) nos mencionan que el turismo se ha caracterizado por el uso intensivo de la información y en la actualidad es una industria basada en los datos, debido a esto los datos se almacenan, combinan y analizan para entender mejor a los clientes, competidores y otros interesados. Por lo tanto, usando las TIC's y el Data Science, la indagación turística no se limita a lo tradicional. (Egger y Yu, 2022) De hecho, la ciencia de datos contiene disciplinas entrelazadas incluyendo la informática, matemática y la estadística. Así mismo el web scraping puede ser utilizado de tal forma en el turismo que permite obtener las valoraciones de los hoteles desde sus sitios web para mejorar los productos a ofrecer, respetando el marco legal que conlleva su uso. (Egger, Kroner y Stockl, 2022).

Los métodos de análisis avanzados incluido los modelos de aprendizaje automático pueden brindar información procesable o conocimientos más exactos sobre los datos, lo que hace que el proceso de cálculo sea automático e inteligente (Sarker, 2021).

En el ambiente digital y de redes sociales que actualmente vivimos, los vínculos están ambientados o gestionados por activos computarizados que ofrecen propuestas personificadas apoyadas en nuestras interrogantes y soluciones (Baldi, 2017). Data Science, que se desprende del análisis de sentimientos, ha contribuido un gran valor de manera especial para el área de marketing, donde la información es vital para comprender y estudiar el mercado; también para reconocer modelos y realizar un rastreo a las ventas entre otros beneficios. Este nuevo campo por explotar sugiere nuevas oportunidades, así como retos para lograr una automatización de la comunicación estratégica. (Valdiviezo ,2020).

El mundo informático posee una enorme cantidad de datos, como los datos del Internet de los objetos (IoT), la data referidas de las empresas, los datos del sector salud, los datos de la tecnología móvil, los datos urbanos entre muchos otros más. La obtención de conocimientos o beneficios útiles generados a partir de la

información es empleada en la toma de dictámenes en los diferentes sectores empresariales (Sarker, 2021).

La justificación teórica de la investigación radica en implementar métodos de Data Science que permitan captar de una forma más óptima la valoración del cliente y el análisis de sentimientos permite la diligencia de la indagación desde una vista informática y estadístico. (Hernández, Duque y Moreno, 2017) Esto sirve de soporte para tomar mejores decisiones más adecuadas para que el negocio tenga mejores rendimiento (Vanegas, Tarazona y Rodríguez, 2020).

Asimismo, como justificación tecnológica sabemos que según Sarker (2021) nos indica que el análisis de sentimientos brinda información más precisa sobre los datos obtenidos utilizando modelos avanzados, entonces se espera que, implementando estos modelos se logre incrementar la gestión comercial de la empresa, estableciendo nuevas formas de lanzamiento de productos y captación de clientes objetivos, garantizando de esa forma un aumento en sus proyecciones estimadas.

Asimismo, como justificación institucional según Romero y Rivera (2019, p.9) la clasificación algorítmica elimina el proceso de juicio de las manos humanas, lo que puede presentar un sesgo característico en medio de elección de arbitrajes. Con el auge de Internet, hay una plétora de formas de influir en las decisiones de compra, y la segmentación es un componente fundamental. Según IAB SPAIN (2017), el 47% de las personas tienen una opinión positiva de la publicidad en las plataformas de medios sociales.

Como justificación económica, según Chen (2020) afirma que la empresa al tener la certeza de que productos, son lo más aceptados mediante las redes sociales, podrán invertir mayor números de recursos en los paquetes turísticos a lanzar y no tener pérdidas al invertir en destinos no muy aceptados por los clientes. De esta manera se estima que la empresa podrá recuperar lo invertido en el proyecto en un lapso de 6 meses ya que al tener la certeza de que productos o paquetes son los más aceptados por los clientes, las inversiones realizadas con los proveedores no

representaran pérdidas y eso se lograra gracias al uso de los métodos de Data Science aplicados.

Según Kassel (2019) investigadores de SPIn señaló 17.792 problemas ocultos en el conglomerado de datos; por datos duplicados causados por los silos de información que no se analizaron, por ende, las empresas han estado optando por automatizar un proceso de control de calidad de datos que antes era manual.

Sobre la problemática a tratar se discutió la incógnita básica y sus detalles de la indagación actual. El enigma general de la investigación es ¿Cuál es el impacto del web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo? De la misma forma se tienen los siguientes problemas específicos:

- **PE1:** ¿Cómo influye el web scraping y análisis de sentimientos en la valoración del cliente hacia los servicios, para mejorar la tasa de crecimiento de ventas de la empresa turística? Según Shahbaz (2021) en esta era de desarrollo tecnológico, todas las empresas quieren equipar a su fuerza de ventas con un plan de mecanización de ventas sostenible para mejorar el rendimiento de las ventas
- **PE2:** ¿Cómo el web scraping y análisis de sentimientos impacta en la valoración del cliente hacia los servicios, para el aumento de la tasa de crecimiento de clientes? Según Shahbaz (2021) este estudio investiga el impacto del análisis de sentimientos, las capacidades de gestión de relaciones con los clientes, el estudio encuentra que las características individuales, como la autoeficacia, el juego y las normas sociales, junto con características organizacionales como la voluntariedad, la intervención de usuarios y el apoyo de la gestión, son predictores positivos de la percepción del análisis de sentimientos.

Según Azwindini (2022) la clasificación manual del análisis de grandes cantidades de información, sistematizó la exploración de información. Se explicaron las características que complementan el antecedente y los estados turísticos modernos

hacia las empresas de ecoturismo sostenible, el objetivo principal es evaluar una estrategia para consolidar el turismo como un instrumento para el desarrollo comunitario local sostenible, aplicando los métodos de data Science en la dinámica de las actualidades turísticas y su gestión, por lo tanto la intención general de este estudio es determinar el impacto del web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo. Y como objetivos específicos:

- **OE1:** Determinar la tasa de crecimiento de ventas en la valoración del cliente hacia los servicios, con la implementación del web scraping y análisis de sentimientos.

Según Hou, Yang y Bai (2018) con la llegada de los métodos del data Science estableció un estilo de valoración de la atracción de bienes turísticos utilizando la información pertinente recopilada en Internet y analizó el patrón especial de la atracción de recursos turísticos, para comprender mejor el comportamiento de los turistas, afirmaron que la analítica de grandes datos podría desarrollar nuevos conocimientos para remodelar la comprensión de la industria hotelera y apoyar la correspondiente toma de decisiones, lo que generaría más ventas de paquetes turísticos.

- **OE2:** Determinar como el web scraping y análisis de sentimientos mejora la tasa de crecimiento de clientes en la valoración del cliente hacia los servicios de una empresa de turismo.

Según Kalvet *et al.* (2020) los macrodatos generados por los usuarios de los medios sociales, las búsquedas en la web y las visitas a sitios web constituyen otra fuente de datos prometedora, ya que a menudo están disponibles públicamente en tiempo real y tienen pocas barreras de uso. A raíz de ello, han surgido nuevos modelos de negocio basados en plataformas de los viajes y el turismo, se genera la data Science, que ayuda a la recopilación inteligente y a su aplicación estratégica.

Chang y Kou (2020) indican a través de los métodos del data Science, despliega un rápido desarrollo de la industria del turismo, cómo construir un nuevo sistema

de servicio de comercialización, y esforzarse por mejorar la capacidad de gestión del mercado, la capacidad de predicción y la capacidad de servicio, que se están empezando a dirigirse a aprovechar las oportunidades para desarrollar el turismo, aprovechando al máximo el uso de los de datos, por lo tanto, la hipótesis general es: Determinar como el web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo. Y como hipótesis específicas tenemos:

- **HE1:** El uso del web scraping y análisis de sentimientos en la valoración del cliente hacia los servicios, mejora la tasa de crecimiento de ventas de la empresa de turismo.

De esta manera Kachniewska (2020) indica que la explosión del análisis de sentimientos, la automatización y el machine learning han dado acceso a las empresas, comprender y predecir el comportamiento humano, lo que permiten predecir las ventas de paquetes turísticos.

- **HE2:** El uso de web scraping y análisis de sentimientos mejora la tasa de crecimiento de clientes en la valoración del cliente hacia los servicios de una empresa de turismo.

De esta manera Syam y Kaul (2021. p.4-166) afirman que el uso de análisis de sentimiento, a través del manejo de información y sus métodos, se ha visto una mayor atracción por parte de los clientes.

II. MARCO TEÓRICO

Pelayo (2022), en su tesis “Ciencia de datos y KNIME, combinación perfecta para el triunfo en la toma de decisiones” desarrollada en el centro de estudio Oberta de Cataluña, para optar por el nivel de Ingeniero. La investigación tiene por finalidad indicar, de una forma simple, los procedimientos a continuar para transformar la data en información y esta misma en entendimiento con la meta de conseguir reducción en el periodo de respuesta debido a la automatización, pronóstico y una toma de decisiones adecuada debido a los beneficios del Data Science y la adopción de sus modelos. Para tal motivo se utilizó un proceso analítico total utilizando un software de licencia libre, KNIME, a través de la información de transacciones del comercio en la ciudad de New York. Se aprecia que este valor es, para el grupo de precios bajos 0,727 y para el grupo de precios altos 0,539. Siendo más efectivo el modelo para el conjunto de precios bajos. Como consecuencia del estudio se definió que se tuvo éxito en aplicar los modelos de Data Science al aplicar modelos de aprendizaje sobre información depurada y pre procesados. Esta tesis tiene un aporte de gran relevancia al demostrar que se puede aplicar análisis de sentimientos, para el progreso de la toma de dictámenes en la gestión comercial.

Asimismo, Van Leeuwen (2020), en su tesis “Analyzing the impact of substitutions in football matches” desarrollada en el centro de estudio Erasmus de Rotterdam, para conseguir el nivel de Magíster. Su tesis brinda un aporte claro sobre el impacto del data Science en el rendimiento de equipos deportivos puesto que los datos indican el impacto de las sustituciones en el resultado de los partidos en diferentes situaciones del marcador, estos datos y su respectivo análisis son útiles tanto para el entrenador como para la dirección general del club. Para su investigación se utilizaron datos de más de 1900 partidos de las principales ligas nacionales de la Premier League, la Serie A italiana, la Liga española, la Bundesliga y la Ligue 1 francesas. La precisión del conjunto de pruebas fue de aproximadamente el 57%, esto último demuestra que es difícil predecir con certeza el resultado de un partido de fútbol a partir de las variables estadísticas disponibles. La contribución de este

estudio fue concebir el interés de utilizar las técnicas de análisis de sentimientos para obtener resultados óptimos.

También tenemos que, Garcés (2019), en su tesis “Análisis de sentimientos en redes sociales orientado a la percepción de la calidad de servicios de internet, redes móviles, tv cable y electricidad” ejecutada en el centro de estudio Andrés Bello, para conseguir el nivel de Ingeniero Civil en informática. Su objetivo fue el de incentivar un enfoque empresarial dirigido a la satisfacción de los usuarios mediante herramientas que realicen análisis automatizados en interacciones dentro de las redes sociales. Para esto se utilizaron más de 20000 Twitts para Enel Chile y una cantidad aproximada a los 10000 Twits para Claro Chile. Los resultados obtenidos indicaron que los comentarios no positivos para Enel Chile fueron entre el 11.56% y el 16.99% en cambio para Claro fueron de 56.34% y 61.87%. Como aporte a la investigación realizada, se obtiene que el análisis de sentimiento nos permite tener una idea clara de cómo nos percibe nuestro público objetivo.

Küçükbatır (2018) en su tesis “THE ADOPTION AND INFLUENCES OF BIG DATA IN TOURISM INDUSTRY IN TURKEY”, ejecutada en el centro de estudio KADİR HAS de Estambul, para adquirir el escalón de magíster. Este estudio tiene como propósito comprender los resultados reales del uso de big data mediante la presentación de los desarrollos de vanguardia de la analítica en la elaboración de los viajes y el turismo. Se realizó una investigación mediante la recopilación y el análisis de índices e informes de las principales empresas que operan en la sección de los viajes y el turismo: Trivago y Metglobal, que utilizan datos de manera muy eficaz y activa para la comercialización y gestión de destinos, y la realización de entrevistas en línea con expertos en la materia. También se concluyó que el acceso a volúmenes de datos cada vez mayores, y la capacidad tecnológica cada vez mayor de los científicos de datos para extraer los datos para obtener información significativa son las fuerzas motrices de esta nueva era. Al examinar el creciente interés por el big data y la data Science, se argumenta que en los próximos años este fenómeno seguirá creciendo aún más. Aunque el país ha tenido que hacer frente a tensiones políticas y a los imprevisibles impactos de los ataques terroristas durante años, el sector turístico ha demostrado su capacidad de recuperación, contribuyendo a un crecimiento directo del PIB del 4,1%. Esta investigación aporta

a entender la relación en la valoración del cliente entre el sector turismo y la recopilación de información y su posterior análisis.

Así mismo, Delgado (2018) en su tesis “Plan de fidelización de clientes enfocado en la mejora de la gestión comercial de los hoteles de primera categoría en la ciudad de Manta” realizada en el centro de estudio politécnica agropecuaria de Manabí Manuel Félix López, para la obtención del nivel de Ingeniero en turismo. Su propósito fue proyectar un programa de fidelización de usuarios que permita prosperar la gestión comercial en sectores turísticos de la capital de Manta. Para esto se identificaron a los hoteles de primera categoría, utilizando una ficha de caracterización que le permitió recopilar la información necesaria; también utilizó la táctica bibliográfica en la exploración de distintas páginas web, libros y boletines científicas. Al concluir la implementación se establecieron estrategias de post ventas, fidelización de clientes mediante recompensas; con el fin de aumentar las ventas y de igual manera amparar a sus usuarios sin que estos dispongan por la competencia. El desenlace descubre que el 53% de los encuestados manifestaron que se encuentran satisfechos; y que el 32% muy satisfecho confirmando que la calidad del servicio de hoteles está por encima del promedio. Esta investigación aporta a entender el indicador de tasa de crecimiento de clientes y a la relación de la captación de la valoración y los usuarios.

También se presentan los siguientes trabajos previos de índole nacional:

Clafoque y Zaravia (2021), en su tesis “función de la ciencia de datos en las sociedades” desarrollada en la universidad peruana de ciencias aplicadas para obtener el nivel de licenciado en Administración de empresas. Como finalidad tuvo la de contrastar las diferentes posturas y valoraciones de diferentes autores en cuanto al uso del Data Science en las organizaciones. Para esto utilizó una metodología descriptiva bibliográfica. Se concluyó que es posible que las organizaciones que basan sus decisiones en datos son más sostenibles y se encuentran más listas para las variaciones tecnológicas y de su entorno, lo cuales son inevitables, para esto los datos deben ser gestionados y analizados eficientemente con un correcto uso de la ciencia de datos. Los resultados son por debajo del 50 % de los asalariados confía en su organización con sus datos y el 44

% no adquiere ninguna información recopilados sobre ellos. Esto aporta a la presente investigación a determinar los métodos de Data Science y su intervención en la toma de decisiones.

Luego Pineda (2021) en su tesis “Táctica para el mejoramiento de la gestión comercial que generen un impacto en la cartera de clientes de la agencia de carga internacional G&S LOGISTICS S.A.C en el año 2019” realizada en la universidad Ricardo Palma para optar por el grado de licenciado en administración de negocios globales. Su propósito fue formular tácticas para modernizar la gestión comercial que tengan una influencia positiva en la base de usuarios de la compañía G&S LOGISTICS S.A.C ante los malos resultados obtenidos en el 2019 para la captación de nuevos clientes. Para esto tomo de población a las empresas cuyas características son similares a las de G&S LOGISTICS S.A.C, en otras palabras, a las sucursales de carga internacional que están en Perú, en cuanto a su muestra fue conformada por una porción determinada de trabajadores, que laboran o laboraron últimamente en alguna de las compañías de la población y que bajo su cargo tenían como responsabilidad a la base de clientes. Los resultados son 9,7% de los encuestados cree que las metas establecidas en el sector empresarial no se fijan en base al potencial real de la cartera de clientes, mientras que el 17,7% no tiene una opinión clara al respecto. Luego el 45,2% de los encuestados está parcialmente de acuerdo con esta premisa, y finalmente el 27,4% está completamente de acuerdo en que las metas se establecen en base al potencial real de los clientes de la empresa. El 64,5% de los encuestados cree que es muy importante y necesario que todo el equipo de ventas comprenda el proceso de planificación y seguimiento de las metas del distrito comercial, mientras que el 27,4% cree que es parcialmente importante. Finalmente, el 8% restante no tenía una opinión clara o no le parecía importante. Se concluyó que existe una colisión provechosa en la carpeta de clientes al comprobarse que existe una asociación con la gestión comercial, todo esto entretanto se ejecuten estrategias vinculadas a promociones y al precio se puede conseguir una propuesta que sea la suficientemente atractiva para captar clientes nuevos. Esta investigación tiene como aporte a determinar la relación entre la variable dependiente y los dos indicadores de esta investigación.

Asimismo, tenemos a Mescua (2020) tesis: “El big data analytics y la competitividad empresarial peruana” realizada en la Universidad Ricardo Palma para la obtención del nivel de Doctor en Administración de negocios globales. Esta investigación se planteó determinar a qué escala el Big Data Analytics puede influir en la rivalidad empresarial peruana. Para esto se tomó como población de estudio a los gerentes y/o directores encargados de las fases de implementación de data Science en los flujos principales del negocio de algunas empresas seleccionadas. Los resultados son de 66.7%, están aplicando el Big data Analytics en conjunto al análisis de sentimientos, significa que los datos integrados, están adquiriendo soluciones favorables para las empresas; por otro lado, un 33.3%, no lo está tomando en cuenta. Como conclusión general se demostró que el big data Analytics brinda una influencia positiva en la competitividad empresarial en el Perú, en valores de rentabilidad y cuota del mercado. Esta investigación tiene como aporte el análisis de datos a través de data Science y sus fases, para generar rentabilidad y en la captación de los clientes.

De la misma forma, Gil (2019) en su tesis “Gestión comercial para el uso de redes sociales en empresas de ventas de ropa de la ciudad de Chiclayo 2019” elaborada en la USAT para adquirir el título de licenciado en administración de empresas. En esta tesis se planteó como objetivo indicar estrategias que apoyen a la mejora de la gestión comercial para explotar los medios sociales en PYMES de ámbito local enfocadas en la venta de vestimentas en la ciudad de Chiclayo para conocer la manera más idónea de tener una impresión en la memoria de los adeptos. Como población se seleccionaron a las siguientes tiendas: “Lizie Moda” con 22239 seguidores, “Dulce Canela Boutique” con 22728 seguidores y “Gis Boutique” con 14104 seguidores, estas empresas fueron distinguidos por ser influyentes y concurridas por el público. Se determinó que al intensificar la interacción con los medios sociales la empresa pudo captar unos 45% más seguidores. De esto se concluye que las verificaciones de interacción tanto como “Likes, Comentarios y Seguidores” son generadas por las empresas en sus redes sociales. Esto aporta a la investigación la noción de que al mejorar la captación de la valoración del cliente, hacia las empresas se puede tener una mayor influencia sobre los consumidores y/o seguidores. Esta investigación tiene como aporte al demostrar el uso de redes

sociales para la mejora de la captación del valor del cliente a través de técnicas de recopilación de datos.

Así mismo, Porras (2019), en su tesis “La administración empresarial eficiente con ayuda del big data en el desarrollo de las microempresas de Lima metropolitana, caso gamarra” desarrollada en la UNFV por seleccionar por el grado de doctor en administración. La finalidad de la investigación fue demostrar si el big data es capaz contribuir de manera eficiente en la dirección gerencial para el crecimiento de las MYPE’s ubicadas en gamarra. La población elegida fue de 25497 personas con participación en las microempresas de gamarra y la muestra es 379 personas. De esto se determinó que el 23.5 % de microempresas ha usado un metodo tecnológico para gestionar a sus usuarios y elaborar un mejor proceso administrativo de su negocio. El 74,2% tiene una preferencia por las redes sociales. Tomando todo esto en cuenta se puede declarar que el utilizar del Big Data en conjunto a la ciencia de datos ayudaría a una eficiente administración en la gestión de las microempresas. El aporte de esta tesis es fundamental puesto se demuestra que el uso de los métodos de Data Science mejora la administración y gestión de las empresas.

A continuación, tenemos las siguientes teorías relacionadas:

Data Science, denominada ciencia de datos, tiene como definición la ciencia encarga extraer ideas de datos desordenados para poder convertirlos en información (Grus, 2019, p.1).

Así mismo, el análisis de sentimientos, es una tarea esencial en el procesamiento del lenguaje natural, que se refiere al hecho de que las máquinas pueden analizar y reconocer las emociones mediante el razonamiento lógico y las operaciones matemáticas. En la fase de pre-procesamiento de datos, las características del texto se extraen utilizando el modelo de lenguaje pre-entrenado BERT (representación de codificadores bidireccionales de transformadores) y se utiliza el BilsTM (memoria bidireccional a corto plazo) para obtener la información interna del audio. En la fase de fusión de datos, la red de fusión de datos multimodal fusiona eficazmente las características multimodales mediante la interacción de

información de texto y audio. Durante la fase de análisis de datos, la red de asociación de datos multimodal analiza los datos explorando la correlación de la información fusionada entre el texto y el audio. En la fase de producción de datos, el modelo genera los resultados del análisis de sentimiento multimodal. (Mingyu. 2022)

Por otro lado, se cree que el big data es una solución milagrosa, capaz de proporcionar una ventaja competitiva y que nunca puede haber demasiados datos. Su recopilación de datos de usuario de cualquier tipo se puede adaptar para cumplir sus fines y obtener una ventaja competitiva. Pero fuera de los gigantes tecnológicos específicos, la mayoría de los macrodatos pueden ser más una responsabilidad de recopilar y almacenar, una pérdida de recursos (tanto financieros como técnicos) y un valor engañoso. El análisis de big data (BDA) es una parte clave del uso de cualquier big data. Los BDA son las técnicas analíticas mejoradas que se utilizan para procesar grandes conjuntos de datos con el fin de descubrir patrones ocultos en los datos que se pueden utilizar para generar conocimiento estratégico que es beneficioso para la organización. (The Big Data Boss, 2021, p20-22)

Inteligencia empresarial, que ha cautivado el interés de los analistas de la industria y los legisladores debido a su potencial para proporcionar más información intelectual para la toma de decisiones y garantizar su continuidad. Las empresas deben usar ERPBI de manera adecuada para prosperar, ya que su mal uso conduce al fracaso. (Zheng y Khalid, 2022)

El objetivo de Business Intelligence es investigar, integrar y recopilar lógicamente un análisis multidimensional de datos de diversas fuentes de información de clientes, entornos, competidores, mercados, etc. para mejorar el rendimiento de las empresas, especialmente las nuevas empresas. (Huang, Savita y Zhong-jie, 2022)

Machine Learning es el estudio del reconocimiento y la predicción de patrones mediante la exploración de la construcción de algoritmos que pueden aprender de los datos y hacer predicciones sobre ellos. Por lo tanto, la predicción del rendimiento basada en el aprendizaje automático ofrece un enorme potencial para resolver los problemas antes mencionados mediante la estimación del rendimiento futuro de los

alumnos en función de los datos del proceso de aprendizaje. (Wang, Zhang y He, 2022)

Python es un lenguaje de programación interpretado que ha sido ampliamente utilizado en muchos campos. La ejecución exitosa de un programa Python depende tanto de la corrección del programa Python como de la corrección del intérprete Python. (Wang, *et. al.* 2022)

Python como herramienta para la automatización de procedimientos no estándar tiene una ventaja absoluta como entorno de desarrollo, caracterizado por la facilidad de aprendizaje, la facilidad de uso, las ricas bibliotecas, una funcionalidad analítica bastante potente. (Dronov y Dronova, 2022)

Tabla 1: Validación de lenguaje de programación

| Experto | Grado | Python | Java | JavaScript |
|---------------------------------|-----------|--------|------|------------|
| Orbezo Llancachagua David Sixto | Ingeniero | 32 | 16 | 19 |
| Cesar Augusto Rosario Rodríguez | Ingeniero | 33 | 22 | 23 |
| Walter Atuncar Céspedes | Ingeniero | 35 | 20 | 22 |
| Total | | 100 | 58 | 64 |

Fuente: Elaboración propia.

El análisis de sentimientos estudia y analiza las respuestas y la aceptación de las personas hacia una entidad (por ejemplo posts, blogs, comentarios, reseñas de libros, vídeos) utilizando el análisis de textos mediante algoritmos computacionales para ayudar a determinar las reacciones textuales, ya sean positivas, negativas o neutras. (Kewsuwun y Kajornkasirat, 2022)

El análisis del sentimiento por categoría de aspecto puede proporcionar más información y más profunda que el análisis del sentimiento a nivel de documento, porque su objetivo es predecir las polaridades de sentimiento de las diferentes categorías de aspecto en el mismo texto. (Wenxiong, Xiuwen y Wei, 2021)

El método de minería de textos puede ser para controlar la calidad del servicio de PT PLN (Persero) clasificando los datos de Twitter con el algoritmo k-Nearest, para

extraer información de datos textuales no estructurados para producir información útil. La clasificación de datos es una aplicación de minería de textos para la recuperación de información. En este estudio, los datos recogidos pasarán por la fase de pre-procesamiento, utilizando el algoritmo k-Nearest Neighbors para clasificar los datos en clases negativas, neutras o positivas. Los resultados se implementan en aplicaciones basadas en la web que se construyen utilizando el lenguaje de programación Python. (Damarta, Hidayat y Abdullah, 2021)

Algoritmo VADER (Valence Aware Dictionary for sEntiment Reasoner) validado por humanos, que puntúa la valencia y la intensidad de los términos individuales y sus elementos semánticos relacionados, como los emoticonos ("(: ") y las abreviaturas comunes en las redes sociales y los foros basados en la web ("lol" y "wtf"). VADER genera puntuaciones de polaridad (positivo-neutral-negativo, en una escala de -1 a +1) para cada cadena de entrada. (Skeen, *et. al.* 2022)

También podemos definir el Data Science como, la ciencia que nos permite aprender de los datos, obteniendo información de estos al manipularlos de manera eficaz y eficiente. (Mailund y Denmark, 2017, p.1).

A continuación, con el siguiente diagrama, damos a conocer que la metodología de Data Science la componen 10 etapas, algunas de las cuales en algunos momentos se pueden repetir más de una vez. (IBM Analytics, 2015).

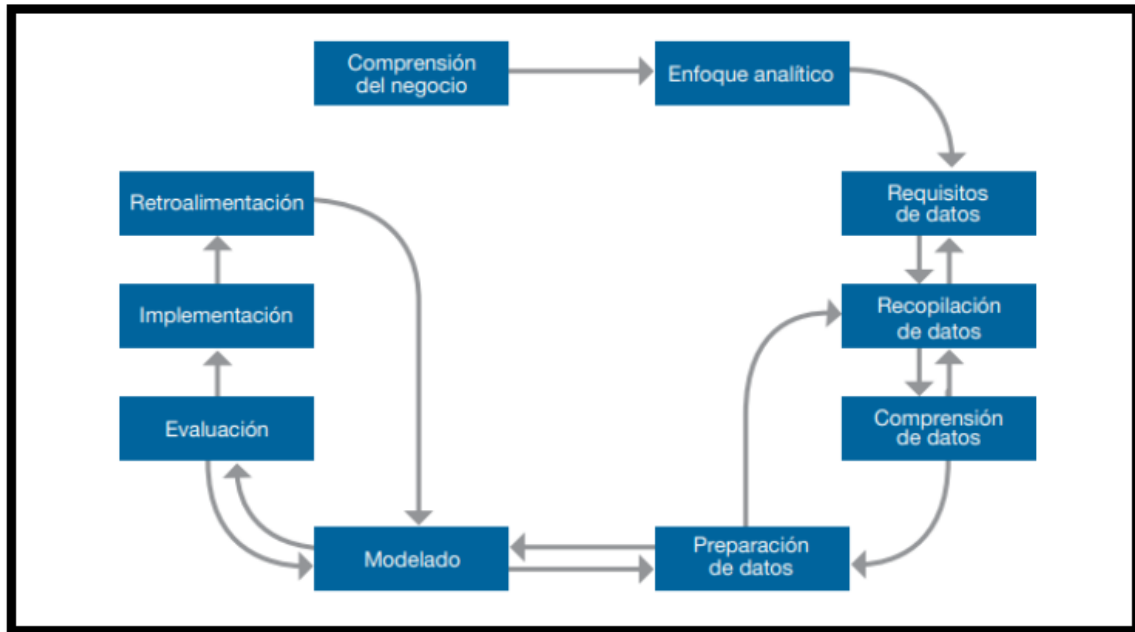


Figura 1 Metodología de 10 etapas Data Science

Fuente: (IBM Analytics, 2015)

Asimismo, se define el web scraping, como la construcción de un agente para descargar, analizar y organizar los datos de la web de forma automatizada, en otras palabras, en lugar que el usuario haga clic en su navegador para copiar y pegar, el web scraping descarga esta información de manera automática que es mucho más rápida que la interacción humana. (Broucke y Baesens, 2018, p.3)

Define el AS es un método de investigación utilizado para analizar la percepción de muestras de texto específicas, a menudo de entornos digitales como plataformas en línea o redes sociales. El análisis de sentimientos es una parte del procesamiento del lenguaje natural (NLP) a través del cual intentamos extraer automáticamente información de sitios web e identificar sentimientos o emociones que el texto pueda contener. (Keenan, 2020. p. 143)

Por otro lado, para gestionar proyectos de Data Science se necesita de forma indispensable un marco de trabajo con la finalidad de tener los lineamientos necesarios para la gestión y planificación del proyecto y de esta manera asegurar su éxito, entre estos marcos encontramos ha: CRISP-DM. (Saura, Reyes y Palos, 2019, p58)

Otro marco de trabajo es del Knowledge Discovery in Databases (KDD), se trata de un proceso que requiere mucho tiempo para detectar patrones válidos, únicos, potencialmente beneficiosos y, finalmente, útiles., comprensible en los datos (ADHIKARI y ADHIKARI, 2015, p.2).

