



UNIVERSIDAD CÉSAR VALLEJO

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

Data mart para la medición de morbilidad y casos de tbc
en el centro de salud pueblo nuevo

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTOR:

Munayco Jacobo, Jeysson José (orcid.org/0000-0002-8271-5282)

ASESOR:

Dr. Ormeño Rojas, Robert Eduardo (orcid.org/0000-0002-8104-9310)

LÍNEA DE INVESTIGACIÓN:

Auditoría de Sistemas y Seguridad de la Información

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Apoyo a la reducción de brechas y carencias en la educación en
todos sus niveles

**LIMA – PERÚ
2021**

DEDICATORIA

Dedico primeramente a Dios y también a mis padres que con sus consejos contribuyeron en mi día a día para superarme como persona y en especial a mi padre que me inculco nociones de mi carrera para mi crecimiento profesional.

AGRADECIMIENTO

Agradezco primeramente a Dios por guiarme y brindarme las oportunidades de mejora en todo aspecto, personal, profesional, económico, entre otros, por lo consiguiente agradezco a mi padre, el cual fue un pilar fundamental en el aprendizaje de mi carrera.

Por último, agradezco a mi asesor Ing. Ormeño Rojas Robert Eduardo por el tiempo, ayuda y dedicación en este trabajo de investigación.

ÍNDICE DE CONTENIDOS

Carátula	i
Dedicatoria	ii
Agradecimiento	iii
Índice de contenidos	iv
Índice de tablas	vi
Índice de figuras	vii
RESUMEN	ix
ABSTRACT	xi
I. INTRODUCCIÓN	1
1.1. Realidad problemática.....	2
1.2. Trabajos previos.....	3
1.3 Teorías relacionadas al tema	4
1.4 FORMULACIÓN DEL PROBLEMA.....	19
1.5 JUSTIFICACIÓN DEL ESTUDIO	19
1.6 HIPÓTESIS	20
1.7 OBJETIVOS	21
II. METODOLOGÍA	21
2.1 DISEÑO DE INVESTIGACIÓN	21
2.2 VARIABLES, OPERACIONALIZACIÓN.....	22
2.3 MÉTODOS DE ANÁLISIS DE DATOS.....	25
2.4 TÉCNICAS: INSTRUMENTOS DE RECOLECCIÓN DE DATOS, VALIDEZ Y CONFIABILIDAD.....	26
2.5 MÉTODOS DE ANÁLISIS DE DATOS.....	31
2.6 ASPECTOS ÉTICOS	35
III. RESULTADOS.....	36
3.1 ANÁLISIS DESCRIPTIVO.....	36

3.2	Análisis Inferencial	38
3.3	Prueba de Hipótesis	42
IV.	DISCUSIÓN	49
V.	CONCLUSIONES	50
VI.	RECOMENDACIONES	51
VI.	Referencias.....	52
ANEXOS		

ÍNDICE DE TABLAS

Tabla N° 1 Factores que afectan a la Data Warehouse	15
Tabla N° 2 Operacionalización de Variables	23
Tabla N° 3 Operacionalización de Indicadores.....	24
Tabla N° 4 Tabla de Expertos	27
Tabla N° 5 Módulo de Coeficientes	29
Tabla N° 6 Estadístico Descriptivo	36
Tabla N° 7 Estadísticos Descriptivos.....	37
Tabla N° 8 Prueba de Normalidad del indicador Productividad de Reportes antes y después del Datamart	39
Tabla N° 9 Prueba de Normalidad del indicador Nivel de Eficiencia antes y después del Datamart.....	41

ÍNDICE DE FIGURAS

Figura N° 1 Base del libro de emergencias	2
Figura N° 2 Data Warehouse	5
Figura N° 3 Gráficos de indicadores.....	5
Figura N° 4 Modelo Copo de Nieve.....	8
Figura N° 5 Cubo OLAP	8
Figura N° 6 Modelo de Estrella.....	9
Figura N° 7 Proceso ETL	10
Figura N° 8 Grafica del proceso ETL.....	11
Figura N° 9 Arquitectura Inmon	13
Figura N° 10 Estructura del Data Warehouse	13
Figura N° 11 Arquitectura Kimball	14
Figura N° 12 Esquema de arquitectura Ralph Kimball	15
Figura N° 13 Coeficiente de Correlación de Pearson.....	28
Figura N° 14 Indicador de Productividad PRE-TEST y RETEST	30
Figura N° 15 Indicador: Nivel de Eficiencia PRE TEST y RETEST	30
Figura N° 16 Contraste de Hipótesis Alfa	32
Figura N° 17 La Varianza	33
Figura N° 18 Desviación Estándar	33
Figura N° 19 Desviación estándar de la población.....	34
Figura N° 20 Fórmula T-Student	34
Figura N° 21 Tabla T-student	35
Figura N° 22 Productividad de Reportes	37
Figura N° 23 Nivel de Eficiencia	38
Figura N° 24 Prueba de normalidad del indicador Productividad de Reportes antes del Datamart.....	40
Figura N° 25 Prueba de normalidad del indicador Productividad de Reportes antes del Datamart.....	40
Figura N° 26 Prueba de normalidad del indicador Nivel de Eficiencia antes del datamart.....	41

Figura N° 27 Prueba de normalidad del indicador Nivel de Eficiencia antes del datamart	42
Figura N° 28 Prueba de normalidad del indicador Productividad de Reportes antes del datamart	43
Figura N° 29 Prueba de normalidad del indicador Productividad de Reportes antes del datamart	44
Figura N° 30 Productividad de Reportes	44
Figura N° 31 Prueba de normalidad de visualización del grado de eficiencia antes del data mart	46
Figura N° 32 Prueba de normalidad de visualización del grado de eficiencia antes del data mart	47
Figura N° 33 Nivel de eficiencia	47

RESUMEN

El presente, es una tesis de investigación a detallar el desarrollo de un datamart en el centro de salud de pueblo nuevo, dentro del área de estadística, la cual se encarga de brindar información confidencial en cuanto a las metas, nivel de productividad y otros reportes que son solicitados por el personal o encargado de turno por estrategia. El fin de la investigación es determinar el nivel de productividad de la cantidad de reportes, acompañado de la eficiencia de los colaboradores que se encuentran laborando en el área de estadística. En el área ya mencionada siempre suelen pedir información en fechas que no se ha está establecido para poder entregar algún informe que requieran, la cual para la realizar un reporte se requiere de un informe para poder proceder con lo solicitado, en la cual se tiene que seguir un proceso de exportación de data, procesamiento de la misma y verificación de campos, para poder entregar un reporte adecuado.

El datamart se realizará con el propósito de realizar todos los reportes que se pueda generar, ya sean ranking de morbilidad, verificar la curva de información en cuanto a la cantidad de pacientes con TBC positivo, el nivel de productividad de los colaboradores que se encuentra laborando en la institución. Para realizar cuya implementación del datamart, utilizaremos el proceso del ETL, que significa Extracción, Transformación y carga de datos, donde nos permitirá realizar mejor el proceso de la información de los datos y analizar de manera rápida los datos almacenados.

En este proyecto utilizaremos la metodología de Kimball donde nos representara el ciclo de vida que para el desarrollo del proyecto. Actualmente el software que utilizan en el área de estadísticas solo es de ingreso de datos (ingresar, eliminar, modificar, buscar), pero no cuenta con una pestaña donde puedan realizar a detalle el tipo de reporte que se pueda solicitar, incluso cuenta con la exportación de un backup, teniendo en cuenta que para cualquier información de primera línea su proceso de obtención de datos es de manera manual, y mediante un data mart optimizaremos este proceso, una vez que obtengamos lo requerimientos para la elaboración procederemos a realizar el datamart de una manera adecuada.

En el presente trabajo se aplicó el tipo y diseño Pre-experimental, para la finalizar la implementación del Datamart en el centro de salud de pueblo nuevo se quiere obtener como resultado la medición de la de la mortalidad y los casos TBC en dicho centro, para que se puedan obtener los reportes que se puedan solicitar de manera eficiente.

Palabras clave: Datamart, indicadores, Power BI, Cie10, Metodología, ETL, herramienta de extracción, transformación, carga de datos, ranking, optimizar.

ABSTRACT

This research work will detail the development of a datamart for the new town health center, for the statistics area, which is in charge of providing confidential information regarding goals, productivity level and other reports that are requested by the personnel or shift manager per strategy. The objective of this research is to determine the productivity level of the number of reports, accompanied by the efficiency of the workers who are working in the statistics area. In the aforementioned area, they always ask for information on dates that have not been established to be able to deliver a report that they require, which to make a report requires a report to proceed with the request, in which you have to Follow a data export process, data processing and field verification, in order to deliver an adequate report.

The datamart will be carried out with the purpose of making all the reports that can be generated, whether they are morbidity ranking, verifying the information curve regarding the number of patients with positive TB, the level of productivity of the personnel who are working in the institution. To implement the datamart, we will use the ETL process, which means Data Extraction, Transformation and Loading, where it will allow us to better process the data information and quickly analyze the stored data.

In this project we will use the Kimball methodology where we will represent the life cycle for the development of the project. Currently the software they use in the statistics area is only for data entry (enter, delete, modify, search), but it does not have a tab where they can make in detail the type of report that can be requested, it even has the Exporting a backup, taking into account that for any first-line information the data collection process is manual, and through a data mart we will optimize this process, once we obtain the requirements for the elaboration we will proceed to carry out the datamart of a proper way.

In this thesis, the Pre-experimental type and design was applied, to finalize the implementation of the Datamart in the New Town health center, it is desired to obtain as a result the measurement of mortality and TB cases in said center, to that the reports that can be requested can be obtained efficiently.

Keywords: Datamart, indicators, Power BI, Cie10, Methodology, ETL, extraction tool, transformation, data loading, ranking, optimize.

I. INTRODUCCIÓN

En la actualidad las empresas del área salud manejan mucho la herramienta Microsoft Excel, la cual es muy buena en la visualización de informes y poder generar reportes como son los indicadores para su estrategia del POI, pero lo que no tienen en cuenta es que pueden mejorar implementando nuevas herramientas que les ayude a la obtención de los datos, tanto sea como los motores de base de datos SQL y Access que son dos herramientas la cual tienen la capacidad de almacenar más datos sin causar ningún congestionamiento del sistema, tal y como sucede con el Excel. Otro de los grandes problemas que se presenta es la implementación de sus sistemas es que solo te permiten realizar los procesos básicos de un sistema, la cual es un inconveniente ya que un sistema debe de contar con una pestaña llamada reportes, para ello es que existe la pestaña exportar ya que la data se trabaja de manera manual y no cuenta con un proceso automatizado como debía realizarse. Lo único rescatable es la opción de poder importar y realizar un backup, en caso de emergencias el ordenador sufra una caída de software.

En la actualidad carecen de información rápida y precisa, ya que el sistema que manejan solo les permite exportar data y no les genera un reporte muy a detalle, solo genera un reporte general, donde no hay mucho que analizar.

Lo único que se puede mostrar de manera instantánea son las cantidad de atenciones que se han realizado dentro de los meses, tanto como la cantidad de atenciones pero separado por grupo etario, la cuales son datos comunes y no precisos para llevar un dato como indicador de la información o para ver si hay un punto de quiebre en la parte de sus pacientes con TBC, así verifican que hay que dar tratamiento a las persona y verifican de igual manera si la población que estaba con contagio a disminuido en un número considerable para poder mostrar en un gráfico de línea la mejora que se ha tenido en cuanto a la atención y a la recuperación de pacientes anteriores que han sido diagnosticados con la enfermedad del TBC.

El datamart que implementaremos nos ayudara a realizar una serie de reportes más a detalle y así poder tener una información donde puedas analizar y ver tus puntos altos de niveles de pacientes positivos en TBC.

Dentro del proyecto de tesis, analizaremos de manera correcta la data la cual nos ayudara a implementar mejor el datamart que queremos realizar para la medición de la morbilidad y los casos TBC, ya sea de manera mensual, anual o trimestral como parte de un indicador para la correcta información de los datos para el MINSA.

1.1. Realidad problemática

En estos tiempos del siglo XXI la tecnología va avanzando y hay que implementar nuevos procesos para automatizar algunas funciones, en este caso se realizara un datamart para poder acelerar el proceso de uno de sus datos informativos que deben proporcionar para el personal que lo requiera.

Por ello uno de los grande problemas en el centro de salud es la desactualización de algunos sistemas, la cual están implementados de una manera la cual no ayuda a poder optimizar algún proceso que se requiera, más en los reportes que se tienen que brindar, cuentan con una interfaz inadecuada no acta para el usuario, con ineficiencia muchas veces de sus actualizaciones, en mayorías de veces el único cambio que se ha observado es en el cambio del año, pero la interfaz y los errores siguen siendo los mismos, de igual manera como generador de información a través de un software para realizar filtros de código lo realizan mediante el open office la cual se refleja que es una herramienta que no es de mucha ayuda, para ello sería recomendable en otra instancia el generador de información (software) comor el Excel para la filtración de códigos y la creación de algunos reportes con datos de etiquetas y gráficos que harán que el reporte tenga un vista más de impacto visual y una buena presentación.

En la Figura N° 01, se muestra algunos datos que forman la base de datos del libro de emergencias.

Figura N° 1 Base del libro de emergencias

RENIPRESS	E_UBIG	FECATE	HORATE	NUMHC	DOC_IDEN	ETNIA	FINANCIA	SEXO	EDAD
0000003419	110207	01/01/2018	08:30	001	179030975	80	02	1	2
0000003419	110207	01/01/2018	08:40	002	176130323	80	01	2	23
0000003419	110207	01/01/2018	08:52	003	121839151	80	02	2	75
0000003419	110207	01/01/2018	09:15	005	147925129	80	01	2	25
0000003419	110207	01/01/2018	09:35	006	144666693	80	01	1	29
0000003419	110207	01/01/2018	09:40	007	121856820	80	01	2	54
0000003419	110207	01/01/2018	10:00	008	144287231	80	01	1	30
0000003419	110207	01/01/2018	10:05	009	148067825	80	01	2	25
0000003419	110207	01/01/2018	10:30	010	179981286	80	02	2	1
0000003419	110207	01/01/2018	11:10	011	170074711	80	02	2	22
0000003419	110207	01/01/2018	11:45	012	161759323	80	01	1	8
0000003419	110207	01/01/2018	11:45	013	121844710	80	01	2	46
0000003419	110207	01/01/2018	12:15	014	171641222	80	02	1	15

Fuente: Centro Salud Pueblo Nuevo

Visualizamos aquí la base principal con la cual trabajaremos para la creación de un datamart.

Finalmente, se llevará a cabo dicho proyecto planteado y ver mejoras en la entrega de información.

1.2. Trabajos previos

Se ha revisado publicaciones e investigaciones, encontrándose los siguientes antecedentes bibliográficos:

Guativa (2018), en su trabajo de la I.U. Politécnico Grancolombiano, en su propuesta de Datamart, aconsejan que esta herramienta no es un gasto sino una herramienta dónde ayudará a la consolidación de datos, optimización y análisis de la organización. En este proyecto se vio como un datamart ayuda mucho en una empresa en la optimización de los reportes, tomaron el área de cobranzas, en la cual ayudo mucho en el análisis de la información. (Tesis de titulación, Institución Universitaria Politécnico GranColombiano. Repositorio Institucional de la Universidad de Guayaquil, <http://repositorio.ug.edu.ec/handle/redug/30763>)

Bustos (2018), en su proyecto Implementación de la solución BI para la empresa Otecel con su control e indicadores de la Esc. Politécnica del Ejército. Como podemos informarnos este proyecto de tesis se realizó un datamart, donde mencionan que utilizaron la metodología Hefesto, la cual les permitió mejorar mucho y entregar los reportes más dinámicos, además que le permite tener indicadores en tiempo real. (Tesis de titulación, Escuela Politécnica del Ejército, Repositorio Dspace, <http://repositorio.espe.edu.ec/xmlui/handle/21000/6305>)

Hernandez (2008), en su tesis construcción y diseño de un datamart para la industria, de U. de Chile. Tiene como objetivo principal contruir una plataforma basada en tecnología y data. En este proyecto se logra ver los resultados obtenidos mediante la implementación de un datamart, pero con resultados actualizados en la nube, mediante una página web. Para que la persona que requiera un reporte y

no se encuentre en la oficina pueda acceder mediante un ordenador y poder visualizar el reporte o la información que se pueda requerir. (Tesis de Titulación, Universidad de Chile, Repositorio Académico de la universidad de Chile, <https://repositorio.uchile.cl/handle/2250/101968>)

En trabajos nacionales obtenemos a Rosales (2009), de la universidad privada católica en su trabajo para la toma de decisiones con datamart para evaluaciones en el sector salud. Su objetivo es implementar esta datamart en las áreas de nutrición y alimentación saludable. Gracias al datamart creado por la autora Carmen Rosales se ve como influyó mucho en la exportación de uno de los reportes, ya que gracias a ello cambiaron los reportes de ser a estáticos a ser unos reportes dinámicos y más sólidos en información. (Tesis de titulación, Universidad Católica del Perú, Repositorio de la PUCP, <https://repositorio.pucp.edu.pe/index/handle/123456789/148982>)

Torres (2016), en su propuesta de BI para el hospital Santa Rosa de la U. Wiener. Para este proyecto vemos que el datamart nos ayuda en las decisiones en cuanto a los reportes para programas de presupuesto. (Tesis de titulación, Universidad Norbert Wiener, Repositorio Institucional Norbert Wiener, <https://repositorio.uwiener.edu.pe/handle/20.500.13053/481?locale-attribute=en>)

Ramos (2018), en su tesis implementando un datamart en Sedachimbote de la U. Vallejo, ve el área de Logística. En este proyecto se ve reflejado con el data mart ayudo en varios procesos de tomar decisiones llegando así a disminuir el costo en el tiempo de elaborar los reportes y un nivel de satisfacción tanto con la gerencia como con el área de logística. (Tesis de Titulación, Universidad Cesar Vallejo, Repositorio UCV, <https://repositorio.ucv.edu.pe/handle/20.500.12692/29061>)

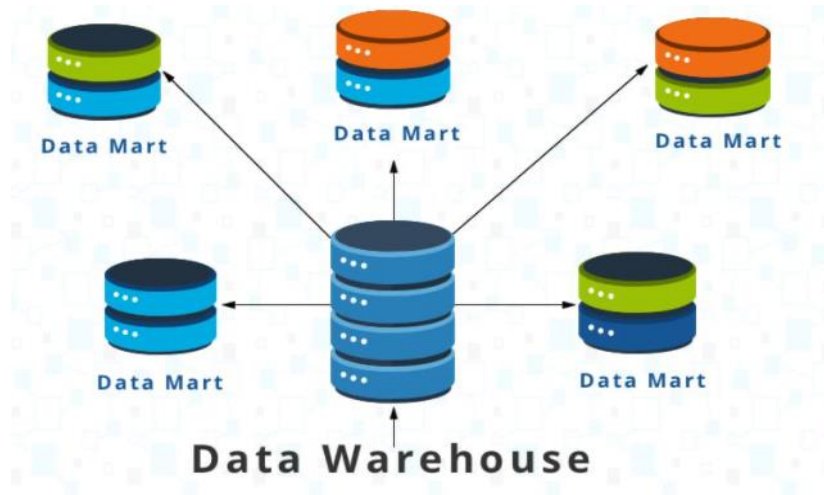
1.3 Teorías relacionadas al tema

1.3.1 Datamart

“Para la Universidad ESAN, en un artículo publicado en el 2020, Los data marts vienen hacer almacenes de datos que se dirigen a áreas específicas, como: B.

Ventas, talento u otras áreas dentro de su organización. Por este motivo, también se denomina base de información del departamento. "

Figura N° 2 Data Warehouse



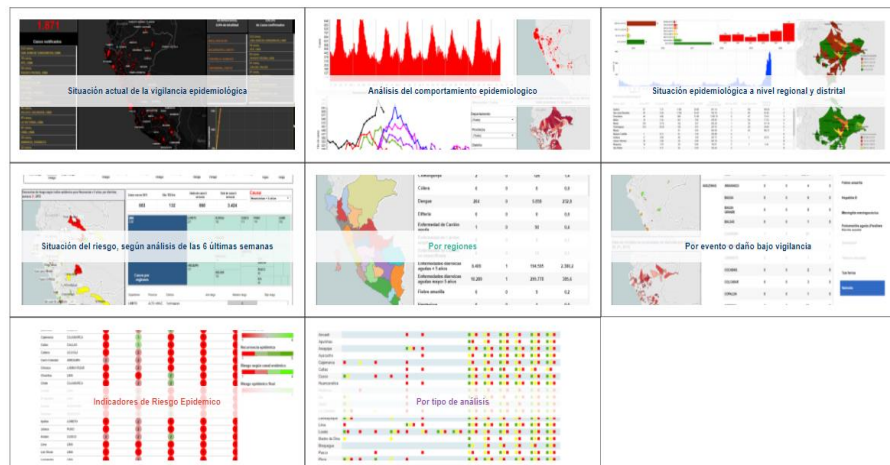
Fuente: Taled

1.3.2 Medición de morbilidad y casos de TBC

La medición de la morbilidad se refiere al número de individuos que son atendidos por emergencias, la cual se les detecta un diagnóstico y dependiendo de la gravedad son atendidos en el establecimiento de la localidad, pero si su diagnóstico es reservado o se complica se le deriva a un hospital con un ambiente más implementado. De la misma manera hablaremos de los casos de TBC, esos diagnósticos están detectados dentro de la localidad la cual si se detecta el diagnóstico en una persona, este lleva un tratamiento de recuperación; la detección se hace mediante pruebas de esputo la cual se le menciona si cuenta con la enfermedad y en qué grado de contagio te encuentras.

Figura N° 3 Gráficos de indicadores

Fuente: Minsa



1.3.3 Tipos de datamart

Hay tres tipos de datamart. La diferencia entre ellos está en cómo se alimentan, es decir, como se hallan los datos y como se insertan en el datamart, en la cual mencionaremos los tres tipos de datamart a continuación:

1. Datamart dependiente: Nace a partir de una data warehouse central, identificando un subconjunto de datos, permitiendo lograr un mejor rendimiento, ya que anteriormente han sido sometidos al proceso de ETL (Extract, Transform y Load), además en este tipo de datamart se incrementa la seguridad, ya que cada departamento cuenta con el control sobre sus datos, y tienen una eficiencia en el seguimiento de los KPI, este caso se le conoce como enfoque Top Down, las cuales se pueden realizar consultas de dos tipos pueden ser de visión lógica o puede ser mediante un subconjunto físico de almacén de datos:

A) Visión Lógica: Se visualiza a través de una tabla, la cual cuenta con una vista virtual separada lógicamente del almacén de datos que se tenga.

B) Subconjunto físico: Es la extracción de una data que comprende la base de datos que están separadas físicamente en el almacén de datos.

2. Datamart independiente: Es un sistema autónomo, que se concentra solamente en la única área del negocio, los datos se extraen de fuentes internas o externas, la cual luego se procesan y luego se cargan al datamart, donde se almacenan hasta que son necesario para algún análisis. Si es

necesario para la organización, los datos se integran a una data warehouse, en la cual se le conoce como enfoque Bottom Up.

- 3. Datamart Híbrido:** Este sistema fusiona datos de un almacén de datos existentes con otros sistemas de fuentes operativas. Combina la velocidad con el énfasis del usuario final dentro de un enfoque llamado Top Down con la ventaja de tener la integración corporativa Bottom up como un buen método.

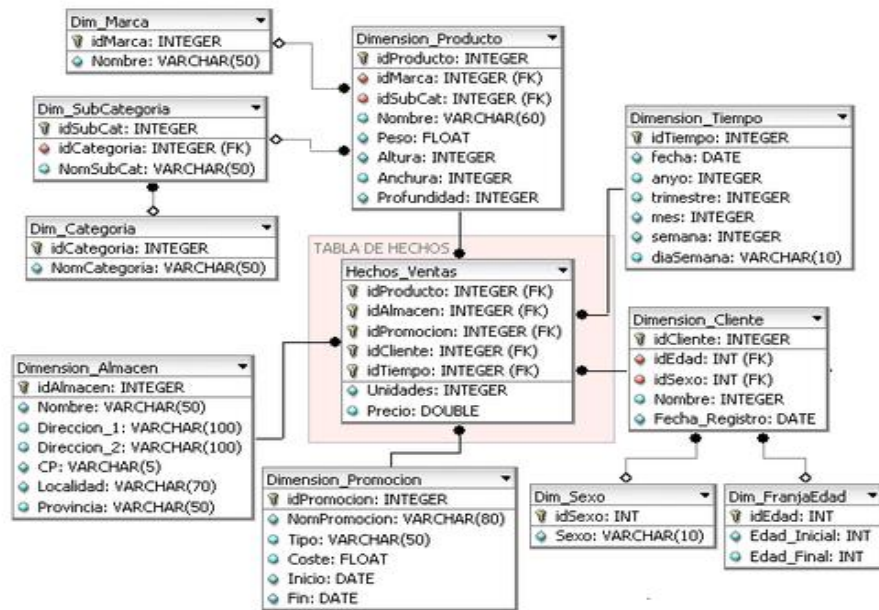
1.3.4 Estructura de un datamart

Un datamart es igual a un almacén de datos, en las cuales se pueden dar en dos tipos de esquema, la cual son los siguientes:

1. Esquema copo de nieve

Un esquema de copo de nieve es un conjunto de tablas normalizadas para una sola entidad de negocio. Sirve para ayudar a reducir la repetición de datos y proteger su integridad. En este tipo de esquema se tiene una tabla central en la que se guardan las medidas que se quiere analizar, en la cual todo ello proviene de un elemento clave que se llama el Cubo OLAP, la cual estos cubos tienen la capacidad de almacenar grandes volúmenes de datos, al normalizar la base de datos para optimizar el espacio es donde obtenemos el esquema de copo de nieve.

Figura N° 4 Modelo Copo de Nieve



Fuente: Wikipedia

2. Cubo OLAP:

“Se construyen agregando las dimensiones y medidas requeridas para cada cubo relacional. La forma de crear, operar y mantener un cubo OLAP depende en gran medida de las herramientas que utilice.”

Figura N° 5 Cubo OLAP



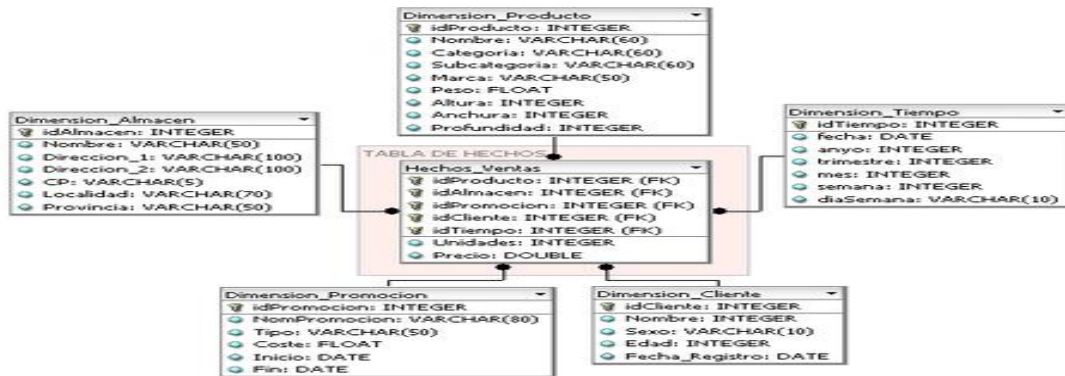
Fuente: Freddy Abad L.

3. Esquema de Estrella

Es un esquema la cual consta de cuadros que pueden ser varios, en las que se indican cuadros de extensiones en una base de datos relacional, una de las mayores ventajas es que cada vez se requieren menos combinaciones entre las tablas al escribir las consultas, esto simplifica el proceso de solicitud de ETL.

También la herramienta ETL es la que nos permite consolidar los distintos datos que poseemos en un mismo almacenamiento, la cual posteriormente se ejecutarán los análisis la cual nos permita extraer datos analíticos.

Figura N° 6 Modelo de Estrella



Fuente: Wikipedia

En este punto procederemos a explicar el proceso del ETL:

- A) El proceso de ETL se encarga de extraer datos de múltiples sistemas para luego transfórmalos y finalmente cargarlos en una base de datos de destino.

Etapas ETL

Extracción: La información es extraída de cualquier fuente de datos, la cual cuenta con un intervalo de tiempo dependiendo de los requisitos y necesidades de la empresa.

Transformación: Se estudian los datos extraídos para transfórmalos en un formato deseado por la empresa, pero manteniendo su integridad, en la cual se llevarán a cabo operaciones como la validación de los datos, cálculos, codificación, filtrado y eliminación de datos duplicados.

Carga: Se suben los datos en un formato sólido para el almacén de destino, en la cual los datos estarán listos para ser explotados.

Figura N° 7 Proceso ETL



Fuente: MediaCloud

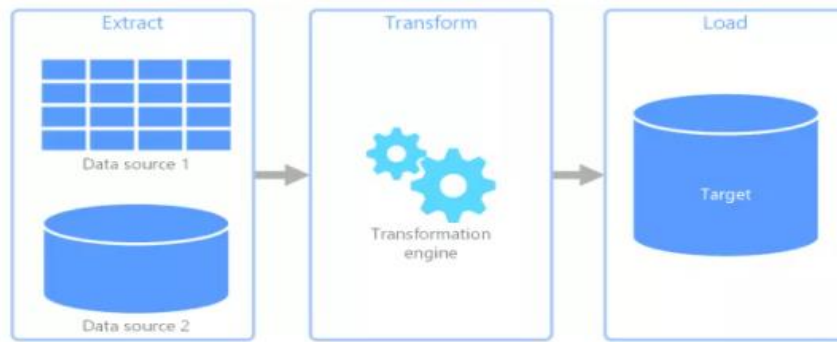
Características

- Complejo
- Continuo
- Critico

Beneficios ETL

- Extrae y consolida datos de múltiples fuentes.
- Proporciona un contexto histórico sobre la empresa.
- Facilita el análisis y reporte de datos de una manera sencilla y eficiente, con una representación visual.
- Aumento de la productividad haciéndolo fácil el trabajo en equipo.
- Permite adaptarse a la evolución de las nuevas tecnologías.
- Permite tomar decisiones de una manera estratégica.

Figura N° 8 Grafica del proceso ETL



Fuente: InetSoft

1.3.5 Características de un Datamart

- **Uso:** Asistente en la toma de decisión de negocio.
- **Objetivo:** Se utiliza principalmente en una división de negocios a nivel de departamentos.
- **Diseño:** fácil de uso.
- **Modelado dimensional:** construido mediante un enfoque en un modelo dimensional.
- **Gestión de datos:** de mucha eficiencia y fácil de utilizar, diseñar e implementar, utilizan cantidades bajas de información.
- **Foco:** dirigida la atención a un área determinada y a nivel de departamentos.
- **Tipo de datos:** Limitados y cortos.
- **Normalización:** No hay preferencia de estructura.
- **Área Temática:** Cuenta con una sola área temática.
- **Almacenamiento de datos:** diseño de esquema para optimizar el funcionamiento de la capa de acceso.
- **Tipo de datos:** Son estructuras de datos las cuales se encargan de las necesidades de informe.
- **Valor de los datos:** Datos transaccionales agrupados directamente desde una data Warehouse.
- **Alcance:** Contiene datos específicos de una empresa, la cual cuenta con uso limitado.
- **Fuente:** Proviene de muy pocas fuentes.

- **Tamaño:** menor a 100 GB.
- **Tiempo de Implementación:** unos pocos meses.

1.3.6 Pasos para implementar un Datamart

- **Diseño:** En esta fase se recopilan los requisitos, las cuales se identifican las fuentes de datos, se selecciona el subconjunto apropiado y se diseña la estructura lógica y física del datamart.
- **Implementación:** Se crea una base de datos en físico y su mapa de almacenamiento asociadas con el datamart; se crean los objetos de los esquemas tales como son las tablas o índices definidos en la parte del diseño y finalmente se determina la mejor manera de configurar las tablas y estructuras para una ejecución óptima de las consultas.
- **Población:** Se inicia el proceso de ETL, la cual también se lleva a cabo la creación y almacenamiento de metadatos.
- **Acceso:** En esta fase se utilizan los datos para realizar consultas, analizarlos, crear informes, cuadros estadísticos y gráficos.
- **Gestión:** Incluye una serie de tareas tales como: Proteja el acceso a la información, gestione el aumento de los datos, ajuste su sistema para mejorar el rendimiento y garantice la disponibilidad de la información en caso de que se produzca un cierre del sistema.

1.3.7 Metodologías para un Datamart

- **Enfoque de Inmon:** La premisa del enfoque de Inmon es que la información se almacene en capacidad máxima de información detallada, la cual permanece sin variación y sin volatilidad. Si ocurriera cambios en la información se registran y no permiten modificación ni eliminación.

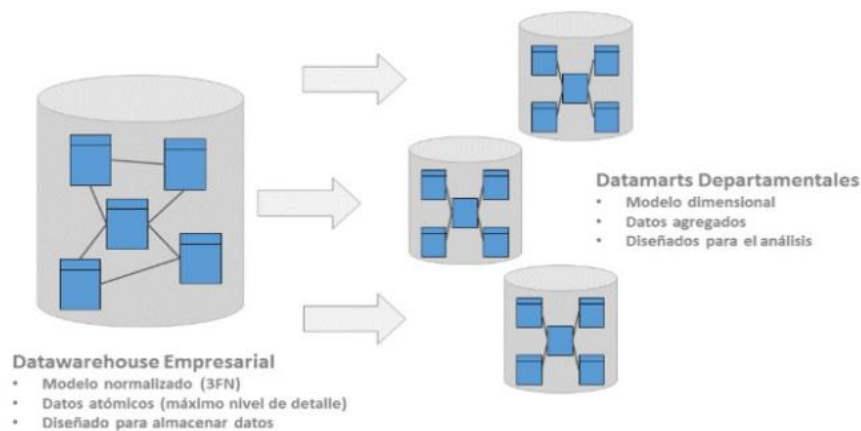
Figura N° 9 Arquitectura Inmon



Fuente: BI Geek

Este encuadre de Inmon a menudo se conoce como el método de trabajo 'Top-Down', dando un enfoque general de la empresa con subdivisiones para la información de los departamentos. Por lo tanto, con esta estructura, todos los datos de Data Marts se encuentran unidos en el data Warehouse.

Figura N° 10 Estructura del Data Warehouse

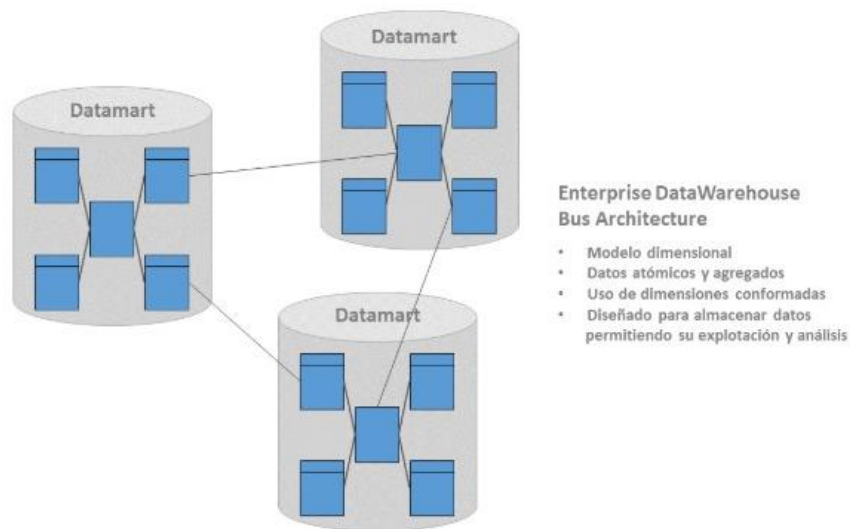


Fuente: BI Geek

En cuanto al ordenamiento interno del Data Warehouse, para Inmon lo esencial es que el cliché de datos esté formado en tercera manera deudo. Por explicar brevemente lo que esto significa, la estimación de combinación consiste en iniciar un inventario de reglas o normas a la hora de establecer, se coteja entre los diferentes objetos interiormente de la almohadilla de datos. Con naciente proceso de fusión se consiguen muchas posesiones, como rehuir el pleonasma de los datos,

sustentar su dignidad referencial, suministrar el sostenimiento del proscenio y debilitar la masa de la colchoneta de datos. Sin requisa, a semejanza de los Data Warehouse desnormalizados, las consultas exigen la tarea de queries en extremo más complejas, lo que dificulta la grafología directa de la noticia y el uso de las herramientas de reporting. De ahí, la urgencia de educar los Data Marts que, como ya comenté, están basados en modelos dimensionales de astro o copo de ampo, diseños desde luego explotables por estas herramientas de explicación de datos.

Figura N° 11 Arquitectura Kimball



Fuente: BI Geek

En cuanto a la BASE sobre la granularidad, el data warehouse presenta los informes agregados de manera consultiva y/o informes que puede generar, Kimball mantiene su teoría de la importancia que las agregaciones deben estar completadas con informes detallados ampliamente.

Los usuarios al ingresar al sistema realizan preguntas irregulares o imprevistos sobre los negocios, por lo que el data warehouse debe dar respuesta creíbles a todas las inquietudes del usuario, por tanto el usuario pueda explorar los informes y navegar por cuadros secuenciales desde datos incorporados hasta información disociadas.

A este tipo de diseño Kimball lo llama "Data Warehouse Bus Architecture" cuya secuencia de construcción es la siguiente:

1. Caracterización detallada de la secuencia del negocio a investigar.
2. Concentración de datos desde el área superior a la inferior.
3. Dimensiones y caracteres.
4. La tipificación de los hechos o métricas.

Figura N° 12 Esquema de arquitectura Ralph Kimball



Fuente: BI Geek

Comparativa

Como se puede apreciar, Kimball e Inmon concuerdan en la importancia de un sistema que almacene informes integrados de almacenamiento la necesidad de establecer un sistema de almacenamiento de datos integrado e invariable garantizando la exploración de los informes y estable que garantice la explotación de la información, facilitando respuestas a todas las interrogaciones de negocio que broten.

Tabla N° 1 Factores que afectan a la Data Warehouse

	Inmon	Kimball
Presupuesto	Coste inicial alto	Coste inicial bajo
Plazos	Requiere más tiempo de desarrollo	Tiempo de desarrollo inferior
Expertise	Equipo con especialización alta	Equipo con especialización media
Alcance	Toda la compañía	Departamentos individuales
Mantenimiento	Fácil Mantenimiento	Mantenimiento más complejo

Fuente: Elaboración Propia

Aspectos a analizar

- Presupuesto para el proyecto disponible para la construcción del Datawarehouse.
- Alcance del Datawarehouse, para obtener los datos de toda la empresa o de determinados departamentos o áreas.
- Complejidad de las labores de mantenimiento

1.3.8 Ventajas de un Datamart

- **Acceso eficiente:** Es la respuesta para acceder a un grupo concreto de datos a efectos de una inteligencia de negocios que permite tener mejor tiempo.
- **Los almacenes de datos y su económica alternativa:** Se da ya que los grupos de datos son más pequeños.
- **Mejora la productividad del almacén de datos:** Los datamart híbridos y/o dependientes pueden mejorar la productividad de un almacenamiento de datos.
- **Mantenimiento de datos:** Los datos son dueños de las áreas a la cual se implementará.
- **Configuración sencilla:** No hay mayor esfuerzo en tanta habilidad de técnica para su configuración.
- **Analítica:** Se realiza un seguimiento fácil de los KPI
- **Entrada sencilla:** Son piezas fundamentales para el almacenamiento de los datos en un futuro.

1.3.9 La morbilidad

A la relación de personas enfermas con personas totales en un tiempo dado y multiplicado por 1000 o 10000 es el índice de morbilidad, la cual es muy importante para verificar el proceso de control (aumento o disminución) de la enfermedad.

Estos registros de morbilidad se extraen de registros de hospitales públicos o privados, consultorios médicos, clínicas y otros.

Con la información obtenida en caso grave como pandemia, el estado puede instaurar políticas apropiadas para salvaguardar la salud de la población.

1.3.10 Morbilidad Materna

Los problemas en la salud materna, como relacionados con el embarazo y el parto pueden ocurrir durante la gestación, durante el parto y después de que finaliza el embarazo. Por ejemplos:

- Enfermedades cardíacas y problemas en los vasos sanguíneos.
- Hipertensión arterial.
- Diabetes.
- Infecciones por cesárea.
- Coágulos de sangre.
- Hemorragia.
- Anemia.
- Náuseas y vómitos.
- Ansiedad y depresión.

1.3.11 Tipos de Tasa de índice de morbilidad

Se tiene los siguientes tipos:

- **La tasa de prevalencia:**

Hace una investigación retrospectiva y actual de cierta enfermedad anómala.

- **Tasa de incidencia:**

Determina el grado de riesgo alto o bajo de contraer una enfermedad.

1.3.12 Principales Causas de morbilidad

- **Enfermedades del aparato cardiovascular**
 - Hipertensión arterial
 - Cardiopatía Isquémica
 - Otras enfermedades del corazón.

- **Enfermedades cerebrovasculares**
 - Hemorrágicas
 - Oclusivas
 - Otras

- **Tumores Malignos**
 - Colon
 - Bronquios y pulmón
 - Mama
 - Próstata.
 - Tráquea

- **Enfermedades Respiratorias**
 - Enfermedad pulmonar obstructiva crónica (EPOC)
 - Bronquiectasia
 - Asma bronquial

- **Enfermedades Digestivas**
 - Úlcera péptica
 - Gastritis crónica
 - Litiasis vesicular

- **Enfermedades Endocrino metabólicas**
 - Bocio Simple
 - Diabetes Mellitus

- **Enfermedades SOMA**
 - Artropatía Crónica
 - Fractura de cadera

- **Enfermedades Psiquiátricas**

- Ansiedad
- Depresión

1.4 FORMULACIÓN DEL PROBLEMA

1.4.1 Problema General

¿De qué manera incide un datamart en la medición de la morbilidad y casos de TBC en el centro de Salud de Pueblo Nuevo?

1.4.2 Problema Específicos

- ¿De qué manera el datamart incide en la productividad de la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo?
- ¿En qué medida el datamart influye en el nivel de eficiencia para la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo?

1.5 JUSTIFICACIÓN DEL ESTUDIO

1.5.1 Justificación institucional

Se necesita un nuevo indicador, o una información rápida y eficiente que nos permita realizar el seguimiento de pacientes con tbc y la cantidad de pacientes que ya se han recuperado, y en el caso de la morbilidad detectar las 10 o 5 causas o diagnósticos muy frecuentes que hay en el centro de salud.

La tesis nos permitió optimizar uno de los procesos de la entrega de información, mediante la automatización de sus bases de datos, para poder realizar un informe más a detalle y siendo más eficiente con la información solicitada. De esta manera llevaremos un control de los pacientes diagnosticados con TBC y se realizara un seguimiento de cifras para compararla por trimestre, obteniendo así una fuente de información más leal en cuanto cifras estadísticas.

Sin olvidarnos que junto a esas cifras se mostraran datos relevantes de la morbilidad, en la cual se realizara un ranking de los diagnósticos más frecuentes para dato de impacto en la sala situacional.

1.5.2 Justificación tecnológica

El estudio conlleva a la elaboración de un datamart donde se pudo lograr mostrar resultados más precisos en cifras estadísticas, para llevar un seguimiento y tener un dato más como indicador dentro de sus informes a elaborar. Para realizar dicho proceso se estudió bien la base de datos y sus diagnósticos precisos en cuanto a los pacientes con TBC, para llegar a una detección de datos más confiable, observando así una mejora en los informes en tiempo real.

En la actualidad, los informes mencionados se han vuelto muy importante ya que con ellos pueden realizar una sala situacional más impactante, con gráficos y datos estadísticos la cual les ayuda a medir su incremento de pacientes.

Un datamart es una concentración de información orientados a un área específica en este caso nos estamos dirigiendo al área de estadística la cual lleva la información y cuentan con la data necesaria para poder llevarse a cabo dicho proyecto de tesis.

1.5.3 Justificación económica

Esta investigación se justifica por la inversión de tiempo que la tesis requiere, beneficiando de una u otra manera al centro de salud de pueblo nuevo en el área de estadísticas con datos muy relevantes y puntuales en tiempo real, tales como; análisis de información, datos relevantes, indicadores óptimos, seguimiento preciso del número de pacientes con TBC.

En un largo plazo se podrá observar como la entrega de estos informes serán de suma importancia para el bien de los trabajadores y para obtener una buena sala situacional que impacte a cualquier persona que entre al centro de salud de pueblo nuevo.

1.6 HIPÓTESIS

1.6.1 Hipótesis General

Un datamart mejora la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo

1.6.2 Hipótesis Específicos

a) Un datamart incrementa la productividad de reportes en la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo

b) Un datamart aumenta el nivel de eficiencia para la medición de la mortalidad y casos de TBC en el centro de salud de pueblo nuevo.

1.7 OBJETIVOS

1.7.1 Objetivo General

a) Determinar cómo influye un datamart en la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo.

1.7.2 Objetivos Específicos

a) Determinar cómo influye un datamart en la productividad de reportes con la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo.

b) Determinar cómo influye un datamart en el nivel de eficiencia para la medición de la morbilidad y casos de TBC en el centro de salud de pueblo nuevo.

II. MÉTODO

2.1 DISEÑO DE INVESTIGACIÓN

2.1.1 Tipo de Investigación

La investigación es del tipo cuantitativo, donde se requiere la ayuda matemática y estadística para explicar los acontecimientos dados durante el experimento o estudio realizado.

2.1.2 Diseño de la Investigación

Este trabajo es experimental, donde se va a controlar las variables en estudio se da en situación de control en la cual se manipulan premeditadamente, donde se obtendrán consecuencias para su análisis e interpretación.

2.1.3 Tipo y diseño de Investigación

El diseño pre experimental es un tipo de un solo grupo cuyo nivel de control es algo mínimo. Mayormente es útil como primera cercanía al problema de investigación en la realidad.

2.2 VARIABLES, OPERACIONALIZACIÓN

2.2.1 Definición conceptual de variables

a) Variable Independiente: Datamart

“Data mart es un sistema de concentración de información con respecto a un negocio. El programa se concentra en un departamento de la empresa almacenando información que se utiliza para un proceso o flujo productivo, que tiene que dar información o solución a las inquietudes del comprado, un grupo de Datamarts anexados entre si se denomina Datawarehouse”. (Source Consulting, 2019).

b) Variable Dependiente: Medición de Morbilidad y caso tbc

La relación de personas enfermas en un espacio teniendo en consideración el tiempo se conoce como índice de morbilidad, del mismo modo es muy útil para estudiar la evolución e involución de las enfermedades a través de un control, esto se aprecia de sobremanera en datos estadísticos.

2.2.2 Definición Operacional

a) Variable Dependiente: Medición de morbilidad y casos de tbc

La medición de la morbilidad y casos tbc nos dará un dato relevante como información primordial para uno de los indicadores que se puede brindar para que se pueda visualizar en su sala situacional.

De esta manera la información proporcionada reflejara de manera trimestral la cual veremos cómo se presenta los casos según los meses que pasan y si hay una mejoría dentro de la población.

2.2.3 OPERACIONALIZACIÓN DE VARIABLES

Tabla N° 2 Operacionalización de Variables

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores	Escala de medición
Variable Independiente	Los data marts son un repositorio de datos históricos sobre una región de la empresa y se utilizan para apoyar el proceso de medición de morbilidad y casos potenciales de tuberculosis. Esta estructura almacena información agregada o integrada que se refleja en herramientas de visualización o análisis de datos.				
Variable dependiente	La morbilidad es el resultado de personas enfermas en un sitio y tiempo dado, se considera también como útil el dato estadístico para estudiar el control y procesos de evolución de las enfermedades. La enfermedad de la tuberculosis es causada por Mycobacterium-tuberculosis, bacteria que mayormente afecta los pulmones, la cual se puede curar y prevenir.	La medición de la morbilidad y de los casos de TBC , se realizaran según los códigos CIE10, ya que cuentan con unos parámetros en el caso de tbc, y para el caso de morbilidad se realizara el proceso de un ranking para la buena presentación de datos para el POI y su sala situacional.	Sistema Productivo	Productividad	Porcentaje
			Análisis de Resultados Alcanzados	Nivel de Eficiencia	Razón

Fuente: Elaboración Propia

Tabla N° 3 Operacionalización de Indicadores

Indicadores	Descripción	Técnica	Instrumento	Unidad de medida	Fórmula
Productividad De Reportes	El número de tipos de informes visitados se clasifica entre el número de tipos de informes solicitados.	Fichaje	Ficha de Registro	Porcentaje	$PR = \left(\frac{R}{\frac{A}{RS}} \right) * 100$ <p>Donde: PR = Productividad de Reportes RA = Reportes Atendidos RS = Reportes Solicitados</p>
Nivel de Eficiencia	Evalúa el tiempo invertido para la realización de los reportes	Fichaje	Ficha de Registro	Razón	$NE = \left(\frac{HT * CE}{RS} \right)$ <p>Donde: NE = Nivel de Eficiencia HT = Horas Trabajador CE = Cantidad de Estadísticos RS = Reportes Solicitados</p>

Fuente: Elaboración propia

2.3 MÉTODOS DE ANÁLISIS DE DATOS

2.3.1 Población

Nuestro proyecto se realizará en el área de Estadística del Centro de Salud de Pueblo Nuevo del distrito de Pueblo Nuevo, ciudad de Chincha Departamento de Ica, en la cual tenemos como proceso de medición la cantidad de reportes que se pueden generar al mes, donde aplicaremos la eficiencia en cada reporte solicitado, y veremos qué tan eficientes son los trabajadores en el área de estadística.

Ante ello los reportes nos informaran el número de atenciones que se dan en el mes, para comparar donde y en qué mes incrementa o disminuye el dato de pacientes atendidos. Generando todos los reportes solicitados y observando la eficiencia de cada trabajador obtendremos un coste de cuanto nos saldría tener a un personal que realice y pueda proporcionar los reportes que se puedan solicitar.

Población 1:

Indicador 1: Productividad de Reportes

Criterio de Conclusión

- Corresponde a generar reportes solicitados y los reportes atendidos que presenta el área de estadísticas.
- El promedio de reportes generados por mes es de 6 reportes que nos indiquen el nivel de productividad en cuanto a los reportes atendidos.

Población 2:

Indicador 2: Nivel de Eficiencia

Criterio de Conclusión

- Corresponde a tiempo que requiere un personal para la elaboración de un reporte solicitado.
- Para este indicador se tomará en cuenta como cantidad de persona a 1 sola, tiempo de trabajo 8 horas jornada laboral, con la cual obtendremos el tiempo para la elaboración de los reportes solicitados.

2.3.2 Muestra

Como la población es menor de 50, la muestra es igual a la población.

La muestra es un pequeño conjunto de la población, para ver reflejados los datos dicha muestra debe ser visualizada y comparada según el dato que nos proporciona la DIRESA cada inicio de año, ya que la población aumenta y disminuye en algunas zonas de la provincia.

2.4 TÉCNICAS: INSTRUMENTOS DE RECOLECCIÓN DE DATOS, VALIDEZ Y CONFIABILIDAD

La investigación requiere de los siguientes instrumentos:

2.4.1 Técnica

Fichaje: hoja de identificación del paciente y que sirve como medio de información para las diferentes áreas de un centro de labor.

2.4.2 Instrumento

Ficha de registro: Es el instrumento que tiene la investigación documental que nos permitirá acumular los informes importantes de las fuentes consultadas.

- **FR01: Ficha de registro para el indicador productividad de Reportes**
En esta ficha se medirá la cantidad de reportes que son atendidos entre los reportes que se solicitan.
- **FR02: Ficha de registro para el indicador Nivel de Eficiencia**
En este punto visualizaremos el nivel de eficiencia de los trabajadores para poder realizar los reportes encomendados.

2.4.3 Validez y confiabilidad

“Son inherentes a la investigación, desde el punto positivista, con el fin de otorgar a la información recaudada y los instrumentos con la consistencia y exactitud necesaria para los hallazgos, determinadas del análisis de los datos en estudio” (Hidalgo 2005).

Validez

“No necesariamente tiene que ser propia del test, sino de usos específicos de las medidas que el instrumento da y las generalizaciones, quiere decir, más que el test en sí mismo, lo que se somete a la validez son los posibles resultados que se vayan a realizar a partir de las respuestas obtenidas” (Prietos 2010).

Según Hurtado mencionara la validez:

Validez de constructo

“La validez de la configuración es un concepto. Por lo tanto, la configuración busca hallar en qué medida se mide el evento en función de cómo el instrumento conceptualiza el evento y la teoría que basa el estudio.” Hurtado (2012, 0.790, 792).

La ficha de registro para esta investigación, es validada por los siguientes especialistas del tema:

Tabla N° 4 Tabla de Expertos

Nº	Expertos	Grado Académico	Puntaje	
			Productividad de Reportes	Nivel de Eficiencia
1	Roberto Roy Saavedra	Magister	85%	85%
2	Henry Paul Bermejo Terrones	Magister	94%	94%

Fuente: Elaboración propia

"La tarjeta flash se envió para su verificación. Fue realizada por tres expertos y el índice de productividad promedio en el informe fue del 85 %, que es un segundo indicador confiable. La eficiencia promedio alcanzada es del 94 %, se muestra un alto nivel de confiabilidad de que el equipo utilizado para medir el indicador es correcto”.

Confiabilidad

“El instrumento de medida se refiere al rango en el que se puede obtener el mismo resultado aplicándolo repetidamente a la misma persona u objeto” (Baptista-Hernández-Fernández 2010, p.200)

Método: Test – Retest

Los procedimientos de prueba y repetición se utilizan para determinar la confiabilidad de los datos. Este procedimiento aplica la encuesta al grupo dos veces dentro de un período de tiempo específico, asegurando así que la encuesta es el medio adecuado para recopilar datos o información.

Técnica: Coef. de correlación de Pearson

“No depende de la unidad de medida de la variable y su valor oscila entre -1 y +1 para resolver el problema anterior. En la práctica, el coeficiente está estandarizado. Es una co-varianza” (G. Olmos, et. Al, 2008 p. 193).

Los valores van de +1 a -1 y lógicamente pasan por cero. Un valor de +1 indica una fuerte relación que existe en la variable bajo investigación, el gráfico se traza linealmente creciente, mientras que un valor -1 es linealmente decreciente. ... Los valores por debajo de +1 y próximos a cero indican una disminución en la proporción de variables y una disociación completa. Lo mismo ocurre con valores desde -1, hasta -0.

El coeficiente de correlación de Pearson se representa por:

Figura N° 13 Coeficiente de Correlación de Pearson

The diagram consists of a rectangular box with a black border. Inside the box, there are two lines of text. The top line reads "Población: $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$ ". The bottom line reads "Muestra: $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$ ".

Fuente: Guardia, et. al., 2008

Donde:

ρ_{xy} = Coeficiente de correlación de Pearson de la Población.

r_{xy} = Coeficiente de correlación de Pearson de la Muestra.

$\sigma_{xy} = S_{xy}$ = Covarianza de x e y.

$\sigma_x = S_x$ = Desviación típica de la variable x.

$\sigma_y = S_y$ = Desviación típica de la variable y.

El método especificado proporciona tres grados de resultados de acuerdo con el valor indicado del valor p de contraste, como se demuestra en la tabla, de acuerdo con las siguientes condiciones:

Tabla N° 5 Módulo de Coeficientes

Coeficiente	Interpretación
1.00	Relación perfecta y +
0.90 – 0.99	relación muy alta
0.70 – 0.89	relación alta
0.40 – 0.69	relación Moderada
0.20 – 0.39	relación baja
0.01 – 0.19	relación muy baja
0	Sin relación

Fuente: Elaboración propia

“Cuando el indicador de razón de Pearson es cercano a 1, la correlación es estable y directa y es una herramienta confiable para mediciones estables y consistentes” (M. Quiñones, 2011, p. 205).

“Si el coef. de correlación de Pearson es cercano a 0, la correlación es baja. Si es bajo, es muy bajo y no correlacionado, y el dispositivo que se está evaluando no es confiable” (M. Quiñones, 2011, p. 205).