CRISP-DM es un modelo de proceso jerárquico que consiste en una serie de tareas representadas en cuatro niveles de abstracción (de lo general a lo específico): fase de tarea genérica, tarea especializada e instancia de proceso. (Perez, 2021, p.2)

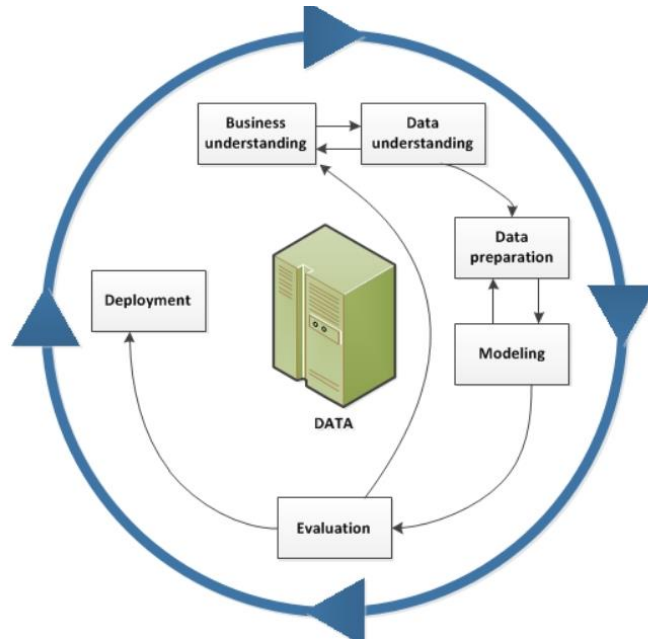
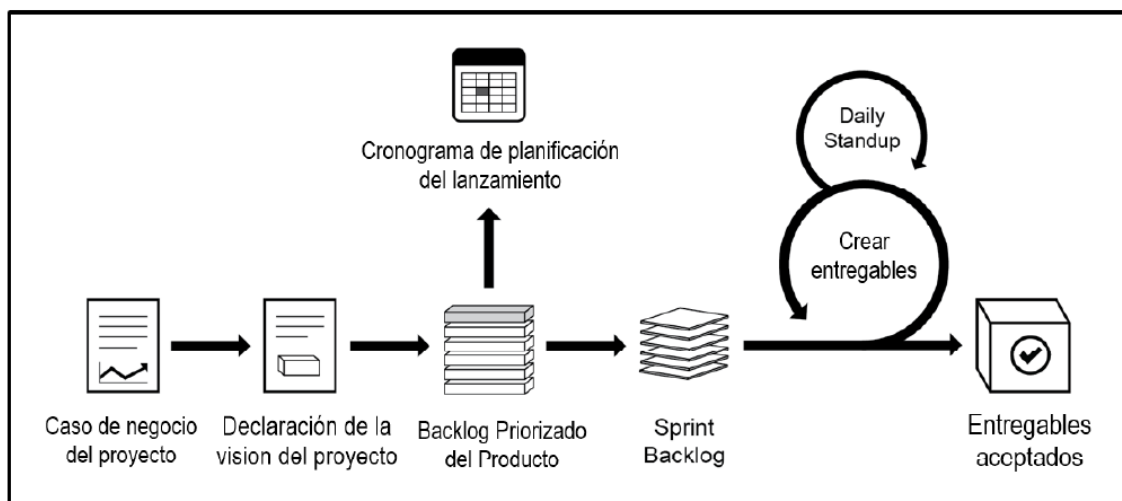


Figura 2 Estructura CRISP-DM

Fuente: IBM: Conceptos básicos de ayuda de CRISP-DM

Otra de los marcos de trabajo más usado es el Scrum, que es una forma tangible de adoptar y arraigar el paradigma ágil. El pequeño conjunto de prescripciones permite una acción inmediata y da como resultado una fructífero y a largo plazo. Con Scrum, las personas desarrollan nuevas formas de trabajar, mediante el descubrimiento, el aprendizaje basado en la experimentación y la colaboración (Verheyen, 2021, p.19).



Anexo Figura 3 Flujo de Scrum para un Sprint

Fuente: Sbok 3° Edición

La ciencia de los datos ágil no se limita a la entrega de software de producción, sino que se trata de alinear mejor la ciencia de los datos con el resto del proceso de la organización. Existe una desalineación crónica entre la ciencia de datos y la ingeniería, donde el equipo de ingeniería a menudo se pregunta qué está haciendo el equipo de ciencia de datos cuando realiza análisis de datos exploratorios e investigación aplicada. La ciencia de datos ágil salva esta brecha entre los dos equipos, creando una alineación más potente de sus esfuerzos. (Jurney, 2017, p.4)

En la investigación, se ejecutó la comprobación por juicio de expertos para definir el marco de trabajo a utilizar.

Tabla 1. Validación para el marco de trabajo realizada por expertos.

| Experto | Grado | KDD | SCRUM | CRISP-DM |
|--------------------------------|-----------|-----|-------|----------|
| Salazar Ávila, Erika Isabel | MBA | 19 | 31 | 22 |
| Regalado Gutiérrez, David | Ingeniero | 25 | 27 | 11 |
| Albert Ronaldo Bernachea Ramos | Ingeniero | 28 | 32 | 25 |
| Total | | 72 | 90 | 58 |

Fuente: Elaboración propia

Con sustento en los puntajes logrados en la tabla 1, se evidencia que el marco de trabajo con puntaje más alto fue SCRUM, por tal motivo, se usó en el proyecto de investigación.

De otro lado, la parte de la variable dependiente, tenemos la valoración del cliente, que la podemos definir como un reglamento que todos los productos están compitiendo para pasar por las cuatro fases: atraer la atención del cliente, el interés, el deseo y finalmente la venta. (Bowlby, 2022. pp.50)

Así mismo la valoración del cliente, que tiene como definición la mayor parte de las administraciones públicas cuenta una gran parte de esto es un proceso computarizado y digitalizado. Así lo ilustran en la información al Usuario y el Sistema de Información del Servicio Estatal de Empleo (SISPE), cada uno con sus propias potencialidades y limitaciones. Sin embargo, estos sistemas funcionan de forma independiente, para la mejora de la planificación, evaluación y toma de decisiones en ventas. (Sordoni, Eransus y Salmerón, 2022, p.34)

Asimismo, tenemos la gestión comercial, la podemos definir como la función encargada de la apertura organizacional al mundo exterior. De esta forma podemos definir las siguientes dimensiones para la gestión comercial: Satisfacción y volumen de clientes, participación en el mercado, incremento de ventas, mejorar las actividades de marketing, capacitación al recurso humano y mejorar el proceso administrativo. (Ponce, 2017, p.55). De la misma forma otros autores mencionan al respecto:

La gestión comercial es la encargada de poder elevar la eficiencia de los negocios convirtiéndose en parte significativa del sistema de direccionamiento en las organizaciones, los que les permite tener un mayor conocimiento sobre el mercado, los competidores y poder adaptarse a los cambios de su entorno, basando su toma de decisiones en las opiniones de los clientes. (Clarke, Cisneros y Yoanis, 2018, p.12)

Asimismo, los elementos de la rentabilidad de las ventas deben tener relación con la supervisión e investigación del entorno competitivo y socioeconómico, con una excelente organización general y comercial teniendo en cuenta las mejores técnicas

de marketing para así formar una adecuada estrategia de ventas. (Artal, 2017, p.32).

Para Portela (2018, p.19), los indicadores en el flujo de ventas indican dimensiones que tienen vínculos, entre sus tipos, entre estas tenemos mediciones presupuestadas, históricas, competidores y sectoriales, las cuales se pueden agrupar en eficiencia, eficacia y productividad.

También podemos definir la dirección de cartera de clientes, como el framework adaptada por una organización que tiene en sus objetivos desarrollar la eficiencia en los trabajos realizados para asociarse a nuevos clientes o mantener a los mismos. (Conexión-ESAN, 2016).

De otro lado para los indicadores nos basamos en lo dicho por Labrunee (2018, p.2) en que una de las medidas más conocidas para realizar una medida del crecimiento económico es la evolución del producto (Se ejemplifica con el PBI, pero puede usarse cualquier variable).

$$\text{TasadecrecimientodelPBIpm} = \frac{\text{PBIpmaño2} - \text{PBIpmaño1}}{\text{PBIpmaño1}} * 100$$

PBI: Producto bruto interno.

Figura 4 Fórmula de la tasa de crecimiento

Fuente: Labrunee. 2018. El crecimiento y desarrollo.

De lo expuesto a continuación, se utilizaron las fórmulas listadas a continuación para la evaluación de nuestros indicadores tasa de crecimiento de ventas y tasa de crecimiento de clientes.

Tasa de crecimiento de ventas:

$$\text{TCV} = \frac{\text{VMA} - \text{VME}}{\text{VME}} * 100$$

VMA: Ventas mes actual.

VME: Ventas mes anterior.

TCV: Tasa de crecimiento de ventas

Tasa de crecimiento de clientes:

$$TCC = \frac{CMA - CME}{CME} * 100$$

CMA: Clientes mes actual.

CME: Clientes mes anterior.

TCC: Tasa de crecimiento de clientes

III. METODOLOGÍA

3.1 Tipo y diseño de investigación

El presente trabajo de estudio buscó determinar el impacto de web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo, la empresa realizaba su gestión comercial basada en conocimientos o experiencias previas, no se realiza un adecuado estudio del mercado con los datos recabados o disponibilidades en sus distintas redes. En este estudio se busca describir los beneficios de la ciencia de datos al aplicar métodos tales como el Web Scraping y el análisis de sentimientos para el progreso en la toma de decisiones que proporcionen a la compañía una adecuada gestión comercial de sus productos y clientes.

De lo ya expuesto, para la presente investigación se elige un enfoque cuantitativo debido a que permite la formulación de la hipótesis y la construcción del sistema de variables. El enfoque cuantitativo para evitar el soslayo, utiliza la medición y a tener en cuenta la objetividad del investigador frente al caso de estudio que plantea. (Galeno, 2021).

El boceto del presente estudio fue Pre-experimental, ya que se aplicarán estudios para ver el efecto que se ocasiona en la variable dependiente, para Chávez, Esparza y Riosvelasco (2020) los pre-experimentales nos aproximan al fenómeno de estudio, gestionando procedimientos para crear hipótesis y luego poder medir una o más variables y así visualizar sus secuelas.

3.2 Variables y operacionalización

Variable dependiente: La valoración del cliente hacia los servicios se compone en una porción significativa de la técnica de mando en las empresas, lo que permite un mayor conocimiento del mercado, la rivalidad y hacer frente a las variables supeditados del entorno, teniendo como misiones la satisfacción al cliente, incremento en las ventas, mejora de procesos, etc. (Bowlby, 2022. pp.50) Así mismo concuerda Starchenkova y Kornienko (2018) que menciona que la valoración del cliente hacia los servicios es un modelo estructural que influyen en la empresa a través del incremento y mejora de procesos dentro de una empresa.

Para la dimensión volumen de ventas, se usó el siguiente indicador: tasa de crecimiento de ventas. (Thomsett ,2017) (Labruneé 2018)

De la misma forma para la dimensión volumen de clientes se tiene como indicador a: tasa de crecimiento de clientes. (Thomsett, 2017) (Labruneé 2018)

Con soporte en las escalas de sondeo, se clasifican en cuatro niveles que son ordinales, nominales, razón y de intervalo.

La escala de razón en las unidades de medida posee un cero real, de tal forma que la proporción de dos unidades es autónomo de la unidad de tamaño. De esta manera los cálculos se hacen con gran amplitud, de la misma forma que las pruebas estadísticas, media geométrica y coeficientes de variaciones. (Cienfuegos y Cienfuegos, 2016, p 10). Por este motivo, en la investigación, se eligió la razón como escala de medición.

3.3 Población, muestra y muestreo

3.3.1 Población

Según Hernández, Sampieri y Mendoza (2018), una población abarca el total de los acontecimientos con características específicas en la mayoría.

En consecuencia, para este proyecto de estudio, se limitó en 30 fichas de registro para el pretest y 30 fichas de registro subsiguientes para el post test, para el indicador tasa de crecimiento de ventas y para el indicador tasa de crecimiento de clientes, en una empresa turística en la localidad de Lima.

3.3.2 Muestra

Está formada de un subconjunto de la población, y es este reducido conjunto el que nos interesa analizar y del que recogeremos los datos esenciales. Estos deben ser un reflejo del público en general. (Hernández Sampieri y Mendoza, 2018)

La demostración se apoya en la racionalidad del indagador, que afirma que, si existe una población de 100 sujetos, hay que seleccionar 30%, es decir 30 individuos. (López y Fachelli, 2015, p.33). Se seleccionó como la población como muestra, dado que el negocio calcula las fichas de registro de forma diaria realizando una consulta a través de un sistema externo a su Base de datos.

Criterios de inclusión

- Que los datos sean de este año.
- Que los datos estén fuera del rango de fechas con índices altos de covid-19.
- Que los datos sean de recientes meses.
- Que los meses que se recojan los datos sean seguidos.

Criterios de exclusión

- No se va a recoger datos de años anteriores.
- No se obtendrá datos que estén en los meses con índices altos de covid-19.
- Que los datos no se vean afectados por festejos nacionales.

3.3.3 Muestreo

Según Brandy, Lin y Mccline (2018) El estudio requiere de matices más específicos de los participantes e identificar un grupo específico, no una población generalizada. Concuenda con Mias (2018) el muestreo en el suceso de obtener una prueba estipular su esencia y complejidad, puede comprometer definir criterios de inclusión y exclusión de la prueba. Asimismo, debe permanecer determinada la población de relevancia de la prueba.

Para el estudio, se utilizó el muestreo no probabilístico, para esto se escogió directa e indirectamente las 30 fichas de registro pre-test y post-test de la compañía turística.

3.4 Técnicas e instrumentos de recolección de datos

Según Ñaupas (2018, p. 311), el fichaje es un método de estudio, subalterno de la compilación de las hojas de datos son los instrumentos utilizados para recoger datos y pueden utilizarse para el registro, el estudio documental o la investigación.

Por otra parte, para filtrar los usos de la especie con los datos recogidos del informante, se elaboró una ficha de registro que consta de dos apartados, los datos generales del informante y los usos de la especie. De este modo, en los formularios de registro se formulan preguntas cerradas de opción múltiple para facilitar la recogida de datos, así como su registro y posterior análisis estadístico. (Peña, et al. 2021)

Además, se extrajeron fichas de registro de la base de datos para el indicador índice de ventas post implementación e índice de clientes post implementación, que se utilizarán para cachear la data, ejecutados a la empresa turística (Atipax Group)

Por procedimiento, en la tabla 2, representa la técnica de recogida de información y la herramienta utilizada en este estudio:

Tabla 3: Técnica e instrumento a la variable dependiente.

| VARIABLE | DIMENSIÓN | INDICADOR | TÉCNICA | INSTRUMENTO |
|---|---------------------|---------------------------------|--------------------------|---------------------------------|
| Variable dependiente: Valoración del cliente | Volumen de Ventas | Tasa de crecimiento de ventas | Observación experimental | Ficha de registro (Ver Anexo 1) |
| | Volumen de clientes | Tasa de crecimiento de clientes | Observación experimental | Ficha de registro (Ver Anexo 2) |

Fuente: Elaboración propia.

La validación, por su parte, garantiza que la herramienta calcula con precisión la variante del análisis.

Estas escalas que tienen una validez conceptual requieren sin embargo estudios de confiabilidad y suelen respaldarse con consultas de expertos. (Mías. 2018).

Según Cabero, Romero y Palacios (2020) el juicio de expertos radica en gestionar a una serie de personas que juzguen un elemento, herramienta o material didáctico, o que den su criterio sobre un aspecto concreto. Por lo tanto, para determinar la fiabilidad del instrumento, los investigadores llevaron a cabo una validación del juicio de expertos en la que participaron tres expertos para determinar la fiabilidad y la veracidad del instrumento. (Ver Anexo 5).

Asimismo, opto por la técnica de recogida de datos: Observación experimental. Estas fueron elaboradas con la intención de conocer las dimensiones para poder analizar y posteriormente su uso estadístico en la empresa turística.

3.5 Procedimientos

Según Sartaj (2020) expresa su preocupación por los factores afectivos del alumno que le ayudan en la marcha de estudio, los estudiantes que posteriormente tengan una gran iniciativa de realizar investigaciones deberán ser avalado por un negocio o empresa donde se realice el trabajo, por ende, debe de tener una previa coordinación con el centro de trabajo.

Por lo tanto, la técnica de acopio de datos fue de la siguiente manera:

En un primer momento, se realizó una reunión con el director general para realizar la reunión. Una vez concluida la reunión, el paso posterior fue fichar el problema de investigación y el elemento de estudio. Después, el director general aceptó que el proyecto de investigación se titula " Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo " y que se realizase en la empresa turística (Atipax). A continuación, se solicitaron los primeros diez días de enero de 2022 para el registro de información de los indicadores, índice de ventas por paquete (Promoción) e índice de volumen de cliente. Los procedimientos para la recolección de datos de la investigación fueron los siguientes:

- Obtener los datos a través de consultas, para adquirir la tasa de crecimiento de clientes y ventas.
- Realizar el juicio de expertos.
- Hacer una limpieza de la data.
- Realizar una prueba de confiabilidad de datos con software SPSS.
- Realizar la prueba piloto con una muestra de 15 datos.
- Realizar ajustes en las consultas de los meses.
- Volver a realizar la prueba piloto.
- Realizar reuniones con los usuarios y expresarles los beneficios de participar en el estudio.
- Obtener el consentimiento informado de los participantes del estudio.
- Aplicar los datos pre-test a la muestra completa.
- Usar el sistema "Gestión de datos"
- Aplicar la data al post-test a la muestra completa.
- Calcular los indicadores para la comprobación de la hipótesis.

- Realizar prueba de hipótesis.

3.6 Método de análisis de datos

La data recogida en los instrumentos de medición se procesó y evaluaron mediante la estadística, y las hipótesis del estudio se confirmaron como resultado.

La prueba de normalidad se utiliza para ver si los efectos del trabajo se amoldan a una distribución normal. Cuando los valores de alguna variable continua se adquieren tras un muestreo es representativo, aparece la distribución normal. (Garrido *et al.*, 2018)

Según Goglin (2020, p.143), si los estudios de Kolmogorov-Smirnov y Shapiro-Wilks están indicadas para probar la normalidad de una distribución, se usa en muestras por debajo de 50 valores, sin evaluar la distancia relativa de la distribución considerada con una adjudicación normal.

Por lo tanto, el proyecto se utilizará la prueba de Shapiro-Wilks, ya que nuestra muestra es inferior de 50.

Asimismo, la distribución Z normalmente está determinada al nivel de que la data adquirida de la investigación, se distancia de la media. Según Entropy (2021) la distancia y la divergencia de las medidas de probabilidad desempeñan un papel central en las estadísticas, el aprendizaje automático y muchos otros campos relacionados, proporcionamos una prueba de este teorema mediante el cálculo, que se utilizará más adelante para la extensión a las distribuciones q-normales, para evaluar en qué medida la regularización de la entropía afecta o no a los resultados de la estimación, de estas se descubre dos variaciones normales a partir de la media.

Asimismo, para Xu (2017) una distribución t de Student en lugar de una distribución normal, dada la capacidad de la distribución de Student para captar más eventos de cola de cola. Sin embargo, no está claro en qué medida la distribución t de Student es superior a la distribución normal. En la figura 5, se visualiza su configuración

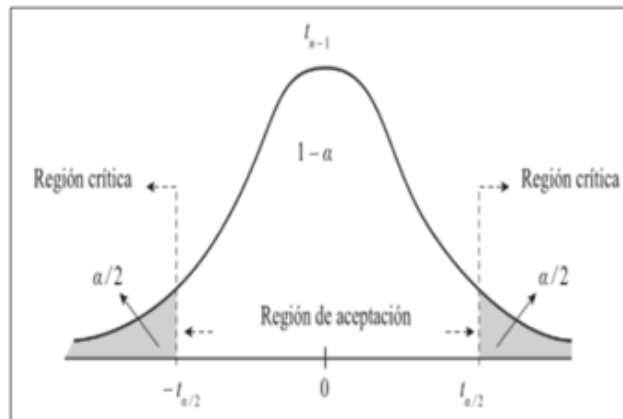


Figura 5 Distribución T-Student

Fuente: Elaboración propia

Por lo tanto, se evaluaron los datos de la ficha de registro, en el software SPSS, para hallar si la muestra es normal y si se va optar por Kolmogorov-Smirnov o Shapiro-Wilks. Por ende, se desarrolla el H_0 y H_a :

b. Criterios de decisión

Si $p < 0.05$ rechazamos la H_0 y acepto la H_a

Si $p \geq 0.05$ aceptamos la H_0 y rechazamos la H_a

Tabla 3: Prueba de normalidad

| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|--------------|---------------------------------|----|-------------------|--------------|----|------|
| | Estadístico | gl | Sig. | Estadístico | gl | Sig. |
| ClienteAbril | ,106 | 25 | ,200 [*] | ,970 | 25 | ,639 |
| ClienteMayo | ,106 | 25 | ,200 [*] | ,950 | 25 | ,245 |
| VentasAbril | ,336 | 25 | ,000 | ,471 | 25 | ,000 |
| VentasMayo | ,266 | 25 | ,000 | ,795 | 25 | ,000 |

Fuente: Elaboración propia.

Se concluye que ventas de abril y mayo tienen datos con una distribución normal, mientras que los datos de clientes, de abril y mayo tienen una distribución no normal.

De otro lado Mao (2022) nos menciona que para estudiar las propiedades asintóticas de la prueba de intención de tratar de Wilcoxon en experimentos aleatorizados con un incumplimiento normal.

3.7 Aspectos éticos

Según Papa de Garcia, Anibal y Isacc (2021) el propósito de esta investigación es definir el grado de conocimiento de los autores sobre la garantía de pertenencia

intelectual, como sus percepciones sobre los aspectos éticos, el riesgo moral y la crisis de valores en la sociedad, especialmente en el ámbito universitario., se debe enfatizar los derechos de toda información no propia. Asimismo, Ross, Iguchi y Panicker (2018) esta sección consagra una visibilidad panorámica de cómo el acceso abierto a los datos de investigación ha impactado en los fundamentos básicos de la ética de la investigación y, a su vez, cómo una reinterpretación de estos principios se traduce en procedimientos para la seguridad de los derechos y el confort de la investigación. Por lo tanto, los autores se aseguraron de que todos los datos proporcionados por la empresa de turismo no están alterados y sólo se utilizará para los fines de este estudio. Además, se mantuvo la total confidencialidad y propiedad intelectual, así como la prudencia del origen y referencias utilizadas. Así mismo, toda la información no investigada es nombrada y referenciada en resolución con la norma editorial ISO como instrumento de ayuda a la transcripción de las referencias (libros, revistas científicas y títulos de obra), lo que permite una redacción exhaustiva y fiable, respetando los derechos y la probidad del ejemplar. Gagñay y Chicaiza (2020), nos comentan que la importancia de la ética en el trabajo de investigación, así como el conocimiento de las normas de citación y referencia, es fundamental para que los investigadores no sean acusados de tomar ideas, frases o artículos que no son suyos.

IV. RESULTADOS

En el capítulo se precisa los resultados por ambos indicadores de la investigación que se nombra en el capítulo 1, los indicadores son: tasa de crecimiento de ventas y tasa de crecimiento de clientes, Para la adquisición de los resultados crecimiento de ventas y crecimiento de clientes, para los cuales se hicieron un estudio antes y después del uso del sistema. Así también para los indicadores de tasa de crecimiento de ventas y clientes, se realizó una ficha de registro.

4.1 Prueba de la hipótesis específica (HE1).

Se detallan los datos estadísticos descriptivos, de la misma forma la obtención de la data de las fichas de registro, que comprende el rango de un mes, a través de la ficha de registro, pre-test y post-test de la investigación, con el propósito de calcular el incremento de la tasa de crecimiento de ventas, a través del uso del sistema por parte de la empresa para la mejora de lanzamientos de paquetes turísticos. Asimismo, se detalla la medida de la prueba que se realizaron antes y después del uso del sistema.

Tabla 5: Incremento en la tasa de crecimiento de ventas

| Descriptivos | | Estadístico | Error típ. |
|--------------|-------|-------------|------------|
| ventasPre | Media | 1,6579 | ,73669 |
| ventasPost | Media | 2,7453 | ,60507 |

La tabla 5 especifica la media de cada muestra, adquiriendo 1.6579 en la prueba inicial y 2,7453 en la prueba final. La media de cada prueba sirve para fijar el incremento porcentual de la tasa de crecimiento de ventas.

Tabla 6: Prueba de normalidad de la tasa de crecimiento de ventas

| | Pruebas de normalidad | | | | | |
|------------|---------------------------------|----|------|--------------|----|------|
| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
| | Estadístico | gl | Sig. | Estadístico | gl | Sig. |
| ventasPre | ,248 | 25 | ,000 | ,692 | 25 | ,000 |
| ventasPost | ,203 | 25 | ,009 | ,780 | 25 | ,000 |

a. Corrección de la significación de Lilliefors

Para este indicador se aplicó el test de shapiro-wilk, dado que la muestra es inferior a 50. La tabla 6 evidencia resultados estadísticos de normalidad obtenidos al emplear este test, los cuales son 0.692 y 0.780 referente a la prueba inicial y final respectivamente.

En el ventasPre se obtuvo un estadístico de 0.692 con un valor de significancia por debajo de 0.05, siendo una distribución no normal.

En el ventasPost se obtuvo un estadístico de 0.780 con un valor de significancia por debajo de 0.05, siendo una distribución no normal.

Hipótesis Específica HE1

Hipótesis nula: el uso del sistema no incrementa las ventas de los paquetes turísticos en cuanto a la tasa de crecimiento de ventas

Hipótesis Alterna: el uso del sistema incrementa las ventas de los paquetes turísticos en cuanto a la tasa de crecimiento de ventas

Prueba de Wilcoxon

Debido a la no normalidad de las pruebas, se realizó la prueba de Wilcoxon en ambas pruebas para comparar los dos valores a continuación.

Tabla 7: Prueba de Wilcoxon-Ventas

| | | Rangos | | |
|------------------------|------------------|-----------------|----------------|----------------|
| | | N | Rango promedio | Suma de rangos |
| ventasPost - ventasPre | Rangos negativos | 6 ^a | 13,50 | 81,00 |
| | Rangos positivos | 19 ^b | 12,84 | 244,00 |
| | Empates | 0 ^c | | |
| | Total | 25 | | |

a. VentasPost < ventasPre

b. VentasPost > VentasPre

c. VentasPost = ventasPre

En la tabla 7 se puede observar la existencia 19 rangos positivos y 6 negativos, el rango promedio de rangos positivos es de 12,84 y el rango promedio de rangos negativos es 13,50, por otro lado, la suma de los rangos positivos es 244 y la suma de los rangos negativos es 81.

Tabla 8: Estadístico de prueba de Wilcoxon-Incremento de ventas

| Estadísticos de contraste ^a | |
|--|---------------------------|
| | ventasPost - ventasPre |
| Z | -2,193 ^b |
| Sig. asintót. (bilateral) | ,028 |

a. Prueba de los rangos con signo de Wilcoxon

b. Basado en los rangos negativos.

En la tabla 8 muestra los estadísticos de prueba con un valor de $Z=-2,193$ donde Z pertenece a la zona de rechazo. También el valor de P está por debajo de 0.05, aceptando la hipótesis alterna con un nivel de confianza de 95%. Por ello se demuestra que el uso del sistema gestión de datos aumenta la tasa de incremento de ventas.

Para la obtención del incremento porcentual del aumento de la tasa de incremento de ventas se tuvo que analizar las medias de cada prueba, teniendo como prueba inicial una media de 1,6579, y en el post una media de 2,7453, la cual indica un incremento de 65.6%. A continuación, el análisis operacional para la obtención del porcentaje.

Tasa de crecimiento de ventas:

$$TCV = \frac{VMA - VME}{VME} * 100$$

VMA: Ventas mes actual.

VME: Ventas mes anterior.

TCV: Tasa de crecimiento de ventas

$$TCV = \frac{2.7453 - 1.6579}{1.6579} * 100 = 65.6\%$$

Después de procesar los datos con el software estadísticos (SPSS) considerando la significancia de los factores fue 0.028, y al ser menor que 0.05 con un nivel de confianza de 95% considerando que la media antes y después fueron significativamente diferentes por lo cual se acepta su hipótesis alterna rechazando la hipótesis nula, se puede afirmar que el uso del sistema para captar la valoración del cliente hacia los servicios de una empresa de turismo, incremento las captación del cliente y por lo tanto el aumento de las ventas.

4.2 Prueba de hipótesis específica (HE2)

Se detallan los datos estadísticos descriptivos, de la misma forma la obtención de la data del grupo de estudio, que comprende el rango de un mes, a través de la ficha de registro, pre-test y post-test de la investigación, con la finalidad de medir el incremento de la tasa de crecimiento de clientes, a través del uso del sistema por parte de la empresa para la mejora de lanzamientos de paquetes turísticos. Asimismo, se detalla la medida de la prueba que se realizaron antes y después del uso del sistema.

Tabla 9: Incremento en la tasa de crecimiento de clientes

| Descriptivos | | Estadístico | Error típ. |
|--------------|-------|-------------|------------|
| clientesPre | Media | 1,1925 | 0,62222 |
| clientesPost | Media | 1,6702 | 0,42638 |

La tabla 9 especifica la media de cada muestra, adquiriendo 1,1925 en la prueba inicial y 1,6702 en la prueba final. La media de cada prueba sirve para fijar el incremento porcentual de la tasa de crecimiento de Clientes.

Tabla 10: Prueba de normalidad de la tasa de crecimiento de Clientes

Pruebas de normalidad

| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|--------------|---------------------------------|----|------|--------------|----|------|
| | Estadístico | gl | Sig. | Estadístico | gl | Sig. |
| clientesPre | ,332 | 25 | ,000 | ,560 | 25 | ,000 |
| clientesPost | ,289 | 25 | ,000 | ,575 | 25 | ,000 |

a. Corrección de la significación de Lilliefors

Para este indicador se aplicó el test de shapiro-wilk, debido a que la muestra es inferior a 50. La tabla 10 evidencia resultados estadísticos de normalidad obtenidos al emplear este test, los cuales son 0.560 y 0.575 referente a la prueba inicial y final respectivamente.

En el pre_test se obtuvo un estadístico de 0.560 con un valor de significancia por debajo de 0.05, siendo una distribución no normal.