INDICADOR: PRODUCTIVIDAD DE REPORTES PRE TEST Y RETEST

Figura N° 14 Indicador de Productividad PRE-TEST y RETEST

		Prod_Report_ PreTest	Prod_Report_ Retest		
Prod_Report_PreTest	Correlación de Pearson	1	,753**		
	Sig. (bilateral)		,001		
	N	15	15		
	Simulación de muestreo ^c	Sesgo	0	-,009	
		Desv. Error	0	,123	
		Intervalo de confianza al 95%	Inferior	1	,476
			Superior	1	,930
Prod_Report_Retest	Correlación de Pearson	,753**	1		
	Sig. (bilateral)	,001			
	N	15	15		
	Simulación de muestreo ^c	Sesgo	-,009	0	
		Desv. Error	,123	0	
		Intervalo de confianza al 95%	Inferior	,476	1
			Superior	,930	1

Fuente: M. Quiñonez, 2011

De la figura N°14 para el indicador “Productividad de Reportes” fue de 0.753, se observa que el grado de confiabilidad del instrumento utilizado en la medición de este indicador y teniendo en consideración la tabla N°05 es alta.

Figura N° 15 Indicador: Nivel de Eficiencia PRE TEST y RETEST

		Niv_Eficiencia_ _PreTest	Niv_Eficiencia_ _a_Retest		
Niv_Eficiencia_PreTest	Correlación de Pearson	1	1,000**		
	Sig. (bilateral)		,000		
	N	15	15		
	Simulación de muestreo ^b	Sesgo	0	,000	
		Desv. Error	0	,000	
		Intervalo de confianza al 95%	Inferior	1	1,000
			Superior	1	1,000
Niv_Eficiencia_Retest	Correlación de Pearson	1,000**	1		
	Sig. (bilateral)	,000			
	N	15	15		
	Simulación de muestreo ^b	Sesgo	,000	0	
		Desv. Error	,000	0	
		Intervalo de confianza al 95%	Inferior	1,000	1
			Superior	1,000	1

Fuente: M. Quiñonez, 2011

De la figura N°15 para el indicador “Nivel Eficiencia” se obtiene un valor de 1.000, es así que, el nivel de confiabilidad de los equipos utilizados según la tabla N°07, es Perfecto y positivo.

2.5 MÉTODOS DE ANÁLISIS DE DATOS

El análisis es la manipulación de números y sucesos para tener información específica de manera que ayude a los gerentes a tomar las decisiones buenas. La idea principal es obtener información concreta, válida y confiable” (Naghi, 2005, p. 254).

"Análisis de contenido cuantitativo es un método para investigar objetiva y sistemáticamente todo tipo de comunicación, categorizar mensajes o contenido en categorías y subconjuntos para el análisis estadístico”.

La encuesta realizó un análisis cuantitativo porque la información se procesa en base a la muestra de datos y las entidades y estadísticas resultantes.

Nivel de Significancia

"Los niveles de significación juegan un papel importante en la prueba de hipótesis porque pueden predeterminar el grado de error que los investigadores están dispuestos a aceptar. Los niveles de significación están asociados con extremos o una pequeña porción de la cola de la distribución de la muestra. Al rechazar la hipótesis nula H_0 queda aceptar la hipótesis alternativa H_1 porque se puede concluir que el evento no puede ocurrir por accidente si el valor calculado se encuentra en esta zona“(Moncada Jiménez, 2005, p.9).

“No existe un único criterio o nivel de importancia universal para probar la hipótesis. Es posible probar la hipótesis en cualquier grado de significancia, se puede probar con una probabilidad aceptable. Nota que elegir el mínimo o grado de significancia también conlleva el riesgo de rechazar la hipótesis nula si es cierta. (Rubín y Lenin, 2004, p. 326)”

En la investigación se utilizó:

$$X = 5\%(\text{error})$$

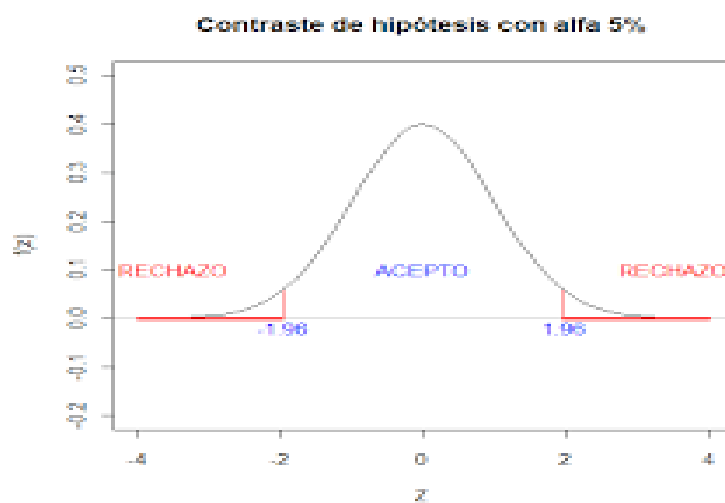
“El nivel de confianza de la estimación se visualiza como un anexo de la probabilidad de error de estimación.” (Vivanco, 2005, p. 60).

$$1-\alpha=0.95$$

Para la confiabilidad, la decisión del 95% es la decisión más común. Esto se cargó a que existe una proporción entre la precisión (reflejada por la latitud del descuido de confianza) y la confiabilidad (expresada por el cota de confianza). (Triola, 2004, p. 301).

La representación gráfica es la siguiente:

Figura N° 16 Contraste de Hipótesis Alfa



Fuente: Triola, 2004

Media Aritmética

Esta medida es el promedio de un grupo de datos, es decir, se suman todos los datos numéricos estudiados y se dividen entre el número de estos datos, la cual se representa por la fórmula:

$$\bar{X} = \frac{\sum_{i=1}^m X_i}{n}$$

Varianza

“Resulta ser el cuadrado de la desviación estándar y se representa por s^2 ” (Hernández, Fernández y Batista, 2010).

Está representada por la fórmula:

Figura N° 17 La Varianza

$$\text{Varianza} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$$

Fuente: M. Quiñonez 2011

Desviación Estándar

Es la desviación (alejamiento o acercamiento) de los datos con respecto a la media, asimismo, se puede indicar que es la raíz cuadrada del valor de la varianza,

Figura N° 18 Desviación Estándar

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Fuente: M. Quiñonez 2011

Figura N° 19 Desviación estándar de la población

Desviación estándar de la población	
$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x - \mu)^2}{N}} = \sqrt{\frac{\sum x^2}{N} - \mu^2}$	[3-13]
donde,	
<ul style="list-style-type: none">• x = observación• μ = media de la población• N = número total de elementos de la población• Σ = suma de todos los valores $(x - \mu)^2$, o todos los valores x^2• σ = desviación estándar de la población• σ^2 = varianza de la población	

Fuente: Hernández, Fernández y Batista, 2010

Fórmula de la desviación estándar

La desviación estándar es la desviación media del resultado de la media. Cuanto más grande es la dispersión de los datos alrededor de la media, más grande es la desviación estándar” (Hernández, Fernández y Batista 2010).

Prueba T-Student

Este test se utiliza cuando los dos indicios son independientes, por lo que las dos muestras se comparan en función del promedio. Se considera que $N \leq 30$.

Es importante mencionar que esta prueba necesita una hipótesis nula y otra alterna. Si el desarrollo de la fórmula da una t muy alta, se rechaza la hipótesis nula.

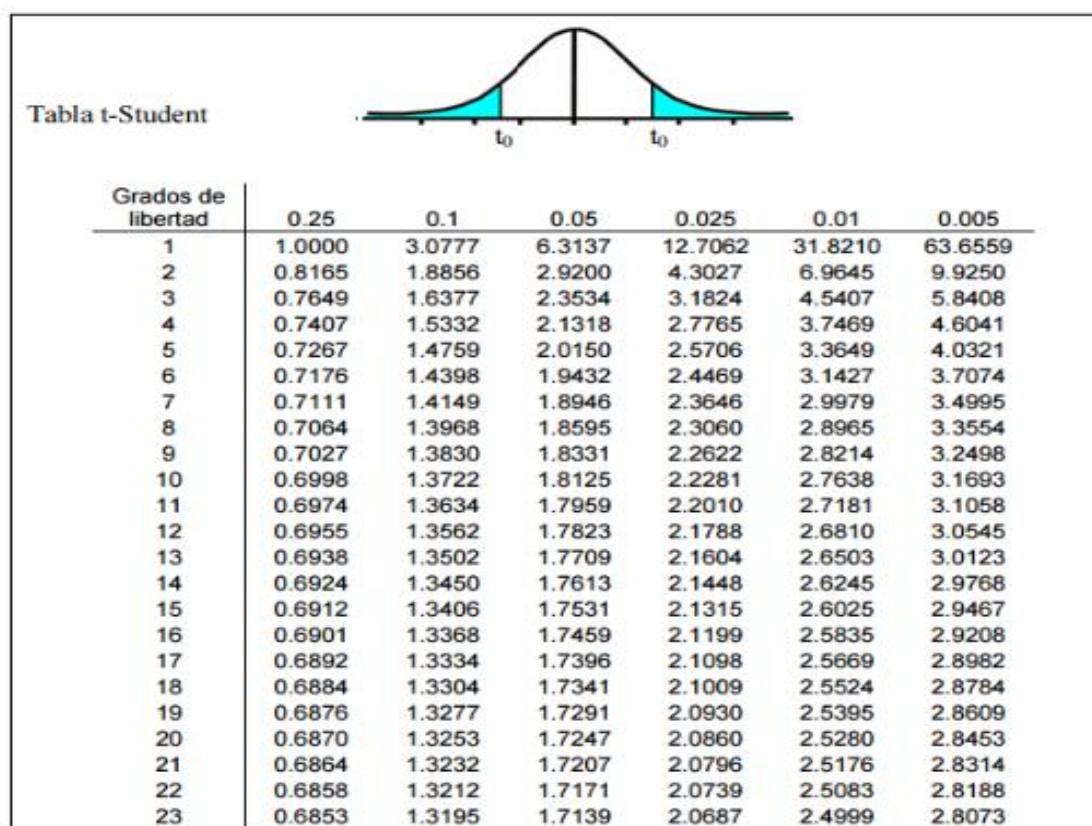
Figura N° 20 Fórmula T-Student

$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$	Donde: <ul style="list-style-type: none">• x es la media muestral,• s es la desviación estándar muestral• N es el tamaño de la muestra.
--	--

Fuente: Molina Quiñonez, 2011

En la imagen vemos los valores que toma T según los niveles de libertad

Figura N° 21 Tabla T-student



Fuente: Hernandez R. Fernandez C y Bapista.P 2010

2.6 ASPECTOS ÉTICOS

Se ha considerado la ética y los valores como la honestidad, el respeto, puntualidad y otros que se enmarca en los lineamientos de la Universidad Cesar Vallejo. Del mismo modo, se ha seguido los protocolos para una investigación con diseño cuantitativo.

Además, se volvió a examinar el juicio de expertos para asesorar sobre metodologías.

Este estudio contiene información proporcionada por la Empresa Centro de Salud Pueblo Nuevo con el fin de desarrollar el proyecto, el cual la empresa y los investigadores han autorizado para su uso.

III. RESULTADOS

3.1 ANÁLISIS DESCRIPTIVO

“En la merienda se aplicó al principio de un Datamart para valorar la productividad de los reportes y el altura de eficiencia, para ello se emplea un pre-examen que permite conservarse las parámetros abreviatura del indicador; después se implementó el datamart y de flamante se registraron los datos de la productividad de los reportes y el altura de eficiencia”. Los resultados del análisis descriptivo de estas medidas se visualizan en las tablas N°06 y N°07.

INDICADOR: Productividad de Reportes

En la tabla N° 06 muestran el resultado de la descripción de los indicadores de productividad en el informe de estas medidas.

Tabla N° 6 Estadístico Descriptivo

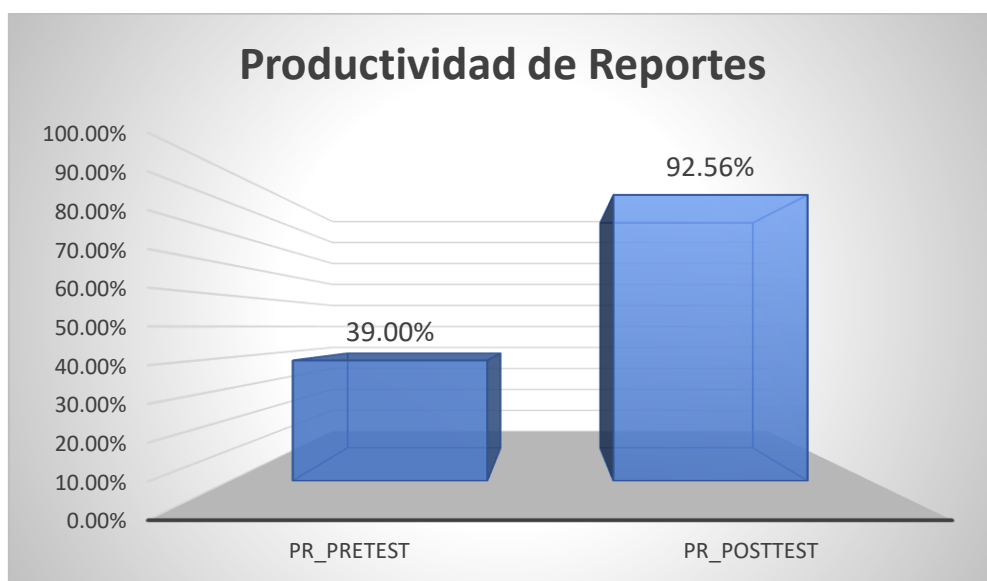
	N	Mínimo	Máximo	Media	Desv. Tip.
PR_PreTest	15	0,00	75,00	39,00	22,74
PR_PostTest	15	66,67	100,00	92,56	13,13

Fuente: Elaboración propia

Medida descriptiva de los indicadores de productividad del informe antes y después de la creación del data mart.

Para los indicadores de productividad del informe, el valor de la prueba previa fue de 22,74 y el valor de la prueba posterior fue de 13,13. Este resultado se puede ver en la Tabla 06. Esto muestra que hay una gran diferencia antes y después de crear la data mart.

Figura N° 22 Productividad de Reportes



Fuente: Elaboración propia

Como podemos apreciar en el gráfico los porcentajes de la productividad de los reportes incrementó luego la creación del datamart, incrementando el cumplimiento de los reportes requeridos.

INDICADOR: Nivel de Eficiencia

Los resultados descriptivos del grado de eficiencia del indicador se encuentran en la tabla 07: Mediciones descriptivas de indicadores de nivel de eficiencia antes y después de la creación del data mart.

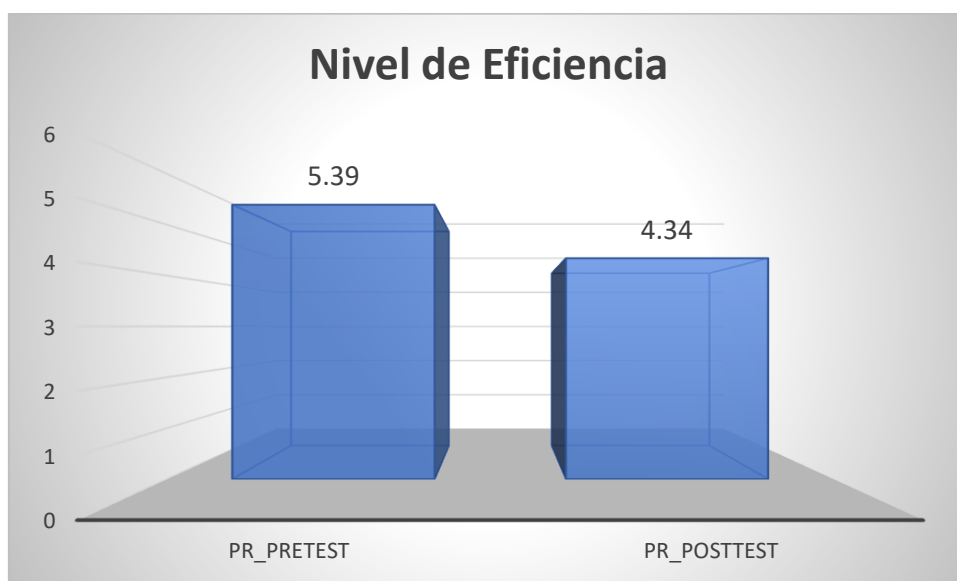
Tabla N° 7 Estadísticos Descriptivos

	N	Mínimo	Máximo	Media	Desv. Tip.
NE_Pretest	15	2,00	8,00	5,39	1,71
NE_Posttest	15	1,50	7,50	4,34	1,68

Fuente: Elaboración propia

Como se visualiza en la Tabla N° 07, el grado de eficiencia se estimó en 1.71 antes de la prueba y 1.68 después de la prueba. Esto muestra una gran diferencia de un antes y un después de la creación de la data mart.

Figura N° 23 Nivel de Eficiencia



Fuente: Elaboración propia

Como podemos apreciar en el gráfico las horas de nivel de eficiencia en la entrega de los reportes disminuyeron luego de la creación del datamart, teniendo un nivel de eficiencia mejor.

3.2 Análisis Inferencial

Pruebas de Normalidad

"Si la muestra es saludable, puede usar métodos estadísticos paramétricos tradicionales, de lo contrario, debe modificar los datos o usar métodos estadísticos no paramétricos u otros para hacer que la muestra de la encuesta sea saludable. La importancia de la verificación de género es la base de las estadísticas (Riesgo, 2003, p.21).

Bernal-Morell señala que existen dos tipos de pruebas:

- **Prueba de Kolmogorov-Smirnov**, análisis > 50 unidades.
- **Prueba de Shapiro-Wilk**, análisis < 50 unidades.

Se utilizó el método ShapiroWilk en ambos casos porque el tamaño de muestra de la métrica de productividad para el informe es de 15 registros y el tamaño de muestra para la métrica de nivel de eficiencia es de 15 registros.

Se realizó introduciendo datos para cada indicador en el software estadístico SPSS26.0.0.0 con un grado de confianza del 95% bajo las siguientes premisas:

Si:

Sig. <0> 0.05 asume una distribución normal

Donde:

Sig. Valor p o valor de contraste crítico

Los resultados son:

INDICADOR: Productividad de Reportes

Los datos se someten a la prueba de hipótesis; con el objetivo de corroborar su distribución, el objetivo de esta comprobación es verificar si la Productividad de reportes tiene distribución normal.

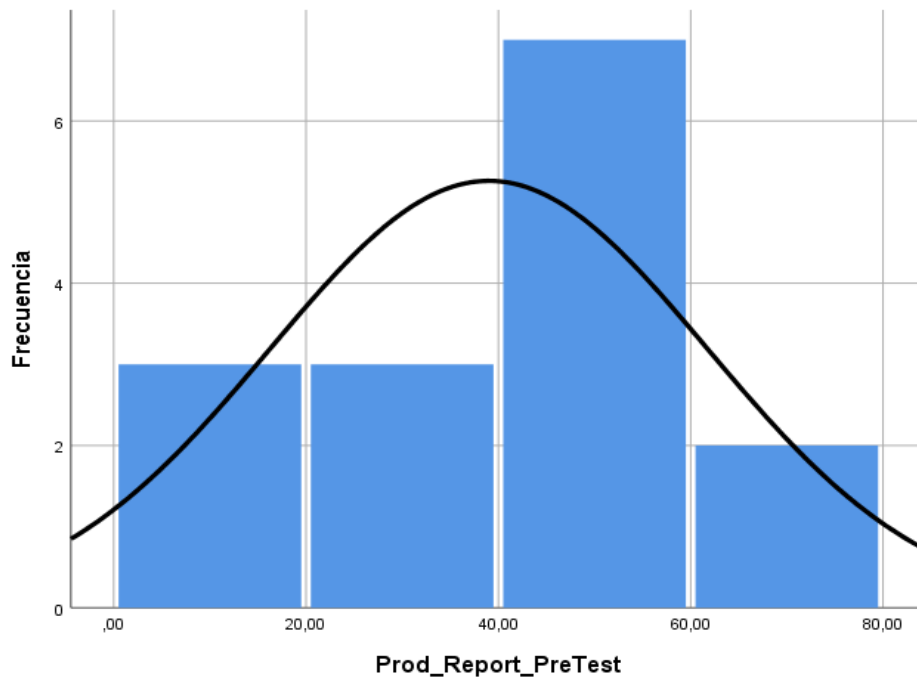
Tabla N° 8 Prueba de Normalidad del indicador Productividad de Reportes antes y después del Datamart

	Shapiro-Wilk		
	Estadístico	Gl	Sig.
PR_PreTest	,835	15	,011
PR_PostTest	,603	15	,000

Fuente: Elaboración Propia

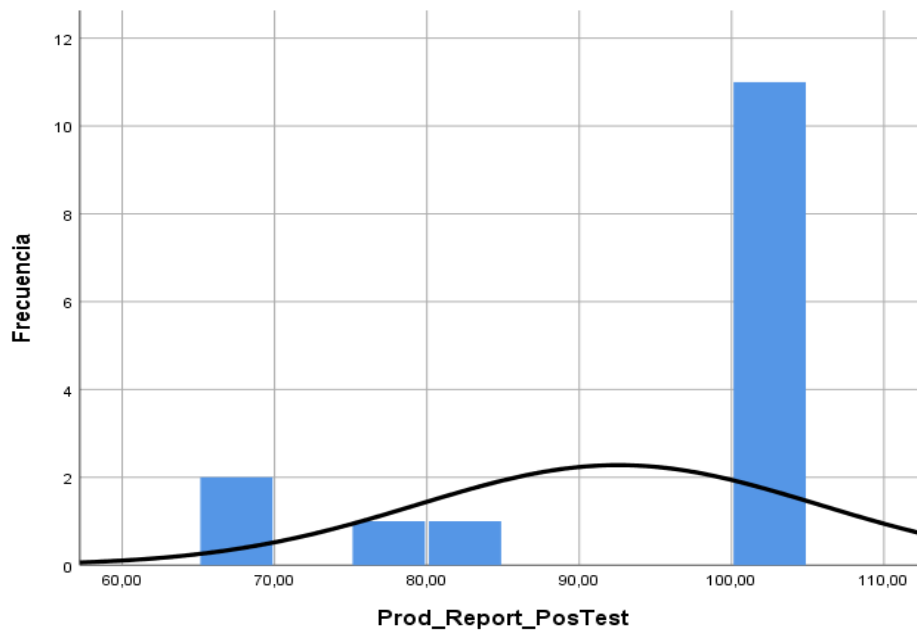
Como se visualiza en la tabla, los resultados de las pruebas muestran que la firma de productividad del informe pre-probado en el Centro de Salud de Pueblo Nuevo es 0.11 y este valor es mayor a 0.05. Entonces, la productividad de los informes se distribuye con normalidad. Los resultados obtenidos de la prueba posterior muestran que la firma de productividad del informe es 0,000, que es inferior a 0,05. Esto demuestra que la productividad del informe no se distribuye normalmente.

Figura N° 24 Prueba de normalidad del indicador Productividad de Reportes antes del Datamart



Fuente: Elaboración Propia

Figura N° 25 Prueba de normalidad del indicador Productividad de Reportes antes del Datamart



Fuente: Elaboración propia

INDICADOR: Nivel de Eficiencia

“Los datos se someten al hipótesis; con el fin de corroborar su repartición, el objetivo de esta comprobación es verificar si el grado de eficiencia tiene repartición normal.”

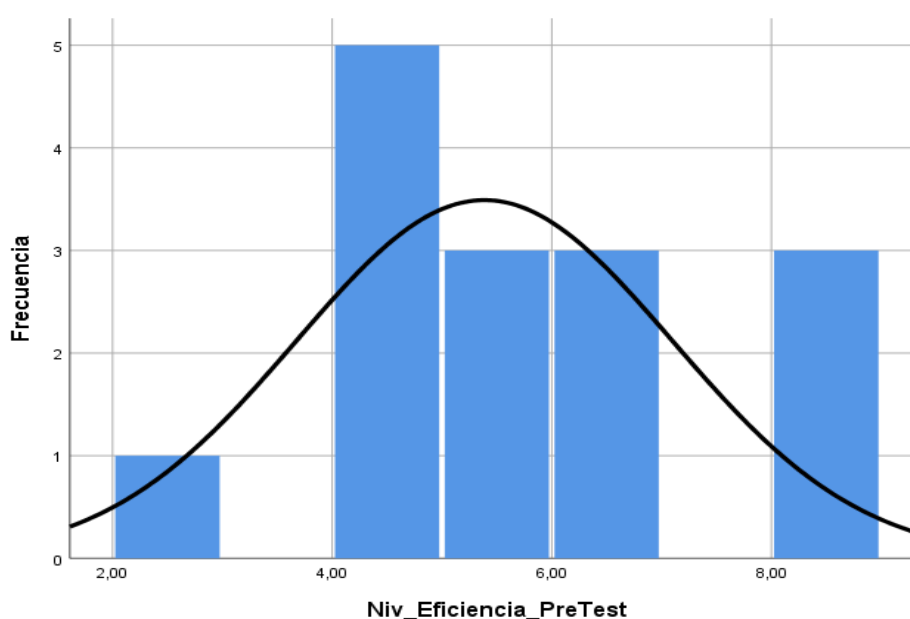
Tabla N° 9 Prueba de Normalidad del indicador Nivel de Eficiencia antes y después del Datamart

	Shapiro-Wilk		
	Estadístico	GI	Sig.
NE_Pretest	,923	15	,215
NE_Posttest	,965	15	,785

Fuente: Elaboración propia

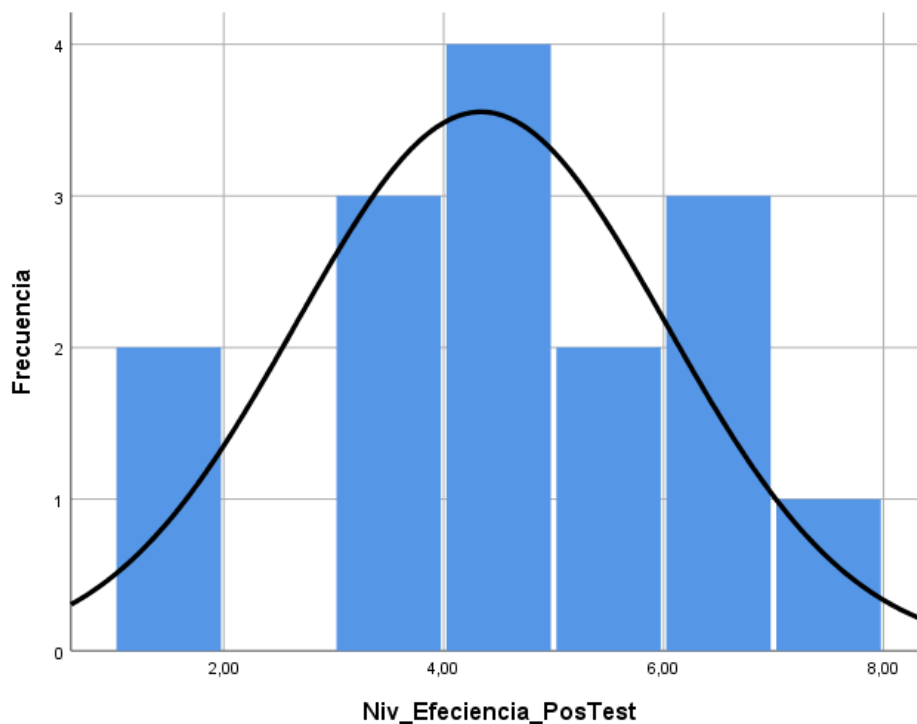
En la tabla N° 09 se visualiza el grado de eficiencia Sig del Centro de Salud Pueblo Nuevo en el pretest. Es 0.215 y los valores van desde a 0.05, indicando una repartición normal de los grados de eficiencia. El resultado de la prueba PostTest tiene un nivel de eficiencia de Sig. Es 0.785, lo que indica que este valor es mayor a 0.05. Esto demuestra que los niveles de eficiencia no se reparten normalmente. Esto confirma la repartición normal de ambos datos de la muestra.

Figura N° 26 Prueba de normalidad del indicador Nivel de Eficiencia antes del datamart



Fuente: Elaboración Propia

Figura N° 27 Prueba de normalidad del indicador Nivel de Eficiencia antes del datamart



Fuente: Elaboración Propia

3.3 Prueba de Hipótesis

Para el indicador de Productividad de Reportes tenemos en su investigación.

a) Hipótesis de Investigación 1(HE1)

Data mart para la medición de morbilidad y casos de tbc en el centro de salud pueblo nuevo

b) Representando de las variables

La1: Productividad de Reportes sin el datamart en la medición de la morbilidad y caso de tbc

Lp1: Productividad de Reportes con el datamart en la medición de la morbilidad y caso de tbc

c) Hipótesis estadística 1

Hipótesis Nula (H01): El datamart no aumenta la productividad de los reportes en la medición de la morbilidad y los casos tbc en el centro de salud de pueblo nuevo.

H01: la1>lp1

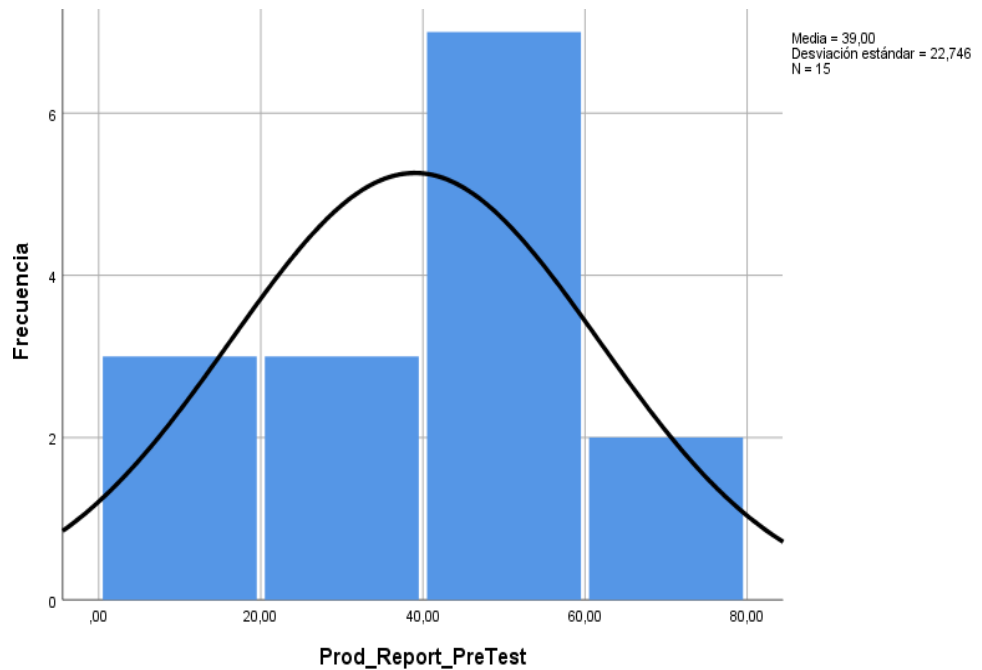
Por lo tanto, la integración del data mart no ha mejorado los indicadores.

Hipótesis Alterna (HA1): El datamart no incrementa la productividad de los reportes en la medición de la morbilidad y los casos tbc en el centro de salud de pueblo nuevo.

HA1: $\mu_1 < \mu_2$

Con la integración de data marts, los indicadores han mejorado. En la Figura N° 24, se puede observar que la productividad del informe (pre-test) es del 39,00%.

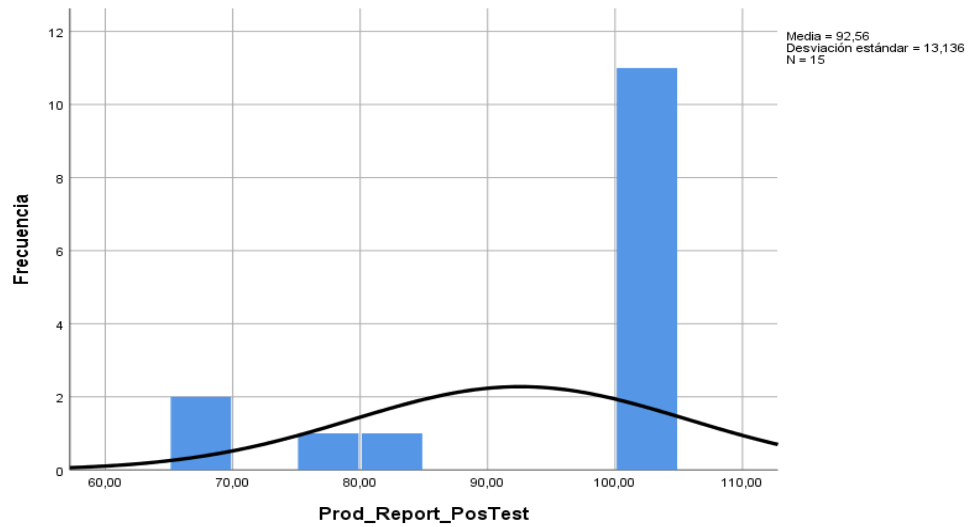
Figura N° 28 Prueba de normalidad del indicador Productividad de Reportes antes del datamart



Fuente: Elaboración Propia

En la figura N° 25 muestra que la productividad del informe (después de la prueba) es del 92,56%.

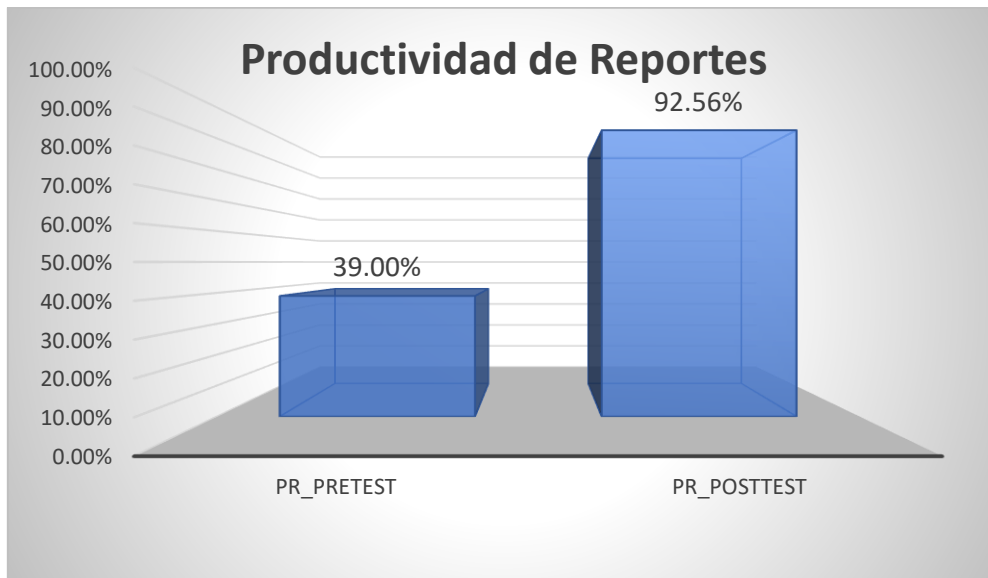
Figura N° 29 Prueba de normalidad del indicador Productividad de Reportes antes del datamart



Fuente: Elaboración Propia

Las figuras N°24 y N°25 muestran las ganancias de productividad para el informe. Esto se puede confirmar comparando los respectivos valores medios en el rango de 39,00 % a 92,56 %.

Figura N° 30 Productividad de Reportes



Fuente: Elaboración Propia

De acuerdo a la figura N°22, se aprecia que existe un aumento importante en la productividad de los reportes en la medición de la morbilidad y casos TBC, el cual se incrementa en 53.56%.

Debido al contraste de hipótesis, se utilizó la prueba de Shapiro Wilk debido a que los datos que se obtuvieron en la encuesta (PreTest y PostTest) se reparten normalmente.

d) Hipótesis estadística 2

Hipótesis Nula (H01): El datamart no incrementa el nivel de eficiencia en la medición de la morbilidad y los casos tbc en el centro de salud de pueblo nuevo.

$$\mathbf{H01: \mu_1 > \mu_2}$$

Por lo tanto, la integración del data mart no ha mejorado los indicadores.

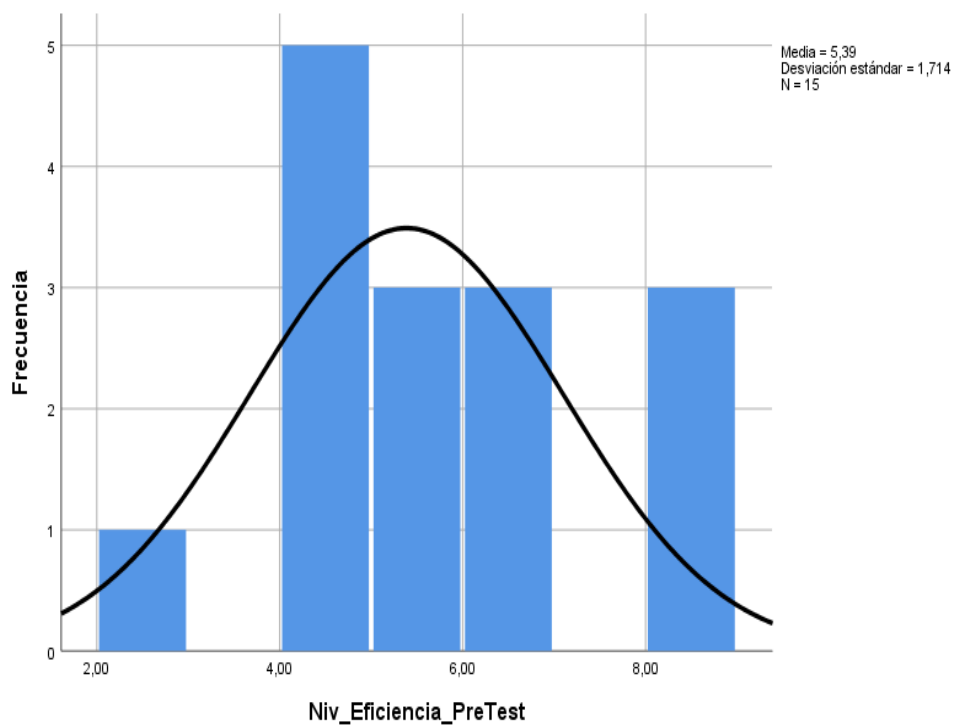
Hipótesis alternativa (HA1): Los data marts no mejoran la morbilidad y la eficiencia de medición de pacientes con TB en los nuevos centros de salud urbanos.

$$\mathbf{HA1: \mu_1 < \mu_2}$$

Se deduce que el índice mejoro con la álaabe del datamart

La figura número 26 muestra que la eficiencia (pretest) es del 5,39%.

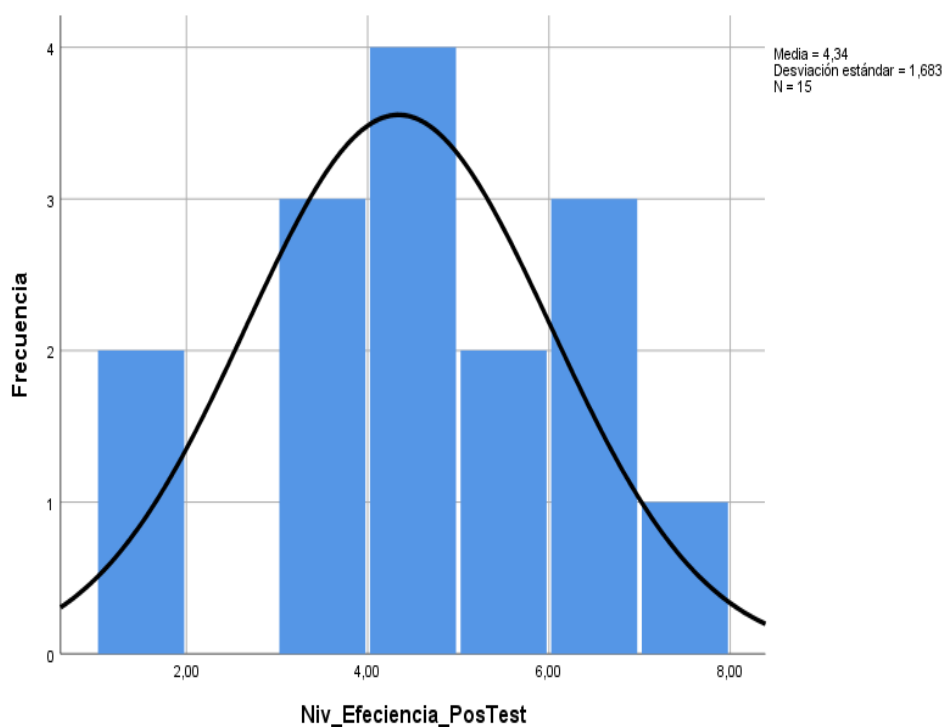
Figura N° 31 Prueba de normalidad de visualización del grado de eficiencia antes del data mart



Fuente: Elaboración Propia

En la figura N°27 muestra que la productividad del informe (después de la prueba) es de 4.34%.

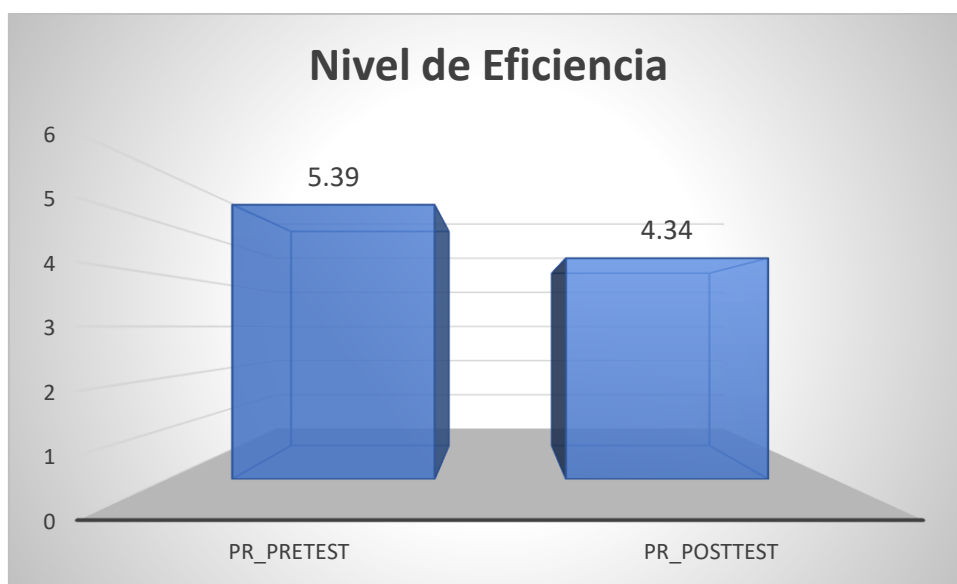
Figura N° 32 Prueba de normalidad de visualización del grado de eficiencia antes del data mart



Fuente: Elaboración Propia

Las figuras N°26 y N°27 muestran la mejora en la eficiencia, lo que se puede confirmar al comparar los valores promedio de cada disminución de 5.39% a 4.34%.

Figura N° 33 Nivel de eficiencia



Fuente: Elaboración Propia

De acuerdo a la figura N°22, observamos un aumento en la productividad de los reportes, el cual se incrementa en 53.56%.

Como respuestas del contraste de hipótesis se utilizó la prueba de ShapiroWilk ya que los datos extraídos en las encuestas (Pre-Test y Post-Test) se reparten normalmente. (Wilk. MB, 1965)

IV. DISCUSIÓN

Como resultado de la investigación, Data Mart aumentó la productividad del informe de 38,99 a 92,55. Esto corresponde a un crecimiento del 53,56%. De manera similar, Carmen Pamela Rosales Sedano concluye en su artículo, "Analizar-diseñar e implementar data mart que apoyen para la toma de decisiones y la evaluación de estrategias de salud", que implementar data mart mejora la calidad del arrozal. Condiciones de validez de los datos más eficientes y rápidas que las que has trabajado.

Con ello comprobamos que el data mart es la de forma de mejorar la realización y la generación los reportes de base de datos amplias, que cuentan con un big data muy valioso.

Como resultado, la eficiencia de envío de informes también se redujo de 5,38 a 4,34, mejorando la eficiencia de envío y preparación de los informes solicitados.

En el centro de salud se ha mejorado mucho en los temas de reportes tanto en productividad, eficiencia y sobre todo presentación que es lo más importante en un reporte.

V. CONCLUSIONES

Primero:

La creación de un data mart para medir morbilidad y pacientes con TB en el Centro de Salud de Pueblo nuevo incrementó la productividad de reportes en un 24.67%, ya que la primera medición sin data mart (pre-test) fue de 42.88%, se puede concluir. El resultado de la medición final (después de la prueba) usando el data mart ya creado es 67.55, "Data mart mejora la productividad del diagnóstico en el centro de salud".

Segundo:

Al instalar un data mart para medir morbilidad y pacientes tuberculosos en el Centro de Salud Ciudad Nueva se obtuvo 4.05 en la primera medición (pre-test) sin el data mart, mejorando la eficiencia de reporte en 1.01, se puede concluir. En la medición final (después de la prueba) lograda y utilizando el data mart ya creado, recibimos 3.04 y confirmamos la hipótesis de que "los data marts mejoran la eficiencia del Centro de Salud.

Tercero:

Para finalizar se concluye que el datamart optimizó varios procesos en el centro de salud de pueblo nuevo pues permitió que la productividad de los reportes sea entregada según la cantidad de reportes que puedan solicitar a diario, y de igual manera mejoro el grado de eficiencia en el otorgamiento de los reportes, ahora esta entrega es más rápida a comparación de la realidad cuando no se había creado el datamart, ante todo ello logramos alcanzar las metas propuestas.

VI. RECOMENDACIONES

Perú sigue utilizando herramientas de visualización que no dejan tanto impacto visual como pueden mostrar. La mayoría de las empresas continúan utilizando las herramientas gratuitas de Microsoft Office Excel, que son herramientas muy útiles para administrar datos, pero necesitan una mejor finalización para visualizar sus gráficos, por ello se recomienda:

Primero: Se recomienda a futuro que en este siglo XXI se sigan actualizando e incorporando nuevas herramientas de visualización de reportes, la cual mediante la nube uno pueda visualizar su reporte de cualquier parte del mundo llevando el control de sí misma para cualquier presentación que se pudiera dar a futuro.

Segundo: Actualmente hay que causar una buena impresión al mostrar sus datos, de una manera muy didáctica en la cual con tal solo observar se pueda interpretar lo que queremos explicar.

Tercero: Del mismo modo se recomienda trabajar con nuevos motores de base de datos la cual realizara un mayor almacenamiento de datos, realizando así un trabajo más ligero y poder así obtener su minería de datos llamado también Big data.

VI. Referencias

Bernabeu, R. (2009).Data Warehousing: Investigación y Sistematización de Conceptos-Hefesto: Metodología para la Construcción de un Data Warehouse
Recuperado de:

http://www.dataprix.com/files/DWH_Metodologia_HEFESTO-V1.0.pdf

BERNABÉU, Ricardo. Hefesto - DataWare Housing [en línea], 2012 [fecha de consulta: 26 de setiembre 2017]. Disponible en:

<https://www.businessintelligence.info/assets/hefesto-v2.pdf>

BERNABEU, Ricardo, Hefesto [en línea] Córdoba Argentina [fecha de consulta 25 de julio 2016] disponible en www.bussonessintelligence.info.docs/hefesto-v2.pdf

HERNANDEZ, Roberto, FERNANDEZ, Carlos y BAPTISTA, Maria. Metodología de la investigación. 6ª ed. México: McGraw-Hill, 2014. 600 pp.

HERNANDEZ, J., FERNANDEZ, U., BAPTISTA, A. como evaluar un proyecto empresarial: una visión práctica. Ed. ediciones diaz de santos ,2010.p.174

Hernández, R. Fernández, C. y Baptista, P. (2010). Metodología de la Investigación (5.a ed.). p. 122México: McGraw-Hill/Interamericana Editores S.A ISBN: 978-968- 460-950-0

INMON, William. Building the Data Warehouse, 2012.

MAZÓN, N., PARDILLO, J. y TRUJILLO, J. (2011). Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional. Recuperado de: http://books.google.com.pe/books?id=E7Aceg-o4oC&hl=es&source=gbs_navlinks_s

Ley Nro. 12426. Diario oficial El Peruano, Lima, Perú, 01 de enero de 2017.

NETTLETON, David. Análisis de datos comerciales [en línea]. Madrid: Ediciones Díaz de Santos S.A., 2003. [Consultado 28 de setiembre 2017] ISBN: 84-7978-593-4. Disponible en:

<https://books.google.com.pe/books?id=GdoCJOaYNUC&pg=PA106&dq=datamart>

ORTEGA, Et. Libre mercado y protección ambiental: intervención y orientación ambiental de las Actividades económicas. Ed.INAP ,2009 p.170

Resolución Ministerial Nro. 1025-2013. Aprueban Plan estratégico de Tecnología de Información 2016-2018 Ministerio del Interior, Lima, Perú, 17 de diciembre de 2017.

Resolución Ministerial Nro. 1361-2016. Aprueban Plan operativo institucional de la Policía Nacional del Perú, Lima, Perú, 17 de diciembre de 2016.

RIVADERA, Gustavo. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses) [en línea], 2013 [fecha de consulta: 15 de Junio 2018]. Disponible en:

<http://www1.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivaderaformateado.pdf>

Rojas, Alejandro. Implementación de un Data Mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el departamento de finanzas de la contraloría general de la república [en línea], 2013 [fecha de consulta: 26 de setiembre 2017]. Disponible en:

http://www.repositorioacademico.usmp.edu.pe/bitstream/usmp/1061/1/rojas_a.pdf

Título importancia de la incorporación temprana a la investigación científica en la universidad de Guadalajara.

Editor Juan Carlos Martínez Coll, 2010

ISBN 8876906967, 9788876906961

Título Introducción al Business Intelligence

Autor Josep Curto Díaz

Editor Editorial UOC, 2012

ISBN 8497889797, 9788497889797

N.º de páginas 238 páginas

Volumen 2 de Trivium Infermeria

Autor Joaquín Tomás-Sábado

Editor Univ. Autònoma de Barcelona, 2010

ISBN 8449026164, 9788449026164

N.º de páginas 146 páginas

ZAMBRANO, A (2011). Análisis, diseño e implementación de un Datamart para el área de mantenimiento y logística de una empresa de transporte público de pasajeros. [En línea] (Tesis para optar el título de ingeniero, Pontificia Universidad

Católica del Perú). [Consulta en setiembre 2015] Disponible en web <http://www.pucp.edu.pe/content/tesis/pucp-5548412>

Blas, David. Análisis de datos comerciales [en línea]. Madrid: Ediciones Díaz de Santos S.A., 2003. [Consultado 28 de setiembre 2017] ISBN: 84-7978-593-4. Disponible en: <https://books.google.com.pe/books?id=GdoCJOaYNUC&pg=PA106&dq=datamart>

SALCEDO, Et. Libre mercado y protección ambiental: intervención y orientación ambiental de las Actividades económicas. Ed.INAP ,2009 p.170

ANEXOS

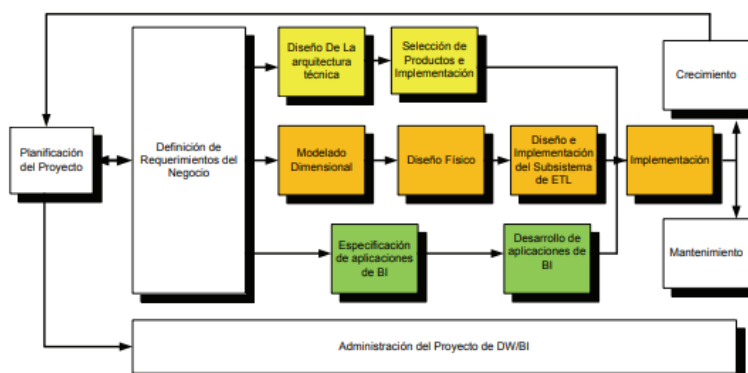
METODOLOGÍA KIMBALL (CICLO DE VIDA) INTRODUCCION

Para este proyecto utilizaremos la metodología de Kimball la cual se basa en un ciclo de vida Dimensional del Negocio (Business Dimensional Lifecycle), este ciclo de vida del proyecto de data warehouse, está basado en cuatro principios básicos:

- **Centrarse en el negocio:** Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos refuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** Crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor del negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.
- **Ofrecer la solución completa:** Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa
- tener un almacén de datos sólido, bien diseñado, con calidad probada y accesible. De igual manera se deberá de entregar herramientas de consultas, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

En este proceso Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. Las tareas de esta metodología se muestran a continuación:

Figura N°17: Proceso Kimball



Como podemos observar en la Figura. Primero, hay que resaltar el rol central de la tarea de definición de requerimientos. Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes. También tiene influencia en el plan del proyecto y en el segundo lugar podemos ver tres rutas en la cual se enfocan en diferentes áreas, a continuación, mencionaremos los pasos a seguir para la metodología Kimball:

- Planificación
- Análisis de requerimiento
- Modelado Dimensional
- Diseño físico
- Diseño del ETL
- Especificaciones y desarrollo de aplicaciones de BI

Principios Básicos

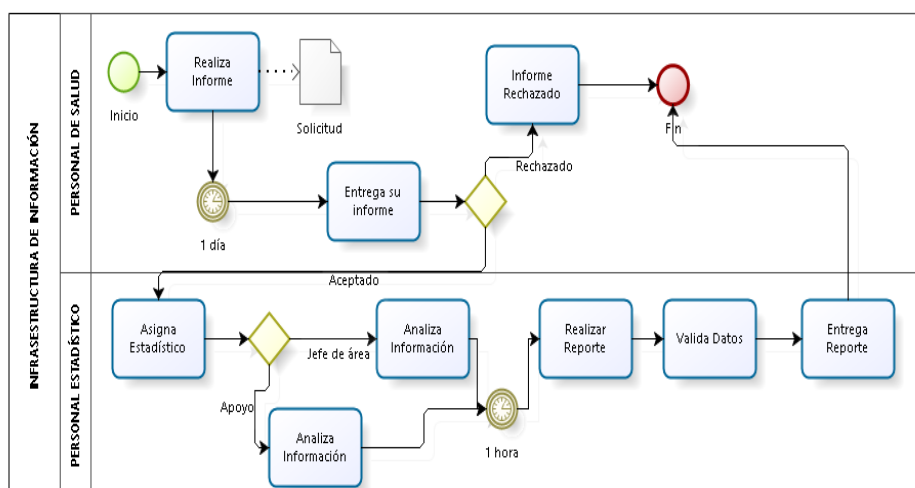
- **Centrarse en el Negocio:** En este punto identificaremos los requerimientos del negocio, para tener un buen análisis y buenas consultas de los implementadores.

Requerimientos

Nº	Requerimientos
R1	Se requiere tener un reporte de atención de pacientes por estrategias
R2	Se requiere un reporte de la atención de pacientes por personal de salud
R3	Se requiere un reporte de atención por grupo etarios
R4	Se requiere un reporte con la cantidad de diagnosticados con el código TBC positivo
R5	Se requiere un reporte donde refleje la cantidad de pacientes por géneros
R6	Se requiere un reporte de la cantidad de atenciones por mes
R7	Se requiere una reporte con los 10 primeros diagnósticos frecuentes por mes.
R8	Se requiere un reporte con la cantidad de pacientes recuperados con TBC
R9	Se requiere un reporte con cantidad de diagnósticos TBC de manera trimestral
R10	Se requiere un reporte con la cantidad de pacientes atendidos en emergencias por el personal medico
R11	Se requiere un reporte con el número de pacientes atendidos por financiador

- **Construir una infraestructura de información adecuada:** En este segundo punto diseñaremos una base unificada, en la cual sea fácil de poder usar, en la cual se reflejará los requerimientos propuestos en el negocio.

Figura N°18: Infraestructura de Información



Powered by
bizagi
Modeler

- **Realizar entregas en incrementos significativos:** En este punto reflejaremos que los requerimientos solicitados sean entregados en un corto plazo, o que sean realizados en el menos tiempo posible dependiendo de la disponibilidad del personal de estadística.