En el pos_test se obtuvo un estadístico de 0.575 con un valor de significancia por debajo de 0.05, siendo una distribución no normal.

Hipótesis Específica HE2

Hipótesis nula: el uso del sistema no incrementa la captación de clientes en cuanto a la tasa de crecimiento de clientes

Hipótesis Alterna: el uso del sistema incrementa la captación de clientes en cuanto a la tasa de crecimiento de clientes

Prueba de Wilcoxon

Debido a la no normalidad de las pruebas, se realizó la prueba de Wilcoxon en ambas pruebas para comparar los dos valores a continuación.

Tabla 11: Prueba de Wilcoxon-Clientes

| | N | Rango promedio | Suma de rangos |
|----------------------------|-----------------|----------------|----------------|
| clientesPost - clientesPre | | | |
| Rangos negativos | 8 ^a | 10,63 | 85 |
| Rangos positivos | 17 ^b | 14,12 | 240 |
| Empates | 0 ^c | | |
| Total | 25 | | |

a. clientesPost < clientesPre

b. clientesPost > clientesPre

c. clientesPost = clientesPre

En la tabla 11 se observa la existencia 17 rangos positivos y 8 negativos, el rango promedio de rangos positivos es de 14,12 y el rango promedio de rangos negativos es 10,63, por otro lado, la suma de los rangos positivos es 240 y la suma de los rangos negativos es 85.

Tabla 12: Estadístico de prueba de Wilcoxon-Incremento de Clientes

| Estadísticos de contraste ^a | |
|--|-------------------------------|
| | clientesPost - clientesPre |
| Z | -2,085 ^b |
| Sig. asintót. (bilateral) | ,037 |

a. Prueba de los rangos con signo de Wilcoxon

b. Basado en los rangos negativos.

En la tabla 12 muestra los estadísticos de prueba con un valor de $Z=-2,085$ donde Z pertenece a la zona de rechazo. También el valor de P está por debajo de 0.05, aceptando la hipótesis alterna con un nivel de confianza de 95%. Por ello se demuestra que el uso del sistema gestión de datos aumenta la tasa de incremento de clientes.

Para la obtención del incremento porcentual del aumento de la tasa de incremento de clientes, se tuvo que analizar las medias de cada prueba, teniendo como prueba inicial una media de 1,1925, y en el post una media de 1,6702, la cual indica un incremento de 40.1%. A continuación, el análisis operacional para la obtención del porcentaje.

Tasa de crecimiento de clientes:

$$TCC = \frac{CMA - CME}{CME} * 100$$

CMA: Clientes mes actual.

CME: Clientes mes anterior.

TCC: Tasa de crecimiento de clientes

$$TCC = \frac{1.6702 - 1.1925}{1.1925} * 100 = 40.1\%$$

Después de procesar los datos con el software estadísticos (SPSS) se obtuvo que la significancia de los factores fue 0.037, y al ser menor que 0.05 con un nivel de confianza de 95% considerando que la media antes y después fueron significativamente diferentes por lo cual se acepta su hipótesis alterna rechazando la hipótesis nula, se puede afirmar que el uso del sistema para captar la valoración del cliente hacia los servicios de una empresa de turismo, incremento las captación del cliente.

V. DISCUSIÓN

Este capítulo de discusión que se comparan los resultados obtenidos a partir de las hipótesis de este estudio. Los resultados se comparan con los antecedentes y las teorías relacionadas de este estudio. Sobre los resultados del uso adicional del sistema de gestión de datos, se lleva a cabo la siguiente discusión.

El sistema gestión de datos elevo la tasa de crecimiento de ventas en un 65.6% en ventas de paquetes a diferencia del pre-test. También se tuvo el incremento en la tasa de crecimiento de clientes del 40.1% en la captación de clientes, a través de las técnicas de análisis de sentimientos y el web scraping, en conjunto a la metodología hibrida CRISP-DM y Scrum.

El sistema gestión de datos cuenta con características similares a otras aplicaciones como las de Pelayo (2022), Van Leeuwen (2020), Garcés (2019), Küçükbatir (2018), Delgado (2018), Clafoque y Zaravia (2021), Pineda (2021), Mescua (2020), Gil (2019), Porras (2019), quienes emplearon los métodos del data Science como web scraping y el análisis de sentimiento, así mismo con el marco de trabajo SCRUM y la metodología CRISP-DM, con la finalidad de incrementar el conocimiento de los beneficios del web scraping y el análisis de sentimientos. Estas técnicas también estuvieron presentes en el sistema gestión de datos.

Según los resultados adquiridos se lograron incrementar la tasa de crecimiento de clientes en 40.1%, siendo superior a los resultados de los estudios de Garcés (2019), quien empleo el análisis de sentimientos en las redes sociales, con dirección a la captación de usuarios a través de Twitter, que obtuvo como resultados obtenidos a través del análisis de sentimientos, que indicaron que los comentarios no positivos para Enel Chile fueron entre el 11.56% y el 16.99%.

Porras (2019), señala en su investigación se determinó que el 23.5 % de microempresas ha usado un metodo tecnológico para gestionar a sus usuarios y elaborar un mejor proceso administrativo de su negocio para aumentar la captación de clientes. Logrando los resultados 74,2% tiene una preferencia por las redes sociales. Tomando todo esto en cuenta se puede

declarar que el utilizar las técnicas de ciencia de datos ayudaría a una eficiente administración en la gestión de las microempresas.

Pineda (2021), tuvo como principales resultados del 9,7% de los encuestados cree que los objetivos marcados en el negocio no se basan en el verdadero potencial de la cartera de clientes, mientras que el 17,7% no tiene una opinión clara al respecto. Finalmente, el 27,4% está muy de acuerdo en que los objetivos se fijan en función del potencial real de los clientes de la empresa.

Van Leeuwen (2020), señala en su investigación del uso de las técnicas de la data Science para evaluar el rendimiento de equipos deportivos, los resultados indican, que la precisión del conjunto de pruebas fue de aproximadamente el 57%, esto último demuestra que es difícil predecir con certeza el resultado de un partido de fútbol a partir de las variables estadísticas disponibles.

Delgado (2018), menciona en su estudio, la exploración de distintas páginas web a través del web scraping. Al concluir la implementación se establecieron estrategias de post ventas, fidelización de clientes mediante recompensas. Los resultados arrojaron que el 53% de los encuestados está satisfecho, mientras que el 32% está muy satisfecho, lo que confirma que la calidad de este tipo de servicio hotelero está por encima de la media y es por ello que los clientes los prefieren porque les interesan sus preferencias.

Küçükbatır (2018), quienes emplearon la adopción e influencias del big data en conjunto a las técnicas de data Science, para la investigación mediante la recopilación y el análisis de índices e informes de las principales empresas que operan en la sección de los viajes y el turismo dirigido a las personas, que como resultados obtuvo el sector turístico ha demostrado su capacidad de recuperación de clientes, contribuyendo a un crecimiento directo del PIB del 4,1%.

Según los resultados adquiridos se lograron incrementar la tasa de crecimiento de ventas en 65.6%, siendo superior a los resultados de los

estudios de Pelayo (2022), quien empleo los modelos del data Science conocidos como análisis de datos, web scraping, en conjunto a otros software, como KNIME, con la meta de conseguir reducción en el periodo debido a la automatización, pronóstico y una toma de decisiones adecuada en cuanto a las ventas y promociones o paquetes. Se aprecia que para el grupo de precios bajos 0,727 y para el grupo de precios altos 0,539. Siendo más efectivo el modelo para el conjunto de precios bajos.

Gil (2019), señala que la aplicación de la gestión comercial para el uso de redes sociales, para aumentar las ventas, se requiere de la captación de clientes a través de las redes sociales, se determinó que al intensificar la interacción con los medios sociales la empresa pudo captar unos 45% más seguidores a lo que proporcionalmente da un aumento en las venta directamente. De esto se concluye que las verificaciones de interacción tanto como “Likes, Comentarios y Seguidores” son generadas por una mejoría en la gestión comercial en el crecimiento de ventas y clientes de las empresas a través de sus redes sociales.

Delgado (2018), tuvo como principal resultados que muestran que el 53% de los encuestados dijeron estar satisfechos; y el 32% se mostró muy satisfecho, lo que confirma que la calidad del servicio de este tipo de hoteles está por encima de la media y por eso los clientes siempre los han preferido. Se establecieron estrategias de post ventas, fidelización de clientes mediante recompensas; con el fin de aumentar las ventas y de igual manera amparar a sus usuarios sin que estos dispongan por la competencia.

Pineda (2021), señala en su estudio que el 64,5% de los encuestados cree que es importante y necesario que todo el equipo de ventas comprenda el proceso de planificación y seguimiento de las metas del territorio de ventas, que a través de un sistema especializado con técnicas de data Science incrementaría las ventas, los resultado obtenidos fue de 27.4% más de las ventas usuales.

Mescua (2020), señala en su estudios obtuvo que los resultados son de 66.7% que genera rentabilidad, aplicando el análisis de sentimientos,

esto significa que los datos completos de implementación brindan resultados favorables para la empresa, mientras que el 33,3% no lo considera.

Clafoque y Zaravia (2021), menciona en cuanto al uso de las técnicas del Data Science en las organizaciones, se obtiene que los resultados son menos del 50 % de los asalariados confía en su empresa con sus datos para realizar estudios sobre el crecimiento de ventas y el 44 % no recibe ninguna información recopilados sobre ellos.

Küçükbatir (2018), señala en su investigación el uso de web scraping para obtener datos de trivago y Merglobal en conjunto de las técnicas de data Science como el análisis de sentimientos, dan como resultado que el sector turístico ha demostrado su capacidad de recuperación, contribuyendo a un crecimiento en ventas del 4.1% del PBI.

VI. CONCLUSIONES

Las resoluciones de la investigación fueron las siguientes:

- A. El sistema de gestión de datos, elevo la tasa de crecimiento de ventas en 65.6% sobre captación de la valoración del cliente hacia los servicios, gracias al marco de trabajo de Scrum en conjunto a la metodología CRISP-DM, que nos da a entender las necesidades del clientes y del negocio.
- B. El sistema de gestión de datos, elevo la tasa de crecimiento de clientes en 40.1%, gracias a las características del sistema de identificar las preferencias de las zonas turísticas, a través de Twitter y Google Places con el análisis de sentimiento y el web scraping.
- C. El web scraping permite adquirir una gran cantidad de información de una manera automática de diferente sitios web, gracias a esto se creó una base de datos aprovechable, para su aplicaciones en el análisis de sentimientos.
- D. El análisis de sentimientos resulta una técnica de alta complejidad, tanto por las implicaciones de procesamiento de datos que involucra, cómo por la naturaleza de los resultados que se pueden obtener y su nivel de precisión dada la ambigüedad que puede contener un estado de ánimo.
- E. Se concluye que el sistema de gestión de datos con métodos de data Science tales como web scraping y análisis de datos, mejoró la captación de la valoración del cliente hacia los servicios de una empresa de turismo, lo que permitió lograr los objetivos del presente estudio.

VII. RECOMENDACIONES

La presente investigación con base a sus resultados y sus conclusiones, se recomienda lo siguiente:

1. Considerando la respuesta positiva de la empresa de investigación en la implementación del sistema de gestión de datos, se recomienda continuar invirtiendo en tecnología para su empresa, al tiempo que introducen técnicas de aprendizaje automático y automatización de procesos.
2. La empresa de investigación han respondido bien a la implementación del sistema de gestión de datos, y con esto en mente, se recomienda que su sistema SATO de registro amplíen el uso de técnicas de ciencia de datos para mejorar aún más su proceso de ventas.
3. Por último, se recomienda, también emplee soluciones de data Science en conjunto a machine learning en los procesos logísticos, que permitan aumentar la experiencia. Por ejemplo, Pelayo (2022), que tiene por finalidad indicar, de una forma simple, los procedimientos a continuar para transformar la data en información y esta misma en entendimiento con la meta de conseguir reducción en el periodo debido a la automatización, pronóstico y una toma de decisiones adecuada.

REFERENCIAS

- ADHIKARI, Animesh & ADHIKARI, Jhimli, 2015. *Advances in Knowledge Discovery in Databases*. ISBN: 978-3-319-13211-2. DOI: 10.1007/978-3-319-13212-9.
- AZWINDINI, I.R., 2022. Musina Municipality Tourism Management and Strategies: A Sustainable-Eco-Tourism Inclusive Business Insights for the Town, Abbuting Peri-Urban and Countryside Existences. *Management of Environmental Quality*, vol. 33, no. 3, pp. 718-738 ProQuest Central. ISSN 14777835. DOI <https://doi.org/10.1108/MEQ-11-2021-0257>.
- BATISTIČ, S. and VAN DER LAKEN, P, 2019. History, Evolution and Future of Big Data and Analytics: A Bibliometric Analysis of Its Relationship to Performance in Organizations. *Brit J Manage*, Vol 30. pp 229-251. ISSN: 1467-8551. Doi: <https://doi.org/10.1111/1467-8551.12340>
- BBVA, 2019. Educación financiera. En: BBVA [en línea]. Disponible en: <https://www.bbva.mx/educacion-financiera/f/financiamiento.html>
- BEYNON-DAVIES, Paul. 2018. *Sistemas de Bases de Datos*. Editorial Reverté, ISBN: 9788429143966. Disponible en: <https://www.proquest.com/legacydocview/EBC/6795626?accountid=37408>
- BOWLBY, Rachel. 2022. *Talking Walking: Essays in Cultural Criticism*. Liverpool University Press. ISBN: 9781845199111. pp. 256. disponible en: <https://www.proquest.com/legacydocview/EBC/5224802?accountid=37408>.
- BRANDY, B.W., LIN, Y. and MCCLINE, R.M., 2018. Q Methodology and Q-Perspectives® Online: Innovative Research Methodology and Instructional Technology. *TechTrends*, 09, vol. 62, no. 5, pp. 450-461 ProQuest Central. ISSN 87563894. DOI <https://doi.org/10.1007/s11528-018-0314-5>.
- BROUCKE, Seppe vanden, BAESSENS, Bart (2018). *Practical Web Scraping for Data Science*. Apress Media LLC: Welmoed Spahr. ISBN: 978-1-4842-3581-2. DOI: <https://doi.org/10.1007/978-1-4842-3582-9>

- CABERO-ALMENARA, J., ROMERO-TENA, R. and PALACIOS-RODRÍGUEZ, A., 2020. Evaluation of Teacher Digital Competence Frameworks through Expert Judgement: The use of the Expert Competence Coefficient. *Journal of New Approaches in Educational Research*, vol. 9, no. 2, pp. 275-0_1,0_2 ProQuest Central. DOI <https://doi.org/10.7821/naer.2020.7.578>.
- CABRERA, S. y VILLAREJO, R. 2019. FACTORS EFFECTING THE ADOPTION OF BIG DATA ANALYTICS IN COMPANIES. *Revista de administración de empresas*, Nov, vol. 59, no.6, pp. 415-429 ProQuest Central. ISSN 00347590. DOI <http://dx.doi.org/10.1590/S0034-759020190607>.
- CHANG, Y. and KOU, M., 2020. Big Data Analysis of Demand for Power Supply Service in Urbanization Development of Tourism Industry in Shaanxi Province. *IOP Conference Series. Earth and Environmental Science*, 02, vol. 440, no. 3 ProQuest Central. ISSN 17551307. DOI <https://doi.org/10.1088/1755-1315/440/3/032076>.
- CHÁVEZ V., S.M., ESPARZA del V., Ó.A. y RIOSVELASCO M., L. (2020). Diseños preexperimentales y cuasiexperimentales aplicados a las ciencias sociales y a la educación. *Enseñanza e Investigación en Psicología*, 2(2), 167-178. ISSN: 2683-2046
- CHEN, N., 2020. Application of Big Data Technology in Smart Tourism. *Journal of Physics: Conference Series*, 10, vol. 1648, no. 4 ProQuest Central. ISSN 17426588. DOI <https://doi.org/10.1088/1742-6596/1648/4/042101>.
- CIENFUEGOS VELASCO, Maria de los Angeles y CIENFUEGOS VELASCO, Adriana, 2016. Lo cuantitativo y cualitativo en la investigación. Un apoyo a su enseñanza. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, vol. 7, núm. 13, Julio-diciembre, 2016, pp. 15-36. ISSN: 2007 – 7467. Disponible en: <https://www.redalyc.org/articulo.oa?id=498155462002>.

CLAFOQUE AREVALO, Wilmer Ulises, ZARAVIA MIOCA, Ana Cecilia. Uso de la ciencia de datos en las organizaciones. 2021. Disponible en: <http://hdl.handle.net/10757/658104>

CLARKE BLOOMFIELD, Merlinda, CISNEROS ARIAS, Yoanis, PANECA GONZALES, Yuriana (2018). Gestión Comercial: diagnóstico del atractivo y rentabilidad del punto de ventas. ISSN: 1027-2127. Disponible en: <https://www.redalyc.org/articulo.oa?id=181557161002>

Conexión-ESAN (2016). Administrar eficientemente una cartera de clientes. Blog Online Apuntes empresariales / Ventas ESAN. Recuperado el 21/05/2022. Disponible en: <https://www.esan.edu.pe/apuntes-empresariales/2016/05/administrar-eficientemente-cartera-clientes/>

DAMARTA, R., HIDAYAT, A. and ABDULLAH, A.S., 2021. The Application of 0RW1S34RfeSDcfkexd09rT2k1RW1S34RfeSDcfkexd09rT2-Nearest Neighbors Classifier for Sentiment Analysis of PT PLN (Persero) Twitter Account Service Quality. Journal of Physics: Conference Series, 01, vol. 1722, no. 1 ProQuest Central. ISSN 17426588. DOI <https://doi.org/10.1088/1742-6596/1722/1/012002>.

DELGADO CARRANZA, Luis Miguel. Plan de fidelización de clientes enfocado en la mejora de la gestión comercial de los hoteles de primera categoría en la Ciudad de Manta. 2018. Disponible en: <https://repositorio.espam.edu.ec/bitstream/42000/942/6/TT87.pdf>

DRONOV, V.Y. and DRONOVA, G.A., 2022. Python as an Automation Tool in IS. Protecting Database Access in Python. Journal of Physics: Conference Series, 03, vol. 2182, no. 1, pp. 012093 ProQuest Central. ISSN 17426588. DOI <https://doi.org/10.1088/1742-6596/2182/1/012093>.

Egger, R., Yu, J. (2022). Data Science and Interdisciplinarity. In: Egger, R. (eds) Applied Data Science in Tourism. Tourism on the Verge. pp. 35-49. Springer, Cham. ISBN: 978-3-030-88389-8. https://doi.org/10.1007/978-3-030-88389-8_3

- EGGER, R., KRONER, M., STÖCKL, A. (2022). Web Scraping. In: Egger, R. (eds) Applied Data Science in Tourism. Tourism on the Verge. Springer, Cham. pp. 67-82. ISBN: 978-3-030-88389-8. https://doi.org/10.1007/978-3-030-88389-8_5
- ENTROPY, Basel. 2021. Entropy-Regularized Optimal Transport on Multivariate Normal and q-Normal Distributions. Entropy, vol. 23, no. 3, pp. 302 ProQuest Central. E-ISSN: 10994300. DOI <https://doi.org/10.3390/e23030302>.
- GAGÑAY, Lisbeth Katerine Inguillay; CHICAIZA, Silvia Lorena Tercero; AGUIRRE, José López. Ética en la investigación científica. Revista Imaginario Social, 2020, vol. 3, no 1. ISSN: 2737-6362
- GALICIA, Liliana, BALDERRAMA, Jorge y NAVARRO, Rubén (2017). Validez de contenido por juicio de expertos: propuesta de una herramienta virtual. ISSN: 1665–6180. Disponible en <http://www.scielo.org.mx/pdf/apertura/v9n2/2007-1094-apertura-9-02-00042.pdf>
- GARCÉS CHAPARRO, Tomás Ignacio. Análisis de sentimientos en redes sociales orientado a la percepción de la calidad de servicios de internet, redes móviles, tv cable y electricidad. 2019. Disponible en: https://repositorio.unab.cl/xmlui/bitstream/handle/ria/17963/a131441_Garces_T_Analisis_de_sentimientos_en_redes_2019_tesis.pdf?sequence=1
- GARCÍA CUSCÓ, Manuel; LÓPEZ LORES, Imirsy, 2014. Definición, clasificación de la variable recursos financieros docentes para la evaluación institucional. Revista Electrónica EduSol. Vol 14, No 47, pp 1-10. ISSN: 1729-8091.
- GARRIDO, Edgardo, GÓNZALES, Laura, CABRERA, Jairo, ROJAS, Jamil, ARIAS, María y TELLO, Matthew. Normal y lognormal: dos distribuciones de frecuencias y una Teoría Neutral Unificada para estudiar los bosques tropicales. ISSN 1727-9933. Fecha de consulta 18 de mayo del 2022.

Disponible

en:

http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1727-99332018000200014

GIL ZUÑIGA, Mindy Francine. Gestion commercial para el uso de redes sociales en empresas de ventas de ropa de la ciudad de Chiclayo 2019. Disponible en: <http://hdl.handle.net/20.500.12423/3212>

GOGLIN, C. (2020). L' affectif et les valeurs dans les choix d'investissement en equity crowdfunding 2. ISTE Editions Ltd. SBN: 9781784056964. ProQuest Ebook Central, <https://www.proquest.com/legacydocview/EBC/6482453?accountid=37408>.

GRUS, Joel (2019). Data Science from Scratch. 2° ed. EEUU: O'Reilly Media, Inc. ISBN: 9781492041139

Hernández-Leal, Emilcy J.; Duque-Méndez, Néstor D.; Moreno-Cadavid, Julián, 2017. Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. Tecno Lógicas, vol. 20, núm. 39. ISSN: 0123-7799

HERNÁNDEZ SAMPIERI, R. y MENDOZA, C., 2018. Metodología de la Investigación: Las rutas de la investigación. S.l.: s.n. ISBN 978-1-4562- 6096-5.

HOU, J., YANG, Y. and BAI, Y., 2018. Evaluation of all-for-One Tourism in Mountain Areas using Multi-Source Data. *Sustainability*, vol. 10, no. 11, pp. 4065 ProQuest Central. E-ISSN: 20711050. DOI <https://doi.org/10.3390/su10114065>.

HUANG, Z., SAVITA, K.S. and ZHONG-JIE, J., 2022. The Business Intelligence Impact on the Financial Performance of Start-Ups. *Information Processing & Management*, 01, vol. 59, no. 1, pp. 1 ProQuest Central. ISSN 03064573. DOI <https://doi.org/10.1016/j.ipm.2021.102761>.

- IAB SPAIN. (2017). Estudio anual de redes sociales 2017. Recuperado de http://iabspain.es/wp-content/uploads/iab_estudiodedessociales_2017_vreducida.pdf
- IBM Analytics. (2015). Metodología Fundamental para la Ciencia de Datos. Recuperado de: <https://www.ibm.com/downloads/cas/6RZMKDN8>
- JULCA MEZA, Emma Carmen, 2020. Pandemias y su efecto en el turismo: lecciones del pasado y nuevos retos. *Cultura*. no. 34, pp. 159 – 166. ISSN: 2224 – 3585. Doi: <https://doi.org/10.24265/cultura.2020.v34.11>
- JURNEY, Russell, 2017. *Agile Data Science 2.0*. Sebastopol: O'Reilly Media, Inc. ISBN: 978-1-491-96011-0
- KACHNIEWSKA, M., 2020. The use of Big Data in Tourism Sales Forecasting. *International Journal of Contemporary Management*, no. 19, pp. 7-35 Coronavirus Research Database; ProQuest Central. ISSN 24498920. DOI <https://doi.org/10.4467/24498939IJCM.20.004.12669>.
- KALVET, T., OLESK, M., TIITS, M. and RAUN, J., 2020. Innovative Tools for Tourism and Cultural Tourism Impact Assessment. *Sustainability*, vol. 12, no. 18, pp. 7470 Coronavirus Research Database; ProQuest Central. E-ISSN: 20711050. DOI <https://doi.org/10.3390/su12187470>.
- KASSEL, R., et al, 2019. The Impact of Data Science as a Liberal Art. *Scholarship and Practice of Undergraduate Research*, Summer, vol. 2, no. 4, pp. 44 ProQuest Central. ISSN 2476101X. DOI <https://doi.org/10.18833/spur/2/4/4>.
- KEENAN, Mark. 2020. Advanced Positioning, Flow, and Sentiment Analysis in Commodity Markets: *Bridging fundamental and technical analysis*. John Wiley & Sons, Incorporated. pp. 283. ISBN: 9781119603740. Disponible en: <https://www.proquest.com/legacydocview/EBC/6001542?accountid=37408>

- KEWSUWUN, N. and KAJORNKASIRAT, S., 2022. A Sentiment Analysis Model of Agritech Startup on Facebook Comments using Naive Bayes Classifier. *International Journal of Electrical and Computer Engineering*, 06, vol. 12, no. 3, pp. 2829-2838 ProQuest Central. DOI <https://doi.org/10.11591/ijece.v12i3.pp2829-2838>.
- KÜÇÜKBATIR, Merve, 2018. THE ADOPTION AND INFLUENCES OF BIG DATA IN TOURISM INDUSTRY IN TURKEY. Tesis Magisteral. Estambul: Universidad Kadir Has. Disponible en: <https://webcache.googleusercontent.com/search?q=cache:a-K3gINDSyYJ:https://academicrepository.khas.edu.tr/bitstream/handle/20.500.12469/2287/0097760MerveKucukbatir.pdf%3Fsequence%3D1%26isAllowed%3Dy&cd=2&hl=es&ct=clnk&gl=pe&client=opera-gx>
- LABRUNÉE, María Eugenia. El crecimiento y el desarrollo. 2018. Disponible en: <http://nulan.mdp.edu.ar/id/eprint/2883>
- LÓPEZ, Pedro y FACHELLI, Sandra. 2015. Metodología de la investigación social cuantitativa. 1era Edición. España Barcelona: Edición Digital Creative Commons, 2015 [fecha de consulta: 17 de mayo del 2022]. Disponible en https://ddd.uab.cat/pub/caplli/2017/185163/metinvsocqua_cap2-4a2017.pdf
- MAILUND, Thomas (2017). *Beginning Data Science in R: Data Analysis, Visualization, and Modelling for the Data Scientist*. New York: Springer Science+Business Media. ISBN: 978-1-4842-2670-4. DOI: 10.1007/978-1-4842-2671-1
- MAO, L., 2022. On the Relative Efficiency of the Intent-to-Treat Wilcoxon–Mann–Whitney Test in the Presence of Noncompliance. *Biometrika*, 09, vol. 109, no. 3, pp. 873-880 ProQuest Central. ISSN 00063444. DOI <https://doi.org/10.1093/biomet/asab053>.
- MARÍN, María Eumelia Galeano. *Investigación cualitativa: Preguntas inagotables*. Fondo Editorial FCSH, 2021. ISBN: 9585157519, 9789585157514

- MESCUA SALHUANA, Moises Carlos. El Big Data Analytics y la Competitividad Empresarial Peruana. 2020. Disponible en : <https://repositorio.urp.edu.pe/handle/URP/3941>
- MIAS, Carlos. 2018. Metodología de investigación, estadística aplicada e instrumentos en neuropsicología: guía práctica para investigación – 1 a ed. – Córdoba: Encuentro Grupo Editor. ISBN 987-987-4078-98-8
- MINGYU, J., Contributed equally to this work with: Ji Mingyu, ZHOU JIAWEI, Z.J. and ZHOU, J., 2022. AFR-BERT: Attention-Based Mechanism Feature Relevance Fusion Multimodal Sentiment Analysis Model. *PLoS One*, 09, vol. 17, no. 9 ProQuest Central. DOI <https://doi.org/10.1371/journal.pone.0273936>.
- ÑAUPAS, Humberto. 2018. Metodología de la investigación cuantitativa - cualitativa y redacción de la tesis. 5ta Edición. Ediciones de la U, [fecha de consulta: 17 de mayo del 2022]. Disponible en <http://www.ebooks7-24.com/stage.aspx?il=&pg=&ed=>
- PELAYO GARCÍA, Bermejo, 2022. Ciencia de datos y KNIME, combinación perfecta para el triunfo en la toma de decisiones. Universitat Oberta de Catalunya.
- PAPA DE GARCÍA, R., ANÍBAL, C.R. and ISACC A., A.S., 2021. Protección a La Propiedad Intelectual Del Autor En Perú En Tiempos De Crisis Moral. *Revista Interamericana De Bibliotecología*, vol. 44, no. 1, pp. e1n1/1-e1n2/13 ProQuest Central. ISSN 01200976. DOI <https://doi.org/10.17533/udea.rib.v44n1e1n2>.
- PEÑA-GUAMÁN, C., GUERRERO-TIPANTUÑA, M., ARCOS-BÓSQUEZ, V. and BAYAS-MOREJÓN, F., 2021. Cosmovisión Etnobiológica De La Cultura Ancestral Florística De La Comunidad Kichwa Quindihua (Ecuador). *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 3, pp. 3109-3130 ProQuest Central. ISSN 20673019.

PÉREZ LÓPEZ, César, 2021. DATA MINING. The CRISP-DM METHODOLOGY. The CLEM language and IBM SPSS MODELER. Lulu Press, Inc. ISBN: 9781008981652

PINEDA CONTRERAS, Jance Erickson. Estrategias para la mejora de la gestión comercial que generen un impacto en la cartera de clientes de la agencia de carga internacional G&G Logistics s.a.c en el año 2019. 2021. Disponible en: <https://repositorio.urp.edu.pe/handle/URP/4193?show=full>

PONCE ANDRADE, Johnny Edison, 2017. La gestión comercial y su influencia en el crecimiento de las pymes hoteleras de Manabí. Revista ECA Sinergia. Facultad de Ciencias Administrativas y Económicas.Vol. 8 N°2. ISSN: 2528-7869.