Tiempo de los requerimientos

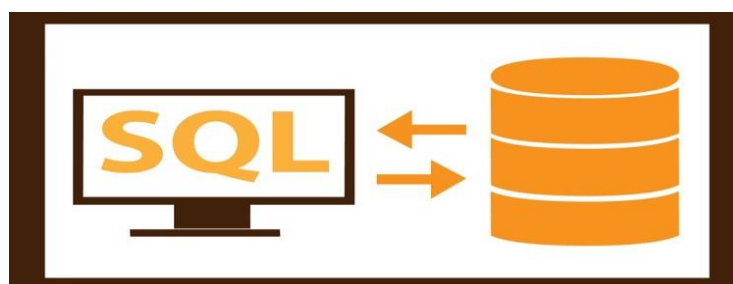
Nº	Requerimientos	Tiempo Min	Tiempo Max
R1	Se requiere tener un reporte de atención de pacientes por estrategias	1 hora	1 día
R2	Se requiere un reporte de la atención de pacientes por personal de salud	1 hora	1 día
R3	Se requiere un reporte de atención por grupo etarios	1 hora	1 día
R4	Se requiere un reporte con la cantidad de diagnosticados con el código TBC positivo	1 hora	1 día
R5	Se requiere un reporte donde refleje la cantidad de pacientes por géneros	1 hora	1 día

R6	Se requiere un reporte de la cantidad de atenciones por mes	1 hora	1 día
R7	Se requiere un reporte con los 10 primeros diagnósticos frecuentes por mes.	1 hora	1 día
R8	Se requiere un reporte con la cantidad de pacientes recuperados con TBC	1 hora	1 día
R9	Se requiere un reporte con cantidad de diagnósticos TBC de manera trimestral	1 hora	1 día
R10	Se requiere un reporte con la cantidad de pacientes atendidos en emergencias por el personal medico	1 hora	1 día
R11	Se requiere un reporte con el número de pacientes atendidos por financiador	1 hora	1 día

- **Ofrecer la solución completa:** Proporcionar los reportes solicitados mediante una fecha propuesta para que tengan los informes completos.

Para el buen valor de los datos utilizaremos un motor de base de datos SQL donde los datos serán almacenados de manera masiva creando un big data.

Figura N°19: Motores de Base de Datos



Tareas del Ciclo de Vida de Kimball

1. Planificación

En este proceso determinaremos el propósito la tesis, los objetivos y el alcance de los reportes, como los principales riesgos y una aproximación inicial a las necesidades de la información requerida, en este proceso tendremos las siguientes acciones:

a) Definir el alcance

- **R1: Reporte de atención de pacientes por estrategia**

En este primer requerimiento nos quiere decir que el reporte solicitado tenga los filtros por estrategias ya sean por: Medicina General, Obstetricia, Nutrición, Psicología, y otras estrategias que se puedan abrir con el pasar del tiempo en el centro de salud.

- **R2: Reporte de atención de pacientes por personal de salud**

En este requerimiento nos referimos a que nuestro filtro para la generación del reporte serán las personas que atienden en el centro de salud de pueblo nuevo, sean en este caso los siguientes: Médicos, Obstetras, Nutricionistas, enfermeras, psicólogas, odontólogas y otros profesionales que puedan llegar.

- **R3: Reporte de atención por grupo etario**

Cabe mencionar que en este requerimiento tenemos como filtros cuatro ítem especificados, en la cual son rangos de edades donde se dividen en: niños, joven, Adulto joven, Adulto Mayor, para ello si surge algún cambio en los grupos etarios nos basaremos a lo que se pueda indicar en el MINSA.

- **R4: Reporte N° Casos de tbc positivo**

En este requerimiento podremos detectar la cantidad de personas que están con diagnósticos positivos, para así ver la curva de contagios.

- **R5: Reporte N° de pacientes por genero**

En este requerimiento procederemos a realizar el filtro por género en este caso serán masculino y femenino, de igual manera, la cantidad será reflejada en un gráfico de barras.

- **R6: Reporte cantidad de pacientes atendidos por mes**

En este requerimiento visualizaremos la cantidad de pacientes atendidos por mes, verificando así la curva de elevación o disminución.

- **R7: Reporte con los 10 primeros diagnósticos más frecuentes**

En este requerimiento nos pide que se muestren los 10 diagnósticos más frecuentes que se presenten dentro del mes, la cuales son atendidos por el personal médico del establecimiento.

- **R8: Reporte N° Pacientes recuperados en TBC**

En este octavo requerimiento mostraremos la cantidad de pacientes recuperados por mes.

- **R9: Reporte N° diagnósticos TBC de manera trimestral**

En el noveno requerimiento se solicita un reporte que nos muestre la cantidad total de los diagnósticos, pero de manera trimestral.

- **R10: Reporte N° de pacientes atendidos en emergencias por personal médico:**

Se requiere la cantidad de pacientes atendidos, la cual el filtro será el personal médico que los atendió.

- **R11: Reporte por tipo de financiador**

En este último requerimiento se solicita la cantidad de pacientes atendidos, pero por financiador para así descubrir y ver quiénes son las personas que aun ni están afiliado al SIS o algún otro seguro.

Identificar las tareas

- **T1: Evaluaremos los procesos**

En este primer ítem de la tarea veremos los procesos a seguir para realizar paso a paso cada requerimiento que se nos ha solicitado.

- **T2: Realizaremos diagramas de procesos**

En la segunda tarea realizaremos diagramas para ver el proceso de cada requerimiento, donde mediremos los tiempos para cada uno.

- **T3: Optimizaremos base de datos**

Realizaremos el proceso de optimización de las tablas para que la base de datos no sea tan amplia.

- **T4: Analizaremos datos**

En este punto analizaremos los datos con sus distintos atributos a considerarse para la buena relación entre las tablas.

- **T5: Gestionaremos Recursos**

En esta tarea gestionaremos los recursos que utilizaremos sea tanto el software como el personal para el diseño de los requerimientos.

- **T6: Realizaremos pruebas**

Una vez gestionado y obtenidos nuestros recursos empezaremos con los avances de los requerimientos realizando pruebas de datos.

- **T7: Validaremos Reportes**

Una vez realizadas las pruebas validaremos los reportes con el personal encargado de la estrategia para la validación correspondiente.

- **T8: Estimaremos tiempos de entrega**

Una vez que tengamos los reportes validados, pasaremos a estimar el tiempo de entrega de los reportes, ya que depende mucho de la data y del tiempo que tenga el personal, ya que en ocasiones se presentan imprevistos o evaluación de otras funciones, tales como mejora en el sistema web, mantenimiento de la web, etc.

- **T9: Reportes Entregados**

En esta última tarea, una vez ya estimado el tiempo de entrega se procederá a entregar los reportes correspondientes según los requerimientos solicitados.

Programar las tareas

Para programar las tareas utilizaremos una herramienta de Microsoft para que se pueda reflejar el tiempo y espacio en cada tramo de las tareas que hemos visualizado anteriormente:

Cronograma del proyecto

Programación de las tareas:

Programación de las tareas con las personas que han intervenido en cada tarea programada:

Programación del estado de los recursos (Programación de las tareas)

- **Planificar el uso de los recursos**

Recursos Humanos:

- **Analista**
Se encargará de ver a detalle cada requerimiento solicitado, y ver las tareas en la cual determinará un tiempo de entrega.
- **Administrador de Base de datos**
Realizará la normalización de las tablas en cuanto a la base de datos proporcionada.
- **Jefe del proyecto**
Su función principal será que todo se trabaje en grupo, así obteniendo resultados favorables y en un corto tiempo.
- **Desarrollador**
Sera el encargado de unirse con el analista y el administrador de la base de datos para proceder con la implementación del datamart y realizar los reportes de acuerdo a los requerimientos.

Recursos Tecnológicos:

Para los recursos tecnológicos utilizaremos de dos a tres herramientas para el proceso del Datamart y estos son:

- **SQL Server**
Con la herramienta SQL Server procederemos a normalizar las tablas de la base de datos, para luego ingresar datos y así enlazar la base con otro sistema para la creación de los reportes, SQL nos permitirá que cada tabla tenga sus atributos y una vez realizada las tablas con sus respectivas relaciones,

tiene un plus que nos permite diagramar las relaciones entre las tablas.



- **Power BI**

Esta es otra de las herramientas que utilizaremos para la creación de los reportes, la ventaja de esta herramienta es que los reportes o gráficos creados los puedes publicar en la nube para que los puedas visualizar de cualquier lugar de donde te encuentres.

Figura N°20: Software para realizar datamart



- **Asignar la carga de trabajo a los recursos**

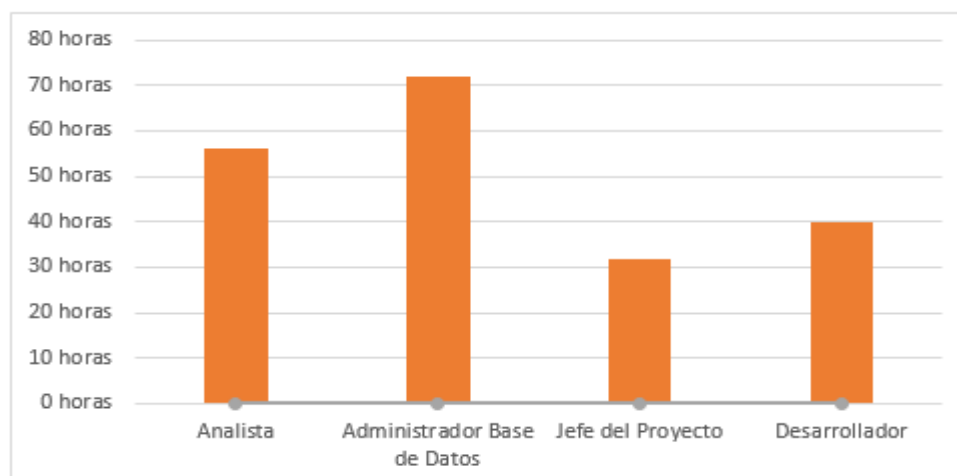
En esta grafica visualizaremos la carga de trabajo que tendrán nuestros recursos solicitados para la implementación del datamart en cuanto a los requerimientos propuestos y la programación de las tareas:

ROLES:

Gráficos N°07: Estadísticas de Recursos

ESTADÍSTICAS DE RECURSOS

Estado de trabajo de todos los recursos de trabajo.



Jefe del proyecto va de inicio a fin

Figura N°21: Estado de los Recursos

ESTADO DE LOS RECURSOS

Resta trabajo para todos los recursos de trabajo

Nombre	Comienzo	Fin	Trabajo restante
Analista	lun 01/02/21	jue 18/02/21	56 horas
Administrador Base de Datos	vie 05/02/21	jue 18/02/21	72 horas
Jefe del Proyecto	vie 12/02/21	mié 24/02/21	32 horas
Desarrollador	lun 15/02/21	vie 19/02/21	40 horas

- **Elaboración de un documento final**

En este último punto se creará un documento final donde estará especificado la función de los reportes y otros datos más como el uso de los recursos.

- **Sub-fase:** En este punto realizaremos la administración o gestión de cómo va el proceso de los requerimientos:
 - **Monitoreo del estado de los procesos y actividades:** Al diagramar los procesos de las tareas o de los requerimientos, estaremos monitoreando continuamente para que no tenga ningún desbalance al realizar cada proceso, realizarlo paso a paso sin saltarse ningún paso ya que ello afectaría y tendríamos problemas a la hora de ejecutar el siguiente proceso.

Cada estado de los procesos y actividades serán evaluados colocando así un puntaje de cómo hemos podido realizar y culminar lo asignado, para tener así los procesos calificados y tener el tiempo que se realizó.

- **Rastreo de problemas**

En caso surja un problema, localizaremos rápidamente donde se encontró la falla para poder corregir y seguir con los demás procesos para ellos seremos bien minuciosos con cada detalle en los procesos.

- **Desarrollo de un plan de comunicación comprensiva que direcciona la empresa y las áreas de TI**

Para esta tercera sub-fase tenemos un plan de comunicación comprensiva con la empresa y el área de TI, ya que en momentos inoportunos se puedan solicitar algunos reportes que ya se le ha proporcionado, pero por motivos secundarios ya no los encuentra, para ello el área de TI, tiene en mente que todos los reportes creadores y validados sean compartidos de momento en la nube, así disminuirémos el riesgo de que un personal no un duplicado de su informe.

1. Análisis de requerimientos

El proceso de análisis de requerimientos tiene principal objetivo entrevistar al personal que manejan las estrategias asignadas, de la misma manera que tenemos que observar los reportes que puedan solicitar a futuro. Parte de este proceso es validar a quien se debe realmente entrevistar, para ello verificaremos el organigrama de la empresa.

A partir de las entrevistas, podemos identificar temas analíticos y procesos de negocio:

Tabla N°14: Tema Analítico de Pacientes con TBC

Tema Analítico	Análisis o requerimiento inferido o pedido	Proceso de negocio de soporte	Comentarios
Pacientes con TBC	Análisis de pacientes con TBC	Tipo de Diagnósticos	Por género, por rango de edades
	Análisis de pacientes recuperados	Tipo de Diagnósticos	Por género, por rango de edades
	Proyección de pacientes recuperados por mes	Tipo de Diagnósticos	La proyección es un proceso la cual los pacientes se van recuperando según el tratamiento que van realizando, y así ver resultados y mostrar la cantidad de pacientes recuperados.

Tabla N°15: Temas analíticos de Top 10 de dx de morbilidad

Tema Analítico	Análisis o requerimiento inferido o pedido	Proceso de negocio de soporte	Comentarios
Top 10 de dx de Morbilidad	Análisis de morbilidad	Tipo de atención	Por género, por rango de edades, por tipo financiador
	Análisis de pacientes atendidos	Tipo de atención	Por género, por rango de edades, por tipo financiador
	Proyección del tipo de financiador	Tipo de atención	La proyección es un proceso en la cual realizaremos un seguimiento a los pacientes que son atendidos y que no cuentan con ningún tipo de seguro para así evaluarlo y afiliarlo al SIS.

Tabla N°16: Tablas de dimensiones

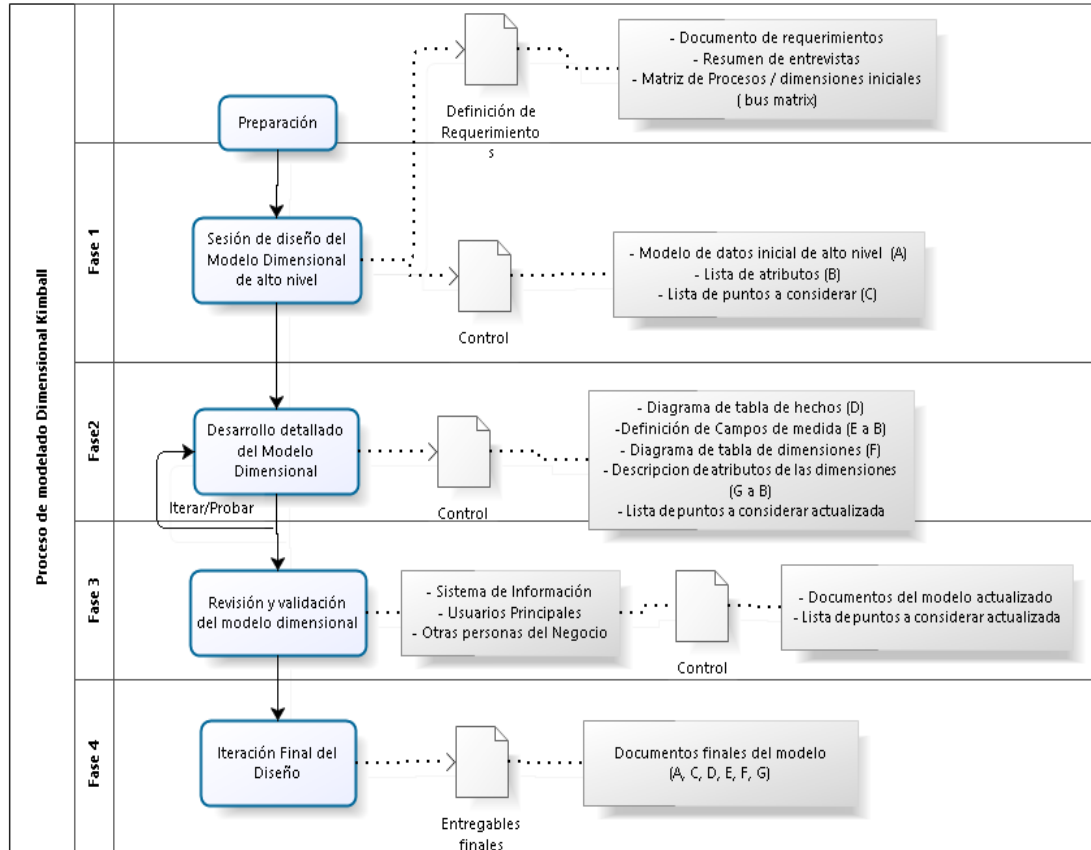
Una vez realizado el tema analítico se puede construir una herramienta metodológica denominada matriz de dimensiones:

Tema Analítico	Dimensiones					
	Financiad or	Diagnóstic os	Gener o	Grup o etare o	Tip o de dx	Profesion al
Proyección de atenciones	X	X	X	X	X	X
Pacientes Recuperados	X	X	X	X	X	X
Control de Pacientes Sintomáticos y Asintomáticos	X	X	X	X	X	X
Top 10 de dx de morbilidad	X	X	X	X	X	X

Matriz de proceso/dimensiones

2. Modelado Dimensional

Figura N°2: Modelado Dimensional



Fase 1: Sesión del diseño Modelo Dimensional de alto nivel

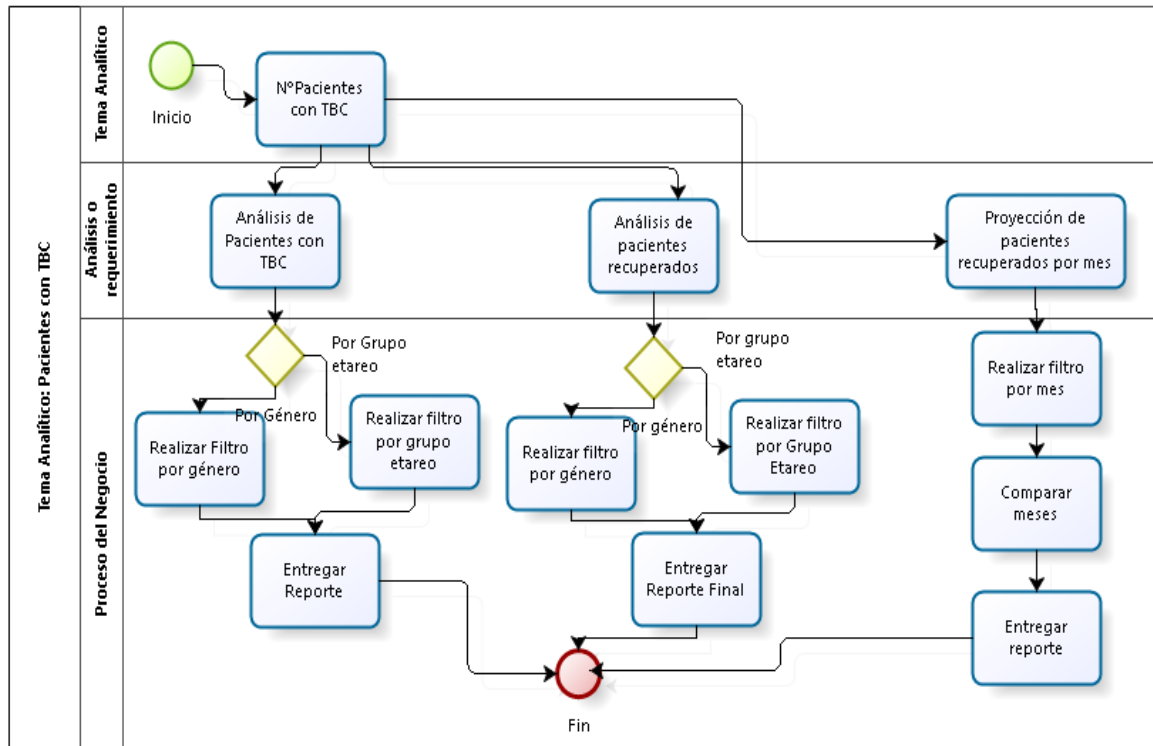
En el proceso del modelado dimensional el diseño comienza con un modelado dimensional de alto nivel obtenido a partir de los procesos priorizados de nuestra matriz, el proceso iterativo consiste en cuatro pasos:

- **Elegir el proceso de Negocio:**

En este proceso vamos a elegir el área a modelizar nuestros temas analíticos mencionados anteriormente:

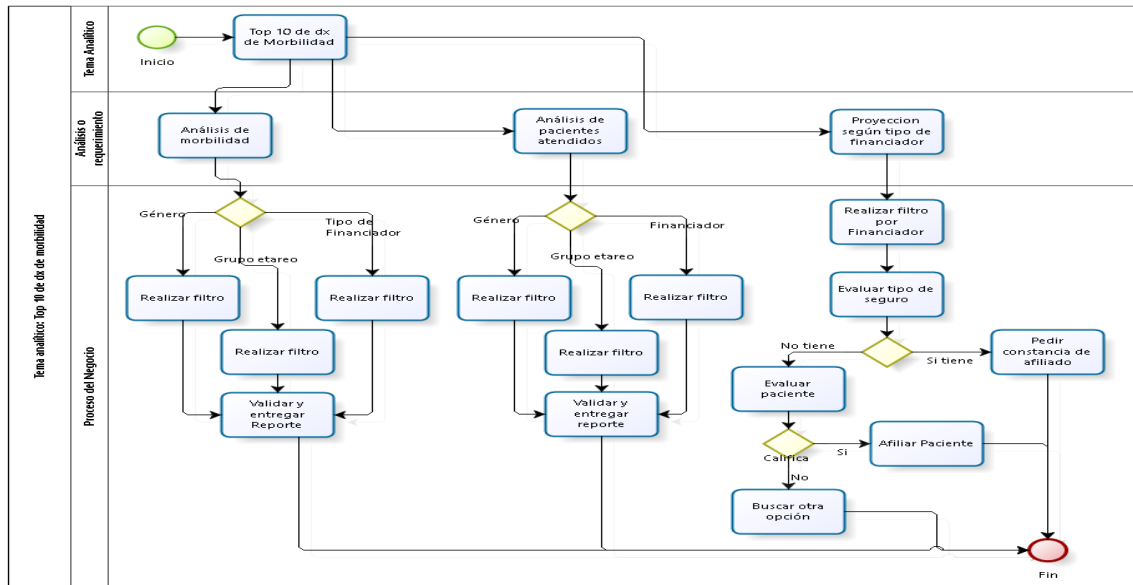
- **Pacientes con TBC:** En este primer tema analítico mencionado anteriormente, veremos varios puntos donde aplicaremos un proceso para la obtención de distintos reportes:

Figura N°23: Tema Analítico de Pacientes con TBC



- **Top 10 de Morbilidad:** En este siguiente proceso queremos visualizar el proceso para la obtención del top 10 de los Dx de morbilidad, para así poder ver cuáles son los diagnósticos más frecuentes por los cuales el paciente llega de emergencias.

Figura N°24: Temas analíticos de Top 10 de dx de morbilidad



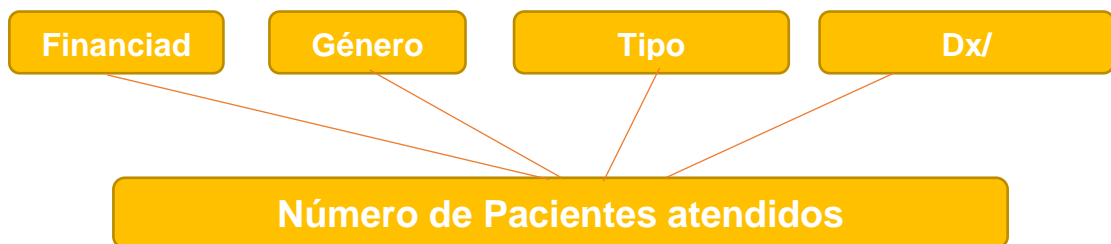
Powered by
bizagi
Modeler

- **Establecer niveles de granularidad**

Inmon (2005, p.41) define la granularidad como el nivel de detalle o somatización de las unidades de data, entre más detalle existe, menor es el nivel de granularidad. Los sistemas operacionales al registrar todas las transacciones de una organización, van a tener el nivel más bajo de granularidad, por ejemplo, una venta, el registro de un cliente, etc.

En el modelo dimensional es importante el término de granularidad, porque esto va a definir los indicadores o métricas dentro de la tabla de hechos. Un bajo nivel de granularidad permite mayor flexibilidad en la creación de reportes y manejo de los datos para generar nuevas agregaciones y análisis. Por tanto, en la figura, medir la cantidad de pacientes atendidos en un centro de salud provenientes con otro tipo de seguro, tipo de género, tipo de documento, Dx diferente.

Figura N°25: Niveles de Granularidad



- **Elegir las dimensiones**

En este proceso daremos a conocer las dimensiones mencionadas anteriormente, para entramos a un debate y vimos más detalle para las tablas dimensionales, así organizaremos mejor las consultas y para realizar el proceso utilizaremos un pequeño Power Pívor, o cubos para la visualización de forma unidimensional o multidimensional.

Por género

Tabla Nº17: Dimensión por género

Atendidos		
-	MASCULINO	FEMENINO
Centro de Salud de Pueblo Nuevo	1041	1404

Por Financiador

Tabla Nº18: Dimensión por financiador

	Centro de Salud Pueblo Nuevo
Sin seguro	2218
SIS	223
ESSALUD	1
Usuario	2
Exonerado	1

Por documento

Tabla Nº19: Dimensión por documento

Tipo de Documento	Cantidad
SIN DOCUMENTO	185
Con DNI	2101

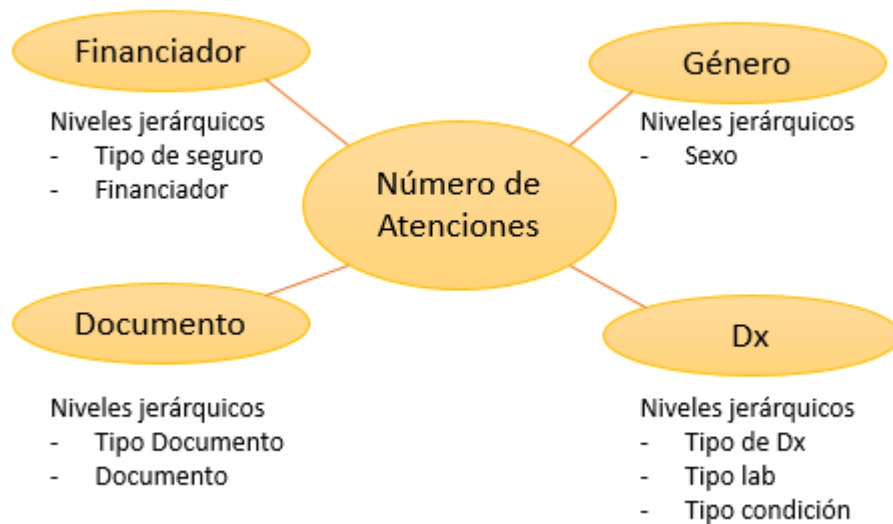
Por dx

Tabla N°20: Dimensión por dx

C-S-P-N	2282
J068	1170
N390	130
A084	106
I10X	95
S518	42
J029	37
T141	30
J209	30

- **Identificar medidas y tablas de hechos**
Como último paso, luego de realizar la granularidad y las dimensiones, se establecerán los hechos, que son más que todas las medidas en la cual nos será útil para un mejor análisis y toma de decisiones.

Figura N°26: Medidas y tablas de hechos



Fase 2: Desarrollo detallado del Modelo Dimensional

En esta segunda fase completaremos cada tabla con sus respectivas listas de atributos bien formados. Para iniciar desarrollaremos tres puntos claves:

- **Identificación de atributos de dimensiones y tablas de hechos**

La identificación de los atributos consta de la descripción de cada una de las tablas de hechos, para ello lo visualizaremos en unas tablas con sus respectivas descripciones:

Tabla N°21: Modelo Dimensional por Mes

Nombre de MES	
Tabla	
Descripción	Esta tabla contiene los meses del año

Nombre Columna	Descripción	Tipo de dato	Valor	Primary key?	Foreign key?	NULL?	Valores de ejemplo
ID_MES	El ítem o código que será la PK	int	()	X		NO	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
DES_MES	Descripción del mes	nvarchar	(11)			SI	Enero, Febrero, Marzo, Abril, Mayo, Junio, Julio, Agosto, Septiembre, Octubre, Noviembre, Diciembre

Tabla N°22: Modelo Dimensional por tipo de documento

Nombre de Tipo_Documento							
Tabla							
Descripción	Esta tabla representa el tipo de documento del paciente						
Nombre Columna	Descripción	Tipo de dato	Valor	Primaria y clave?	Foreign key?	NUL L?	Valores de ejemplo
ID_TIP_DOC	El ítem o código que será la PK	int	()	X		NO	1, 2, 3, 4, 5, 6
DES_TIP_DOC	Descripción del tipo de documento con el cual el paciente se ha registrado	nvarchar	(12)			SI	DNI, CE, PASS, DIE, S/DOCUMENTO, CNV

Tabla N°23: Modelo Dimensional por diagnostico

Nombre de Diagnóstico	
Tabla	
Descripción	La tabla contiene una infinidad de código de diagnósticos del CIE10

Nombre Columna	Descripción	Tipo de dato	Valor	Pri mar y key ?	Foreign key?	NUL L?	Valores de ejemplo
ID_DX	El ítem o código que será la PK	nvarchar	(11)	X		NO	99208, 99401, etc
DES_DX	Descripción de los código de CIE10	nvarchar	(350)			SI	Atención de planificación familiar, Consejería integral

Tabla N°24: Modelo Dimensional por tipo de diagnostico

Nombre de Tipo de Diagnóstico	
Tabla	
Descripción	La tabla contiene el tipo de diagnósticos según su estado de salud

Nombre Columna	Descripción	Tipo de dato	Valor	Primaria y clave?	Foreign key?	NUL L?	Valores de ejemplo
ID_TIP_DX	El ítem o código que será la PK	Int	()	X		NO	1, 2, 3
DES_TIP_DX	Describe el tipo de diagnóstico del cual se encuentra el paciente	nvarchar	(11)			SI	D, P, R

Tabla N°25: Modelo Dimensional por condición

Nombre de Condición	Tabla
Descripción	Esta tabla o entidad contiene el dato de la condición en que se encuentra el paciente

Nombre Columna	Descripción	Tipo de dato	Valor	Primaria y clave?	Foreign key?	NUL L?	Valores de ejemplo
ID_CON	El ítem o código	int	()	X		NO	1, 2, 3

	que será la PK						
DES_CON	Se describe el tipo de condición del paciente	nvarchar	(13)			SI	N, C, R

Tabla N°26: Modelo Dimensional por financiador

Nombre de Financiador	
Tabla	
Descripción	Esta tabla o entidad contiene el financiador o más conocido como el tipo de seguro con el cual cuenta el paciente

Nombre Columna	Descripción	Tipo de dato	Valor	Pri mar y key ?	Foreign key?	NUL L?	Valores de ejemplo
ID_CON	El ítem o código que será la PK	int	()	X		NO	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
DES_CON	El tipo de seguro con el que cuenta el paciente	nvarchar	(16)			SI	Usuario, SIS, ESSALUD, SOAT, SANIDAD FAP, SANIDAD NAVAL, ETC

Tabla N°27: Modelo Dimensional por UPS

Nombre de UPS (Unidad Productora de Servicio)	
Tabla	
Descripción	Esta entidad tenemos la descripción de las unidades productoras con las que cuenta el centro de salud

Nombre Columna	Descripción	Tipo de dato	Valor	Pri mar y key ?	Foreign key?	NUL L?	Valores de ejemplo
ID_UPS	El ítem o código que será la PK	char	(6)	X		NO	301612 303203 303304 302303 301203
DES_UPS	El tipo de seguro con el que cuenta el paciente	nvarchar	(100)			SI	Planificación Familiar, Obstetricia, Odontología General, Medicina General, Enfermería,

Tabla N°28: Modelo Dimensional por Personal

Nombre de Personal	
Tabla	
Descripción	Esta entidad tenemos a los trabajadores la cual prestan la atención al paciente, entre ellos médicos, enfermeros y más.

Nombre Columna	Descripción	Tipo de dato	Valor	Primaria y clave?	Foreign key?	NUL L?	Valores de ejemplo
DNI_PER	El ítem o código que será la PK	char	(8)	X		NO	40410169 43083034 21880774
NOM_APE	Se visualiza a los nombres del personal de atención	nvarchar	(90)			SI	AIDA GLADIS LEANDRO RAMIREZ, ALEXANDER REATEGUI GUERRA, AMELIA FERNANDA SORIA SARAVIA
PROFESION	Describe al área o estrategia la cual prestan sus servicios en la salud	nvarchar	(60)			SI	BIOLOGO, BIOLOGO, ODONTOLOGO

Tabla N°29: Modelo Dimensional por Paciente

Nombre de Paciente	
Tabla	
Descripción	Esta entidad pertenece a la lista de pacientes que son atendidos en el centro de salud

Nombre Columna	Descripción	Tipo de dato	Valor	Pri mar y key ?	Foreign key?	NUL L?	Valores de ejemplo
DNI_PAC	El ítem o código que será la PK	Nvarchar	(17)	X		NO	73124852
APE_PAT	Se visualizar a el apellido paterno del paciente	Nvarchar	(50)			SI	Munayco
APE_MAT	Se visualizar a el apellido materno del paciente	Nvarchar	(50)			SI	Jacobo
NOMBRES	Se visualizar a el	nvarchar	(70)			SI	Jeysson José

	ambos nombres del paciente						
FEC_NAC	La fecha de nacimiento del paciente	Nvarchar	(11)			SI	06/03/1996

Tabla N°30: Modelo Dimensional por atenciones

Nombre de Atenciones	
Tabla	
Descripción	En esta entidad podemos adjuntar todas las tablas mencionadas anteriormente, visualizando aquí las PK y las FK

Nombre Columna	Descripción	Tipo de dato	Valor	Pri mar y key ?	Foreign key?	NUL L?	Valores de ejemplo
RENI	Es el código de establecimiento de salud	Nvarchar	(11)			SI	000003419
FEC_ATE N	La fecha en que el paciente es atendido	Nvarchar	(11)			NO	DD/MM/AAA A

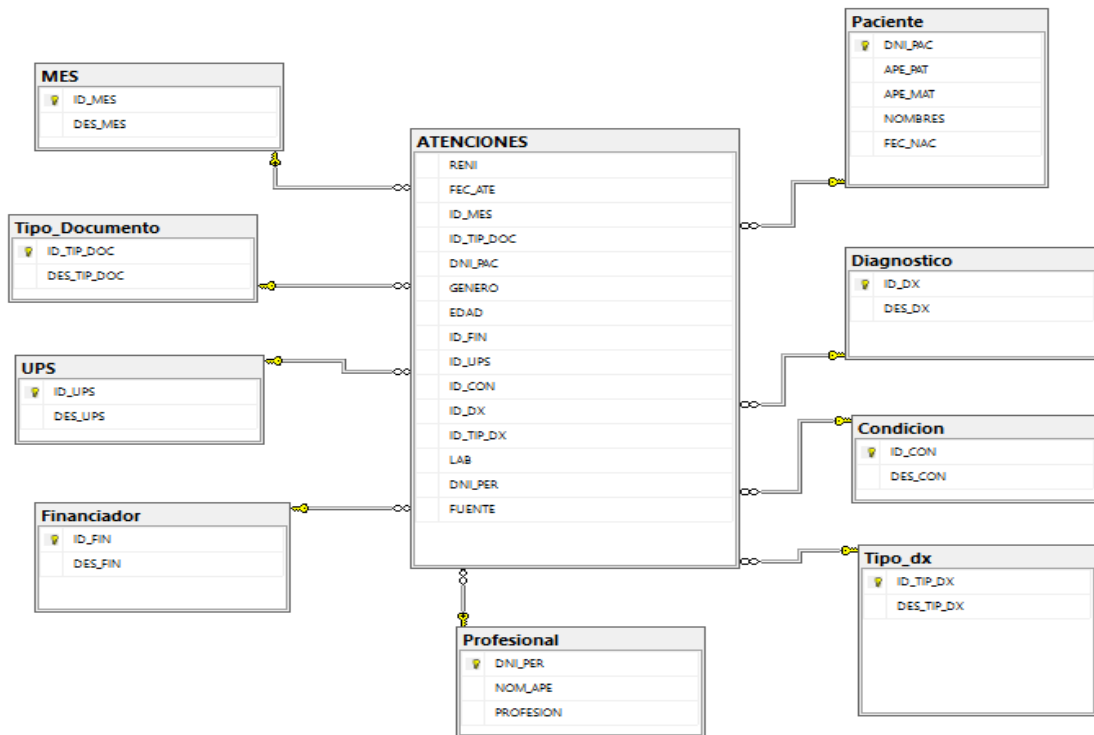
ID_MES	Código del mes según la fecha de atención	Int	()		X	NO	Tabla N°
ID_TIP_DOC	Código único del tipo de documento del paciente	Int	()		X	NO	Tabla N°
DNI_PAC	Código único de identidad del paciente	Nvarchar	(17)			NO	90909090
GENERO	Código para ver el género del paciente	Nvarchar	(10)		X	NO	Tabla N°
EDAD	Edad actual del paciente	Int	()			NO	AÑOS
ID_FIN	Código del financiador o tipo de seguro del paciente	Int	()		X	NO	Tabla N°
ID_UPS	Código de la	char	(6)		X	NO	Tabla N°

	Unidad de Productora de Servicio						
ID_CON	Código de la condición del paciente	Int	()		X	NO	Tabla N°
ID_DX	Código del diagnóstico a encontrar	Nvarchar	(11)		X	NO	Tabla N°
ID_TIP_DX	Código del tipo de diagnóstico del paciente	Int	()		X	NO	Tabla N°
LAB	Tipo de lab	Nvarchar	(17)			SI	Tabla N°
DNI_PER	Documento de identidad del personal que atendió al paciente	Char	(8)		X	NO	Tabla N°
FUENTE	Fuente de donde	nvarchar	(12)			NO	HIS

	se exporto la data						EMERGENCIAS
--	--------------------------	--	--	--	--	--	-------------

- Implementar el modelo dimensional detallado**
 Este punto completaremos la información del paso anterior, en la cual se mostrará el modelo dimensional ya con sus entidades, atributos y con sus relaciones indicadas, mostrados las PK (Primary key) y FK (Foreign Key)

Figura N°27: Modelo BD dimensional

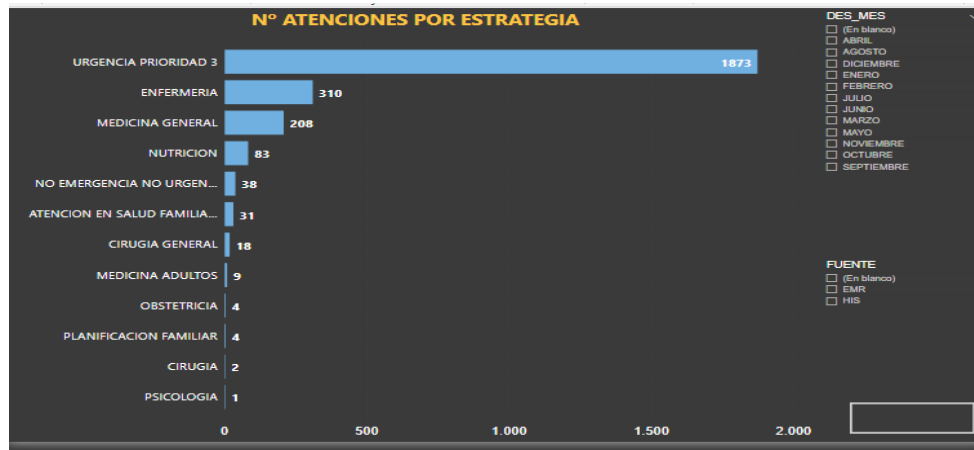


- Prueba del Modelo**
 En este punto como ya tenemos estable nuestro modelo dimensional, visualizaremos algunas pruebas según los requerimientos mencionados anteriormente, dando prueba de que nuestra tabla de

hechos está realizada de una manera correcta con datos para luego poder otorgar los requerimientos solicitados:

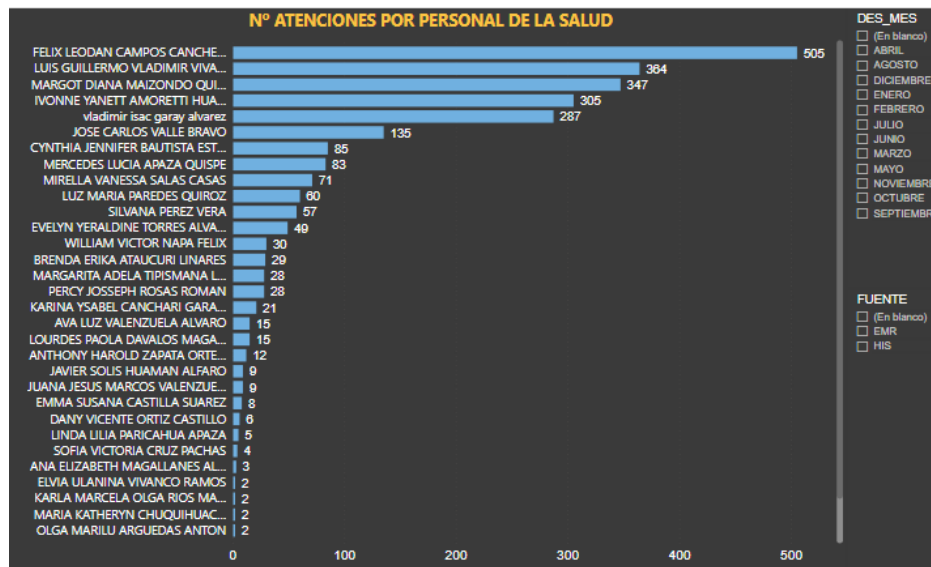
R1: Reporte de Atención de pacientes por estrategias

Figura N°28: Reporte N° de atenciones



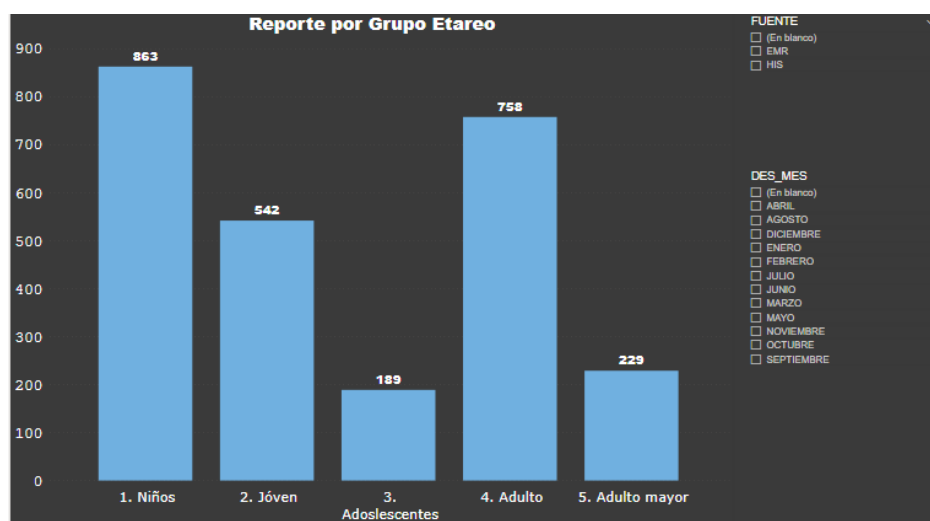
R2: Reporte de atención de pacientes por personal de salud

Figura N°29: Reporte N° de atenciones por personal



R3: Reporte de atención por grupo etáreos

Figura N°30: Reporte por Grupo etareo



Fase 3: Revisión y validación Modelo Dimensional

En este tercer punto revisaremos y validaremos el modelo dimensional, para ello lo dividiremos en tres partes las cuales son:

- **Sistemas Información**

El sistema de información es un conjunto de datos que interactúan entre sí para obtener un resultado, en la cual vamos a administrar, recolectar, recuperar, procesar, almacenar y distribuir información importante para los procesos que se puedan presentar.

Para un sistema de información tenemos 4 componentes que los conforman las cuales son:

La entrada: Este punto nos menciona por donde se alimentarán nuestros datos, en este caso mencionaremos la exportación de la data web del Minsa la cual nos será útil para el procesado de los datos.

El proceso: Para el proceso utilizaremos la herramienta de Power bi como visualizador de reportes.

Salida: La salida serán los reportes que vamos a generar mediante el modelo dimensional que hemos creado en base a los datos exportados del Minsa.

La retroalimentación: La retroalimentación serán los datos que exportamos del Minsa, pero de manera consecutiva o mensual ya que

el SQL será un motor de base de datos fundamental para la cantidad de datos almacenados que puedan existir.

- **Usuarios Principales**

En este punto se mencionará la lista de los usuarios principales la cual estarán activamente interactuando con el datamart que sea creado:

- Jefe de estadística
- Auxiliares estadísticos

- **Otras personas del negocio**

Como personas secundarias del negocio estarán interactuando solo el jefe del establecimiento a cargo, ya que los reportes estarán subidos a la web y así de donde este podrá visualizar los reportes para cualquier información que pueda requerir de manera inmediata.

Fase 4: Iteración final del diseño

En esta última fase del modelado dimensional se dará un resumen de los documentos finales a presentarse, en la cual mencionaremos el orden en que se presentará:

- **Modelo de datos inicial de alto nivel**

Figura N°31: Modelo de datos inicial de alto nivel



- **Lista de atributos**

Tabla N°31: Atributos mes y UPS

MES		UPS	
ID_MES		ID_UPS	
DES_MES		DES_UPS	

Tabla N°32: Atributos financiador y tipo de documento

FINANCIADOR		Tipo Documento	
ID_FIN		ID_TIP_DOC	
DES_FIN		DES_TIP_DOC	

Tabla N°33: Atributos Paciente y Profesional

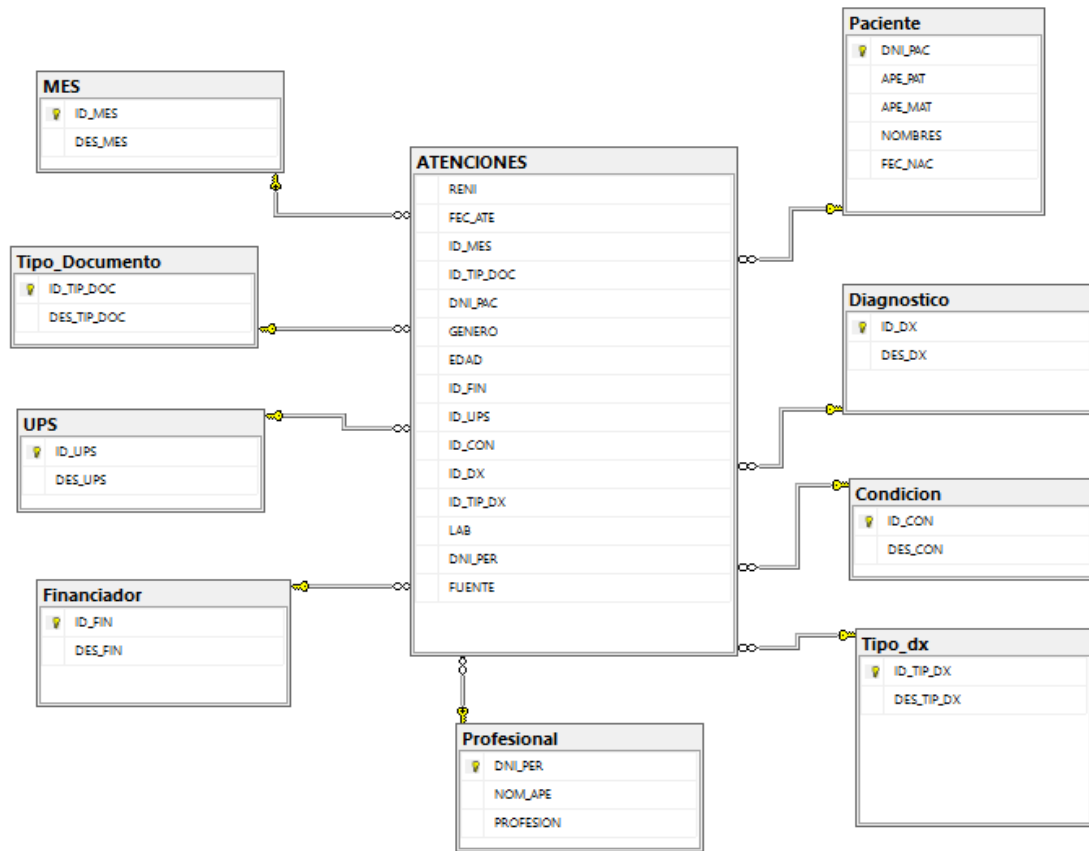
PACIENTE		PROFESIONAL	
DNI_PAC		DNI_PER	
APE_PAT		DES_TIP_DOC	
APE_MAT		NOM_APE	
NOMBRES		PROFESIÓN	
FEC_NAC			

Tabla N°34: Atributos Condición, Diagnostico y tipo de diagnostico

CONDICIÓN		DIAGNOSTICO		TIPO_DX	
ID_CON		ID_DX		ID_TIP_DX	
DES_CON		DES_DX		DES_TIP_DX	

- **Diagrama de tabla de hechos**

Figura N°32: Diagrama de tabla de hechos



- **Definición de campos de medida**

- **Int:** El tipo de dato entero alberga números enteros pequeños, en la cual tiene un valor de 32 bits y pueden representarse en números enteros positivos o negativos.
- **Char:** La medida char, almacena datos de carácter fijo, donde los datos también pueden ser alfanuméricos. Ejemplo: Si colocamos char (8) en este caso tomaremos los números del DNI que son exactamente 8 dígitos, el campo debe contener los 8 dígitos exacto, que pueden ser entre letras y números.
- **Nvarchar:** El tipo de medida nvarchar almacena datos de carácter en un campo de longitud variable, donde los datos pueden ser alfanuméricos. Ejemplo: Si tenemos un atributo llamado nombres, y le colocamos nvarchar (20),

- Diagrama de tablas de dimensiones

Figura N°33: Diagrama de tablas de dimensiones



- Descripción de los atributos de las dimensiones

- **Dimensión 1: Mes**

- **ID_MES**: Hace referencia al número según el mes.
- **DES_MES**: Nos muestra la descripción del número del mes.

- **Dimensión 2: Tipo documento**

- **ID_TIP_DOC**: Hace referencia a un numero la cual es relacionada con el tipo de documento.
- **DES_TIP_DOC**: Es la descripción del tipo de documento según el ID que se nos muestre.

- **Dimensión 3: UPS**

- **ID_UPS**: Nos muestra un código de 6 dígitos que es la identidad del UPS.
- **DES_UPS**: Nos muestra la descripción según el ID_UPS agregado.

- **Dimensión 4: Financiator**

- **ID_FIN**: Este campo nos da un valor entero la cual mediante ello representa el tipo de financiator al que pertenece el paciente.
- **DES_FIN**: Muestra la descripción del Tipo de financiator que pueda ser el paciente que es atendido.

- **Dimensión 5: Tipo_dx**
 - **ID_TIP_DX:** Este es un dato de valor entero la cual nos representa el mediante un número el tipo de diagnóstico.
 - **DES_TIP_DX:** Nos muestra la descripción del tipo de diagnóstico.
- **Dimensión 6: Condición**
 - **ID_CON:** Este atributo también es un valor entero la cual está representado por tres números que hace referencia a las condiciones.
 - **DES_CON:** Este atributo nos muestra la descripción de la condición según el número mostrado.
- **Dimensión 7: Profesional**
 - **DNI_PER:** Es un numero de 8 dígitos la cual representa la identidad del profesional a cargo en la atención.
 - **NOM_APE:** Representa al nombre y apellido del personal en un solo campo.
 - **PROFESIÓN:** Nos muestra a que profesión pertenece el personal que está en su momento atendiendo.
- **Dimensión 8: Paciente**
 - **DNI_PAC:** Nos representa la identidad del paciente, pero este puede variar sus dígitos ya que hay personas que llegan con otro tipo de documento y constan de más de 8 dígitos.
 - **APE_PAT:** Muestra el apellido paterno de la persona.
 - **APE_MAT:** Muestra el apellido materno de la persona.
 - **NOMBRES:** Muestra los nombres que pueda tener la persona.
 - **FEC_NAC:** Este atributo nos muestra la fecha de nacimiento de la persona atendida.
- **Dimensión 9: Diagnostico**
 - **ID_DX:** Es el código que hace referencia al diagnóstico que se le detecta al paciente.
 - **DES_DX:** Es la descripción del código del diagnóstico.
- **Matriz DW (o DW Bus Matrix) Completa**

Tabla N°35: Matriz Dw o bus

PROCESOS DEL NEGOCIO	DIMENSIONES								
	MES	TIP DOC	UPS	FINANCIADOR	TIPO DX	CONDICIÓN	PROFESIONAL	PACIENTE	DIAGNÓSTICO
Nº de Pacientes	X	X		X					X
Nº de atenciones	X		X				X		
Nº de afiliados	X			X					
Nº de pacientes nuevos	X		X			X			
Top 10 Dx general	x						x		x
Top 10 Dx	x		x				x		x

1. Diseño Físico

En esta tarea nos realizaremos las siguientes preguntas:

- **¿Cómo puede determinar cuán grande será el sistema del datamart?**

La medida para verificar que tan grande pueda ser el datamart que realizaremos es visualizar las áreas específicas del negocio, en la cual observaremos donde se almacenan los datos que queremos obtener para la elaboración del datamart.

El datamart será alimentado de una fuente de información web.

- **¿Cuáles son los factores que llevaran a una configuración más grande y más compleja?**

Los factores son los siguientes:

- Cantidad de Información almacenada
- Fuente de información
- Áreas específicas del negocio
- Herramienta a utilizar

- **¿Cómo se debe configurar el sistema?**

Para nuestro datamart la configuración del sistema seguirá un procedimiento factible para el uso del usuario en la cual los puntos son los siguientes:

- Enlazar SQL con Power BI
- Reportes automatizados

- **¿Cuánta memoria y servidores se necesitan? ¿Qué tipo de almacenamiento y procesadores?**

En temas de equipamiento físico Hardware se necesitará lo siguiente para solo un servidor:

- Un ordenador HP
- Una memoria RAM 8GB
- Disco Duro de 500 GB
- Procesador Mayor a Intel i5 de decima generación

- **¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?**

Para ello necesitaremos de dos softwares que son el SQL y el Power BI, mencionaremos paso a paso la instalación de cada uno:

SQL

- Descargaremos de la web el SQL de su preferencia
- Ejecutaremos el software para su instalación correspondiente
- Seguiremos lo pasos a seguir
- Configuraremos los usuarios y contraseñas
- Ejecutaremos el sistema

Power BI

- Descargaremos el software desde la página oficial para el modo escritorio
- Una vez terminada la descarga procederemos a la instalación de la misma
- Doble click instalar
- Verificar cada paso y presionar siguiente
- Luego Finalizo y ya está el programa para trabajar

- **¿Qué necesitan instalar los diferentes miembros del equipo de DW/BI en sus estaciones de trabajo?**

No se va a requerir instalar nada ya que los reportes que vamos a generar serán colgados en la nube la cual los usuarios lo podrán utilizar de manera rápida la visualización de algún producto.

- **¿Cómo convertir el modelo de datos lógico en modelo de datos físico en la base de datos relacional?**

Para convertir el modelo lógico primero debemos de definir bien nuestras dimensiones con sus respectivos atributos, una vez culminado la obtención de las dimensiones, pasaremos a realizar la tabla de hechos, la cual tenga una unión con las dimensiones ya creada.

- **¿Cómo conseguir un plan de indexación inicial?**
Para ello debemos visualizar las cantidades de tablas que tendremos para poder realizar nuestra base de datos, asignando de manera correcta los Primary Key y las Foreign Key

2. Diseño del ETL

Este proceso ETL, consta en la Extracción, transformación y carga de los datos que se puedan obtener, su principal indicador es la calidad del dato. Para realizar el proceso del ETL pasaremos a desglosarlo para dar una mayor explicación modelándolo al diseño de nuestro datamart. Para la implementación seguiremos tres sencillos pasos:

Paso 1: Extracción

El primer paso antes de comenzar a organizar, es extraer los datos de las fuentes donde se encuentren almacenada la información, las fuentes pueden ser obtenidas de base de datos ya preinstaladas, almacenes de datos en la nube, información de un sistema web o escritorio y cualquier otra fuente de la que queremos obtener información a través del procesamiento analítico.