PORRAS CERRÓN, Jaime Carlos, La administración empresarial eficiente con ayuda del big data en el desarrollo de las microempresas de lima metropolitana, caso Gamarra. Disponible en: <http://repositorio.unfv.edu.pe/bitstream/handle/UNFV/3728/PORRAS%20CERRON%20JAIME%20CARLOS%20-%20DOCTORADO.pdf?sequence=1&isAllowed=y>

Promperú, 2022. Perspectivas de la recuperación turística en el Perú. Turismoin investiga innova. Disponible: <https://www.promperu.gob.pe/turismoin/sitio/VisorDocumentos?titulo=Perspectivas%20de%20la%20recuperaci%C3%B3n%20tur%C3%ADstica%20en%20el%20Per%C3%BA&url=Uploads/infografias/1109/Perspectivas%20de%20la%20recuperaci%C3%B3n%20tur%C3%ADstica%20en%20el%20Per%C3%BA.pdf&nombObjeto=BibliotecaReportes&back=/TurismoIn/&issuuid=0>

ROMERO L. y RIVERA D.2019. La comunicación en el escenario digital: Actualidad, retos y perspectivas. Pearson Educación de Perú, S.A. 2019. ISBN: 978-607-32-4859-4.

- ROSS, M.W., IGUCHI, M.Y. and PANICKER, S., 2018. Ethical Aspects of Data Sharing and Research Participant Protections. *The American Psychologist*, Feb, vol. 73, no. 2, pp. 138 ProQuest Central. ISSN 0003066X.
- SARKER, I.H. (2021) Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, Vol 2, No. 160. ISSN: 2662-995X
Doi: <https://doi.org/10.1007/s42979-021-00592-x>
- SARTAJ, S. and SYED RAZAQ, A.S., 2020. Determining the Factors that Intensify Oral Communicative Anxiety: A Case Study of Students of IELL, University of Sindh, Jamshoro. *International Research Journal of Arts and Humanities*, vol. 48, no. 48, pp. 173 ProQuest Central. ISSN 10169342.
- SAURA.J. REYES-MENENDEZ, A. and PALOS-SANCHEZ, P., 2019. Are Black Friday Deals Worth it? Mining Twitter Users' Sentiment and Behavior Response. *Journal of Open Innovation : Technology, Market, and Complexity*, vol. 5, no. 3, pp. 58 ProQuest Central.e-ISSN21998531
DOI <https://doi.org/10.3390/joitmc5030058>.
- SHAHBAZ, M., et al, 2021. Impact of Big Data Analytics on Sales Performance in Pharmaceutical Organizations: The Role of Customer Relationship Management Capabilities. *PLoS One*, 04, vol. 16, no. 4 ProQuest Central. e-ISSN: 19326203. DOI <https://doi.org/10.1371/journal.pone.0250229>.
- SKEEN, S.J., JONES, S.S., CRUSE, C.M. and HORVATH, K.J., 2022. Integrating Natural Language Processing and Interpretive Thematic Analyses to Gain Human-Centered Design Insights on HIV Mobile Health: Proof-of-Concept Analysis. *JMIR Human Factors*, 07, vol. 9, no. 3 ProQuest Central. e-ISSN: 22929495. DOI <https://doi.org/10.2196/37350>.
- SORDONI, L.M., ERANSUS, B.P. and SALMERÓN, V.S., 2022. POTENCIALIDADES Y LÍMITES DE LA INTEGRACIÓN DE SERVICIOS SOCIALES Y DE EMPLEO LA VALORACIÓN DE LOS ACTORES IMPLICADOS. *Prisma Social*, 04, no. 37, pp. 315-339 Coronavirus Research Database; ProQuest Central. Disponible en: e-ISSN- 19893469

- STARCHENKOVA, Elena y KORNIENKO, Anzhelika. 2018. Structural model of personal qualities influence on professional health of sales managers. Vol. 39, no. 34. pp30. ISSN 0798 1015.
- SYAM, N., & KAUL, R. (2021). *Machine learning and artificial intelligence in marketing and sales: Essential reference for practitioners and data scientists*. Emerald Publishing Limited. ISBN: 9781800438811. <https://www.proquest.com/legacydocview/EBC/6503991?accountid=37408>.
- THE BIG DATA BOOST (2021). Using Proprietary Big Data to Boost Competitive Advantage. *Strategic Direction*, vol. 37, no. 5, pp. 20-21 ProQuest Central. ISSN 02580543. DOI <https://doi.org/10.1108/SD-03-2021-0027>.
- THOMSETT, M. C. 2017. Stock Market Math: Essential Formulas for Selecting and Managing Stock and Risk, Walter de Gruyter GmbH, Boston. ISBN: 9781501515811- 9781501507427. Available from: ProQuest Ebook Central, <https://www.proquest.com/legacydocview/EBC/5159135?accountid=37408>.
- VAN LEEUWEN, Quint. Master Thesis: Data science and Marketing Analytics. 2020. Disponible en: <https://thesis.eur.nl/>
- VANEGAS, D.A., TARAZONA-BERMÚDEZ, G.M., RODRÍGUEZ-ROJAS, L.A., 2020. Mejora de la toma de decisiones en ciclo de ventas del subsistema comercial de servicios en una empresa de IT. *Revista Científica*, no. 38, pp. 174-183. ISSN: 2344-2350. <https://doi.org/10.14483/23448350.15241>
- VALDIVIEZO ABAD, Cesibel. 2021. Uso de big data y data mining en los procesos de automatización de la comunicación de las organizaciones. GIGAPP Estudios Working Papers. Vol.8. Núm. 199, págs. 128-142. ISSN: 2174-9515.
- VERHEYEN, Gunther, 2021. Scrum: A Pocket Guide: A Smart Travel Companion. 3° ed. Netherlands: Van Haren Publishing, ISBN: 978940180734 0

- WANG, X., ZHANG, L. and HE, T., 2022. Learning Performance Prediction-Based Personalized Feedback in Online Learning Via Machine Learning. *Sustainability*, vol. 14, no. 13, pp. 7654 Coronavirus Research Database; ProQuest Central. DOI <https://doi.org/10.3390/su14137654>.
- WANG, Z., et al, 2022. An Empirical Study on Bugs in Python Interpreters. *IEEE Transactions on Reliability*, vol. 71, no. 2, pp. 716-734 ProQuest Central. ISSN 00189529. DOI <https://doi.org/10.1109/TR.2022.3159812>.
- WENXIONG, L., BI, Z., XIUWEN, Y. and WEI, P., 2021. An Improved Aspect-Category Sentiment Analysis Model for Text Sentiment Analysis Based on RoBERTa. *Applied Intelligence*, 06, vol. 51, no. 6, pp. 3522-3533 ProQuest Central. ISSN 0924669X. DOI <https://doi.org/10.1007/s10489-020-01964-1>.
- XU, S., 2017. A VaR Assuming Student Distribution Not significantly Different from a VaR Assuming Normal Distribution. *Risk Management*, 08, vol. 19, no. 3, pp. 189-201 ProQuest Central. ISSN 14603799. DOI <https://doi.org/10.1057/s41283-017-0017-9>.
- ZHENG, J. and KHALID, H., 2022. The Adoption of Enterprise Resource Planning and Business Intelligence Systems in Small and Medium Enterprises: A Conceptual Framework. *Mathematical Problems in Engineering*, vol. 2022 Coronavirus Research Database; ProQuest Central. ISSN 1024123X. DOI <https://doi.org/10.1155/2022/1829347>.

ANEXOS

Anexo 1: Ficha de registro de ventas antes de la implementación de web scraping y análisis de sentimientos.

Tasa de crecimiento de ventas: Este indicador nos muestra en cuando varían las ventas por periodo transcurrido durante el estudio, lo cual nos proporciona la tasa de crecimiento de ventas de los 30 últimos días.

Unidad de medida: unidades

Fórmula: Tasa de crecimiento de ventas = $((\text{Ventas mes Actual} - \text{Ventas mes Anterior}) / \text{Ventas mes Anterior}) * 100$

Tabla 13: Tasa de crecimiento de ventas

| TEST | | | | |
|------------|----------------|--------------|----------------|-----------------------------------|
| Mes Actual | Ventas Totales | Mes Anterior | Ventas Totales | Tasa de crecimiento de ventas (%) |

Anexo 2: Ficha de registro de clientes antes de la implementación de web scraping y análisis de sentimientos.

Tasa de crecimiento de clientes: Este indicador nos muestra en cuando varía la cantidad de clientes por periodo transcurrido durante el estudio, lo cual nos proporciona la tasa de crecimiento de clientes de los 30 últimos días.

Unidad de medida: unidades

Fórmula: Tasa de crecimiento de clientes = $((\# \text{ Clientes Mes Actual} - \# \text{ Clientes Mes Anterior}) / \# \text{ Clientes Mes Anterior}) * 100$

Tabla 14: Tasa de crecimiento de clientes

| TEST | | | | |
|------------|--------------------|--------------|--------------------|-------------------------------------|
| Mes Actual | # Clientes Totales | Mes Anterior | # Clientes Totales | Tasa de crecimiento de clientes (%) |

Anexo 3

Tabla 15: Matriz de consistencia

| PROBLEMA | OBJETIVO | HIPOTESIS | Variable de estudio | Dimensiones | Indicadores | Formula | Metodología |
|--|---|--|---|---------------------|---|--|---|
| General | General | General | | | | | |
| ¿Cuál es el impacto del web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo? Kassel (2019) | Determinar el impacto del web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo. Azwindini (2022) | Determinar web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo. Chang y Kou (2020) | INDEPENDIENTE: web scraping y análisis de sentimientos | - | - | - | |
| Específicos | Específicos | Específicas | | | | | |
| ¿Cómo influye el web scraping y análisis de sentimientos en la valoración del cliente hacia los servicios, para mejorar la tasa de crecimiento de ventas de la empresa turística? Shahbaz (2021) | Determinar la tasa de crecimiento de ventas en la valoración del cliente hacia los servicios, después de la implementación del web scraping y análisis de sentimientos. Hou, Yang y Bai (2018) | HE1: El uso del web scraping y análisis de sentimientos en la valoración del cliente hacia los servicios, mejora la tasa de crecimiento de ventas de la empresa de turismo. Kachniewska (2020) | DEPENDIENTE: valoración del cliente | Volumen de ventas | Tasa de crecimiento de ventas Labroneé (2018) y Thomsett (2017) | $TCV = \frac{VPA - VPE}{VPE} * 100$ VPA: Ventas periodo actual. VPE: Ventas periodo anterior. TCV: Tasa de crecimiento de ventas Labroneé (2018) y Thomsett (2017) | Tipo de Investigación: Aplicada Enfoque: Cuantitativa Tipo de Estudio: Experimental Diseño de estudio: Pre experimental Población: 30 registros de venta(indicador de ventas) 30 registros de venta(indicador de clientes) |
| ¿Cómo el web scraping y análisis de sentimientos impacta en la valoración del cliente hacia los servicios, para el aumento de la tasa de crecimiento de clientes? Shahbaz (2021) | Determinar como el web scraping y análisis de sentimientos mejora la tasa de crecimiento de clientes en la valoración del cliente hacia los servicios de una empresa de turismo. Kalvet et al. (2020) | HE2: El uso web scraping y análisis de sentimientos mejora la tasa de crecimiento de clientes en la valoración del cliente hacia los servicios de una empresa de turismo. Syam y Kaul (202) | | Volumen de clientes | Tasa de crecimiento de clientes Labroneé (2018) y Thomsett (2017) | $TCC = \frac{CPA - CPE}{CPE} * 100$ CPA: Clientes periodo actual. CPE: Clientes periodo anterior. TCC: Tasa de crecimiento de clientes. Labroneé (2018) y Thomsett (2017) | |

Anexo 4

Tabla 16: Matriz de operacionalización de variables

| VARIABLE (puede ser título) | DEFINICIÓN CONCEPTUAL | DEFINICIÓN OPERACIONAL | DIMENSIONES | INDICADORE S | INSTRUMENTO | ESCALA DE MEDICIÓN/ DESCRIPCIÓN |
|--|--|--|--------------------------------|---|-------------------|---------------------------------------|
| Variable independiente: web scraping y análisis de sentimientos | Es el conjunto de herramientas, en su mayoría derivadas de las matemáticas, machine learning, estadísticas, conjunto de datos, programación, etc. Que en conjunto al conocimiento experto del dominio generan un análisis sistematizado y metodológico de datos obtenidos a partir de diferentes fuentes.(Sepúlveda, 2021, p. 79) | Conjunto de metodologías que permite a la empresa turística acceder a las métricas necesarias para la toma de decisiones durante las horas que lo solicite (24x7). Grus, 2019, p.1 | - | - | | - |
| Variable dependiente: Valoración del cliente | La valoración del cliente toma información de los usuarios para la explotación y la planificación y las ventas.(Sordoni, Eransus y Salmerón, 2022, p.34). Asi tambien, que la podemos definir como un reglamento que todos los productos están compitiendo para pasar por las cuatro fases: atraer la atención del cliente, el interés, el deseo y finalmente la venta. (Bowlby, 2022. pp.50) | Proceso importante en la empresa turística, donde se generan las estrategias de captación de clientes y aumento de volumen de ventas. (Clarke, Cisneros y Yoanis, 2018, p.12) | ventas (Bowlby, 2022. pp.50) | Tasa de crecimiento de ventas. Labruneeé (2018) y Thomsett (2017) | Ficha de registro | Razón |
| | | | clientes (Bowlby, 2022. pp.50) | Tasa de crecimiento de clientes. Labruneeé (2018) y Thomsett (2017) | Ficha de registro | Razón |

Anexo 5: Validación de Instrumentos

I. Datos Generales:

- a. Apellidos y nombres: Ivan Petrlik
- b. Nombre del motivo de evaluación: Ficha de Observación (Tasa de crecimiento de ventas).
- c. Título de la investigación: Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo
- d. Autores : Apaza Ramos, Gilberto y Ñamo Alayo, Edinson

II. Semblantes de la validación

Tabla 17: Validación del instrumento de ventas

| Indicadores | Criterios | Deficiente 0 – 20% | Regular 21– 50% | Bueno 51- 70% | Muy Bueno 71- 80% | Excelente 81–100% |
|-----------------|--|-----------------------|-----------------------|---------------------|----------------------------|----------------------|
| CLARIDAD | Está formado con el lenguaje apropiado. | | | | | 91% |
| ORGANIZACIÓN | Es conveniente el avance de la ciencia y tecnología. | | | | 78% | |
| CAPACIDAD | Comprende los semblantes de cantidad y calidad | | | | 78% | |
| INTENCIONALIDAD | Preciso para calificar aspectos del sistema | | | | 78% | |

| | | | | | | |
|-------------|---|--|--|--|-----|--|
| | metodológico y científico | | | | | |
| COHERENCIA | Entre los índices, indicadores y dimensiones | | | | 78% | |
| PERTINENCIA | El instrumento es correcto al tipo de investigación | | | | 78% | |


Valor promedio: 80.2%

Orden de aplicabilidad:

(x) El instrumento está en la condiciones de ser aplicado de la forma como está elaborado.

() El instrumento necesita ser mejorado antes de su aplicación.

Firma



Firma del Experto



Ing. Iván PETRUK A. ZASACHNE
CIP 91445

Validación de Instrumento tasa de crecimiento de ventas

I. Datos Generales:

- a. Apellidos y nombres: **CRISPIN SANCHEZ, IVAN**
- b. Nombre del motivo de evaluación: Ficha de Observación (Tasa de crecimiento de ventas).
- c. Título de la investigación: web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.
- d. Autores : Apaza Ramos, Gilberto y Ñamo Alayo, Edinson

II. Semblantes de la validación

| Indicadores | Criterios | Deficiente 0 – 20% | Regular 21– 50% | Bueno 51- 70% | Muy Bueno 71- 80% | Excelente 81–100% |
|-----------------|---|-----------------------|-----------------------|---------------------|----------------------------|----------------------|
| CLARIDAD | Está formado con el lenguaje apropiado. | | | | | 90% |
| ORGANIZACIÓN | Es conveniente el avance de la ciencia y tecnología. | | | | 75% | |
| CAPACIDAD | Comprende los semblantes de cantidad y calidad | | | | 75% | |
| INTENCIONALIDAD | Preciso para calificar aspectos del sistema metodológico y científico | | | | 75% | |

| | | | | | | |
|-------------|---|--|--|--|-----|--|
| COHERENCIA | Entre los índices, indicadores y dimensiones | | | | 75% | |
| PERTINENCIA | El instrumento es correcto al tipo de investigación | | | | 75% | |

Valor promedio: 77.5%

Orden de aplicabilidad:

(X) El instrumento está en la condiciones de ser aplicado de la forma como está elaborado.

() El instrumento necesita ser mejorado antes de su aplicación.

Firma



Mg. Ivan Crispin Sanchez

CIP.58591

Validación de Instrumento

I. Datos Generales:

- a. Apellidos y nombres: Ivan Petrlik
- b. Nombre del motivo de evaluación: Ficha de Observación (Tasa de crecimiento de clientes).
- c. Título de la investigación: web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.
- d. Autores : Apaza Ramos, Gilberto y Ñamo Alayo, Edinson

II. Semblantes de la validación

| Indicadores | Criterios | Deficiente 0 – 20% | Regular 21– 50% | Bueno 51- 70% | Muy Bueno 71- 80% | Excelente 81–100% |
|-----------------|---|-----------------------|-----------------------|---------------------|----------------------------|----------------------|
| CLARIDAD | Está formado con el lenguaje apropiado. | | | | | 93% |
| ORGANIZACION | Es conveniente el avance de la ciencia y tecnología. | | | | 79% | |
| CAPACIDAD | Comprende los semblantes de cantidad y calidad | | | | 79% | |
| INTENCIONALIDAD | Preciso para calificar aspectos del sistema metodológico y científico | | | | 79% | |

| | | | | | | |
|-------------|---|--|--|--|-----|--|
| COHERENCIA | Entre los índices, indicadores y dimensiones | | | | 79% | |
| PERTINENCIA | El instrumento es correcto al tipo de investigación | | | | 79% | |

Valor promedio: 81.3 %

Orden de aplicabilidad:

(x) El instrumento está en la condiciones de ser aplicado de la forma como está elaborado.

() El instrumento necesita ser mejorado antes de su aplicación.

Firma



Firma del Experto



Ing. Ivan PETRLÍK APASACHE
CIP 91445

Validación de Instrumento tasa de crecimiento de clientes

I. Datos Generales:

- a. Apellidos y nombres: **CRISPIN SANCHEZ, IVAN**
- b. Nombre del motivo de evaluación: Ficha de Observación (Tasa de crecimiento de clientes).
- c. Título de la investigación: web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo. Autores : Apaza Ramos, Gilberto y Ñamo Alayo, Edinson

II. Semblantes de la validación

Tabla 18: Validación de instrumento de cliente

| Indicadores | Criterios | Deficiente 0 – 20% | Regular 21– 50% | Bueno 51- 70% | Muy Bueno 71- 80% | Excelente 81–100% |
|-----------------|---|-----------------------|-----------------------|---------------------|----------------------------|----------------------|
| CLARIDAD | Está formado con el lenguaje apropiado. | | | | | 90% |
| ORGANIZACIÓN | Es conveniente el avance de la ciencia y tecnología. | | | | 77% | |
| CAPACIDAD | Comprende los semblantes de cantidad y calidad | | | | 77% | |
| INTENCIONALIDAD | Preciso para calificar aspectos del sistema metodológico y científico | | | | 77% | |

| | | | | | | |
|-------------|---|--|--|--|-----|--|
| COHERENCIA | Entre los índices, indicadores y dimensiones | | | | 77% | |
| PERTINENCIA | El instrumento es correcto al tipo de investigación | | | | 77% | |

Valor promedio: 79.2 %

Orden de aplicabilidad:

(X) El instrumento está en la condiciones de ser aplicado de la forma como está elaborado.

() El instrumento necesita ser mejorado antes de su aplicación.

Firma



Mg. Ivan Crispin Sanchez

CIP.58591

Validación de Instrumento

I. Datos Generales:

- a. Apellidos y nombres: Rivera Crisostomo, Renee
- b. Nombre del motivo de evaluación: Ficha de Observación (Tasa de crecimiento de ventas).
- c. Título de la investigación: web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.
- d. Autores : Apaza Ramos, Gilberto y Ñamo Alayo, Edinson

II. Semblantes de la validación

| Indicadores | Criterios | Deficiente 0 – 20% | Regular 21– 50% | Bueno 51- 70% | Muy Bueno 71- 80% | Excelente 81–100% |
|-----------------|--|-----------------------|-----------------------|---------------------|----------------------------|----------------------|
| CLARIDAD | Está formado con el lenguaje apropiado. | | | 70% | | |
| ORGANIZACIÓN | Es conveniente el avance de la ciencia y tecnología. | | | 70% | | |
| CAPACIDAD | Comprende los semblantes de cantidad y calidad | | | 70% | | |
| INTENCIONALIDAD | Preciso para calificar aspectos del sistema | | | | 79% | |

| | | | | | | |
|-------------|---|--|--|-----|--|--|
| | metodológico y científico | | | | | |
| COHERENCIA | Entre los índices, indicadores y dimensiones | | | 70% | | |
| PERTINENCIA | El instrumento es correcto al tipo de investigación | | | 70% | | |

Valor promedio: 71.5%

Orden de aplicabilidad:

(x) El instrumento está en la condiciones de ser aplicado de la forma como está elaborado.

() El instrumento necesita ser mejorado antes de su aplicación.



Firma

RIVERA CRISOSTOMO, RENEE

LE/DNI 08554321

Validación de Instrumento tasa de crecimiento de clientes

I. Datos Generales:

- a. Apellidos y nombres: Rivera Crisostomo, Renee
- b. Nombre del motivo de evaluación: Ficha de Observación (Tasa de crecimiento de clientes).
- c. Título de la investigación: web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.
- d. Autores : Apaza Ramos, Gilberto y Ñamo Alayo, Edinson

II. Semblantes de la validación

| Indicadores | Criterios | Deficiente 0 – 20% | Regular 21– 50% | Bueno 51- 70% | Muy Bueno 71- 80% | Excelente 81–100% |
|--------------|--|-----------------------|-----------------------|---------------------|----------------------------|----------------------|
| CLARIDAD | Está formado con el lenguaje apropiado. | | | 70% | | |
| ORGANIZACIÓN | Es conveniente el avance de la ciencia y tecnología. | | | | 75% | |
| CAPACIDAD | Comprende los semblantes | | | 70% | | |

| | | | | | | |
|-----------------|---|--|--|-----|-----|--|
| | de cantidad y calidad | | | | | |
| INTENCIONALIDAD | Preciso para calificar aspectos del sistema metodológico y científico | | | | 78% | |
| COHERENCIA | Entre los índices, indicadores y dimensiones | | | 70% | | |
| PERTINENCIA | El instrumento es correcto al tipo de investigación | | | | 74% | |

Valor promedio: 72.83%

Orden de aplicabilidad:

(x) El instrumento está en las condiciones de ser aplicado de la forma como está elaborado.

() El instrumento necesita ser mejorado antes de su aplicación.



Firma

RIVERA CRISOSTOMO, RENEE

LE/DNI08554321

Anexo 6: Validación del marco de trabajo de desarrollo de software

TABLA DE EVALUACION DE EXPERTOS

(Marcos de trabajo de desarrollo de software para web scraping y análisis de sentimientos)

Nombres y Apellidos del experto: Albert Ronaldo Bernachea Ramos

Título y/o Grado: Ing. Sistemas e Informática

Fecha: 25/05/2022

Investigación

Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.

Mediante la tabla de evaluación de expertos, usted tiene la potestad de calificar los marcos de trabajos involucrados, mediante un conjunto de preguntas con puntuación indicadas al final de la tabla.

Asimismo, se invita a la corrección de los elementos señalando sus observaciones y/o sugerencias, con la finalidad de mejorar la coherencia de las preguntas.

Tabla 19: Evaluación de Marco de trabajo

| ITEM | PREGUNTAS | MARCOS DE TRABAJO | | | OBSERVACIONES |
|------|--|-------------------|-------|----------|---------------|
| | | KDD | SCRUM | CRISP-DM | |
| 1 | Marco de trabajo para proyectos a largo plazo | 3 | 5 | 4 | |
| 2 | Marco de trabajo que utiliza flujos de trabajo o disciplinas | 4 | 4 | 3 | |
| 3 | Marco de trabajo que cada una de sus fases termina con un hito | 4 | 5 | 3 | |

| | | | | | |
|---|---|---|---|---|--|
| 4 | Marco de trabajo centrada en la arquitectura | 5 | 4 | 4 | |
| 5 | Es un Marco de trabajo interactivo e incremental | 4 | 4 | 4 | |
| 6 | Transforma los requisitos de los usuarios en un sistema de software | 4 | 5 | 4 | |
| 7 | Sus fases están relacionadas con asuntos del negocio | 4 | 5 | 3 | |

Utilizar la siguiente valoración:

1. Malo, 2. Regular, 3. Bueno, 4. Muy Bueno, 5. Excelente.

Firma del experto



Validación del marco de trabajo de desarrollo de software

TABLA DE EVALUACION DE EXPERTOS

(Marcos de trabajo de desarrollo de software para web scraping y análisis de sentimientos)

Nombres y Apellidos del experto: David Regalado Gutiérrez

Título y/o Grado: Ingeniero de Sistemas - Co-Founder at Data Engineering LATAM

LinkedIn: <https://www.linkedin.com/in/davidregaladodataengineer/?originalSubdomain=pe>

Fecha: 30-04-2022

Investigación

Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.

Mediante la tabla de evaluación de expertos, usted tiene la potestad de calificar los marcos de trabajos involucrados, mediante un conjunto de preguntas con puntuación indicadas al final de la tabla.

Asimismo, se invita a la corrección de los elementos señalando sus observaciones y/o sugerencias, con la finalidad de mejorar la coherencia de las preguntas.

| ITEM | PREGUNTAS | MARCOS DE TRABAJO | | | |
|------|--|-------------------|-------|----------|---------------|
| | | KDD | SCRUM | CRISP-DM | OBSERVACIONES |
| 1 | Marco de trabajo para proyectos a largo plazo | 5 | 3 | 1 | |
| 2 | Marco de trabajo que utiliza flujos de trabajo o disciplinas | 1 | 5 | 3 | |
| 3 | Marco de trabajo que cada una de | 5 | 3 | 1 | |

| | | | | | |
|---|---|---|---|---|--|
| | sus fases termina con un hito | | | | |
| 4 | Marco de trabajo centrada en la arquitectura | 5 | 3 | 1 | |
| 5 | Es un Marco de trabajo interactivo e incremental | 1 | 5 | 3 | |
| 6 | Transforma los requisitos de los usuarios en un sistema de software | 5 | 3 | 1 | |
| 7 | Sus fases están relacionadas con asuntos del negocio | 3 | 5 | 1 | |

Utilizar la siguiente valoración:

1. Malo, 2. Regular, 3. Bueno, 4. Muy Bueno, 5. Excelente.

Firma del experto



Validación del marco de trabajo de desarrollo de software

TABLA DE EVALUACION DE EXPERTOS

(Marcos de trabajo de desarrollo de software para web scraping y análisis de sentimientos)

Nombres y Apellidos del experto: Erika Isabel, Salazar Avila

Título y/o Grado: MBA tecnologías de Información de sistemas

Fecha: 25/05/2022

Investigación

Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.

Mediante la tabla de evaluación de expertos, usted tiene la potestad de calificar los marcos de trabajos involucrados, mediante un conjunto de preguntas con puntuación indicadas al final de la tabla.