Una vez que hallamos consolidado la información, notaremos que cada fuente tiene su estructura en diferentes formatos, en este proceso debemos organizar los datos para adaptarse al proceso de transformación.

Pasos extracción ETL

- **Compilar datos de fuentes relevantes**
En este caso nuestras fuentes será de un sistema web y un sistema de escritorio la cual ambos tienen distintas estructuras.
- **Organiza los datos para que sean consistentes**

En este paso realizaremos la unificación de las estructuras para poder trabajar con ambas en un solo archivo.

Proceso Extracción BD HIS

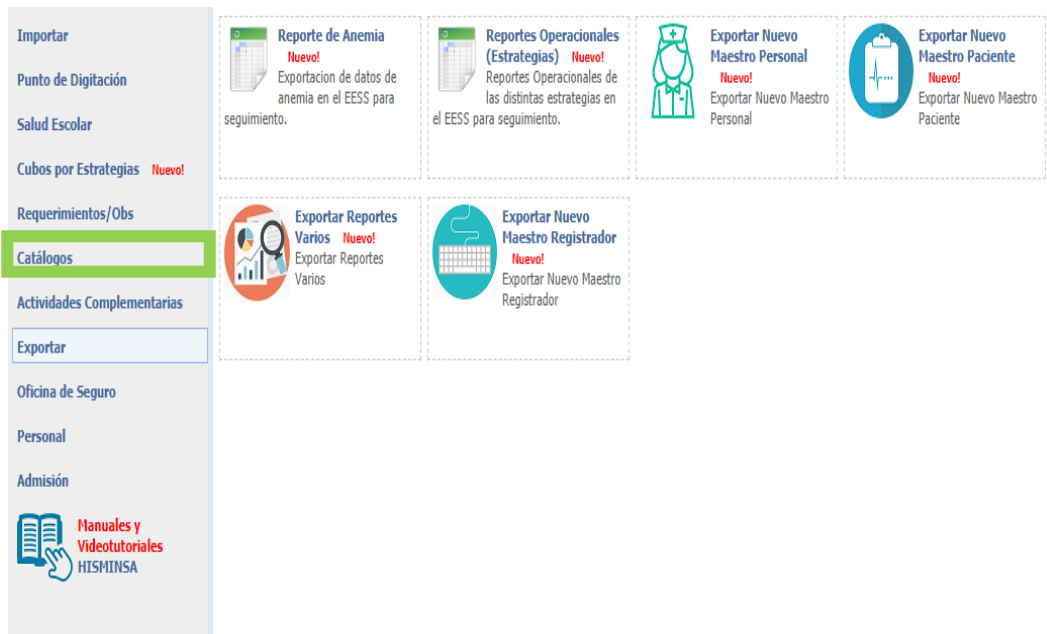
1. Nos dirigimos a la página web de HISMINSA.

Figura N°34: Interfaz de HISMINSA



2. Iniciamos sesión y nos dirigimos hacia la pestaña exportar

Figura N°35: Modulo de catálogos



3. Exportamos y obtenemos el archivo

Figura N°36: Datos del HISMINSA

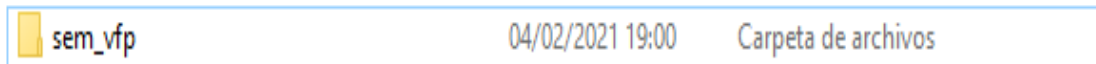
ID	Cita	Año	Mes	Día	Fecha	Atenció	Lote	Páam	Re	Id	Ups	Descripción	Sector	Ipcior	Descripción	Red	Descripción	MicroRed	Codigo	Unico mbre	Establecimie	fipo	Doc	Po	Docum
443782833	2020	12	31	2020-12-31	242	14	2	301612				PLANIFICACION	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			71432040	
443782833	2020	12	31	2020-12-31	242	14	2	301612				PLANIFICACION	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			71432040	
437641830	2020	12	11	2020-12-11	WAT	11	1	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			47888303	
437641830	2020	12	11	2020-12-11	WAT	11	1	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			47888303	
437641830	2020	12	11	2020-12-11	WAT	11	1	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			47888303	
437641830	2020	12	11	2020-12-11	WAT	11	1	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			47888303	
443323462	2020	12	28	2020-12-28	186	6	13	303304				ODONTOLOGIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			77768316	
443323462	2020	12	28	2020-12-28	186	6	13	303304				ODONTOLOGIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			77768316	
443323462	2020	12	28	2020-12-28	186	6	13	303304				ODONTOLOGIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			77768316	
443323462	2020	12	28	2020-12-28	186	6	13	303304				ODONTOLOGIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			77768316	
441704776	2020	12	22	2020-12-22	228	6	11	302303				MEDICINA GENEF	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			42827249	
441896123	2020	12	28	2020-12-28	WAT	11	11	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			44879762	
441896123	2020	12	28	2020-12-28	WAT	11	11	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			44879762	
441896123	2020	12	28	2020-12-28	WAT	11	11	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			44879762	
441896123	2020	12	28	2020-12-28	WAT	11	11	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			44879762	
441896123	2020	12	28	2020-12-28	WAT	11	11	303203				OBSTETRICIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			44879762	
437859736	2020	12	12	2020-12-12	CAR	13	6	301203				ENFERMERIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			50442340	
437859736	2020	12	12	2020-12-12	CAR	13	6	301203				ENFERMERIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			50442340	
437859736	2020	12	12	2020-12-12	CAR	13	6	301203				ENFERMERIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			50442340	
437859736	2020	12	12	2020-12-12	CAR	13	6	301203				ENFERMERIA	GOBIERNO REGIONAL	ICA	CHINCHA - PISCO	PUEBLO NUEVO			000003419	PUEBLO NUEVO	DNI			50442340	

Figura N°37: Proceso Extracción BD EMERGENCIAS

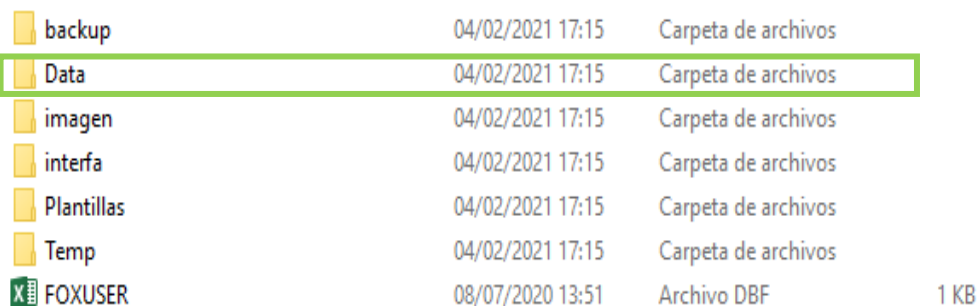
1. Nos dirigimos a la unidad C



2. En la unidad C buscaremos la carpeta Sem_Vfp



3. Abrimos y nos dirigimos a la carpeta Data



4. Seleccionamos el archivo llamado Emergencias DBF

<input checked="" type="checkbox"/>	colegios	16/07/2015 9:51	Archivo DBF	1 KB
<input checked="" type="checkbox"/>	cond_egre	18/12/2015 14:18	Archivo DBF	1 KB
<input checked="" type="checkbox"/>	condiciones	22/02/2004 19:26	Archivo DBF	1 KB
<input type="checkbox"/>	config.BAK	14/12/2015 18:10	Archivo BAK	1 KB
<input checked="" type="checkbox"/>	config	14/08/2020 13:13	Archivo DBF	2 KB
<input checked="" type="checkbox"/>	destinos	26/01/2016 14:24	Archivo DBF	1 KB
<input checked="" type="checkbox"/>	disa	08/01/2021 8:23	Archivo DBF	2 KB
<input type="checkbox"/>	disa	08/01/2021 8:23	SQL Server Replica...	1 KB
<input checked="" type="checkbox"/>	dist	08/01/2021 8:23	Archivo DBF	243 KB
<input type="checkbox"/>	dist	08/01/2021 8:23	SQL Server Replica...	20 KB
<input checked="" type="checkbox"/>	dpto	14/04/2016 9:10	Archivo DBF	2 KB
<input type="checkbox"/>	dpto	08/01/2021 8:23	SQL Server Replica...	1 KB
<input checked="" type="checkbox"/>	egresos	08/01/2021 8:23	Archivo DBF	22 KB
<input checked="" type="checkbox"/>	egresos imp	09/03/2016 11:46	Archivo DBF	2 KB
<input type="checkbox"/>	emergencias.BAK	14/08/2020 13:16	Archivo BAK	710 KB
<input checked="" type="checkbox"/>	emergencias	04/02/2021 15:37	Archivo DBF	1.292 KB
<input checked="" type="checkbox"/>	emergencias_imp	07/04/2016 18:22	Archivo DBF	3 KB
<input checked="" type="checkbox"/>	ESTABLEC	15/12/2015 8:35	Archivo DBF	1.198 KB
<input checked="" type="checkbox"/>	etareo	06/01/2016 11:24	Archivo DBF	2 KB
<input type="checkbox"/>	etareo	08/01/2021 8:23	SQL Server Replica...	1 KB
<input checked="" type="checkbox"/>	etnias	15/09/2015 12:04	Archivo DBF	5 KB
<input type="checkbox"/>	Etnias	08/01/2021 8:23	SQL Server Replica...	1 KB
<input checked="" type="checkbox"/>	etnias_grp	15/09/2015 12:04	Archivo DBF	1 KB

5. Abrimos el archivo y nos mostrara los datos

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
RENIPRESS	E_UBIG	CC	CC	CC	FECATE	HORA	NUMHC	DOC_IDEN	ETI	FI	SIEDA	T	NOMB		APELL	
0000003419	110207	15	02	02	#####	08	15	001	178456520	50	01	1	26	1	LUIS	DIRECC
0000003419	110207	15	02	02	#####	09	15	002	178929902	50	01	2	4	1	LUHANA	LOS LAU
0000003419	110207	15	02	02	#####	09	20	003	145260027	50	01	2	31	1	REGINA	HUATA MARTINEZ
0000003419	110207	15	02	02	#####	10	25	004	0	50	01	1	2	1	MAX	AV PRIN
0000003419	110207	15	02	02	#####	10	02	005	171836930	50	01	2	53	1	ANGELICA	A H VIL
0000003419	110207	15	02	02	#####	10	10	006	179484647	50	01	2	3	1	MARY	SEBAST
0000003419	110207	15	02	02	#####	10	15	007	121876377	50	01	1	43	1	JOSE	A H MIC
0000003419	110207	15	02	02	#####	10	20	008	145688458	50	01	2	35	1	ROSELI	A H MIC
0000003419	110207	15	02	02	#####	10	30	009	163058790	50	01	1	8	1	JOSE	A H SEB
0000003419	110207	15	02	02	#####	10	35	010	161298345	50	01	1	11	1	LEONARD	BARRIO
0000003419	110207	15	02	02	#####	10	40	011	145741858	50	01	1	30	1	DANIEL	C P CON
0000003419	110207	15	02	02	#####	10	50	012	178370960	50	01	2	6	1	LUHANA	HUAYA
0000003419	110207	15	02	02	#####	11	00	013	115357898	50	01	1	47	1	JAIME	A H LOS
0000003419	110207	15	02	02	#####	11	05	014	190982976	50	01	2	1	1	KIARA	JR CABE
0000003419	110207	15	02	02	#####	20	00	015	191590536	50	01	1	1	2	MAICHELL	AV PRIN
0000003419	110207	15	02	02	#####	20	10	016	144387905	50	01	2	35	1	KATHERIN	TASAYCO SANCHEZ
0000003419	110207	15	02	02	#####	20	15	017	178401902	50	01	1	27	1	CARLOS	BENAVI
0000003419	110207	15	02	02	#####	20	20	018	170340350	50	01	1	26	1	ERIKSON	JR CABE
0000003419	110207	15	02	02	#####	20	25	019	179599042	50	01	1	3	1	DAVID	MICHAEL
0000003419	110207	15	02	02	#####	20	30	020	121857054	50	01	2	56	1	GLADYS	TUPAC
															VALENZUELA ALVARO	PROGR
																BENAVI

Paso 2: Transformación

En este proceso del ETL en el almacén de datos incluye los siguientes pasos durante la transformación de los datos:

- **Convertir datos de acuerdo a los requerimientos del Negocio**
Para ello procederemos a convertir los datos en un solo Excel, ya sean valores enteros o valores de variables. Para ello verificaremos ambas bases de datos y verificaremos todos los campos para unificar los datos en los archivos obtenidos.
- **Reformatee los datos convertidos a un formato estándar para compatibilidad**

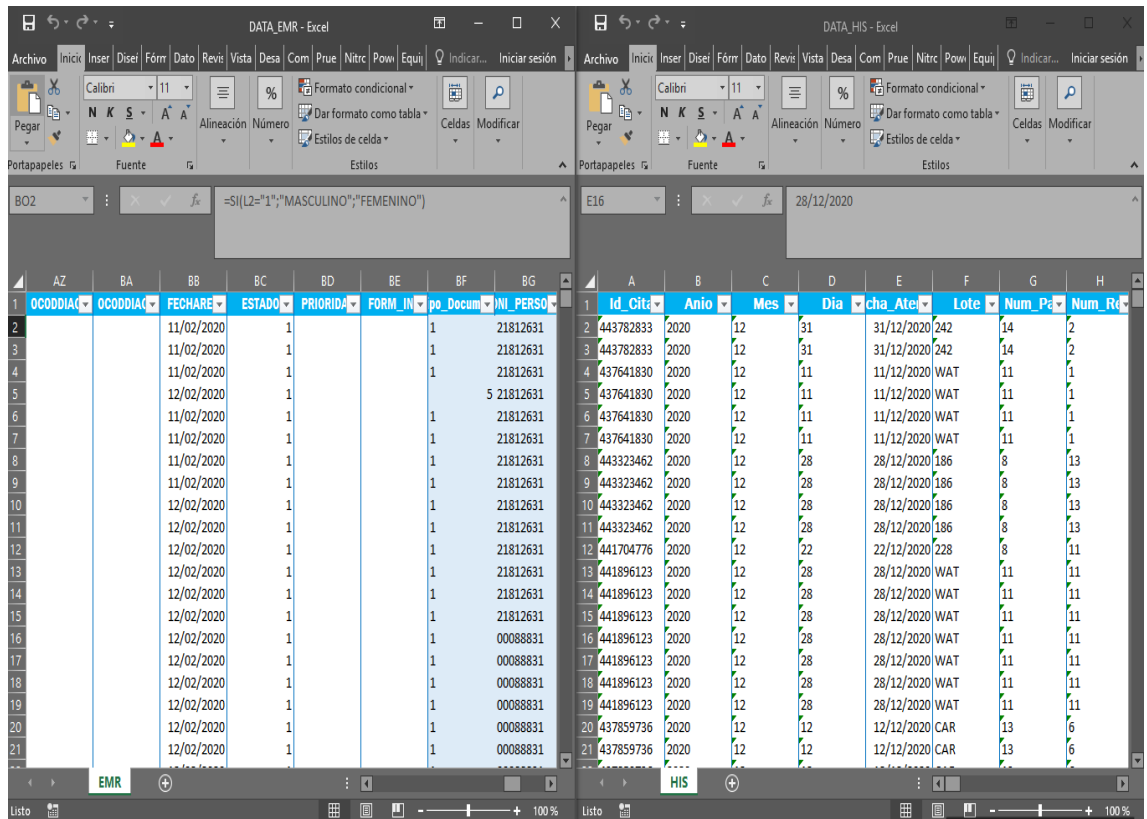
En este proceso es muy importante reformatear los datos ya que en Excel los datos al pasarlo a un sistema se modifican para ellos realizaremos en un Query la unificación de los datos.

- **Limpiar los datos irrelevantes de los conjuntos de datos**
En este punto eliminaremos campos que no se utilizaran para el proceso del Datamart en la cual seguiremos tres simples pasos.
 - **Ordenar y filtrar datos**
Ordenaremos y filtraremos los datos dependiendo de la información que vamos a obtener.
 - **Borrar información**
Eliminaremos los campos que no utilizaremos para realizar el proceso de la unificación.
 - **Traducir cuando sea necesario**
Si el dato o los campos no cuenta con una unificación debemos tener una plantilla para poder leer los datos que vamos a seguir exportando

Figura N°38: PROCESO DE TRANSFORMACIÓN

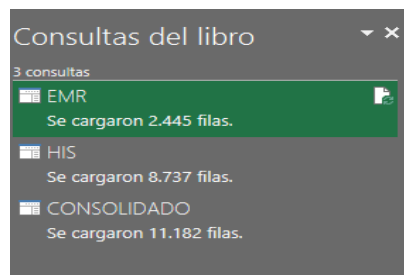
1. Mostraremos los Excel a transformar

Ambos Excel procederemos a transformar en un solo Excel



2. Unificaremos los encabezados

En este paso unificaremos los datos donde solo queremos visualizar los datos más relevantes

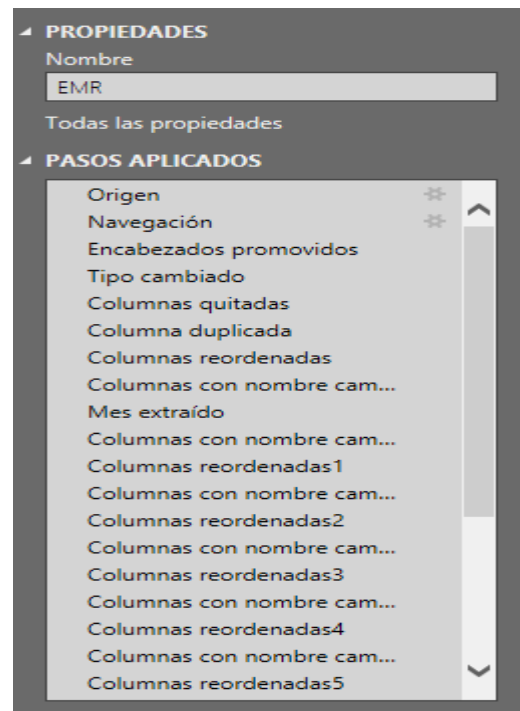
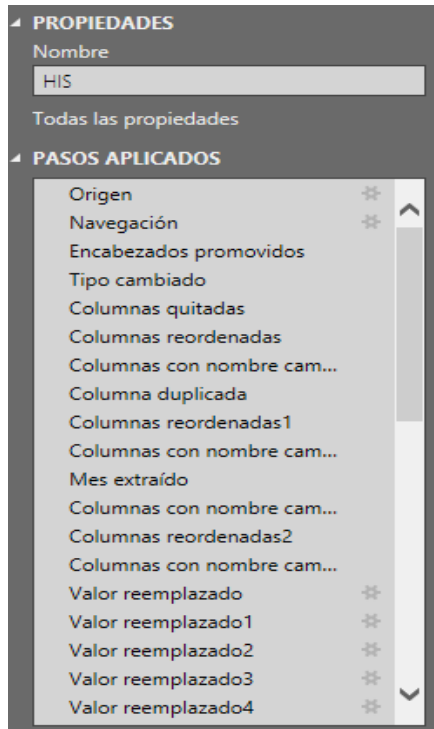


RENI	FEC_ATE	ID_MES	ID_TIP_DOC	DNI_PAC	GENERO	EDAD	ID_FIN	ID_UPS	ID_CON	ID_DX	ID_TIP_DX	LAB	DNI_PER	FUENTE
3419	01/01/2020	1	1	78456520	MASCULINO	26	1	230500	1	J068	1		21812631	EMR
3419	01/01/2020	1	1	78929902	FEMENINO	4	1	230500	1	J068	1		21812631	EMR
3419	01/01/2020	1	1	45260027	FEMENINO	31	1	230500	1	J029	1		21812631	EMR
3419	01/01/2020	1	5	4	MASCULINO	2	1	230500	1	J068	1		21812631	EMR
3419	01/01/2020	1	1	71836930	FEMENINO	53	1	230500	1	J068	1		21812631	EMR
3419	01/01/2020	1	1	79484647	FEMENINO	3	1	230500	1	J039	1		21812631	EMR
3419	01/01/2020	1	1	21876377	MASCULINO	43	1	230500	1	R509	1		21812631	EMR

3. Realizaremos la eliminación y filtros en cada archivo

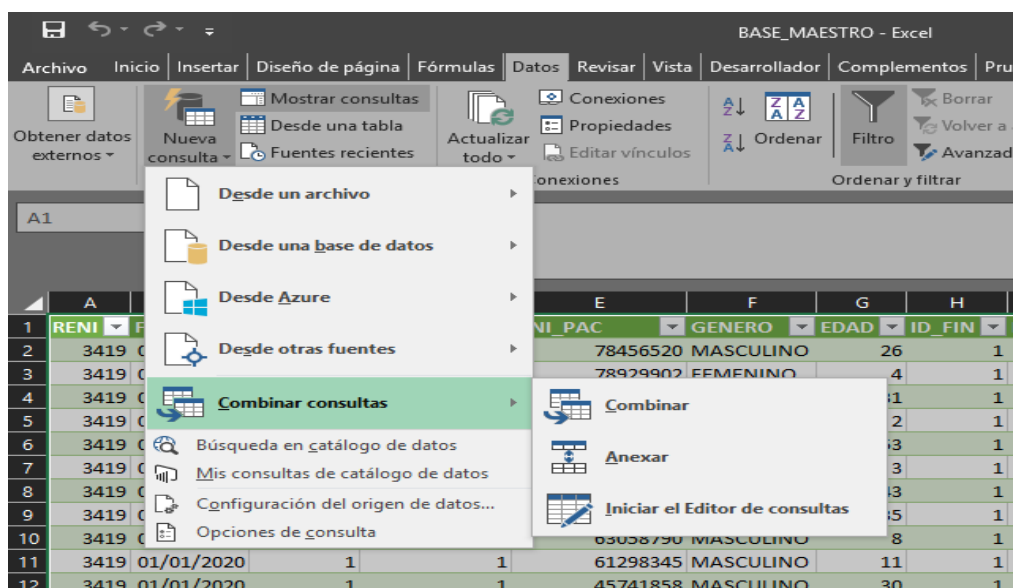
Aquí visualizaremos los datos y los filtros que vamos a realizar para la unificación del archivo.

HIS

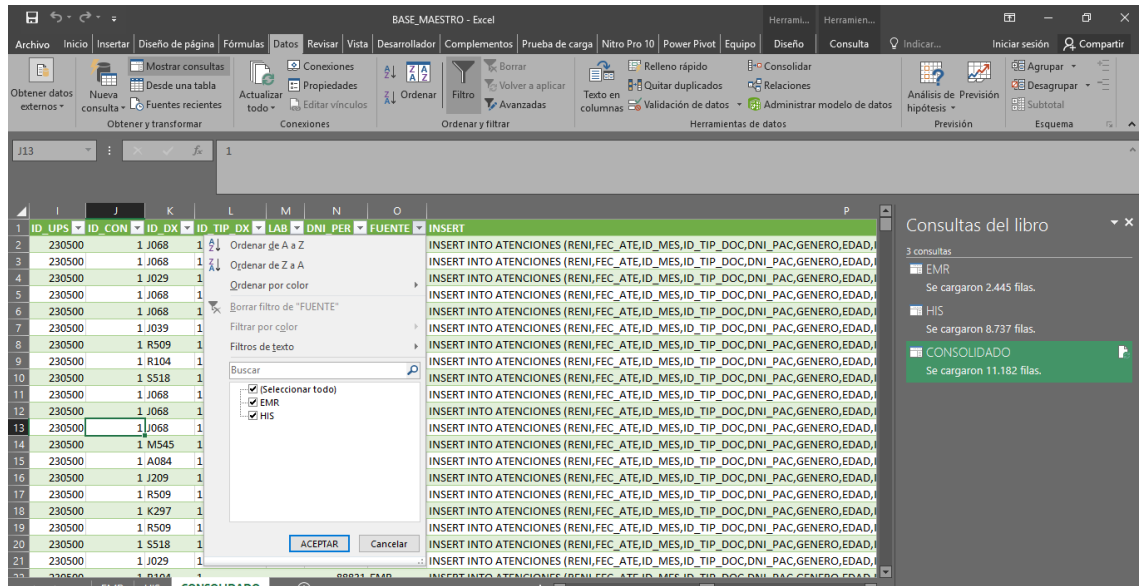


COMBINAMOS

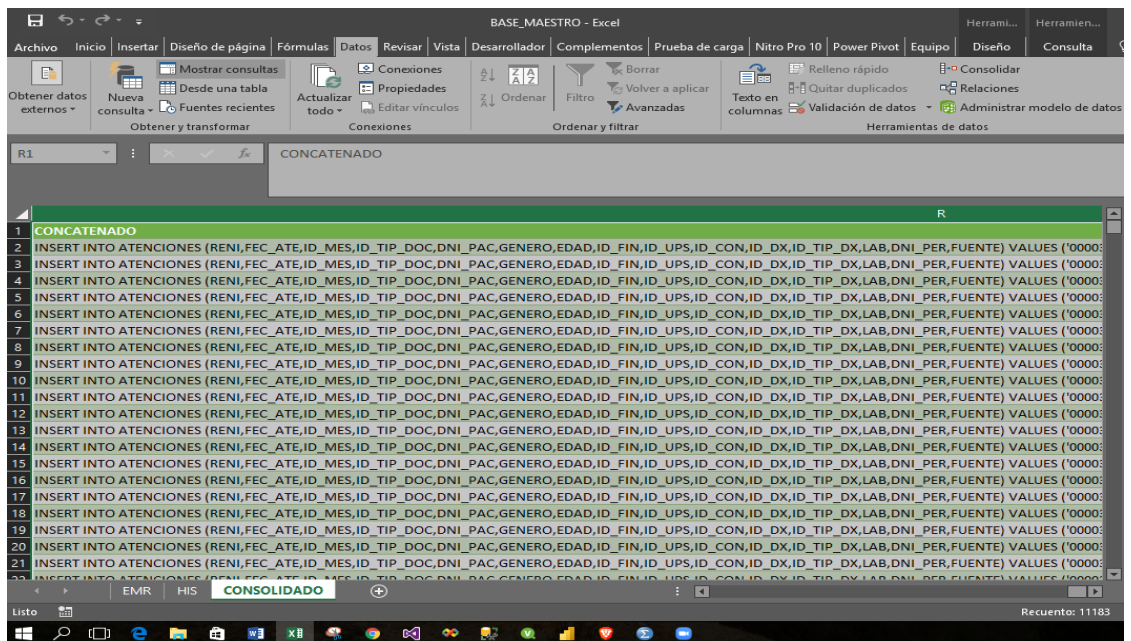
Combinamos ambos Excel consolidando los datos



EXCEL CONSOLIDADO



4. Una vez unificada realizamos el código SQL



Paso 3: Load – Carga

El paso final es el acto de cargar los conjuntos de datos de las fuentes que se han extraído y transformado anteriormente, para ello realizaremos dos formas de hacerlo:

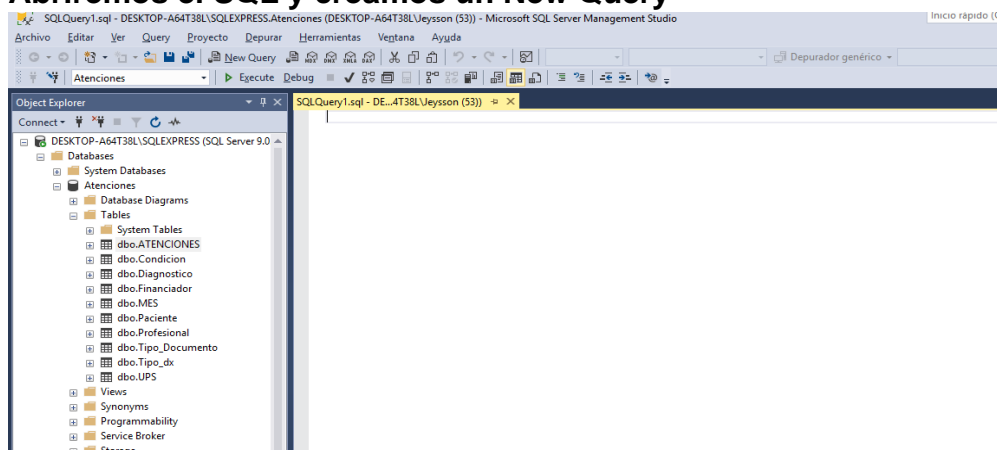
- **Realizar una inserción de SQL**
Implica realizar la inserción de manera manual cada registro en la tabla de la base de datos.
- **Carga Masiva**
Se realizará cuando tenemos conjuntos de datos que están seguro que no contengan errores.