Asimismo, se invita a la corrección de los elementos señalando sus observaciones y/o sugerencias, con la finalidad de mejorar la coherencia de las preguntas.

| ITEM | PREGUNTAS | MARCOS DE TRABAJO | | | |
|------|--|-------------------|-------|----------|---------------|
| | | KDD | SCRUM | CRISP-DM | OBSERVACIONES |
| 1 | Marco de trabajo para proyectos a largo plazo | 1 | 4 | 5 | |
| 2 | Marco de trabajo que utiliza flujos de trabajo o disciplinas | 3 | 5 | 3 | |
| 3 | Marco de trabajo que cada una de sus fases termina con un hito | 2 | 4 | 2 | |
| 4 | Marco de trabajo centrada en la arquitectura | 4 | 3 | 4 | |
| 5 | Es un Marco de trabajo interactivo e incremental | 3 | 5 | 3 | |

| | | | | | |
|---|---|---|---|---|--|
| 6 | Transforma los requisitos de los usuarios en un sistema de software | 2 | 5 | 3 | |
| 7 | Sus fases están relacionadas con asuntos del negocio | 4 | 5 | 2 | |

Utilizar la siguiente valoración:

1. Malo, 2. Regular, 3. Bueno, 4. Muy Bueno, 5. Excelente.

Firma del experto



MBA. Salazar Avila, Erika

Anexo 7

Tabla 20: Análisis estadístico de SPSS

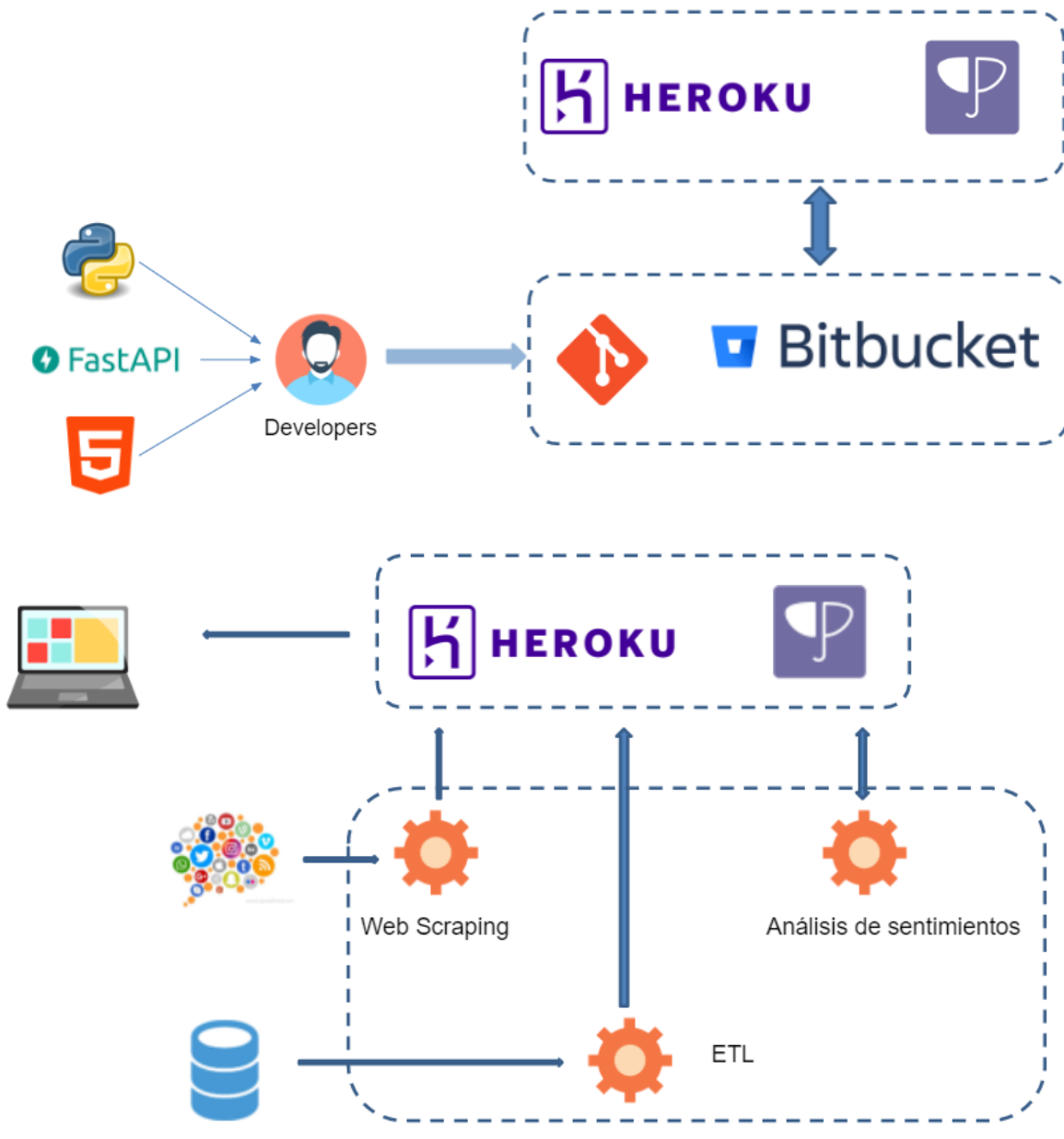
| | | | Estadístico | Error típ. |
|--------------|---|-------------------------|-----------------|-----------------|
| ClienteAbril | Media | | 29,6400 | 3,22174 |
| | Intervalo de confianza para la media al 95% | Límite inferior | 22,9907 | |
| | | Límite superior | 36,2893 | |
| | Media recortada al 5% | | 28,9667 | |
| | Mediana | | 32,0000 | |
| | Varianza | | 259,490 | |
| | Desv. típ. | | 16,10869 | |
| | Mínimo | | 3,00 | |
| | Máximo | | 71,00 | |
| | Rango | | 68,00 | |
| | Amplitud intercuartil | | 20,50 | |
| | Asimetría | | ,448 | ,464 |
| | Curtosis | | ,459 | ,902 |
| ClienteMayo | Media | | 35,0400 | 3,74899 |
| | Intervalo de confianza para la media al 95% | Límite inferior | 27,3025 | |
| | | Límite superior | 42,7775 | |
| | Media recortada al 5% | | 34,2556 | |
| | Mediana | | 32,0000 | |
| | Varianza | | 351,373 | |
| | Desv. típ. | | 18,74495 | |
| | Mínimo | | 9,00 | |
| | Máximo | | 76,00 | |
| | Rango | | 67,00 | |
| | Amplitud intercuartil | | 26,00 | |
| | Asimetría | | ,598 | ,464 |
| | Curtosis | | -,221 | ,902 |
| VentasAbril | Media | | 807531012,7200 | 380521992,88167 |
| | Intervalo de confianza para la media al 95% | Límite inferior | 22172218,9438 | |
| | | Límite superior | 1592889806,4962 | |
| | Media recortada al 5% | | 460472495,1222 | |
| | Mediana | | 58168002,0000 | |
| | Varianza | 3619924676665847800,000 | | |
| | Desv. típ. | 1902609964,40833 | | |
| | Mínimo | | 1046,00 | |
| | Máximo | | 9,20E+009 | |
| | Rango | | 9204061840,00 | |
| | Amplitud intercuartil | | 923956668,50 | |
| | Asimetría | | 3,902 | ,464 |
| | Curtosis | | 16,971 | ,902 |
| Media | | 765436043,0800 | 192671983,67005 | |
| VentasMayo | Intervalo de confianza para la media al 95% | Límite inferior | 367780613,1174 | |
| | | Límite superior | 1163091473,0426 | |
| | Media recortada al 5% | | 683737885,6222 | |
| | Mediana | | 328777964,0000 | |
| | Varianza | 928062332283833730,000 | | |
| | Desv. típ. | 963359918,35027 | | |
| | Mínimo | | 855,00 | |
| | Máximo | | 3,15E+009 | |
| | Rango | | 3145648773,00 | |
| | Amplitud intercuartil | | 1699268545,50 | |
| | Asimetría | | 1,003 | ,464 |
| | Curtosis | | -,163 | ,902 |

Anexo 8

Tabla 21: Prueba de Normalidad

| | Pruebas de normalidad | | | | | |
|--------------|---------------------------------|----|-------|--------------|----|------|
| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
| | Estadístico | gl | Sig. | Estadístico | gl | Sig. |
| ClienteAbril | ,106 | 25 | ,200* | ,970 | 25 | ,639 |
| ClienteMayo | ,106 | 25 | ,200* | ,950 | 25 | ,245 |
| VentasAbril | ,336 | 25 | ,000 | ,471 | 25 | ,000 |
| VentasMayo | ,266 | 25 | ,000 | ,795 | 25 | ,000 |

Anexo 9: Arquitectura de desarrollo



Anexo 10:

Tabla 22: Correlación de Pearson Pre-Test

Estadísticos descriptivos

| | Media | Desviación típica | N |
|--------------|--------------------|----------------------|----|
| ClienteAbril | 29,6400 | 16,10869 | 25 |
| VentasAbril | 807531012,720 0 | 1902609964,40 833 | 25 |

Correlaciones de Abril

| | | CME | VME |
|-----|------------------------|--------|--------|
| | Correlación de Pearson | 1 | ,711** |
| CME | Sig. (bilateral) | | ,000 |
| | N | 25 | 25 |
| | Correlación de Pearson | ,711** | 1 |
| VME | Sig. (bilateral) | ,000 | |
| | N | 25 | 25 |

** . La correlación es significativa al nivel 0,01 (bilateral).

Estadísticos descriptivos

| | Media | Desviación típica | N |
|-------------|--------------------|----------------------|----|
| ClienteMayo | 35,0400 | 18,74495 | 25 |
| VentasMayo | 765436043,080 0 | 963359918,350 27 | 25 |

Correlaciones de Mayo

| | | CMA | VMA |
|-----|------------------------|--------|--------|
| | Correlación de Pearson | 1 | ,822** |
| CMA | Sig. (bilateral) | | ,000 |
| | N | 25 | 25 |
| | Correlación de Pearson | ,822** | 1 |
| VMA | Sig. (bilateral) | ,000 | |
| | N | 25 | 25 |

** . La correlación es significativa al nivel 0,01 (bilateral).

Anexo 11

Tabla 23: Pre-test de Clientes

| FICHA DE REGISTRO | | | | |
|------------------------------|---|-------------|--|-------|
| Investigadores | Apaza Ramos, Gilberto; Ñamo Alayo, Edinson | | Tipo de prueba: Pre-Test | |
| Empresa de estudio | ATIPAX GROUP | | | |
| Ubicación | Av. Arequipa 3750 San Isidro I27 Lima, Perú | | | |
| Proceso | Volumen de clientes | | | |
| Fecha de inicio | 15/06/2022 | Fecha final | 28/06/2022 | |
| Indicador | Descripción | Medida | Técnica | |
| Tasa crecimiento de clientes | Tasa crecimiento de clientes | unidad | Observación experimental | |
| Fórmula | $TCC = \frac{CMA - CME}{CME} * 100$ | | CMA: Clientes mes actual. CME: Clientes mes anterior. TCC: Tasa de crecimiento de clientes | |
| Ítem | Fechas-Abril/Mayo | CMA | CME | TCC |
| 1 | 02-Abril-Mayo | 76 | 7 | 986% |
| 2 | 03-Abril-Mayo | 23 | 32 | -28% |
| 3 | 04-Abril-Mayo | 71 | 36 | 97% |
| 4 | 05-Abril-Mayo | 44 | 19 | 132% |
| 5 | 06-Abril-Mayo | 55 | 27 | 104% |
| 6 | 07-Abril-Mayo | 10 | 34 | -71% |
| 7 | 09-Abril-Mayo | 18 | 7 | 157% |
| 8 | 10-Abril-Mayo | 26 | 15 | 73% |
| 9 | 11-Abril-Mayo | 29 | 42 | -31% |
| 10 | 12-Abril-Mayo | 21 | 54 | -61% |
| 11 | 13-Abril-Mayo | 47 | 27 | 74% |
| 12 | 14-Abril-Mayo | 9 | 45 | -80% |
| 13 | 16-Abril-Mayo | 37 | 21 | 76% |
| 14 | 17-Abril-Mayo | 32 | 29 | 10% |
| 15 | 18-Abril-Mayo | 49 | 36 | 36% |
| 16 | 19-Abril-Mayo | 49 | 49 | 0% |
| 17 | 20-Abril-Mayo | 66 | 43 | 53% |
| 18 | 21-Abril-Mayo | 12 | 71 | -83% |
| 19 | 23-Abril-Mayo | 40 | 3 | 1233% |

| | | | | |
|----|---------------|-----|-----|------|
| 20 | 24-Abril-Mayo | 35 | 8 | 100% |
| 21 | 25-Abril-Mayo | 38 | 32 | 19% |
| 22 | 26-Abril-Mayo | 26 | 33 | -21% |
| 23 | 27-Abril-Mayo | 25 | 20 | 25% |
| 24 | 28-Abril-Mayo | 10 | 33 | -70% |
| 25 | 30-Abril-Mayo | 28 | 8 | 250% |
| | TOTAL | 876 | 741 | 18% |

Anexo 12

Tabla 24: Pre-test de Ventas

| FICHA DE REGISTRO | | | | |
|------------------------------|---|-------------|--|-------|
| Investigadores | Apaza Ramos, Gilberto; Namo Alayo, Edinson | | Tipo de prueba: Pre-Test | |
| Empresa de estudio | ATIPAX GROUP | | | |
| Ubicación | Av. Arequipa 3750 San Isidro I27 Lima, Perú | | | |
| Proceso | Volumen de ventas | | | |
| Fecha de inicio | 15/06/2022 | Fecha final | 28/06/2022 | |
| Indicador | Descripción | Medida | Técnica | |
| Tasa crecimiento de clientes | Tasa crecimiento de ventas | unidad | Observación experimental | |
| Fórmula | $TCV = \frac{VMA - VME}{VME} * 100$ | | VMA: Ventas mes actual. VME: Ventas mes anterior. TCV: Tasa de crecimiento de ventas | |
| Ítem | Fechas-Abril/Mayo | VMA | VME | TCV |
| 1 | 02-Abril-Mayo | 17355.0778 | 1372.47 | 1165% |
| 2 | 03-Abril-Mayo | 3544 | 19762.88 | -82% |
| 3 | 04-Abril-Mayo | 24245.9178 | 12445.2392 | 95% |
| 4 | 05-Abril-Mayo | 20709.239 | 8818.32 | 135% |
| 5 | 06-Abril-Mayo | 31456.4963 | 5816.8002 | 441% |
| 6 | 07-Abril-Mayo | 2727.66 | 10324.4 | -74% |
| 7 | 09-Abril-Mayo | 6999.19996 | 1461 | 379% |
| 8 | 10-Abril-Mayo | 4043.19988 | 4929.60994 | -18% |
| 9 | 11-Abril-Mayo | 5214.6 | 23583.2297 | -78% |
| 10 | 12-Abril-Mayo | 5517.55992 | 20651.56 | -73% |

| | | | | |
|----|---------------|------------|------------|-------|
| 11 | 13-Abril-Mayo | 16631.2993 | 8338 | 99% |
| 12 | 14-Abril-Mayo | 2500 | 24943.1952 | -90% |
| 13 | 16-Abril-Mayo | 17702.5091 | 5089.6 | 248% |
| 14 | 17-Abril-Mayo | 7949 | 13434.199 | -41% |
| 15 | 18-Abril-Mayo | 20799.819 | 6231.1596 | 234% |
| 16 | 19-Abril-Mayo | 32877.7964 | 6036.29904 | 445% |
| 17 | 20-Abril-Mayo | 20901.53 | 15715.8771 | 33% |
| 18 | 21-Abril-Mayo | 892 | 92040.6289 | -99% |
| 19 | 23-Abril-Mayo | 15115.7 | 1046 | 1345% |
| 20 | 24-Abril-Mayo | 11441.12 | 5927.46 | 93% |
| 21 | 25-Abril-Mayo | 15478.22 | 10808.92 | 43% |
| 22 | 26-Abril-Mayo | 4847.3 | 11518.2862 | -58% |
| 23 | 27-Abril-Mayo | 11988.53 | 19180.9388 | -37% |
| 24 | 28-Abril-Mayo | 85.5 | 8612 | -99% |
| 25 | 30-Abril-Mayo | 10569.1593 | 4410.67902 | 140% |
| | TOTAL | 311592.434 | 342498.752 | -9% |

Anexo 13: Prototipo

Login

Username

Password

Register?

Tus búsquedas

Lista de tus búsquedas

Información de las ultimas búsquedas realizadas

| # | Tipo | Fec.Busqueda | Filtro |
|---|---------|--------------|----------|
| 1 | Google | 22-10-07 | Medellín |
| 2 | Twitter | 22-11-08 | Lima |

Realizar nueva búsqueda

Fuente: Criterio:

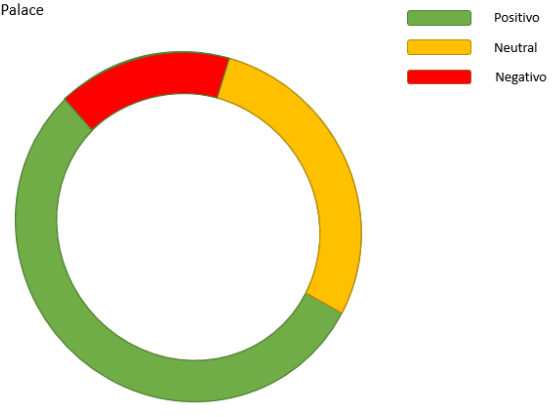
Realizar nueva búsqueda

Fuente: Criterio:

Cantidad:

Búsqueda y Extracción Búsqueda

Análisis de búsqueda : Palace



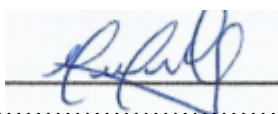
Anexo 14: Declaratoria de autenticidad del (de los) autor(es)

DECLARATORIA DE AUTENTICIDAD DE LOS AUTORES

Nosotros, Apaza Ramos, Gilberto Romer y Ñamo Alayo, Edinson Manuel, alumnos de la Facultad de Ingeniería y Arquitectura; Escuela Profesional de Ingeniería de Sistemas de la Universidad César Vallejo Lima-Este declaramos bajo juramento que todos los datos e información que acompañan al Trabajo de Tesis titulado “web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.” son:

1. De nuestra autoría
2. La presente Tesis no ha sido plagiado ni total, ni parcialmente.
3. La Tesis no ha sido publicado ni presentado anteriormente.
4. Los resultados presentados en la presente Tesis son reales, no han sido falseados, ni duplicados, ni copiados.

Lima, 13 de Setiembre de 2022



.....
Ñamo Alayo Edinson Manuel
DNI: 46122297



.....
Apaza Ramos Gilberto Romer
DNI: 73010514

Anexo 15: Compromiso de Confidencialidad

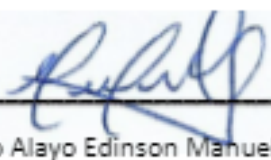
COMPROMISO DE CONFIDENCIALIDAD

Conste por el presente Compromiso de Confidencialidad que suscribe **ÑAMO ALAYO EDINSON MANUEL**, identificado(a) con DNI N° **46122297**, con domicilio en **MZ Z2 LT 13 AMPLIACIÓN BAYOVAR, SAN JUAN DE LURIGANCHO** y **APAZA RAMOS GILBERTO ROMER**, identificado(a) con DNI N° **73010514**, con domicilio en **MZ D LT 4 AA.HH CERRITO FELIZ, SAN JUAN DE LURIGANCHO**, en adelante **LOS INVESTIGADORES**, lo siguiente:

PRIMERO: **LOS INVESTIGADORES** se obliga a guardar absoluta confidencialidad respecto a toda información, data o documentación que no sea de acceso público, a la que tenga acceso directa o indirectamente como consecuencia de su participación en la **TESIS “web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios en la empresa turística ATIPAX GROUP.”** para optar el grado de Ingeniero en Sistemas en la **Universidad Privada César Vallejo**. De la misma manera, se comprometen a no reproducir, transformar, distribuir, ni comunicar a terceros, la información, data, ni documentos e instrumentos que utilice y que se generen para efectos de la evaluación. Asimismo, los investigadores se comprometen a remitir las bases de datos con los resultados de los análisis que se obtengan a la **Universidad Privada César Vallejo**, para posibles nuevos estudios o investigaciones. En caso los investigadores consideren realizar actividades para la difusión de los resultados de evaluación (publicación académica, resumen de política educativa, presentaciones, etc.) deberá coordinar previamente con la empresa **ATIPAX GROUP**, remitiendo además una copia del documento a ser difundido a la **Universidad Privada César Vallejo**. Para efectos del presente compromiso, la información confidencial puede incluir, pero no de manera limitativa: bases de datos, información comercial, información financiera, propiedad intelectual, secretos comerciales, productos, procesos, técnicas, conocimiento científico, materiales, secuencias, inventos, máquinas, datos, fórmulas, muestras, modelos, sistemas, redes, planes de negocio, requerimiento de clientes, software, códigos, diseños, dibujos, esquemas, esbozos, fotografías, outputs digitales, especificaciones, documentos, reportes, y/o estudios generados por la parte divulgadora.]

SEGUNDO: Los Investigadores se comprometen a impedir que los documentos y data con carácter de confidencialidad que son entregados por **ATIPAX GROUP** sean observados, reproducidos o manipulados por personas no autorizadas por la **Universidad Privada César Vallejo**.

TERCERO: Los Investigadores se comprometen a cumplir con lo de arriba detallado y luego de concluir la evaluación.



Ñamo Alayo Edinson Manuel

46122297



Apaza Ramos Gilberto Romer

73010514

Anexo 16: Solicitud para realizar la Tesis en la empresa

SEÑORA JAISSY VELIT CARNERO
GERENTE GENERAL DE ATIPAX GROUP

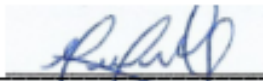
Nosotros, **NAMO ALAYO EDINSON**, identificada con DNI N. ° 46122297, con domicilio MZ z2 LT 13 Bayóvar Ampliación, San Juan de Lurigancho y **APAZA RAMOS GILBERTO ROMER** identificado con DNI N° 73010514, con domicilio MZ D LT 4 AA.HH Cerrito Feliz, San Juan de Lurigancho. Ante Ud. respetuosamente nos presentamos y exponemos:

Solicitamos a Ud. permiso para realizar trabajo de Investigación en su Institución sobre **“web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios en la empresa turística Atipax.”** para optar el grado de Ingeniero en Sistemas.

POR LO EXPUESTO:

Ruego a usted acceder a nuestra solicitud.

Lima, 18 de Junio del 2022



NAMO ALAYO EDINSON
DNI N. ° 46122297



APAZA RAMOS GILBERTO
DNI N. ° 73010514

Anexo 17: Matriz de verificación de originalidad

Tabla 25: Matriz de verificación de originalidad

| Título: Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo. | | |
|--|--|--|
| Matriz de verificación de originalidad de una propuesta de investigación o innovación | | |
| Referencia de la solución tecnológica | Aspectos funcionales, técnicos, metodológicos, algoritmo o estadísticos de la solución tecnológica de la referencia | Aspectos a incluir en la nueva solución propuesta |
| VERHEYEN, Gunther, 2021. Scrum: A Pocket Guide: A Smart Travel Companion. 3° ed. Netherlands: Van Haren Publishing, ISBN: 978940180734 0 | Marco de trabajo Scrum | x |
| | Gestión de proyectos | |
| SAURA.J. REYES-MENENDEZ, A. and PALOS-SANCHEZ, P., 2019. Are Black Friday Deals Worth it? Mining Twitter Users' Sentiment and Behavior Response. Journal of Open Innovation: Technology, Market, and Complexity, vol. 5, no. 3, pp. 58 ProQuest Central.e-ISSN21998531 DOI https://doi.org/10.3390/joitmc5030058. | Metodología CRISP-DM | x |
| | Análisis de sentimientos | x |
| | Redes sociales Twitter | x |
| WANG, X., ZHANG, L. and HE, T., 2022. Learning Performance Prediction-Based Personalized Feedback in Online Learning Via Machine Learning. <i>Sustainability</i> , vol. 14, no. 13, pp. 7654 Coronavirus Research Database; ProQuest Central. DOI https://doi.org/10.3390/su14137654. | Machine Learning | |
| | Predicción a base de datos | |
| | Extracción de información | x |
| BROUCKE, Seppe vanden, BAESENS, Bart (2018). Practical Web Scraping for Data Science. Apress Media LLC: Welmoed Spahr.ISBN: 978-1-4842-3581-2. DOI: https://doi.org/10.1007/978-1-4842-3582-9 | Web Scraping | x |
| | Data Science | |
| DAMARTA, R., HIDAYAT, A. and ABDULLAH, A.S., 2021. The Application of 0RW1S34RfeSDcfkexd09rT2k1RW1S34RfeSDcfkexd09rT2-Nearest Neighbors Classifier for Sentiment Analysis of PT PLN (Persero) Twitter Account Service Quality. Journal of Physics: Conference Series, 01, vol. 1722, no. 1 ProQuest Central. ISSN 17426588. DOI https://doi.org/10.1088/1742-6596/1722/1/012002. | Python | x |
| | algoritmo k-Nearest | |
| SKEEN, S.J., JONES, S.S., CRUSE, C.M. and HORVATH, K.J., 2022. Integrating Natural Language Processing and Interpretive Thematic Analyses to Gain Human-Centered Design Insights on HIV Mobile Health: Proof-of-Concept Analysis. JMIR Human Factors, 07, vol. 9, no. 3 ProQuest Central. e-ISSN: 22929495. DOI https://doi.org/10.2196/37350. | | |
| | Algoritmo VADER | x |

Anexo 18: Consentimiento informado

CONSENTIMIENTO INFORMADO

Yo Jaissy Velit Carnero identificado(a) con DNI (Carné de extranjería o pasaporte) N°40702728 he sido informado(a) sobre el procedimiento de la investigación titulada **"Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo"**, cuyos autores con **ÑAMO ALAYO EDINSON MANUEL**, identificado(a) con DNI N° 46122297, con domicilio en **MZ Z2 LT 13 AMPLIACIÓN BAYOVAR, SAN JUAN DE LURIGANCHO** y **APAZA RAMOS GILBERTO ROMER**, identificado(a) con DNI N° 73010514, con domicilio en **MZ D LT 4 AA.HH CERRITO FELIZ-SAN JUAN DE LURIGANCHO**, y se me ha entregado una copia de este consentimiento informado, fechado y firmado.

Además, se me ha explicado las características y el objetivo del estudio, así como los posibles beneficios de este. He contado con el tiempo y la oportunidad para realizar preguntas y plantear las dudas que poseía. Todas las preguntas fueron respondidas a mi entera satisfacción.

Se me ha asegurado que se mantendrá la confidencialidad de mis datos. Mi consentimiento lo otorgó de manera voluntaria y sé que soy libre de retirarme del estudio en cualquier momento, por cualquier razón de fuerza mayor. Por lo tanto, en forma consciente y voluntaria doy mi consentimiento para ser parte de esta investigación.

Lima, 03 de octubre del 2022



Jaissy Velit Carnero
Gerente General Atipax Group

Anexo 19: Aspectos Administrativos

Recursos y Presupuesto

Para García y López (2014) un recurso se define como un conglomerado de objetos disponibles para solucionar o cubrir una carencia o sacar adelante una empresa.

Recursos Humanos

- 01 asesor Metodológico
- 02 investigadores
- Gerente General de la empresa Atipax Group.
- 01 administrador de la empresa Atipax Group.
- 01 vendedor de la empresa Atipax Group.

Tabla 26: Presupuesto de proyecto

| CLASIFICADOR DE GASTOS | DESCRIPCIÓN | CANTIDAD | COSTO S/. |
|------------------------------------|--|-------------|-----------|
| 2. | GASTOS PRESUPUESTARIOS | | |
| 2.3. | BIENES Y SERVICIOS | | |
| 2.3.1. | Equipos y Bienes Duraderos | | |
| | Servidor de Aplicaciones | | 7999.00 |
| | Laptop | 01 unidad | 1649.00 |
| | Impresora | 01 unidad | 859.00 |
| Equipos de la empresa ya obtenidos | | | 10,507 |
| 2.3.2 | Materiales y útiles | | |
| 2.3.2.1 | De oficina | | |
| 2.3.2.1.1 | Papelería en General, útiles y Materiales de Oficina | | |
| | Papel bond A4 | 02 Millares | 24.00 |
| | Lapiceros | 1/2 Docena | 6.00 |
| | Engrapador | 01 Unidad | 20.00 |
| | Grapas | 01 caja | 4.00 |
| | Folder manila | 5 unidades | 10.50 |

| | | | |
|---------------------------|-----------------------------------|------------|-----------------|
| | Perforadora | 01 Unidad | 25.00 |
| | Resaltador | 2 Unidades | 6.00 |
| | USB | 2 unidades | 60.00 |
| | Cd's | 4 unidades | 4.00 |
| 2.3.3 | Contratación de Servicios | | |
| 2.3.3.1 | Viajes | | |
| 2.3.3.2 | Viajes domésticos | | |
| 2.3.3.3 | Otros Gastos | | |
| | Movilidad Local | | 40.00 |
| | Imprevistos | | 100.00 |
| | Espacio en la nube para BD | | \$50 x mes |
| | postgres | | \$19.90 x mes |
| | Licencia Pycharm | | |
| 2.3.3.4 | Servicios Básicos, | | |
| | comunicaciones, Publicidad | | |
| 2.3.3.5 | y Difusión | | |
| 2.3.3.6 | Servicios de Telefonía e Internet | | |
| | Servicio de Internet | | |
| | Internet | | 120.00 |
| /ANS/. TOTAL | | | 1,740.61 |
| 2.3.3.7 | Servicios Básicos, | | |
| | comunicación, Publicidad y | | |
| 2.3.3.8 | Difusión | | |
| | Servicios de publicidad, | | |
| | Impresiones, Difusión e | | |
| 2.3.3.9 | Imagen Institucional | | |
| | Servicios de impresiones, | | |
| | encuadernación y | | |
| | Empastado | | |
| | Impresiones | 2 Millares | 300.00 |
| | Anillados | 6 unidades | 24.00 |
| | Empastado | 4 Unidades | 80.00 |

| | | | |
|------------------|------------|------------|-----------------|
| | Fotocopias | 400 copias | 40.00 |
| S/. TOTAL | | | 2,184.61 |

Fuente: Elaboración Propia

4.1. Financiamiento

Se define el financiamiento como el proceso que provee capital a un ente para usarlo en un negocio o proyecto, medios como crédito y dinero para que lleve a cabo sus fines. (BBVA, 2019).

Para la investigación la empresa cubre los gastos de bienes puestos ya cuenta con ellos, adicional a esto asumirá los costos de espacios en la nube y licencias a utilizar. El resto de los gastos será cubierto por los investigadores.