PASOS DE CARGA ETL

- Cargar conjuntos de datos bien transformados a través de la carga a granel
- Cargar conjuntos de datos cuestionables a través de inserciones SQL

Figura N°39: PROCESO CARGA

1. Abriremos el SQL y creamos un New Query



2. Abriremos el Excel unificado y copiaremos el campo **CONCATENADO** y ejecutamos

```
1 CONCATENADO
2 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
3 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
4 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
5 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
6 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
7 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
8 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
9 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
10 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
11 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
12 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
13 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
14 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
15 INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:
```



Object Explorer: DESKTOP-A64T38L\SQLEXPRESS (SQL Server 9.0)

- Databases
 - System Databases
 - Atenciones
 - Database Diagrams
 - Tables
 - System Tables
 - dbo.ATENCIONES
 - dbo.Condicion
 - dbo.Diagnostico
 - dbo.Financiodor
 - dbo.MES
 - dbo.Paciente
 - dbo.Profesional
 - dbo.Tipo_documento
 - dbo.Tipo_dx
 - dbo.UPS
 - Views
 - Synonyms
 - Programmability
 - Service Broker
 - Storage
 - Security
 - Data_Maestro
 - PACIENTES_ATTENDIDOS
 - Pacientes_CSPN
 - persona
- Security
- Server Objects
- ... (rest of the tree)

SQLQuery1.sql - DE...4T38L\Jeysson (53) * [X]
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;
INSERT INTO ATENCIONES (RENI,FEC_ATE,ID_MES,ID_TIP_DOC,DNI_PAC,GENERO,EDAD,ID_FIN,ID_UPS,ID_CON,ID_DX,ID_TIP_DX,LAB,DNI_PER,FUENTE) VALUES ('0000:;

100 % | Connected. (1/1) | DESKTOP-A64T38L\SQLEXPRESS ... | DESKTOP-A64T38L\Jeysson ... | Atenciones | 00:00:00 | 0 rows

3. Realizamos la consulta en SQL y verificaremos que se hallan cargado los datos

Select*from ATENCIONES

100 %

Results Messages

	RENI	FEC_ATE	ID_MES	ID_TIP_DOC	DNI_PAC	GENERO	EDAD	ID_FIN	ID_UPS	ID_CON	ID_DX	ID_TIP_DX	LAB
1	00003419	04-01-2020	1	1	90215101	FEMENINO	2	1	230500	1	J068	1	
2	00003419	06-01-2020	1	1	21827711	FEMENINO	80	1	230500	1	R509	1	
3	00003419	09-01-2020	1	1	81540747	FEMENINO	6	1	230500	1	J068	1	
4	00003419	12-01-2020	1	1	77031196	MASCULINO	18	1	230500	1	J068	1	
5	00003419	16-01-2020	1	1	25855653	MASCULINO	2	1	230500	1	R509	1	
6	00003419	21-01-2020	1	1	70605612	FEMENINO	25	1	230500	1	R104	1	
7	00003419	25-01-2020	1	5	518	FEMENINO	3	1	230500	1	R509	1	
8	00003419	28-01-2020	1	1	21786826	FEMENINO	56	1	230500	1	J039	1	
9	00003419	01-02-2020	2	1	81167890	MASCULINO	6	1	230101	1	J068	1	
10	00003419	03-02-2020	2	1	75449034	MASCULINO	19	1	230500	1	N390	1	
11	00003419	05-02-2020	2	1	21000750	MASCULINO	55	1	230500	1	J068	1	

Query executed successfully, DESKTOP-A64T38L\SQLEXPRESS ... DESKTOP-A64T38L\Jeyso... Atenciones 00:00:00 2581 rows

TABLA DE EVALUACIÓN DE EXPERTOS

Figura N°40: Tabla de Evaluaciones



TABLA DE EVALUACIÓN DE EXPERTOS

Apellidos y Nombres del Experto: SAAVEDRA JIMENEZ ROBERT ROY

Título y Grado: MAGISTER EN DIRECCION Y GESTION DE EMPRESAS

Universidad que labora: UCV

Fecha: 20/03/2021

TÍTULO DEL PROYECTO DE INVESTIGACIÓN:

DATA MART PARA LA MEDICIÓN DE MORBILIDAD Y CASOS DE TBC EN EL CENTRO DE SALUD PUEBLO NUEVO

Evaluación de Metodologías de Desarrollo del Datamart

Mediante la tabla de evaluación de experto, usted tiene la facultad de calificar las metodologías consideradas para el desarrollo de software, mediante una serie de criterios en una escala del 1 al 5, siendo:

(1) MUY MALO, (2) MALO, (3) REGULAR, (4) BUENO, (5) MUY BUENO

ÍTEMS	CRITERIOS	METODOLOGÍAS		
		HEFESTO	BILL INMON	RALPH KIMBALL
1.	Es independiente de las herramientas que se utilicen para su implementación.	4	3	5
2.	Logra impacto positivo sobre los procesos de toma de decisiones.	4	3	5
3.	Se basa en el ciclo de vida dimensional del Negocio	4	3	5
4.	No requiere que los miembros del equipo de trabajo estén altamente capacitados.	4	3	5
5.	Las fases han sido diseñadas para ser desarrolladas en paralelo o en forma secuencial	4	3	5
6.	Se basa en la implementación de tablas para la representación de las dimensiones y hechos.	4	3	5
TOTAL		24	18	30

Observaciones y/o Sugerencias:

Firma del Experto

TABLA DE EVALUACIÓN DE EXPERTOS

 Apellidos y Nombres del Experto: ___BERMEJO TERRONES HENRY PAÚL___

 Título y Grado: ___MAESTRO EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TI___

 Universidad que labora: ___UNIVERSIDAD CÉSAR VALLEJO___

 Fecha: ___11___ / ___03___ / ___2021___
TÍTULO DEL PROYECTO DE INVESTIGACIÓN:
DATA MART PARA LA MEDICIÓN DE MORBILIDAD Y CASOS DE TBC EN EL CENTRO DE SALUD PUEBLO NUEVO
Evaluación de Metodologías de Desarrollo del Datamart

Mediante la tabla de evaluación de experto, usted tiene la facultad de calificar las metodologías consideradas para el desarrollo de software, mediante una serie de criterios en una escala del 1 al 5, siendo:

(1) MUY MALO, (2) MALO, (3) REGULAR, (4) BUENO, (5) MUY BUENO

ÍTEMS	CRITERIOS	METODOLOGÍAS		
		HEFESTO	BILL INMON	RALPH KIMBALL
1.	Es independiente de las herramientas que se utilicen para su implementación.	4	3	5
2.	Logra impacto positivo sobre los procesos de toma de decisiones.	4	4	5
3.	Se basa en el ciclo de vida dimensional del Negocio	4	4	4
4.	No requiere que los miembros del equipo de trabajo estén altamente capacitados.	4	3	4
5.	Las fases han sido diseñadas para ser desarrolladas en paralelo o en forma secuencial	3	3	4
6.	Se basa en la implementación de tablas para la representación de las dimensiones y hechos.	4	3	5
TOTAL		23	20	27

Observaciones y/o Sugerencias:



 Firma del Experto

Figura N°41: FICHAS DE REGISTROS

	objetivos a lograr					
Pertenencia	El instrumento es adecuado al tipo de investigación					85%
Promedio		85%				



VALIDACION DE INSTRUMENTO

Título de Tesis : Data Mart para la medición de morbilidad y casos de TBC en el Centro de Salud Pueblo Nuevo

Autora : Munayco Jacobo, Jeysson José

Nombre del Instrumento de Evaluación: Ficha de Registro

Indicador : Nivel de eficiencia

Datos del Experto:

1. **Apellido y Nombres** : SAAVEDRA JIMENEZ ROBERT ROY
2. **Cargo** : DOCENTE DE INVESTIGACION
3. **Título y/o Grado** : MAGISTER EN GESTION Y DIRECCION DE EMPRESAS
4. **Fecha** : 05/03/2021

Indicadores	Criterio	Deficiente 1% - 20%	Regular 21% - 40%	Bueno 41% - 60%	Muy Bueno 61% - 80%	Excelente 81% - 100%
Claridad	Esta formulado con lenguaje apropiado					85%
Objetividad	Esta expresado en conducta observable					85%
Actualidad	Es adecuado al avance de la ciencia y tecnología					85%
Organización	Existe una organización lógica					85%
Suficiencia	Comprende los aspectos de cantidad y calidad					85%
Intencionalidad	Comprende los aspectos del sistema metodológicos y científico					85%

Título de Tesis : Data Mart para la medición de morbilidad y casos de TBC en el Centro de Salud Pueblo Nuevo

Autora : Munayco Jacobo, Jeysson José

Nombre del Instrumento de Evaluación: Ficha de Registro

Indicador : Productividad de Reportes

Datos del Experto:

1. **Apellido y Nombres** : SAAVEDRA JIMENEZ ROBERT ROY
2. **Cargo** : DOCENTE DE INVESTIGACION
3. **Título y/o Grado** : MAGISTER EN GESTION Y DIRECCION DE EMPRESAS
4. **Fecha** : 05/03/2021

Indicadores	Criterio	Deficiente	Regular	Bueno	Muy Bueno	Excelente
		1% - 20%	21% - 40%	41% - 60%	61% - 80%	81% - 100%
Claridad	Esta formulado con lenguaje apropiado					85%
Objetividad	Esta expresado en conducta observable					85%
Actualidad	Es adecuado al avance de la ciencia y tecnología					85%
Organización	Existe una organización lógica					85%
Suficiencia	Comprende los aspectos de cantidad y calidad					85%
Intencionalidad	Comprende los aspectos del sistema metodológicos y científico					85%
Consistencia	Está basado en aspectos teóricos y científicos					85%
Coherencia	Entre los índices e indicadores					85%
Metodología	Responde al propósito del trabajo bajos los					85%

Ficha de Registro: Nivel de Eficiencia

Ficha de Registro		
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba
Institución Investigada	Centro de Salud Pueblo Nuevo	Pos Test
Dirección	Av. Oscar R Benavides 717	
Proceso Observado	Análisis de resultados Esperados	
Fecha		

Indicador	Técnica	Medida	Instrumento	Fórmula
Nivel de Eficiencia	Registro	Razón	Ficha de Registro	$NE = (HT * CE) / RS$

Nº Estadísticos	Unidad	Horas Trabajadas	Reporte Solicitados	Fórmula $NE = (HT * CE) / RS$
3	Hora	6	4	4,50
3	Hora	6	5	3,60
3	Hora	5	2	7,50
3	Hora	6	4	4,50
1	Hora	6	4	1,50
1	Hora	6	4	1,50
1	Hora	6	2	3,00
1	Hora	6	2	3,00
2	Hora	5	2	5,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
1	Hora	6	1	6,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
2	Hora	5	2	5,00
Total		87	42	60,07

Ficha de Registro: Productividad de Reportes

FICHA DE REGISTRO			
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba	Pos Test
Institución Investigada	Centro de Salud Pueblo Nuevo	Dirección	Av. Oscar R Benavides 717
Proceso de Observación	Productividad de Reportes	Fórmula	PR = (RA/RS)*100
Fecha de Inicio		Fecha Fin	

Ítem	Fecha de Inicio	Fecha Fin	Nº Reportes Atendidos	Nº Reportes Solicitados	Productividad Reportes = (RA/RS) * 100
1	01/02/2021	01/02/2021	4	4	100,00
2	02/02/2021	02/02/2021	4	5	80,00
3	03/02/2021	03/02/2021	2	2	100,00
4	04/02/2021	04/02/2021	4	4	100,00
5	05/02/2021	05/02/2021	4	4	100,00
6	06/02/2021	06/02/2021	3	4	75,00
7	08/02/2021	08/02/2021	2	2	100,00
8	09/02/2021	09/02/2021	2	2	100,00
9	10/02/2021	10/02/2021	2	2	100,00
10	11/02/2021	11/02/2021	2	2	100,00
11	12/02/2021	12/02/2021	2	3	66,67
12	13/02/2021	13/02/2021	1	1	100,00
13	15/02/2021	15/02/2021	2	2	100,00
14	16/02/2021	16/02/2021	2	3	66,67
15	17/02/2021	17/02/2021	2	2	100,00
Total			38	42	90,48

Ficha de Registro: Nivel de Eficiencia

Ficha de Registro		
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba
Institución Investigada	Centro de Salud Pueblo Nuevo	Re Test
Dirección	Av. Oscar R Benavides 717	
Proceso Observado	Análisis de resultados Esperados	
Fecha		

Indicador	Técnica	Medida	Instrumento	Fórmula
Nivel de Eficiencia	Registro	Razón	Ficha de Registro	$NE = (HT * CE) / RS$

Nº Estadísticos	Unidad	Horas Trabajadas	Reporte Solicitados	Fórmula $NE = (HT * CE) / RS$
3	Hora	6	4	4,50
3	Hora	6	5	3,60
3	Hora	6	4	4,50
3	Hora	6	4	4,50
1	Hora	6	4	1,50
1	Hora	6	2	3,00
1	Hora	6	2	3,00
1	Hora	6	2	3,00
2	Hora	6	2	6,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
1	Hora	6	2	3,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
2	Hora	6	3	4,00
Total		90	44	59,32

Ficha de Registro: Productividad de Reportes

FICHA DE REGISTRO			
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba	Re Test
Institución Investigada	Centro de Salud Pueblo Nuevo	Dirección	Av. Oscar R Benavides 717
Proceso de Observación	Productividad de Reportes	Fórmula	PR = (RA/RS)*100
Fecha de Inicio		Fecha Fin	

Ítem	Fecha de Inicio	Fecha Fin	Nº Reportes Atendidos	Nº Reportes Solicitados	Productividad Reportes = (RA/RS) * 100
1	01/02/2021	01/02/2021	4	4	100,00
2	02/02/2021	02/02/2021	4	5	80,00
3	03/02/2021	03/02/2021	3	4	75,00
4	04/02/2021	04/02/2021	3	4	75,00
5	05/02/2021	05/02/2021	3	4	75,00
6	06/02/2021	06/02/2021	1	2	50,00
7	08/02/2021	08/02/2021	2	2	100,00
8	09/02/2021	09/02/2021	1	2	50,00
9	10/02/2021	10/02/2021	2	2	100,00
10	11/02/2021	11/02/2021	2	2	100,00
11	12/02/2021	12/02/2021	2	3	66,67
12	13/02/2021	13/02/2021	1	2	50,00
13	15/02/2021	15/02/2021	1	2	50,00
14	16/02/2021	16/02/2021	2	3	66,67
15	17/02/2021	17/02/2021	2	3	66,67
Total			33	44	75,00

Ficha de Registro: Nivel de Eficiencia

Ficha de Registro		
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba
Institución Investigada	Centro de Salud Pueblo Nuevo	Pre Test
Dirección	Av. Oscar R Benavides 717	
Proceso Observado	Análisis de resultados Esperados	
Fecha		

Indicador	Técnica	Medida	Instrumento	Fórmula
Nivel de Eficiencia	Registro	Razón	Ficha de Registro	$NE = (HT * CE) / RS$

Nº Estadísticos	Unidad	Horas Trabajadas	Reporte Solicitados	Fórmula $NE = (HT * CE) / RS$
3	Hora	8	4	6,00
3	Hora	8	5	4,80
3	Hora	8	4	6,00
3	Hora	8	4	6,00
1	Hora	8	4	2,00
1	Hora	8	2	4,00
1	Hora	8	2	4,00
1	Hora	8	2	4,00
2	Hora	8	2	8,00
2	Hora	8	2	8,00
2	Hora	8	3	5,33
1	Hora	8	2	4,00
2	Hora	8	2	8,00
2	Hora	8	3	5,33
2	Hora	8	3	5,33
Total		120	44	79,09

Ficha de Registro: Productividad de Reportes

FICHA DE REGISTRO			
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba	Pre Test
Institución Investigada	Centro de Salud Pueblo Nuevo	Dirección	Av. Oscar R Benavides 717
Proceso de Observación	Productividad de Reportes	Fórmula	PR = (RA/RS)*100
Fecha de Inicio		Fecha Fin	

Ítem	Fecha de Inicio	Fecha Fin	Nº Reportes Atendidos	Nº Reportes Solicitados	Productividad Reportes = (RA/RS) * 100
1	01/02/2021	01/02/2021	3	4	75,00
2	02/02/2021	02/02/2021	3	5	60,00
3	03/02/2021	03/02/2021	2	4	50,00
4	04/02/2021	04/02/2021	2	4	50,00
5	05/02/2021	05/02/2021	2	4	50,00
6	06/02/2021	06/02/2021	0	2	0,00
7	08/02/2021	08/02/2021	1	2	50,00
8	09/02/2021	09/02/2021	0	2	0,00
9	10/02/2021	10/02/2021	1	2	50,00
10	11/02/2021	11/02/2021	1	2	50,00
11	12/02/2021	12/02/2021	1	3	33,33
12	13/02/2021	13/02/2021	0	2	0,00
13	15/02/2021	15/02/2021	1	2	50,00
14	16/02/2021	16/02/2021	1	3	33,00
15	17/02/2021	17/02/2021	1	3	33,00
Total			19	44	43,18

Consistencia	Está basado en aspectos teóricos y científicos					85%
Coherencia	Entre los índices e indicadores					85%
Metodología	Responde al propósito del trabajo bajos los objetivos a lograr					85%
Pertenencia	El instrumento es adecuado al tipo de investigación					85%
Promedio		85%				

[Handwritten signature]

Título de Tesis : Data Mart para la medición de morbilidad y casos de TBC en el Centro de Salud Pueblo Nuevo

Actora : Munayco Jacobo, Jeysson José

Nombre del Instrumento de Evaluación: Ficha de Registro

Indicador : Nivel de eficiencia

Datos del Experto:

1. **Apellido y Nombres** : BERMEJO TERRONES, HENRY PAÚL
2. **Cargo** : DOCENTE TIEMPO COMPLETO - DTC
3. **Título y/o Grado** : MAGISTER
4. **Fecha** : __23__/_02__/_2021__

Indicadores	Criterio	Deficiente 1% - 20%	Regular 21% - 40%	Buena 41% - 60%	Muy Buena 61% - 80%	Excelente 81% - 100%
Claridad	Esta formulado con lenguaje apropiado					95%
Objetividad	Esta expresado en conducta observable					95%
Actualidad	Es adecuado al avance de la ciencia y tecnología					95%
Organización	Existe una organización lógica					90%
Suficiencia	Comprende los aspectos de cantidad y calidad					90%
Intencionalidad	Comprende los aspectos del sistema metodológicos y científico					95%
Consistencia	Está basado en aspectos teóricos y científicos					95%
Coherencia	Entre los índices e indicadores					95%

Título de Tesis : Data Mart para la medición de morbilidad y casos de TBC en el Centro de Salud Pueblo Nuevo

Autora : Munayco Jacobo, Jeysson José

Nombre del Instrumento de Evaluación: Ficha de Registro

Indicador : Productividad de Reportes

Datos del Experto:

1. **Apellido y Nombres** : BERMEJO TERRONES, HENRY PAÚL
2. **Cargo** : DOCENTE TIEMPO COMPLETO - DTC
3. **Título y/o Grado** : MAGISTER
4. **Fecha** : 23 / 02 / 2021

Indicadores	Criterio	Deficiente 1% - 20%	Regular 21% - 40%	Bueno 41% - 60%	Muy Bueno 61% - 80%	Excelente 81% - 100%
Claridad	Esta formulado con lenguaje apropiado					95%
Objetividad	Esta expresado en conducta observable					90%
Actualidad	Es adecuado al avance de la ciencia y tecnología					95%
Organización	Existe una organización lógica					95%
Suficiencia	Comprende los aspectos de cantidad y calidad					95%
Intencionalidad	Comprende los aspectos del sistema metodológicos y científico					95%
Consistencia	Está basado en aspectos teóricos y científicos					90%
Coherencia	Entre los índices e indicadores					95%

	objetivos a lograr					
Pertenencia	El instrumento es adecuado al tipo de investigación					95%
	Promedio	94%				



Ficha de Registro: Nivel de Eficiencia

Ficha de Registro		
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba
Institución Investigada	Centro de Salud Pueblo Nuevo	Pos Test
Dirección	Av. Oscar R Benavides 717	
Proceso Observado	Análisis de resultados Esperados	
Fecha		

Indicador	Técnica	Medida	Instrumento	Fórmula
Nivel de Eficiencia	Registro	Razón	Ficha de Registro	$NE = (HT * CE) / RS$

Nº Estadísticos	Unidad	Horas Trabajadas	Reporte Solicitados	Fórmula $NE = (HT * CE) / RS$
3	Hora	6	4	4,50
3	Hora	6	5	3,60
3	Hora	5	2	7,50
3	Hora	6	4	4,50
1	Hora	6	4	1,50
1	Hora	6	4	1,50
1	Hora	6	2	3,00
1	Hora	6	2	3,00
2	Hora	5	2	5,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
1	Hora	6	1	6,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
2	Hora	5	2	5,00
Total		87	42	60,07

Ficha de Registro: Productividad de Reportes

FICHA DE REGISTRO			
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba	Pos Test
Institución Investigada	Centro de Salud Pueblo Nuevo	Dirección	Av. Oscar R Benavides 717
Proceso de Observación	Productividad de Reportes	Fórmula	PR = (RA/RS)*100
Fecha de Inicio		Fecha Fin	

Ítem	Fecha de Inicio	Fecha Fin	Nº Reportes Atendidos	Nº Reportes Solicitados	Productividad Reportes = (RA/RS) * 100
1	01/02/2021	01/02/2021	4	4	100,00
2	02/02/2021	02/02/2021	4	5	80,00
3	03/02/2021	03/02/2021	2	2	100,00
4	04/02/2021	04/02/2021	4	4	100,00
5	05/02/2021	05/02/2021	4	4	100,00
6	06/02/2021	06/02/2021	3	4	75,00
7	08/02/2021	08/02/2021	2	2	100,00
8	09/02/2021	09/02/2021	2	2	100,00
9	10/02/2021	10/02/2021	2	2	100,00
10	11/02/2021	11/02/2021	2	2	100,00
11	12/02/2021	12/02/2021	2	3	66,67
12	13/02/2021	13/02/2021	1	1	100,00
13	15/02/2021	15/02/2021	2	2	100,00
14	16/02/2021	16/02/2021	2	3	66,67
15	17/02/2021	17/02/2021	2	2	100,00
Total			38	42	90,48

Ficha de Registro: Nivel de Eficiencia

Ficha de Registro		
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba
Institución Investigada	Centro de Salud Pueblo Nuevo	Re Test
Dirección	Av. Oscar R Benavides 717	
Proceso Observado	Análisis de resultados Esperados	
Fecha		

Indicador	Técnica	Medida	Instrumento	Fórmula
Nivel de Eficiencia	Registro	Razón	Ficha de Registro	$NE = (HT * CE) / RS$

Nº Estadísticos	Unidad	Horas Trabajadas	Reporte Solicitados	Fórmula $NE = (HT * CE) / RS$
3	Hora	6	4	4,50
3	Hora	6	5	3,60
3	Hora	6	4	4,50
3	Hora	6	4	4,50
1	Hora	6	4	1,50
1	Hora	6	2	3,00
1	Hora	6	2	3,00
1	Hora	6	2	3,00
2	Hora	6	2	6,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
1	Hora	6	2	3,00
2	Hora	6	2	6,00
2	Hora	6	3	4,00
2	Hora	6	3	4,00
Total		90	44	59,32

Ficha de Registro: Productividad de Reportes

FICHA DE REGISTRO			
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba	Re Test
Institución Investigada	Centro de Salud Pueblo Nuevo	Dirección	Av. Oscar R Benavides 717
Proceso de Observación	Productividad de Reportes	Fórmula	PR = (RA/RS)*100
Fecha de Inicio		Fecha Fin	

Ítem	Fecha de Inicio	Fecha Fin	Nº Reportes Atendidos	Nº Reportes Solicitados	Productividad Reportes = (RA/RS) * 100
1	01/02/2021	01/02/2021	4	4	100,00
2	02/02/2021	02/02/2021	4	5	80,00
3	03/02/2021	03/02/2021	3	4	75,00
4	04/02/2021	04/02/2021	3	4	75,00
5	05/02/2021	05/02/2021	3	4	75,00
6	06/02/2021	06/02/2021	1	2	50,00
7	08/02/2021	08/02/2021	2	2	100,00
8	09/02/2021	09/02/2021	1	2	50,00
9	10/02/2021	10/02/2021	2	2	100,00
10	11/02/2021	11/02/2021	2	2	100,00
11	12/02/2021	12/02/2021	2	3	66,67
12	13/02/2021	13/02/2021	1	2	50,00
13	15/02/2021	15/02/2021	1	2	50,00
14	16/02/2021	16/02/2021	2	3	66,67
15	17/02/2021	17/02/2021	2	3	66,67
Total			33	44	75,00

Ficha de Registro: Nivel de Eficiencia

Ficha de Registro		
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba
Institución Investigada	Centro de Salud Pueblo Nuevo	Pre Test
Dirección	Av. Oscar R Benavides 717	
Proceso Observado	Análisis de resultados Esperados	
Fecha		

Indicador	Técnica	Medida	Instrumento	Fórmula
Nivel de Eficiencia	Registro	Razón	Ficha de Registro	$NE = (HT * CE) / RS$

Nº Estadísticos	Unidad	Horas Trabajadas	Reporte Solicitados	Fórmula $NE = (HT * CE) / RS$
3	Hora	8	4	6,00
3	Hora	8	5	4,80
3	Hora	8	4	6,00
3	Hora	8	4	6,00
1	Hora	8	4	2,00
1	Hora	8	2	4,00
1	Hora	8	2	4,00
1	Hora	8	2	4,00
2	Hora	8	2	8,00
2	Hora	8	2	8,00
2	Hora	8	3	5,33
1	Hora	8	2	4,00
2	Hora	8	2	8,00
2	Hora	8	3	5,33
2	Hora	8	3	5,33
Total		120	44	79,09

Ficha de Registro: Productividad de Reportes

FICHA DE REGISTRO			
Investigador	Munayco Jacobo, Jeysson José	Tipo de Prueba	Pre Test
Institución Investigada	Centro de Salud Pueblo Nuevo	Dirección	Av. Oscar R Benavides 717
Proceso de Observación	Productividad de Reportes	Fórmula	PR = (RA/RS)*100
Fecha de Inicio		Fecha Fin	

Ítem	Fecha de Inicio	Fecha Fin	Nº Reportes Atendidos	Nº Reportes Solicitados	Productividad Reportes = (RA/RS) * 100
1	01/02/2021	01/02/2021	3	4	75,00
2	02/02/2021	02/02/2021	3	5	60,00
3	03/02/2021	03/02/2021	2	4	50,00
4	04/02/2021	04/02/2021	2	4	50,00
5	05/02/2021	05/02/2021	2	4	50,00
6	06/02/2021	06/02/2021	0	2	0,00
7	08/02/2021	08/02/2021	1	2	50,00
8	09/02/2021	09/02/2021	0	2	0,00
9	10/02/2021	10/02/2021	1	2	50,00
10	11/02/2021	11/02/2021	1	2	50,00
11	12/02/2021	12/02/2021	1	3	33,33
12	13/02/2021	13/02/2021	0	2	0,00
13	15/02/2021	15/02/2021	1	2	50,00
14	16/02/2021	16/02/2021	1	3	33,00
15	17/02/2021	17/02/2021	1	3	33,00
Total			19	44	43,18

Metodología	Responde al propósito del trabajo bajos los objetivos a lograr					95%
Pertenencia	El instrumento es adecuado al tipo de investigación					95%
Promedio		94%				





ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Declaratoria de Autenticidad del Asesor

Yo, ORMEÑO ROJAS , ROBERT EDUARDO , docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "DATA MART PARA LA MEDICIÓN DE MORBILIDAD Y CASOS DE TBC EN EL CENTRO DE SALUD PUEBLO NUEVO ",

cuyo autores son MUNAYCO JACOBO JEYSSON JOSÉ, constato que la investigación cumple con el índice de 20% similitud establecido, y 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, 16 de mayo del 2022

Apellidos y Nombres del Asesor	Firma
ORMEÑO ROJAS , ROBERT EDUARDO DNI: 44439590 ORCID: 0000-0003-2300-1017	