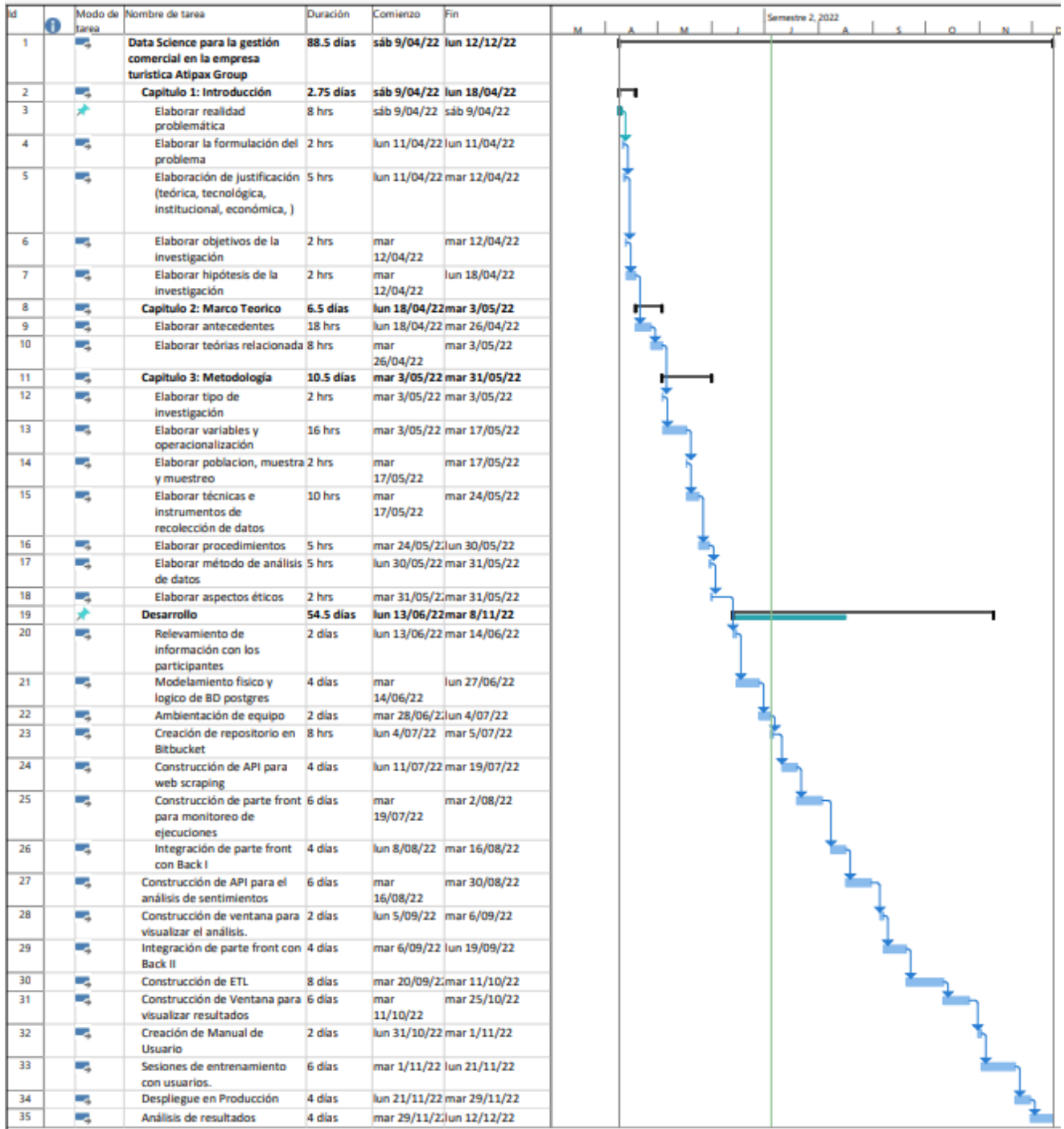
Tabla 27: Detalle de financiamiento

| Entidad financiadora | Monto |
|----------------------|----------------|
| Atipax Group | \$ 89.99 x mes |
| Investigadores | S/. 483.80 |

Fuente: Elaboración propia

Anexo 20: Cronograma de ejecución

Figura 6: Cronograma de actividades



Fuente: Elaboración propia

Anexo 27: Validación de lenguaje de programación

Tabla 28: Juicio de expertos para lenguaje de programación

Nombres y Apellidos del experto: Cesar Augusto Rosario Rodríguez

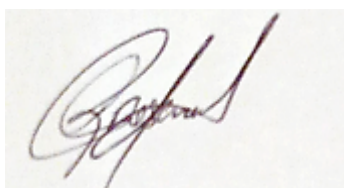
Título y/o Grado: Ingeniero de software

Mediante la tabla de evaluación de expertos, usted tiene la potestad de calificar los lenguajes de programación involucrados, mediante un conjunto de preguntas con puntuación indicadas al final de la tabla.

| ITEM | PREGUNTAS | Lenguajes de programación | | | |
|------|--|---------------------------|------|------------|---------------|
| | | Python | Java | JavaScript | OBSERVACIONES |
| 1 | Lenguaje de programación más flexible | 5 | 3 | 4 | |
| 2 | Lenguaje de programación con repositorio mejor adaptados a análisis de sentimientos | 4 | 2 | 3 | |
| 3 | Lenguaje de programación que proporciona librerías especialidad a análisis de sentimientos | 5 | 2 | 3 | |
| 4 | Lenguaje de programación que permite el aprendizaje automático | 4 | 2 | 2 | |
| 5 | Lenguaje de programación orientado a objetos de código abierto | 5 | 4 | 3 | |
| 6 | Lenguaje de programación mejor adaptado para conjunto de datos masivos | 5 | 4 | 3 | |
| 7 | Lenguaje de programación con más accesibilidad a sus herramientas | 5 | 5 | 5 | |

Utilizar la siguiente valoración:

1. Malo, 2. Regular, 3. Bueno, 4. Muy Bueno, 5. Excelente.



Nombres y Apellidos del experto: Orbezo Llancachagua

David Sixto

Título y/o Grado: INGENIERO DE SISTEMAS

Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.

Mediante la tabla de evaluación de expertos, usted tiene la potestad de calificar los lenguajes de programación involucrados, mediante un conjunto de preguntas con puntuación indicadas al final de la tabla.

| ITEM | PREGUNTAS | Lenguajes de programación | | | |
|------|--|---------------------------|------|------------|---------------|
| | | Python | Java | JavaScript | OBSERVACIONES |
| 1 | Lenguaje de programación más flexible | 5 | 3 | 3 | |
| 2 | Lenguaje de programación con repositorio mejor adaptados a análisis de sentimientos | 4 | 2 | 3 | |
| 3 | Lenguaje de programación que proporciona librerías especialidad a análisis de sentimientos | 5 | 2 | 2 | |
| 4 | Lenguaje de programación que permite el aprendizaje automático | 5 | 2 | 3 | |
| 5 | Lenguaje de programación orientado a objetos de código abierto | 4 | 3 | 3 | |
| 6 | Lenguaje de programación mejor adaptado para conjunto dedatos masivos | 4 | 2 | 3 | |
| 7 | Lenguaje de programación con más accesibilidad sus herramientas | 5 | 2 | 2 | |



Firma

Nombres y Apellidos del experto: Walter Atuncar Céspedes

Título y/o Grado: Ingeniero de sistemas

Web scraping y análisis de sentimientos para captar la valoración del cliente hacia los servicios de una empresa de turismo.

Mediante la tabla de evaluación de expertos, usted tiene la potestad de calificar los lenguajes de programación involucrados, mediante un conjunto de preguntas con puntuación indicadas al final de la tabla.

| ITEM | PREGUNTAS | Lenguajes de programación | | | OBSERVACIONES |
|------|--|---------------------------|------|------------|---------------|
| | | Python | Java | JavaScript | |
| 1 | Lenguaje de programación más flexible | 5 | 3 | 4 | |
| 2 | Lenguaje de programación con repositorio mejor adaptados a análisis de sentimientos | 5 | 2 | 2 | |
| 3 | Lenguaje de programación que proporciona librerías especialidad a análisis de sentimientos | 5 | 2 | 3 | |
| 4 | Lenguaje de programación que permite el aprendizaje automático | 5 | 2 | 2 | |
| 5 | Lenguaje de programación orientado a objetos de código abierto | 5 | 2 | 3 | |
| 6 | Lenguaje de programación mejor adaptado para conjunto de datos masivos | 5 | 4 | 3 | |
| 7 | Lenguaje de programación con más accesibilidad a sus herramientas | 5 | 5 | 5 | |

Utilizar la siguiente valoración:

1. Malo, 2. Regular, 3. Bueno, 4. Muy Bueno, 5. Excelente.



Walter Alexander, Atuncar Céspedes
DNI: 21877203

Anexo 21: Metodología Híbrida Scrum- CRISP-DM

El marco de trabajo que más se adapta a las necesidades de nuestro proyecto, y a través de un juicio de expertos, es del tipo ágil que será administrado bajo el marco de trabajo Scrum en conjunto a la metodología CRISP-DM.

Se optó por este marco de trabajo por el juicio de expertos y por su flexibilidad, la cual permite la generación de iteraciones para cada sprint programado para hacer cada vez más funcional al sistema de gestión de datos, también nos ayudara a corregir errores antes de pasar a la siguiente funcionalidad. Se definirá los requisitos del proyecto creando nuestro Backlog para luego ser particionado en Sprints donde las reuniones serán diarias y los entregables del Sprints de forma semanal.

Para la validación del marco de trabajo ágil se aplicó en esta investigación, se usó el software de apoyo Jira, donde creamos tres listas las cuales fueron llamadas: "Tareas Por Hacer", "En Curso" y "Finalizada". En el cual definimos nuestras tareas, roles y fechas de acuerdo con nuestro product backlog para cumplir con los tiempo estimados.

Figura 7 Tablero Jira



Fuente: Jira Software

Planificación inicial o sprint 0

En este sprint se definirá el objetivo del trabajo que realizaremos, también se elegirán las herramientas que se va a ir utilizando y definiremos roles del equipo de trabajo. Este sprint tendrá la siguiente estructura:

a) Técnicas e instrumentos

Las técnicas e instrumentos nos ayudarán para recolectar la información y dar respuesta a nuestros objetivos.

- Técnicas:

La técnica la utilizaremos para la recolección de información relacionada con el tema de estudio.

- Instrumentos:

Son los medios físicos y digitales para almacenar o registrar la información. Se estimará el tiempo para el cronograma de actividades, el cual deberá ser cumplido por los autores de la presente investigación. Se utilizará materiales de escritorio, equipos de cómputo y servicios de electricidad e internet para estos equipos. Se usará software de apoyo para el análisis, diseño, desarrollo y pruebas del modelo propuesto.

b) Roles de Scrum

En la Tabla 29, se listará los roles de nuestra investigación basado en la metodología Scrum, así mismo se describirá la persona a cargo y las actividades que realizará.

Tabla 29: Role Scrum

| ROL | ACTIVIDAD |
|---------------|---|
| Product Owner | ➤ Determinar las características funcionales de alto y bajo nivel en donde se implementará el modelo propuesto. |

| | |
|------------------------------------|---|
| <p>Scrum Máster</p> | <p>Este rol será llevado a cabo por el asesor Juan Brues Lee Chumpe Agosto</p> <p>El Scrum Máster asegura un ambiente laboral adecuado al equipo Scrum, con el objetivo de alcanzar una máxima productividad a lo largo del desarrollo del proyecto, el cual realizará las siguientes actividades:</p> <ul style="list-style-type: none"> - Asegurar que el desarrollo del modelo propuesto sea completado de manera exitosa. - Detectar y facilitar la eliminación de problemas que puedan presentarse en el desarrollo del proyecto. - Monitorear el correcto empleo y evolución de Scrum. |
| <p>Equipo de Desarrollo</p> | <p>Este rol será llevado a cabo por las personas que realizaron la siguiente investigación: Edinson Manuel Ñamo Alayo y Gilberto Romer Apaza Ramos.</p> <p>El rol del equipo de desarrollo en Scrum es auto organizado. Es el mismo equipo quien determina la forma de trabajar y resolver el problema. El equipo estará conformado por las autoras de esta investigación, las cuáles organizarán la construcción del modelo de la siguiente manera:</p> <ul style="list-style-type: none"> - Evaluación de la estructura HTML de las páginas a extraer. - Desarrollo de sistema de Web Scraping en Python. - Creación del sistema y conexión con la base de datos MongoDB. - Integración del sistema con Google Place y Twitter. - Pruebas funcionales del Sistema y corrección de errores. |

c) Product Backlog

Es una lista donde se prioriza las historias de usuario, funcionalidades que el cliente necesita, descritas usando la terminología del cliente.

El Product Backlog cuenta con tres Sprints, que a su vez tienen

funcionalidades por cada Sprint, dentro de las funcionalidades contiene tareas diarias con un tiempo de 4 horas, donde están designadas a un responsable que deberá cumplirlas horas programadas.

Tabla 30: Sprint Backlog

| SPRINT BACKLOG | | |
|-----------------------|------|---|
| Sprint 1 | H 1 | Entendimiento del Negocio- Definición de necesidades del cliente |
| | H 2 | Entendimiento de la Data- Estudio y comprensión de los datos |
| | H 3 | Preparación de la Data-Análisis de los datos y selección de características |
| Sprint 2 | H 4 | Modelado Propuesto |
| | H 5 | Evaluación de resultados y visualización de los mismos |
| | H 6 | Desarrollo de prototipo (Mockup) de propuesta de solución. |
| Sprint 3 | H 7 | Desarrollo del servicio scraping en Google Place y Twitter |
| | H 8 | Implementación de routes |
| | H 9 | Servicios de Autenticación de usuario |
| | H 10 | Desarrollo parte Front de Autenticación |
| | H 11 | Servicio de Análisis |
| | H12 | Despliegue (Paso de producción) |

Fuente: Elaboración propia.

d) Sprint Backlog

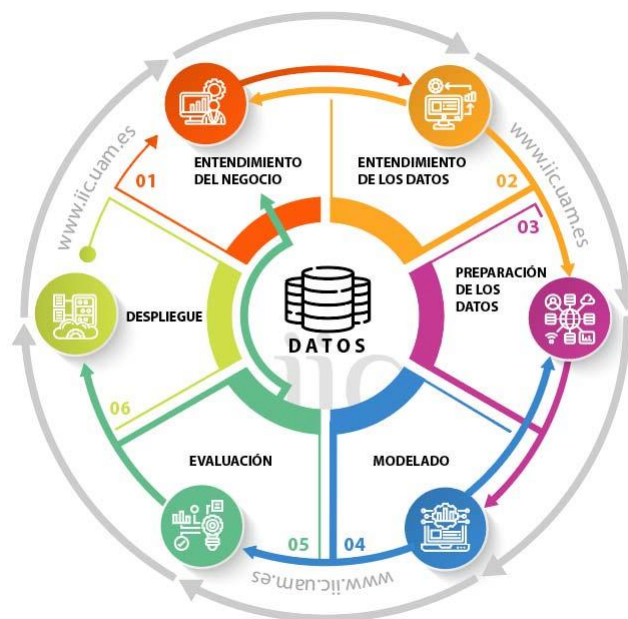
Es un conjunto de historias seleccionadas del Product Backlog, que cumplan una funcionalidad para trabajar en ellas durante un sprint, aquí se definen las tareas que

se realizarán para crear un incremento funcional de la aplicación. En la Tabla 8 observamos un resumen del Sprint Backlog que cuenta con 3 sprints y 12 historias.

Sprint 1

El sprint 1 tuvo 3 historias, las cuales se basaron en las fases de la metodología CRISP-DM así como se muestra en la figura 8, ya que permite entender el contexto del proyecto y resolver las dudas sobre el negocio.

Figura 8: Fases de metodología CRISP-DM



Fuente: Instituto de ingeniería del conocimiento

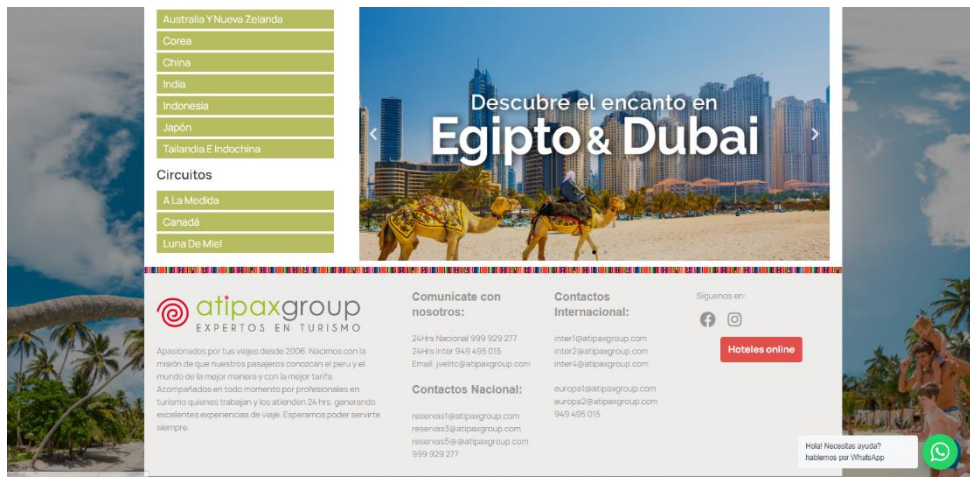
H1: Entendimiento del Negocio- Definición de necesidades del cliente

Para entender el negocio y conocer las necesidades, nos situamos a sus redes sociales, principalmente en su página web figura 9 y en su perfil de Facebook figura 10. Donde entendemos un poco más del rubro, a sus necesidades, ya que por el tema de la pandemia se ha visto afectada, como hemos podido observar en la figura 11, donde vemos que la captación de los clientes ha ido en caída.

Link de Página: <https://www.atipaxgroup.com>

Link de Facebook: https://www.facebook.com/atipaxperumundo/videos_by

Figura 9: Pagina de Atipax



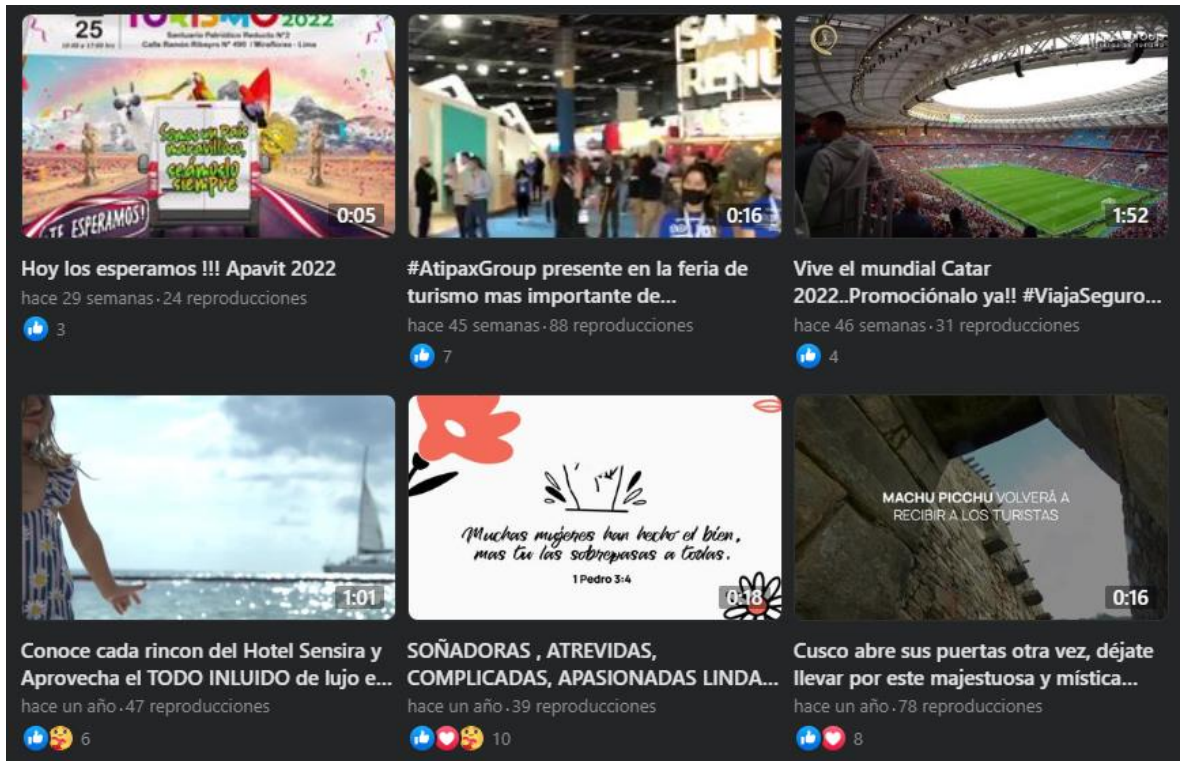
Fuente: Atipax Group

Figura 10: Facebook de Atipax



Fuente: Atipax Group

Figura 11: Captación de clientes desde Facebook



Fuente: Atipax Group.

Ya teniendo las necesidades de la empresa y estudiando el proceso ASIS para la gestión de ventas y clientes que maneja la empresa se procede a la siguiente historia.

H2: Entendimiento de la Data- Estudio y comprensión de los datos

Una vez realizado la reunión con el representante de la empresa para poder realizar un software en beneficio de ambas partes y haber estudiado la empresa nos centramos en la data proporcionada por la empresa (Figura 12) para su posterior análisis, la cual es las ventas de Abril (figura 13) y Mayo (figura 14).

Figura 12: Entendimiento de la data



Fuente: Jira

Figura 13: Ventas de Abril

| A | B | C | D | E | J | K | AG | AH | AI | AJ | AK | AL | AM | AN | AO | OB |
|--------|-----------|-----------|---------|-----------|----------------------------------|--------------|------------|-----------|-------------|------------|-------------|--------|------------|----------|-----------|-----|
| Codigo | Codigo000 | Codigo001 | CodUser | Fecha | Pax | NUM_Clientes | mFacturado | LiqVuelos | LiqProveedo | Itinerario | mTicketAere | mComtk | NotasCount | MarkUp10 | RefCotiza | OB |
| 21716 | 503201 | 3201 | 5 | 1/04/2022 | IDIAZ CLEMENTE/ALDO MANUEL X 02 | 2 | 156 | | | NULL | 0 | 0 | NULL | 0 | NULL | CEI |
| 21717 | 3702992 | 2992 | 37 | 1/04/2022 | SILVAX2 | 2 | 2066 | | | NULL | 0 | 0 | NULL | 0 | NULL | FIJ |
| 21718 | 3702993 | 2993 | 37 | 1/04/2022 | HERMOZAX3 | 3 | 315 | | | NULL | 0 | 0 | NULL | 0 | NULL | 01/ |
| 21719 | 3702994 | 2994 | 37 | 1/04/2022 | FERNANDEZ1 | 1 | 1290 | | | NULL | 0 | 0 | NULL | 0 | NULL | FIJ |
| 21720 | 2103204 | 3204 | 21 | 1/04/2022 | GIOVANNA MAZZINI X01 | 1 | 891 | | | NULL | 0 | 0 | NULL | 0 | NULL | 01/ |
| 21721 | 3702995 | 2995 | 37 | 1/04/2022 | RODRIGUEZ2 | 2 | 1156 | | | NULL | 0 | 0 | NULL | 0 | NULL | 11/ |
| 21722 | 2103205 | 3205 | 21 | 1/04/2022 | LILIANA RUIZ X02 | 2 | 1544 | | | NULL | 0 | 0 | NULL | 0 | NULL | FAI |
| 21723 | 3702996 | 2996 | 37 | 1/04/2022 | GARATEX2 | 2 | 696 | | | NULL | 0 | 0 | NULL | 0 | NULL | USI |
| 21724 | 3702997 | 2997 | 37 | 1/04/2022 | ALBINOX4 | 4 | 2633,88 | | | NULL | 0 | 0 | NULL | 0 | NULL | 02/ |
| 21725 | 2103206 | 3206 | 21 | 1/04/2022 | MOLINA IRIS X02 | 2 | 3920 | | | NULL | 0 | 0 | NULL | 0 | NULL | FAI |
| 21726 | 2103207 | 3207 | 21 | 1/04/2022 | ANA HI PAZOS X01 | 1 | 195 | | | NULL | 0 | 0 | NULL | 0 | NULL | FAI |
| 21727 | 2103208 | 3208 | 21 | 1/04/2022 | MONICA MAYER X10 | 10 | 4900 | | | NULL | 0 | 0 | NULL | 0 | NULL | FAI |
| 21728 | 3702998 | 2998 | 37 | 2/04/2022 | BUJANDAX2 | 2 | 216 | | | NULL | 0 | 0 | NULL | 0 | NULL | USI |
| 21729 | 3702999 | 2999 | 37 | 2/04/2022 | BUENDIA X2 | 2 | 53 | | | NULL | 0 | 0 | NULL | 0 | NULL | TAI |
| 21730 | 503202 | 3202 | 5 | 2/04/2022 | PATRICIA INES PAEZ ESPINOZA X 03 | 3 | 1103,47 | | | NULL | 0 | 0 | NULL | 0 | NULL | 03/ |
| 21731 | 2402979 | 2979 | 24 | 4/04/2022 | ANA PAULA PANDO X 03 | 3 | 793,8 | | | NULL | 0 | 0 | NULL | 0 | NULL | 04/ |
| 21732 | 2103209 | 3209 | 21 | 4/04/2022 | BLOISE GIOVANNI X02 | 2 | 386 | | | NULL | 0 | 0 | NULL | 0 | NULL | FAI |
| 21733 | 2103210 | 3210 | 21 | 4/04/2022 | XIMENA DELGADO X02 | 2 | 81 | | | NULL | 0 | 0 | NULL | 0 | NULL | PAI |
| 21734 | 2103211 | 3211 | 21 | 4/04/2022 | JULIA HAFNER X04 | 4 | 2180 | | | NULL | 0 | 0 | NULL | 0 | NULL | FAI |
| 21735 | 3202774 | 2774 | 32 | 4/04/2022 | CALDERON X1 | 1 | 255,1 | | | NULL | 0 | 0 | NULL | 0 | NULL | 04/ |
| 21736 | 503203 | 3203 | 5 | 4/04/2022 | PERCY JIMENEZ X 02 | 2 | 106 | | | NULL | 0 | 0 | NULL | 0 | NULL | CEI |
| 21737 | 3202775 | 2775 | 32 | 4/04/2022 | PERALTA X1 | 1 | 53,1 | | | NULL | 0 | 0 | NULL | 0 | NULL | 02- |
| 21738 | 3202776 | 2776 | 32 | 4/04/2022 | MARRUFO X1 | 1 | 105 | | | NULL | 0 | 0 | NULL | 0 | NULL | 06- |
| 21739 | 503204 | 3204 | 5 | 4/04/2022 | MILAGROS TOVAR X 02 | 2 | 476,15924 | | | NULL | 0 | 0 | NULL | 0 | NULL | 06/ |
| 21740 | 503205 | 3205 | 5 | 4/04/2022 | CECILIA TEJEDA ANYOSA X 02 | 2 | 949,18 | | | NULL | 0 | 0 | NULL | 0 | NULL | 05/ |
| 21741 | 5702928 | 2928 | 57 | 4/04/2022 | ERIC CABA X1 | 1 | 396,9 | | | NULL | 0 | 0 | NULL | 0 | NULL | FEC |
| 21742 | 3703000 | 3000 | 37 | 4/04/2022 | CAMAX3 | 3 | 3297 | | | NULL | 0 | 0 | NULL | 0 | NULL | USI |

Fuente: Atipax Group

Figura 14: Ventas de Mayo

| | | | | | | | | | | | | | | | | |
|-------|---------|------|----|------------|-----------------------|--|----|----------------|------------|------------|---|---|---|---|---|---|
| 15448 | 501769 | 1769 | 5 | 9/05/2019 | LUIS CALDERON X 02 | | | 2 SOLO ALOJAI | 16/05/2019 | 18/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15449 | 6602137 | 2137 | 66 | 9/05/2019 | ZEIDA VALDIVIEZO | | | 1 TICKET AERE | 22/07/2019 | 22/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15450 | 3201963 | 1963 | 32 | 9/05/2019 | TANTALEAN X1 | | | 1 AEREO LIMA | 13/08/2019 | 30/08/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15451 | 2301958 | 1958 | 23 | 9/05/2019 | BRAULIO QUISEP X1 | | | 1 BOLETO AERI | 13/06/2019 | 29/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15452 | 5702150 | 2150 | 57 | 9/05/2019 | HUGO ARRIETA | | | 4 COMISION | 11/01/2020 | 17/01/2020 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15453 | 6602138 | 2138 | 66 | 9/05/2019 | HUGO ARRIETA | | | 1 MEXICO ESEF | 8/08/2019 | 12/08/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15454 | 501770 | 1770 | 5 | 9/05/2019 | XX | | | 1 CHICLAYO | 28/06/2019 | 30/06/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15455 | 5702151 | 2151 | 57 | 9/05/2019 | GUILHERMO | | | 1 TOURS EN CA | 11/08/2019 | 11/08/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15456 | 6400051 | 51 | 64 | 9/05/2019 | MARIA GLADYS VASQUEZ | | | 1 PLAYAS | 29/06/2019 | 1/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15457 | 3201964 | 1964 | 32 | 9/05/2019 | VALADOLID | | | 1 ALOJAMIENT | 14/05/2019 | 15/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15458 | 5702152 | 2152 | 57 | 9/05/2019 | CABEZAS JOSE X4 | | | 4 TICKET A CAI | 11/01/2020 | 17/01/2020 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15459 | 2301959 | 1959 | 23 | 9/05/2019 | CARRASCO X3 | | | 3 BOLETOS AER | 5/07/2019 | 8/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15460 | 3702129 | 2129 | 37 | 9/05/2019 | JAISY VELITX15 | | 15 | | 5/05/2019 | 10/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15461 | 6400052 | 52 | 64 | 9/05/2019 | CARLOS CHAVERRIA | | | 1 CUSCO | 14/05/2019 | 17/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15462 | 3201965 | 1965 | 32 | 9/05/2019 | RAMIREZ X 2 | | | 2 PAQUETE PA | 23/05/2019 | 5/06/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15463 | 6400053 | 53 | 64 | 9/05/2019 | LUIS LANTIGUA | | | 1 CUSCO | 24/05/2019 | 26/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15464 | 3201966 | 1966 | 32 | 9/05/2019 | CAMACHO X 2 | | | 2 TESOROS DE | 12/09/2019 | 21/09/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15465 | 3201967 | 1967 | 32 | 9/05/2019 | CAMACHO X 2 | | | 2 PROGRAMA | 21/09/2019 | 24/09/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15466 | 3702130 | 2130 | 37 | 9/05/2019 | CLAUDIA SALAS | | | 1 | 14/05/2019 | 15/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15467 | 6602139 | 2139 | 66 | 9/05/2019 | LUIS HUAYTA | | | 1 HOTEL REGEF | 5/10/2019 | 10/10/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15468 | 5702153 | 2153 | 57 | 9/05/2019 | RUTH PACHECO X6 | | | 6 ALOJAMIENT | 12/10/2019 | 16/10/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15469 | 6400054 | 54 | 64 | 9/05/2019 | FERNANDEZ GERARDO | | | 1 CUSCO | 10/05/2019 | 12/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15470 | 5702154 | 2154 | 57 | 10/05/2019 | CRUZ MERCEDES X2 | | | 2 TOUR A PAN | 14/05/2019 | 14/05/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15471 | 2101691 | 1691 | 21 | 10/05/2019 | LUZ ELENA LEON X03 | | | 3 MANUAL | 15/07/2019 | 19/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15472 | 5702155 | 2155 | 57 | 10/05/2019 | DUEÑAS EMMA | | | 1 TICKET A CAI | 21/08/2019 | 25/08/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15473 | 2101692 | 1692 | 21 | 10/05/2019 | CARLOS CARRASCO X02 | | | 2 MANUAL | 27/07/2019 | 29/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15474 | 501771 | 1771 | 5 | 10/05/2019 | JUAN CARLOS LOCK X 02 | | | 2 PUNO | 26/07/2019 | 30/07/2019 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15475 | 2301960 | 1960 | 23 | 10/05/2019 | PAUL RIOS X4 | | | 4 EUROPA 17 | 23/07/2019 | 8/08/2019 | 0 | 0 | 0 | 0 | 0 | 0 |

Fuente: Atipax Group

Con la data obtenida se procede a realizar un análisis con el fin de obtener los datos más relevantes para poder realizar la preparación de la Data-Análisis de los datos y selección de características.

H3: Preparación de la Data-Análisis de los datos y selección de características.

Una vez analizado la data desde una consulta, se procede a seleccionar la información a utilizar, haciendo filtros (Figura 15), en la data de ambos meses. (Figura 16 y 17).

Figura 15: Preparación de la Data

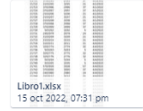
Preparación de la Data

Adjuntar Vincular incidencia

Descripción

Añadir una descripción...

Archivos adjuntos (11)



Actividad

Mostrar: Todo Comentarios Historial

Más recientes primero

Finalizada

✓ Listo

Detalles

Responsable: **RR** romer apaza ramos
 Etiquetas: Ninguno
 Desarrollo: [Crear rama](#), [Crear confirmación](#)
 Informador: **M** ManuZM

Creado 15 de septiembre de 2022, 15:16

Actualizado anteyear

Resuelto anteyear

Fuente: Jira

Figura 15: Filtro de Abril

| A | B | C | D | E | J | K | AG | AJ | AJ | AK | AL | AM | JACAP | AQ | AR | AS | AT | AU | AV | AW | AX | AY | AZ |
|----|-------|-------|-------|--------------|-----------------------------|-----|-----|------|----|----|----|----|-------|----|-----------|----|----|------|--------------|------|------------|------------|----|
| 1 | Codi | Codi | Codi | Fecha | Pax | NUM | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li |
| 2 | ##### | ##### | ##### | 5/1/04/2022 | IDIAZ CLEMENTE/ALDO MANU | 2 | 156 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 5 | 4324 | 5441558856 | 19 | USER-PC | APTO-F2204 | 0 |
| 3 | ##### | ##### | ##### | 37/1/04/2022 | SILVAX2 | 2 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 37 | 2603 | FILE 2204002 | 8 | USER-PC | AM-F220400 | 0 |
| 4 | ##### | ##### | ##### | 37/1/04/2022 | HERMOZAK3 | 3 | 315 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 41 | 4354 | | 1033 | EVELIN | AM-F220400 | 0 |
| 5 | ##### | ##### | ##### | 37/1/04/2022 | FERNANDEZ1 | 1 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 37 | 2603 | FILE 2204010 | 8 | USER-PC | AM-F220400 | 0 |
| 6 | ##### | ##### | ##### | 21/1/04/2022 | GIOVANNA MAZZINI X01 | 1 | 891 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 41 | 4339 | | 24 | EVELIN | APTO-F2204 | 2 |
| 7 | ##### | ##### | ##### | 37/1/04/2022 | RODRIGUEZ2 | 2 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 41 | 4355 | | 1037 | EVELIN | AM-F220400 | 0 |
| 8 | ##### | ##### | ##### | 21/1/04/2022 | LILIANA RUIZ X02 | 2 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 21 | 4339 | AUT X ELUSA | 750 | DESKTOP-KD | APTO-F2204 | 0 |
| 9 | ##### | ##### | ##### | 37/1/04/2022 | GARATEX2 | 2 | 696 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 37 | 2281 | | 1037 | USER-PC | AM-F220400 | 0 |
| 10 | ##### | ##### | ##### | 37/1/04/2022 | ALBINOX4 | 4 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 41 | 4355 | | 1012 | EVELIN | AM-F220400 | 0 |
| 11 | ##### | ##### | ##### | 21/1/04/2022 | MOLINA IRIS X02 | 2 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 21 | 4339 | AUT X PATRI | 176 | DESKTOP-KD | APTO-F2204 | 0 |
| 12 | ##### | ##### | ##### | 21/1/04/2022 | ANA HI PAZOS X01 | 1 | 195 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 21 | 4507 | FILE TRAVEX | 750 | DESKTOP-KD | APTO-F2204 | 0 |
| 13 | ##### | ##### | ##### | 21/1/04/2022 | MONICA MAYER X10 | 10 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 1/04/2022 | 0 | 21 | 4321 | AUT X ELUSA | 750 | DESKTOP-KD | APTO-F2204 | 0 |
| 14 | ##### | ##### | ##### | 37/2/04/2022 | BUJANDAX2 | 2 | 216 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 2/04/2022 | 0 | 37 | 207 | | 1037 | USER-PC | AM-F220400 | 0 |
| 15 | ##### | ##### | ##### | 37/2/04/2022 | BUENDIA X2 | 2 | 53 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 2/04/2022 | 0 | 37 | 4355 | | 936 | USER-PC | AM-F220400 | 0 |
| 16 | ##### | ##### | ##### | 5/2/04/2022 | PATRICIA INES PAEZ ESPINOZA | 3 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 2/04/2022 | 0 | 5 | 4324 | | 995 | USER-PC | APTO-F2204 | 0 |
| 17 | ##### | ##### | ##### | 24/4/04/2022 | ANA PAJUA PANDO X 03 | 3 | 794 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 41 | 4355 | VIENE DE FIL | 7 | EVELIN | AM-F220400 | 0 |
| 18 | ##### | ##### | ##### | 21/4/04/2022 | BLOSE GIOVANNI X02 | 2 | 386 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 21 | 4516 | FILE VC 2204 | 8 | DESKTOP-KD | APTO-F2204 | 0 |
| 19 | ##### | ##### | ##### | 21/4/04/2022 | XIMENA DELGADO X02 | 2 | 81 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 21 | 4330 | VITA X CERRA | 24 | DESKTOP-KD | APTO-F2204 | 0 |
| 20 | ##### | ##### | ##### | 21/4/04/2022 | JULIA HAFFNER X04 | 4 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 21 | 4339 | AUT X MARLU | 176 | DESKTOP-KD | APTO-F2204 | 0 |
| 21 | ##### | ##### | ##### | 32/4/04/2022 | CALDERON X1 | 1 | 255 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 32 | 653 | | 389 | LAPTOP-SCU | AM-F220401 | 0 |
| 22 | ##### | ##### | ##### | 5/4/04/2022 | PERCY JIMENEZ X 02 | 2 | 106 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 5 | 4499 | EMD 5441556 | 14 | USER-PC | APTO-F2204 | 0 |
| 23 | ##### | ##### | ##### | 32/4/04/2022 | PERALTA X1 | 1 | 53 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 32 | 653 | | 1012 | LAPTOP-SCU | AM-F220401 | 0 |
| 24 | ##### | ##### | ##### | 32/4/04/2022 | MARBUFO X1 | 1 | 103 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 32 | 653 | | 215 | LAPTOP-SCU | AM-F220401 | 0 |
| 25 | ##### | ##### | ##### | 5/4/04/2022 | MILAGROS TOVAR X 02 | 2 | 476 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 41 | 4320 | | 1013 | EVELIN | APTO-F2204 | 0 |
| 26 | ##### | ##### | ##### | 5/4/04/2022 | CECILIA TEJEDA ANYOSA X 02 | 2 | 949 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 41 | 4325 | | 157 | EVELIN | APTO-F2204 | 0 |
| 27 | ##### | ##### | ##### | 57/4/04/2022 | ERIC CABA X1 | 1 | 397 | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 41 | 3822 | | 484 | EVELIN | AM-F220401 | 0 |
| 28 | ##### | ##### | ##### | 37/4/04/2022 | CAMAX3 | 3 | ### | NULL | 0 | 0 | 0 | 0 | 0 | 0 | 4/04/2022 | 0 | 41 | 4355 | | 1037 | CONTABILID | AM-F220401 | 0 |

Fuente: Elaboración propia

Figura 16: Filtro de Mayo

| A | B | C | D | E | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG | AH | AI | AJ | AK | AL | AM | AN | AO | AP | AQ | AR | |
|----|-------|-------|-------|-------|--------------|----------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | Cx | Cx | Co | Co | Fecha | Pax | UNIDADES | FE | FE | N | N | N | N | N | N | N | N | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | Li | |
| 2 | ##### | ##### | ##### | ##### | 2/05/2022 | ACHO CHÁVEZ K7 | 7 | EXC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 3 | ##### | ##### | ##### | ##### | 2/05/2022 | JOSE VILCA X 3 | 3 | EXC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 4 | ##### | ##### | ##### | ##### | 2/05/2022 | JUAN CARLOS GUZMAN X02 | 2 | SOI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 5 | ##### | ##### | ##### | ##### | 2/05/2022 | ALEJANDRO WONG X02 | 2 | FUL | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 6 | ##### | ##### | ##### | ##### | 2/05/2022 | VANESSA RUSSO X03 | 3 | ADI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 7 | ##### | ##### | ##### | ##### | 2/05/2022 | BLATTERX2 | 2 | PAI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 8 | ##### | ##### | ##### | ##### | 24/05/2022 | MILAGROS CALDERON X 02 | 2 | TOI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 9 | ##### | ##### | ##### | ##### | 2/05/2022 | CARMELO VENERO X04 | 4 | MA | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 10 | ##### | ##### | ##### | ##### | 2/05/2022 | GIOVANNA MAZZINI X02 | 2 | REF | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 11 | ##### | ##### | ##### | ##### | 2/05/2022 | VILLALVA ALCAS/FELIX X 02 | 2 | DEC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 12 | ##### | ##### | ##### | ##### | 28/02/2022 | YI CHANG X 30 | 30 | REF | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 13 | ##### | ##### | ##### | ##### | 37/2/05/2022 | PONTEKX7 | 7 | PAI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 14 | ##### | ##### | ##### | ##### | 37/2/05/2022 | MEDRANO X2 | 2 | ALC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 15 | ##### | ##### | ##### | ##### | 2/05/2022 | HUGHES MICHAEL MURPHY X 02 | 2 | FUL | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 16 | ##### | ##### | ##### | ##### | 37/2/05/2022 | CHAVEZ LYNCHX2 | 2 | ALC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 17 | ##### | ##### | ##### | ##### | 24/05/2022 | MARILYN BENAVIDES BEJARANO | 1 | PRC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 18 | ##### | ##### | ##### | ##### | 2/05/2022 | STEPHAN LARSSON | 1 | BOI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 19 | ##### | ##### | ##### | ##### | 2/05/2022 | SHANE O'CONNOR / X 02 | 2 | FUL | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 20 | ##### | ##### | ##### | ##### | 3/05/2022 | EMERSON AYALA X 02 | 2 | FUL | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 21 | ##### | ##### | ##### | ##### | 3/05/2022 | LOPEZX14 PAX2 | 14 | EXC | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 22 | ##### | ##### | ##### | ##### | 37/05/2022 | DONAYREX2 | 2 | PAI | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### | ### |
| 23 | ##### | ##### | ##### | ##### | 3/05/2022 | LORENA CABRERA X02 | 2 | SOI | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

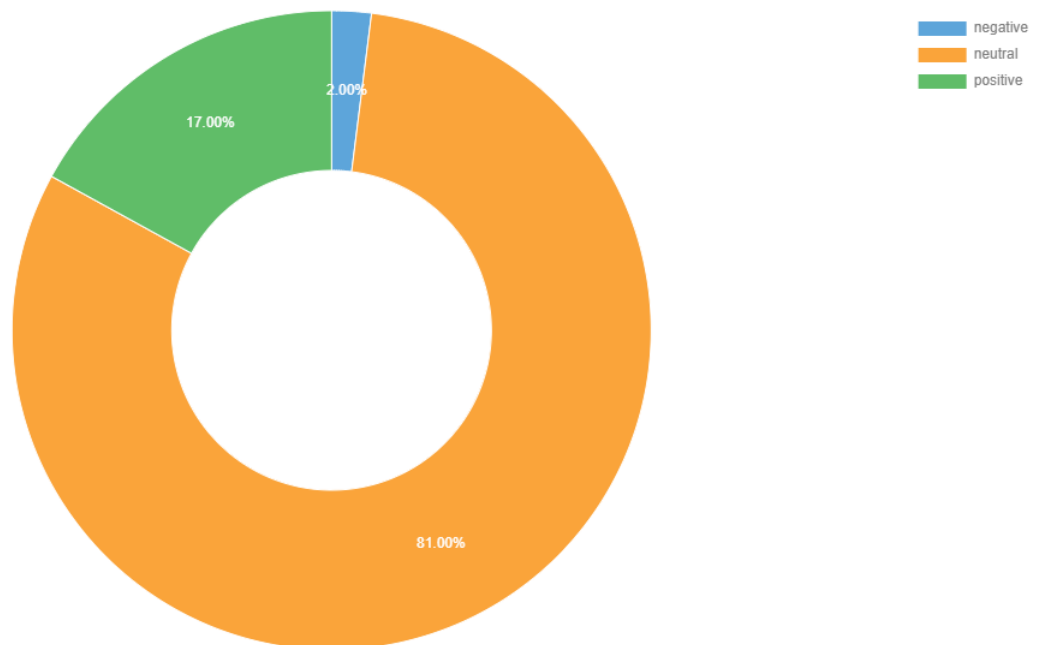
Sprint 2

El sprint 2 contempla de 3 historia, la cual es el modelamiento de prototipo del software, o mockup de solución, el modelo propuesto del análisis de sentimientos y la evaluación de resultados y visualización de sí mismos.

H4: Modelado Propuesto.

Aplicación de Análisis de sentimientos a los resultados obtenidos, una vez se obtenga los datos del web scraping a través de las plataformas de Twitter o Google Place, se realizara un análisis de sentimientos a la data obtenida con fin de obtener una leyenda donde visualicemos los datos tanto positivos como negativos y neutrales. Figura 37.

Figura 37: Modelado de análisis de sentimiento



Fuente: Elaboración propia.

H5: Evaluación de propuesta

Primera reunión con área administrativa de Atipax para coordinación de propuesta de solución, se llevó a cabo con los representantes, la Sra. Jaissy Velit Carnero gerente general de la empresa de turismo Atipax Group, donde se coordinó los términos y condiciones para llevar a cabo la investigación.

Figura 35: Evidencia 1



Fuente: Elaboración propia.

Figura 36: Evidencia 2



Fuente: Elaboración propia.

Figura 37: Evidencia 3



Fuente: Elaboración propia.

H6: Se realiza el mockup de la solución, que se realizó en dos partes, para poder dar una revisión más objetiva, la cual quedo en 5 vistas. Como primera vista tenemos el Login, el cual contiene la parte de autenticar al usuario a través de usuario y contraseña, para la segunda vista tenemos el Scraping de publicaciones, que se centra en escoger la plataforma en la que se a buscar, en este caso (Twitter y Google Place), así mismo, el filtrado que consiste en la ubicación, en caso de Twitter se escogerá el número de Twitts para iniciar el proceso. La tercera parte se basara en visualización del análisis, que se realizara a través de un cuadro estadístico para su posterior interpretación, dar una mejor toma de decisión.

Figura 17: Primera parte de prototipo

Añadir epic / PTESIS-5

Archivos adjuntos (3)

- image-20220916... 758.png
16 sept 2022, 06:48 pm
- image-20220916... 813.png
16 sept 2022, 06:48 pm
- image-20220916... 805.png
16 sept 2022, 06:48 pm

Incidencias secundarias

Ordenar por

PTESIS-7 Desarrollo de Mockup de propuesta de solución

Finalizada Listo

Detalles

Responsable ManuZM
[Asignarme a mí](#)

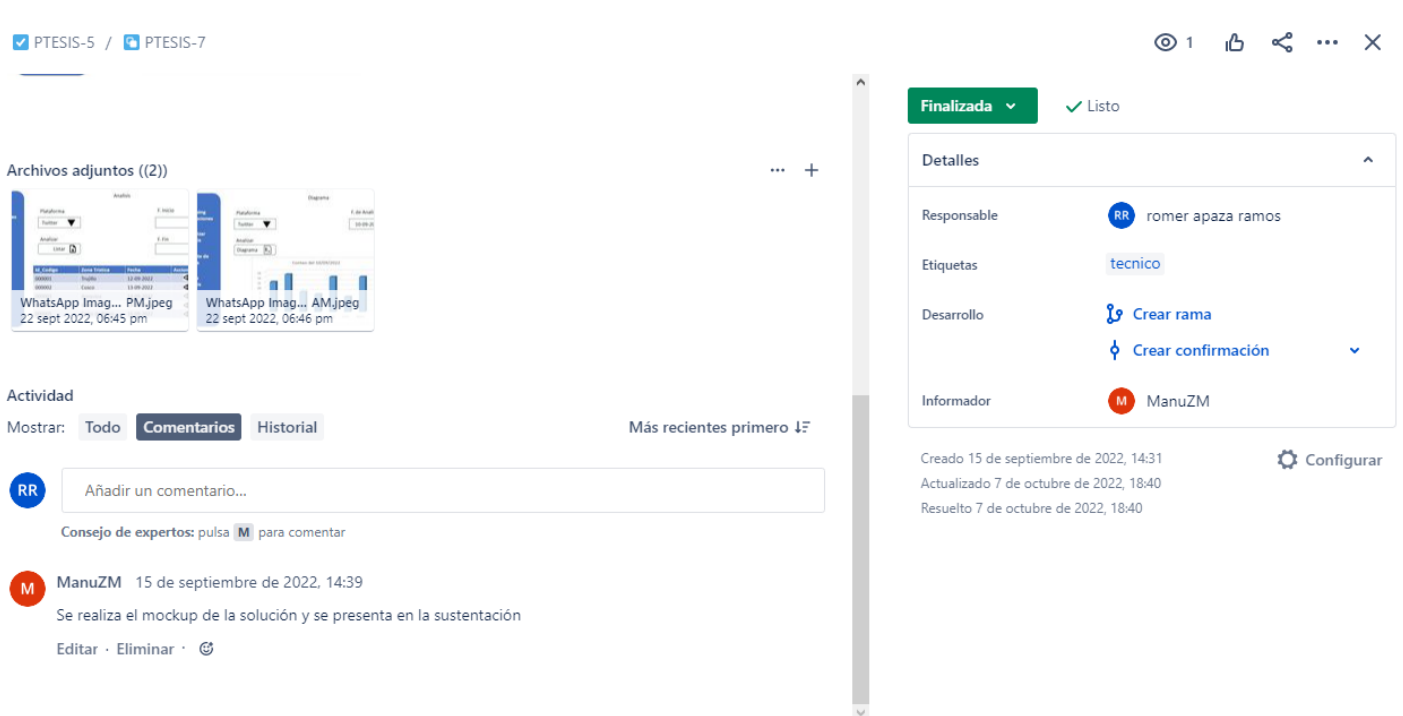
Etiquetas [tecnico](#)

Desarrollo [Crear rama](#)
[Crear confirmación](#)

Informador ManuZM

Fuente: Jira

Figura 18: Segunda parte del prototipo



Fuente: Jira

Sprint 3

El sprint 3 contempla 6 historias, centradas en programación del software, podremos visualizar el web scraping de Twitter en la figura 19 y el de Google Place en la figura 20, así mismo el front y servicios.

H7: Desarrollo del servicio scraping en Google Place y Twitter

Se optó por hacer el web scraping en estas dos plataformas debido a su flexibilidad y facilidad de poder acceder a su información, a diferencia de otras redes sociales se tiene que tener credenciales extra.

Figura19: Web Scraping de Twitter y Google Place


```
scraping.py x auth.py x database.py x models.py x requirements.txt x main.py x
Package requirements 'aiofiles==0.8.0', 'alembic==1.7.7', 'anyio==3.5.0', 'asgiref==3.5.0', 'bcrypt==3.2.0', 'certifi==2021.10.8', 'cffi==1.15.0', 'charset-normalizer==2.0... Install requireme

61
62 @router.post("/add-busqueda", response_class=HTMLResponse)
63 async def create_todo(request: Request, tipo_busqueda: str = Form(...), filtro: str = Form(...),
64 cant_respuestas: Optional[int] = Form(...), db: Session = Depends(get_db)):
65     user = await get_current_user(request)
66     if user is None:
67         return RedirectResponse(url="/auth", status_code=status.HTTP_302_FOUND)
68
69     tipo = tipo_busqueda
70     if(tipo == "Google"):
71         gmaps = googlemaps.Client(key='AIzaSyBj8uzVCn8FvDy1kbnONCXETowVa0DsRV8')
72         place_name = filtro
73         places_result = gmaps.places(place_name)
74         place_id = places_result['results'][0]['place_id']
75         place = gmaps.place(place_id=place_id)
76         reviews = place['result']['reviews']
77
78     if (tipo == "Twitter"):
79         consumer_key = "WWLZ4mhEcSMOXkdJbDB8wSDAT"
80         access_token_secret = "0gYR0kNs1C7unStKJbTenFgBFnAwhJFZ8tntFb0hPIw0sxqo3X"
81         bearer_token = "AAAAAAAAAAAAAAAAAAADvThqEAAAAAutxMOCF1LN8o1w6NLVd%2F2QTjxAs%3DTI5x0DoY3qqXdavCsZjzcQB2DB"
82         client = tweepy.Client(bearer_token=bearer_token,
83                               consumer_key=consumer_key,
84                               access_token_secret=access_token_secret,
85                               return_type=requests.Response,
86                               wait_on_rate_limit=True)
87
88         place_name = filtro
89         tweets = client.search_recent_tweets(query=place_name,
90                                             tweet_fields=['author_id', 'created_at'],
```

Fuente: Elaboración propia.

Se puede observar que a través de token, nos da la facilidad para poder trabajar con el web scraping en twitter. Así mismo con la llave para googlemaps (Google Places).

Figura 20: Recuperación de la data

```
scrapping.py x auth.py x database.py x models.py x requirements.txt x main.py x
Package requirements 'aiofiles==0.8.0', 'alembic==1.7.7', 'anyio==3.5.0', 'asgiref==3.5.0', 'bcrypt==3.2.0', 'certifi==2021.10.8', 'cffi==1.15.0', 'charset-normalizer==2.0... Install requireme

91     tweets_dict = tweets.json()
92     tweets_data = tweets_dict['data']
93     reviews = tweets_data
94
95     scrap_model = models.Scrapping()
96     scrap_model.tipo_busqueda = tipo_busqueda
97     scrap_model.filtro = filtro
98     scrap_model.cant_respuestas = cant_respuestas
99     scrap_model.user_id = user.get("id")
100    scrap_model.fecha_busqueda = str(date.today())
101    scrap_model.resultados = reviews
102    db.add(scrap_model)
103    db.commit()
104
105    return RedirectResponse(url="/busquedas", status_code=status.HTTP_302_FOUND)
106
107 def successful_response(status_code: int):
108     return {
109         'status': status_code,
110         'transaction': 'Successful'
111     }
112
113
114 def http_exception():
115     return HTTPException(status_code=404, detail="Todo not found")
```

Fuente: Elaboración propia.

Se puede visualizar que toda la información obtenida por el web scraping que se le aplico a twitter y google places, se está guardando todo en un json, para su posterior uso en el análisis de sentimientos.

H8: Implementación de routers

En la implementación de routers, se crea las rutas que se van a utilizar para el web scraping, tanto de twitter y google places, así mismo es la parte donde se va a centralizar todo el programa como vamos a poder visualizar en la figura 21 y 22.

Figura 21: programación de routers

```

@router.post("/token")
async def login_for_access_token(response: Response, form_data: OAuth2PasswordRequestForm = Depends(), db: Session = Depends(get_db)):
    user = authenticate_user(form_data.username, form_data.password, db)
    if not user:
        return False
    token_expires = timedelta(minutes=60)
    token = create_access_token(user.username, user.id, expires_delta=token_expires)
    response.set_cookie(key="access_token", value=token, httponly=True)
    return True

Edinson
@router.get("/", response_class=HTMLResponse)
async def authentication_page(request: Request):
    return templates.TemplateResponse("login.html", {"request": request})

Edinson
@router.post("/", response_class=HTMLResponse)
async def login(request: Request, db: Session = Depends(get_db)):
    try:
        form = LoginForm(request)
        await form.create_auth_form()
        response = RedirectResponse(url="/busquedas", status_code=status.HTTP_302_FOUND)

        validate_user_cookie = await login_for_access_token(response=response, form_data=form, db=db)

    if not validate_user_cookie:
        msg = "Incorrect Username or Password"
        return templates.TemplateResponse("login.html", {"request": request, "msg": msg})

```

Fuente: Elaboración propia

Se observa en la figura 21, las rutas, y parte de los servicios como el .get, .post, que están guardando el token para los accesos, así mismos algunas validaciones, también vemos la parte de la autenticación de la página, y posteriormente el Login.

Figura 22: programación de routers 1

```

@router.get("/logout")
async def logout(request: Request):
    msg = "Logout Successful"
    response = templates.TemplateResponse("login.html", {"request": request, "msg": msg})
    response.delete_cookie(key="access_token")
    return response

Edinson
@router.get("/register", response_class=HTMLResponse)
async def register(request: Request):
    return templates.TemplateResponse("register.html", {"request": request})

Edinson
@router.post("/register", response_class=HTMLResponse)
async def register_user(request: Request, email: str = Form(...), username: str = Form(...),
    firstname: str = Form(...), lastname: str = Form(...),
    password: str = Form(...), password2: str = Form(...),
    db: Session = Depends(get_db)):
    validation1 = db.query(models.Users).filter(models.Users.username == username).first()
    validation2 = db.query(models.Users).filter(models.Users.email == email).first()

    if password != password2 or validation1 is not None or validation2 is not None:
        msg = "Invalid registration request"
        return templates.TemplateResponse("register.html", {"request": request, "msg": msg})

```

Fuente: Elaboración propia.

En la figura 22 observamos la obtención del logueo, y sus acciones, como el mensaje, así mismo el de registrar, que te pedirá credenciales del usuario, para posteriormente validarlo con la historia 10, de servicios de autenticación del usuario.

H 9: Servicios de Autenticación de usuario

En la historia 9 se contempla la autenticación de usuario Figura 23 y las validaciones de la autenticación figura 24.

Figura 23: Servicio de Autenticación

```
SECRET_KEY = "LA CLAVE SECRETA"
ALGORITHM = "HS256"

templates = Jinja2Templates(directory="templates")

# Edinson
class CreateUser(BaseModel):
    username: str
    email: Optional[str]
    first_name: str
    last_name: str
    password: str

bcrypt_context = CryptContext(schemes=["bcrypt"], deprecated="auto")

models.Base.metadata.create_all(bind=engine)

oauth2_barer = OAuth2PasswordBearer(tokenUrl="token")

# app = FastAPI()
router = APIRouter(
    prefix="/auth",
    tags=["auth"],
    responses={401: {"user": "Not authorized"}}
)
```

Fuente:

elaboración propia

En la figura 23 observamos la parte de creación de usuario donde se le pide el username, email, first_name, last_name y password, se está encriptando la contraseña por temas de seguridad y pasándolo por la ruta para su posterior almacenamiento en la base de datos, así mismo si se llega a ingresar con credenciales erróneas, saldrá un aviso que no autorizado.

Figura 24: Obtención del usuario

```
def __init__(self, request: Request):
    self.request: Request = request
    self.username: Optional[str] = None
    self.password: Optional[str] = None

    Edinson
    async def create_auth_form(self):
        form = await self.request.form()
        self.username = form.get("email")
        self.password = form.get("password")

    Edinson
    def get_db():
        try:
            db = SessionLocal()
            yield db
        finally:
            db.close()

    Edinson
    def get_password_hash(password):
        return bcrypt_context.hash(password)

    Edinson
    def verify_password(plain_password, hashed_password):
        return bcrypt_context.verify(plain_password, hashed_password)
```

Fuente: Elaboración propia.

En la figura 24 se está llevando a cabo la creación del usuario con la función `init`, que está pasando como parámetro al `request`, lo que contiene el `username` y `password`, posteriormente la creación y la obtención del ingreso a través del `email` y `password`.

Figura 25: Validaciones del usuario

```
def authenticate_user(username: str, password: str, db):
    user = db.query(models.Users).filter(models.Users.username == username).first()

    if not user:
        return False
    if not verify_password(password, user.hashed_password):
        return False
    return user

    Edinson
    def create_access_token(username: str, user_id: int, expires_delta: Optional[timedelta] = None):
        encode = {"sub": username, "id": user_id}
        if expires_delta:
            expire = datetime.utcnow() + expires_delta
        else:
            expire = datetime.utcnow() + timedelta(minutes=15)
        encode.update({"exp": expire})
        return jwt.encode(encode, SECRET_KEY, algorithm=ALGORITHM)

    Edinson
    async def get_current_user(request: Request):
        try:
            token = request.cookies.get("access_token")
            if token is None:
                return None
            payload = jwt.decode(token, SECRET_KEY, algorithms=[ALGORITHM])
            username: str = payload.get("sub")
```

Fuente: Elaboración propia.

En la figura 25, se observa la autenticación del usuario a base de validaciones en la función `authenticate_user` que va a realizar una consulta del usuario, si no es retornara falso, así mismo validaciones para el token para la posterior validación en la obtención del usuario.

H 10: Desarrollo parte Front de Autenticación

La historia 10, que se centra en el desarrollo por parte del Front de autenticación, observaremos, algunas validaciones en Json en la figura 26 y 27 que son fundamentales para el front y autenticar las variables. Así mismo observaremos en la figura 28 el código del HTML.

Figura 26: Validación de variables

```
typeof exports === 'object' && typeof module !== 'undefined' ? factory(exports, require('jquery'), require('popper.js')) :
typeof define === 'function' && define.amd ? define(['exports', 'jquery', 'popper.js'], factory) :
(global = global || self, factory(global.bootstrap = {}, global.jQuery, global.Popper));
}(this, function (exports, $, Popper) { 'use strict';

$ = $ && $.hasOwnProperty('default') ? $['default'] : $;
Popper = Popper && Popper.hasOwnProperty('default') ? Popper['default'] : Popper;

function _defineProperties(target, props) {
  for (var i = 0; i < props.length; i++) {
    var descriptor = props[i];
    descriptor.enumerable = descriptor.enumerable || false;
    descriptor.configurable = true;
    if ("value" in descriptor) descriptor.writable = true;
    Object.defineProperty(target, descriptor.key, descriptor);
  }
}

function _createClass(Constructor, protoProps, staticProps) {
  if (protoProps) _defineProperties(Constructor.prototype, protoProps);
  if (staticProps) _defineProperties(Constructor, staticProps);
  return Constructor;
}

function _defineProperty(obj, key, value) {
  if (key in obj) {
    Object.defineProperty(obj, key, {
      value: value,
      enumerable: true,

```

Fuente: Elaboración propia.

En la figura 26 se puede observar distintas funciones, de definición de propiedades y que estas cumplan con una normativa, después de corroborar las validaciones se procederá al guardado de información.

Figura 27: Validaciones de ejecución

```

(function($) {
  "use strict"; // Start of use strict
  // Toggle the side navigation
  $("#sidebarToggle, #sidebarToggleTop").on('click', selector: function(e) {
    $("body").toggleClass("sidebar-toggled");
    $(".sidebar").toggleClass("toggled");
    if ($("#sidebar").hasClass("toggled")) {
      $(".sidebar .collapse").collapse("hide");
    }
  });
  // Close any open menu accordions when window is resized below 768px
  $(window).resize(function() {
    if ($(window).width() < 768) {
      $(".sidebar .collapse").collapse("hide");
    }
    // Toggle the side navigation when window is resized below 480px
    if ($(window).width() < 480 && !$(".sidebar").hasClass("toggled")) {
      $("body").addClass("sidebar-toggled");
      $(".sidebar").addClass("toggled");
      $(".sidebar .collapse").collapse("hide");
    }
  });
  // Prevent the content wrapper from scrolling when the fixed side navigation hovered over
  $('body.fixed-nav .sidebar').on('mouseover', selector: function(e) {
    if ($(window).width() > 768) {
      var e0 = e.originalEvent,
          delta = e0.wheelDelta || -e0.detail;
      this.scrollTop += (delta < 0 ? 1 : -1) * 30;
      e.preventDefault();
    }
  });
}

```

Fuente: Elaboración propia.

En la figura 27 se observa las validaciones en cuanto a diseño de la vista, en la primera parte tenemos la función para activar la navegación lateral, así mismos, cerrar cualquier menú abierto cuando la ventana se redimensiona por debajo del a abierta y por ultimo de evitar que la envoltura de contenido se desplace cuando se pasa por encima de la navegación lateral fija.

Figura 28: HTML del front

```

<div class="card-header">
  Login
</div>
<div class="card-body">
  {% if msg %}
  {% if msg == 'Logout Successful' or msg == 'User successfully created' %}
  <div class="alert alert-success" role="alert">
    {{msg}}
  </div>
  {% else %}
  <div class="alert alert-danger" role="alert">
    {{msg}}
  </div>
  {% endif %}
  {% endif %}
  <form method="POST" action="/auth">
    <div class="form-group">
      <label>Username</label>
      <input type="text" class="form-control"
        name="email" required>
    </div>
    <div class="form-group">
      <label>Password</label>
      <input type="password" class="form-control" name="password" required>
    </div>
    <button type="submit" class="btn btn-primary">Login</button>
  </form>
</div>
<div class="card-footer text-muted">
  <a href="/register"> Register? </a>

```

Fuente: Elaboración propia.

En la figura 28, se observa el HTML, que es el lenguaje de marcas de Hipertexto que es un campo fundamental, el cual nos muestra las vistas o pantallas de las páginas web, en este caso podemos observar las entradas que se le va dar a la página y también algunas validación, como el caso de logout.

H11: Servicio de Análisis

En la historia 11 de servicio del análisis de sentimientos que se le va a dar a los datos obtenidos del web scraping de Twitter y Google, se va a utilizar el algoritmo VADER que nos ayudad a analizar las polaridades entre negativo, neutral y positivo.

Figura: 38 Obtención de datos y limpieza

```
def get_db():
    try:
        db = SessionLocal()
        yield db
    finally:
        db.close()

Edinson
def limpieza_textos(tweet):
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|(^0-9A-Za-z \t)|(\w+:\w+\s+)", "", tweet).split())
```

Fuente: Elaboración propia

En la figura 38, observamos la obtención de la data desde la base de datos, para posteriormente hacer una limpieza de la data, ya que esta al hacer el web scraping se guarda directamente la data en un json, donde la data viene con otros elementos, y por esta razón antes de hacer el análisis de sentimientos se requiere de hacer una limpieza de la data para posteriormente analizarla

Figura 39: Servicio del análisis y creación de grafico


```

def get_sentimientos_textos(tweet):
    analysis = TextBlob(limpieza_textos(tweet))

    if analysis.sentiment.polarity > 0:
        return 'positive'
    elif analysis.sentiment.polarity == 0:
        return 'neutral'
    else:
        return 'negative'

Edinson

def data_grafico( analisis_model ):
    dfResultados = pd.json_normalize( analisis_model.resultados )
    dfCount = dfResultados.groupby(['sentiment']).count()
    arrayCabeceras = [row.name for index, row in dfCount.iterrows()]
    arrayValues = [row[0] for index, row in dfCount.iterrows()]
    return arrayCabeceras, arrayValues

```

Fuente: Elaboración propia.

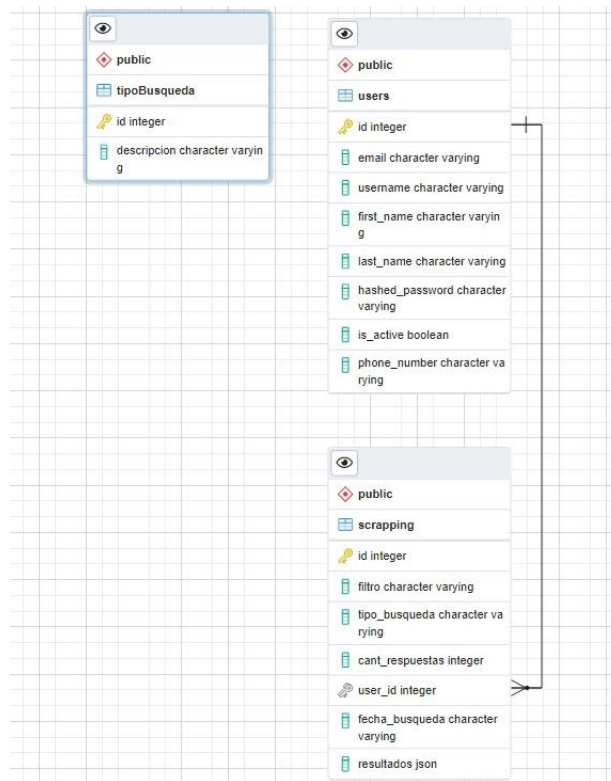
En la figura 39, se puede visualizar la obtención de la data ya limpiada y se procede a realizar en análisis, según la polaridad de comentario o tweet, se reagrupa entre positivo si es mayor a 0, neutral si es igual a 0 y negativo por defecto si es menor a 0.

Así mismo observamos los resultados obtenidos y posteriormente con los datos obtenidos, realizamos un gráfico para la mejor visualización del usuario.

H12: Despliegue pase a producción

Una vez finalizado la primera fase, se hace un pase a producción, en el cual tocaremos el deslgué de la solución en repositorio de Bitbucket, creación de la Base de Datos PostgreSQL en Heroku figura 29, repositorio de la rama master en la figura 30 y el despliegue del sistema.

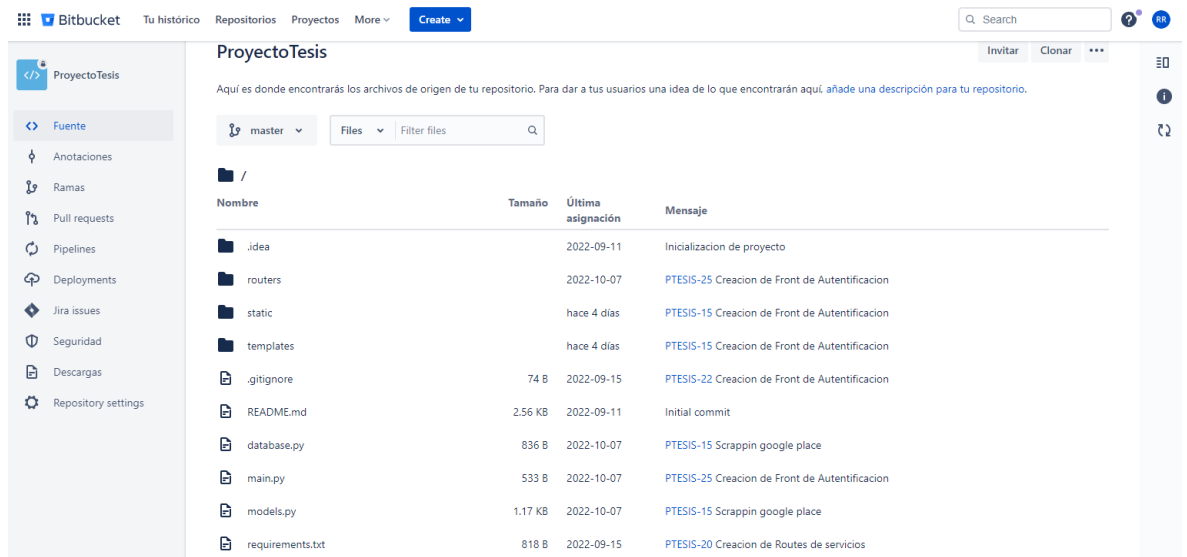
Figura 29: Base de datos



Fuente: Elaboración propia.

En la figura 29 se puede observar tres tablas, las cuales son `tipoBusqueda`, `Users`, `Scraping`, las tres con sus respectivas variables donde almacenaran la data recopilada, tanto de los usuarios, las búsquedas que se hacen a través del web scraping de los lugares o zonas turísticas.

Figura 30: Rama master del proyecto

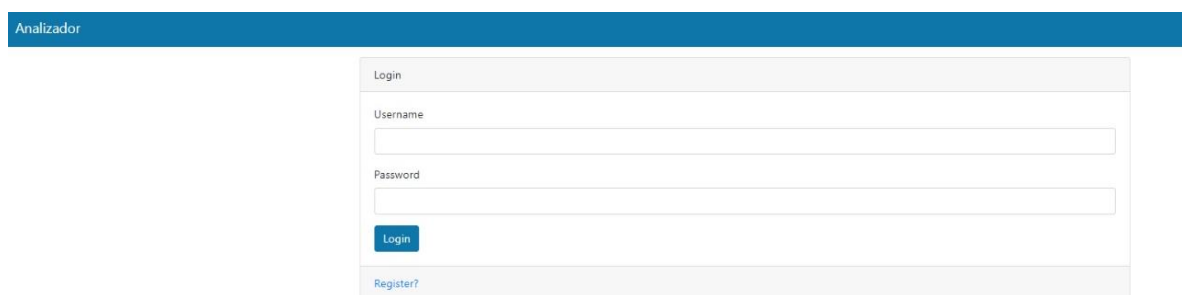


Fuente: Elaboración propia.

En la figura 30 observamos el repositorio que sirve para el almacenamiento virtual de tu proyecto, te permite guardar versiones del código a las que puedes acceder, en nuestro caso se utilizó Bitbucket, donde guardamos el sistema.

En las posteriores figuras 31, 32, 33 y 34, se podrá observar el despliegue de la fase 1 de pase a producción de la página web.

Figura 31: Login del sistema



Fuente: Elaboración Propia

En la figura 31 se puede observar la primera parte de front, una vez ejecutado el programa, pasará al login, donde se podrá ingresar solo si el user y password sean correctos, en caso de ser nuevo, tendrá un botón de registro, una vez pasado todas las validaciones de procederá a ingresar a la lista de las búsquedas anteriores.

Figura 32: Lista de Búsquedas

| # | Tipo | Fec. Búsqueda | Filtro |
|----|---------|---------------|------------------------|
| 1 | Google | 2022-10-07 | Siete sopa surquillo |
| 2 | Twitter | 2022-10-07 | Montaña siete colores |
| 3 | Google | 2022-10-07 | Tanta Jockey Plaza |
| 4 | Twitter | 2022-10-07 | Medellin |
| 5 | Google | 2022-10-07 | Madam Tusan Real Plaza |
| 6 | Twitter | 2022-10-07 | Lima |
| 7 | Google | 2022-10-07 | Hotel Bolivar Lima |
| 8 | Google | 2022-10-15 | Black Llama Hotel |
| 9 | Google | 2022-10-15 | Ayenda la paz hotel |
| 10 | Google | 2022-10-15 | Westin San Isidro |

Fuente: Elaboración propia.

En la figura 32 se observa la lista de las búsquedas que anteriormente se hayan realizado, obteniendo la información desde la base de datos, y listándolo en orden por número de búsqueda, por tipo si es por Google Places o por Twitter, seguido de la fecha de búsqueda y el filtro que se le aplica.

Figura 33: Búsqueda por Twitter

Realizar nueva búsqueda

Fuente: Criterio:

Cantidad:

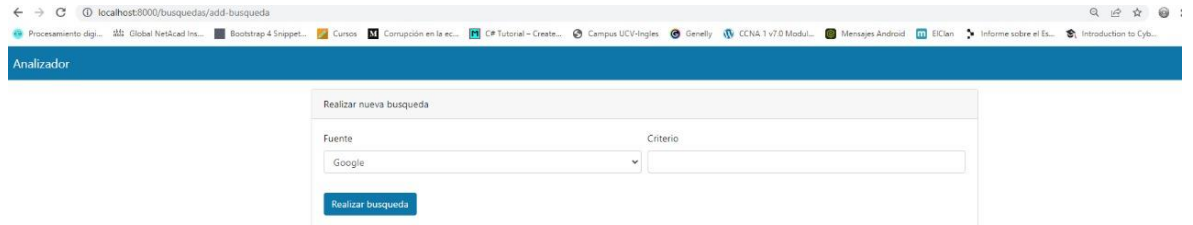
Realizar búsqueda

Fuente: Elaboración propia.

En la figura 32 observamos el proceso del web scraping para la búsqueda a través de twitter, donde definimos cuando Twists vamos a requerir, pero por

defecto obtendremos al menos 10, después de esto seleccionaremos el criterio el cual va ser la zona turística.

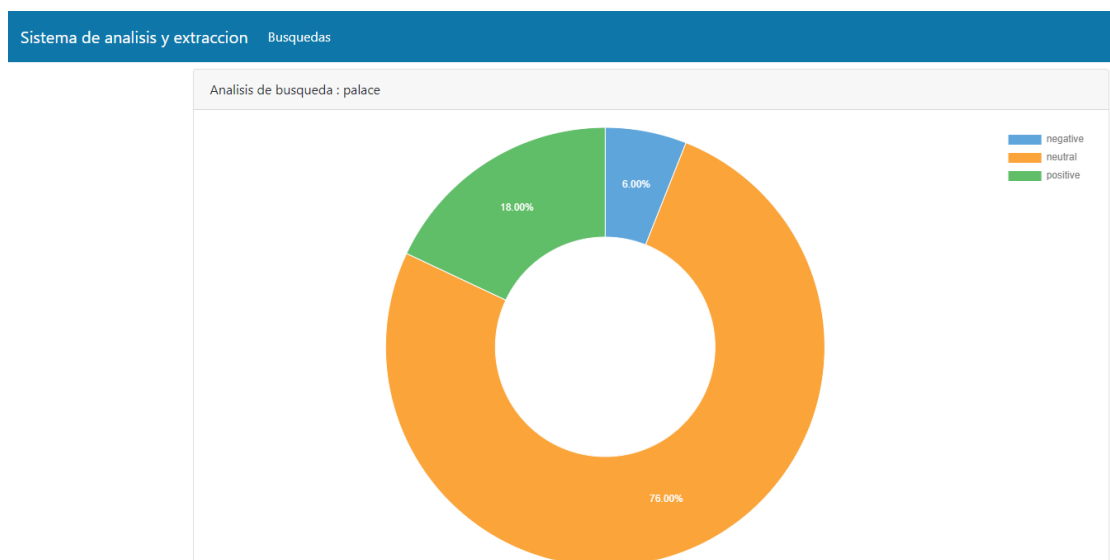
Figura 34: Búsqueda por Google



Fuente: Elaboración propia.

En la figura 34 se puede observar el proceso del web scraping para la búsqueda a través de google, después de esto seleccionaremos el criterio el cual va ser la zona turística, una vez presionado el botón se obtendrá la información de reseñas del sitio turísticos, a través de un json en la base de datos para su posterior análisis.

Figura 40: Visualización de análisis



Fuente: Elaboración propia

En la figura 40, se puede observar el análisis de web scraping a través del análisis de sentimientos con el algoritmo VADER.

Anexo 22

Tabla 31: Análisis estadístico de SPSS Post-Test

| Descriptivos | | | Estadístico | Error típ. |
|--------------|--------------------------------|-----------------|--------------|------------|
| ClienteO | Media | | 44,84 | 3,317 |
| | Intervalo de confianza para la | Límite inferior | 37,99 | |
| | media al 95% | Límite superior | 51,69 | |
| | Media recortada al 5% | | 45,51 | |
| ClienteN | Curtosis | | -,185 | ,902 |
| | Media | | 43,80 | 2,943 |
| | Intervalo de confianza para la | Límite inferior | 37,73 | |
| | media al 95% | Límite superior | 49,87 | |
| | Media recortada al 5% | | 44,19 | |
| | Mediana | | 43,00 | |
| | Varianza | | 216,500 | |
| | Desv. típ. | | 14,714 | |
| | Mínimo | | 12 | |
| | Máximo | | 69 | |
| VentasO | Media | | 13609,3101 | 1634,72504 |
| | Intervalo de confianza para la | Límite inferior | 10235,4034 | |
| | media al 95% | Límite superior | 16983,2168 | |
| | Media recortada al 5% | | 13105,2097 | |
| | Mediana | | 11711,4955 | |
| | Varianza | | 66808149,075 | |
| | Desv. típ. | | 8173,62521 | |
| | Mínimo | | 3382,00 | |
| VentasN | Máximo | | 34129,42 | |
| | Media | | 13859,3101 | 1634,72504 |
| | Intervalo de confianza para la | Límite inferior | 10485,4034 | |
| | media al 95% | Límite superior | 17233,2168 | |
| | Media recortada al 5% | | 13355,2097 | |
| | Mediana | | 11961,4955 | |
| | Varianza | | 66808149,075 | |
| | Desv. típ. | | 8173,62521 | |

Anexo 23

Tabla 32: Prueba de normalidad Post-Test

| Pruebas de normalidad | | | | | | |
|-----------------------|---------------------------------|----|-------|--------------|----|------|
| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
| | Estadístico | gl | Sig. | Estadístico | gl | Sig. |
| ClienteO | ,112 | 25 | ,200* | ,952 | 25 | ,273 |
| ClienteN | ,103 | 25 | ,200* | ,975 | 25 | ,778 |
| VentasO | ,142 | 25 | ,200* | ,935 | 25 | ,116 |
| VentasN | ,142 | 25 | ,200* | ,935 | 25 | ,116 |

*. Este es un límite inferior de la significación verdadera.

a. Corrección de la significación de Lilliefors

Anexo 24

Tabla 33: Correlación de Pearson Post-Test

| Correlaciones | | | |
|---------------|------------------------|----------|---------|
| | | ClienteO | VentasO |
| ClienteO | Correlación de Pearson | 1 | ,486* |
| | Sig. (bilateral) | | ,014 |
| | N | 25 | 25 |
| VentasO | Correlación de Pearson | ,486* | 1 |
| | Sig. (bilateral) | ,014 | |
| | N | 25 | 25 |

*. La correlación es significativa al nivel 0,05 (bilateral).

| Correlaciones | | | |
|---------------|------------------------|----------|---------|
| | | ClienteN | VentasN |
| ClienteN | Correlación de Pearson | 1 | ,482* |
| | Sig. (bilateral) | | ,015 |
| | N | 25 | 25 |
| VentasN | Correlación de Pearson | ,482* | 1 |
| | Sig. (bilateral) | ,015 | |
| | N | 25 | 25 |

*. La correlación es significativa al nivel 0,05 (bilateral).

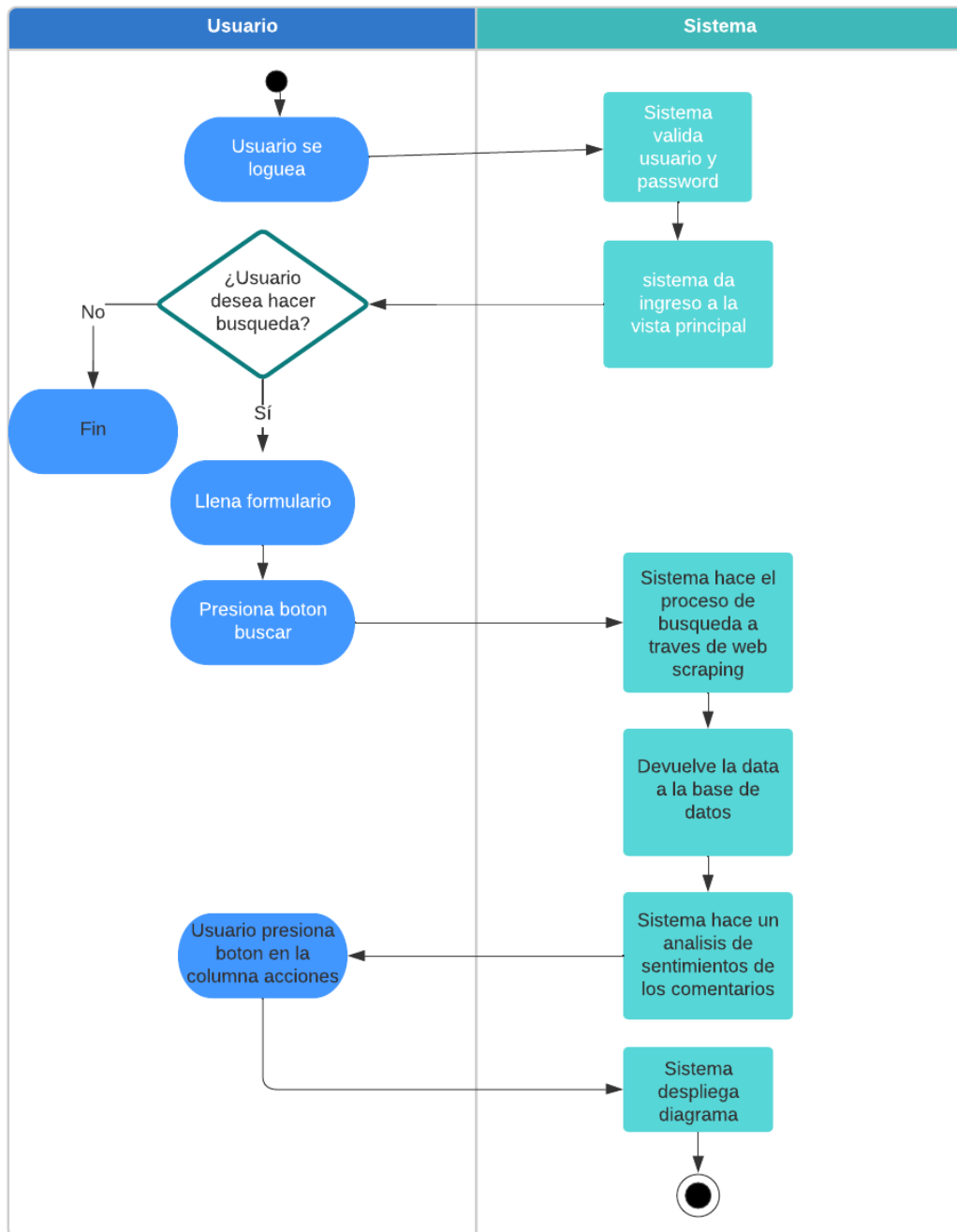
Anexo 25: Instrumento Crecimiento de cliente Post-Test
 Tabla 34: Post-Test clientes

| FICHA DE REGISTRO | | | | |
|------------------------------|---|-------------|--|-------------|
| Investigadores | Apaza Ramos, Gilberto; Ñamo Alayo, Edinson | | Tipo de prueba: Post-Test | |
| Empresa de estudio | ATIPAX GROUP | | | |
| Ubicación | Av. Arequipa 3750 San Isidro l27 Lima, Perú | | | |
| Proceso | Volumen de clientes | | | |
| Fecha de inicio | 1/09/2022 | Fecha final | 31/10/2022 | |
| Indicador | Descripción | Medida | Técnica | |
| Tasa crecimiento de clientes | Tasa crecimiento de clientes | unidad | Observación experimental | |
| Fórmula | $TCC = \frac{CMA - CME}{CME} * 100$ | | CMA: Clientes mes actual. CME: Clientes mes anterior. TCC: Tasa de crecimiento de clientes | |
| Ítem | Fechas-Abril/Mayo | CMA | CME | TCC |
| 1 | 01-Set-Oct | 109 | 63 | 0,73015873 |
| 2 | 02-Set-Oct | 73 | 53 | 0,377358491 |
| 3 | 03-Set-Oct | 82 | 57 | 0,438596491 |
| 4 | 05-Set-Oct | 90 | 64 | 0,40625 |
| 5 | 06-Set-Oct | 96 | 59 | 0,627118644 |
| 6 | 07-Set-Oct | 80 | 46 | 0,739130435 |
| 7 | 08-Set-Oct | 106 | 49 | 1,163265306 |
| 8 | 10-Set-Oct | 67 | 12 | 4,583333333 |
| 9 | 11-Set-Oct | 103 | 70 | 0,471428571 |
| 10 | 12-Set-Oct | 92 | 46 | 1 |
| 11 | 13-Set-Oct | 108 | 64 | 0,6875 |
| 12 | 14-Set-Oct | 97 | 43 | 1,255813953 |
| 13 | 15-Set-Oct | 83 | 21 | 2,952380952 |
| 14 | 17-Set-Oct | 112 | 29 | 2,862068966 |
| 15 | 19-Set-Oct | 95 | 36 | 1,638888889 |
| 16 | 20-Set-Oct | 82 | 47 | 0,744680851 |
| 17 | 21-Set-Oct | 85 | 55 | 0,545454545 |
| 18 | 22-Set-Oct | 98 | 30 | 2,266666667 |
| 19 | 24-Set-Oct | 87 | 45 | 0,933333333 |
| 20 | 25-Set-Oct | 93 | 8 | 10,625 |
| 21 | 26-Set-Oct | 92 | 40 | 1,3 |
| 22 | 27-Set-Oct | 98 | 26 | 2,769230769 |
| 23 | 28-Set-Oct | 108 | 64 | 0,6875 |
| 24 | 29-Set-Oct | 89 | 52 | 0,711538462 |
| 25 | 30-Set-Oct | 94 | 42 | 1,238095238 |
| | TOTAL | 2319 | 1121 | 4175% |

Anexo 26: Instrumento Crecimiento de ventas Post-Test
 Tabla 35: Post-Test ventas

| FICHA DE REGISTRO | | | | |
|------------------------------|---|-------------|--|-------------|
| Investigadores | Apaza Ramos, Gilberto; Ñamo Alayo, Edinson | | Tipo de prueba: Post-Test | |
| Empresa de estudio | ATIPAX GROUP | | | |
| Ubicación | Av. Arequipa 3750 San Isidro l27 Lima, Perú | | | |
| Proceso | Volumen de ventas | | | |
| Fecha de inicio | 1/09/2022 | Fecha final | 31/10/2022 | |
| Indicador | Descripción | Medida | Técnica | |
| Tasa crecimiento de clientes | Tasa crecimiento de ventas | unidad | Observación experimental | |
| Fórmula | $TCV = \frac{VMA - VME}{VME} * 100$ | | VMA: Ventas mes actual. VME: Ventas mes anterior. TCV: Tasa de crecimiento de ventas | |
| Ítem | Fechas-Abril/Mayo | VMA | VME | TCV |
| 1 | 01-Set-Oct | 22207,9794 | 16723,9794 | 0,327912386 |
| 2 | 02-Set-Oct | 50964,42 | 34129,42 | 0,493269443 |
| 3 | 03-Set-Oct | 32969,2888 | 8352,28884 | 2,947335811 |
| 4 | 05-Set-Oct | 18199,9 | 8712,9 | 1,088845275 |
| 5 | 06-Set-Oct | 111156,24 | 17832,24 | 5,233442349 |
| 6 | 07-Set-Oct | 36598,73 | 20850,73 | 0,755273317 |
| 7 | 08-Set-Oct | 36417,7575 | 23109,7575 | 0,575860651 |
| 8 | 10-Set-Oct | 31394,84 | 4557,84 | 5,888096116 |
| 9 | 11-Set-Oct | 110346 | 12795,9996 | 7,623476348 |
| 10 | 12-Set-Oct | 31335,882 | 17824,882 | 0,757985382 |
| 11 | 13-Set-Oct | 75407,7302 | 22297,7302 | 2,381856782 |
| 12 | 14-Set-Oct | 10249,7682 | 9286,76816 | 0,103695923 |
| 13 | 15-Set-Oct | 77715,35 | 11331,35 | 5,858436991 |
| 14 | 17-Set-Oct | 95090,4 | 6682,4 | 13,22997725 |
| 15 | 19-Set-Oct | 34758,2192 | 8377,2192 | 3,149135694 |
| 16 | 20-Set-Oct | 86766,6894 | 22549,6894 | 2,847799755 |
| 17 | 21-Set-Oct | 28102,0402 | 14552,0402 | 0,931140914 |
| 18 | 22-Set-Oct | 20237 | 3726 | 4,431293612 |
| 19 | 24-Set-Oct | 54407,1587 | 28031,1587 | 0,940952899 |
| 20 | 25-Set-Oct | 7796 | 3382 | 1,305144885 |
| 21 | 26-Set-Oct | 7152,72 | 6189,72 | 0,155580543 |
| 22 | 27-Set-Oct | 22575,7479 | 8354,74788 | 1,702145918 |
| 23 | 28-Set-Oct | 28995,0988 | 15485,0988 | 0,872451648 |
| 24 | 29-Set-Oct | 25221,4955 | 11711,4955 | 1,153567452 |
| 25 | 30-Set-Oct | 16513,2973 | 3385,29726 | 3,877946009 |
| | TOTAL | 1072579,75 | 340232,753 | 6863% |

Anexo 28: Flujo del sistema





UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, JUAN BRUES LEE CHUMPE AGESTO, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, asesor de Tesis titulada: "WEB SCRAPING Y ANÁLISIS DE SENTIMIENTOS PARA CAPTAR LA VALORACIÓN DEL CLIENTE HACIA LOS SERVICIOS DE UNA EMPRESA DE TURISMO", cuyos autores son ÑAMO ALAYO EDINSON MANUEL, APAZA RAMOS GILBERTO ROMER, constato que la investigación tiene un índice de similitud de 17.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 03 de Diciembre del 2022

| Apellidos y Nombres del Asesor: | Firma |
|---|--|
| JUAN BRUES LEE CHUMPE AGESTO DNI: 44824114 ORCID: 0000-0001-7466-9872 | Firmado electrónicamente por: JCHUMPEA el 03-12- 2022 17:05:41 |

Código documento Trilce: TRI - 0470035