



**UNIVERSIDAD NACIONAL DE CHIMBORAZO
FACULTAD DE INGENIERÍA
CARRERA DE SISTEMAS Y COMPUTACIÓN**

**SISTEMA MULTIDIMENSIONAL Y CUADRO DE MANDO
INTEGRAL CON FUENTES DE INFORMACIÓN DEL SISTEMA
MATERNO NEONATAL DEL MINISTERIO DE SALUD PÚBLICA**

**Trabajo de Titulación para optar al título de Ingeniero en Sistemas y
Computación**

Autores:

**Enríquez Chamba, Lidia Esperanza
Moina Pinzón, Clinton Jamil**

Tutor:

Ing. Jorge Edwin Delgado Altamirano

Riobamba, Ecuador. 2023

DERECHOS DE AUTORÍA

Nosotros, Lidia Esperanza Enríquez Chamba con cédula 1105494692; Clinton Jamil Moina Pinzón con cédula 0605709690, autores del trabajo de investigación titulado: “SISTEMA MULTIDIMENSIONAL Y CUADRO DE MANDO INTEGRAL CON FUENTES DE INFORMACIÓN DEL SISTEMA MATERNO NEONATAL DEL MINISTERIO DE SALUD PÚBLICA”, certifico que la producción, ideas, opiniones, criterios, contenidos y conclusiones expuestas son de mí exclusiva responsabilidad.

Asimismo, cedemos a la Universidad Nacional de Chimborazo, en forma no exclusiva, los derechos para su uso, comunicación pública, distribución, divulgación y/o reproducción total o parcial, por medio físico o digital; en esta cesión se entiende que el cesionario no podrá obtener beneficios económicos. La posible reclamación de terceros respecto de los derechos de los autores de la obra referida será de mi entera responsabilidad; librando a la Universidad Nacional de Chimborazo de posibles obligaciones.

En Riobamba, 22 de marzo de 2023.



Lidia Esperanza Enríquez Chamba
C.I: 1105494692



Clinton Jamil Moina Pinzón
C.I: 0605709690

DICTAMEN FAVORABLE DEL TUTOR Y MIEMBROS DE TRIBUNAL

Quienes suscribimos, catedráticos designados Tutor y Miembros del Tribunal de Grado para la evaluación del trabajo de investigación “SISTEMA MULTIDIMENSIONAL Y CUADRO DE MANDO INTEGRAL CON FUENTES DE INFORMACIÓN DEL SISTEMA MATERNO NEONATAL DEL MINISTERIO DE SALUD PÚBLICA”, presentado por Lidia Esperanza Enríquez Chamba, con cédula de identidad número 1105494692 y Clinton Jamil Moina Pinzón con cédula de identidad número 0605709690, certificamos que recomendamos la APROBACIÓN de este con fines de titulación. Previamente se ha asesorado durante el desarrollo, revisado y evaluado el trabajo de investigación escrito y escuchada la sustentación por parte de su autor; no teniendo más nada que observar.

De conformidad a la normativa aplicable firmamos, en Riobamba 22 de marzo de 2023



Milton López. Mgs

PRESIDENTE DEL TRIBUNAL DE GRADO



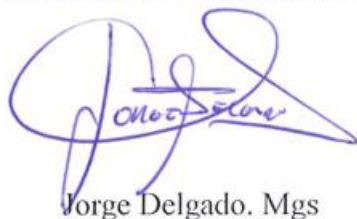
Ana Congacha. Mgs

MIEMBRO DEL TRIBUNAL DE GRADO



Pamela Buñay. Mgs

MIEMBRO DEL TRIBUNAL DE GRADO



Jorge Delgado. Mgs

TUTOR

DEDICATORIA

Dedico este proyecto de investigación a mi familia, quienes me han apoyado y han creído en mí, a mis Padres Marlene Chamba y Klever Enríquez que son el pilar fundamental y mi motivación para lograr todos mis éxitos y a mis hermanos por todo el apoyo brindado durante toda mi etapa universitaria.

Lidia Esperanza Enríquez Chamba

Dedico a mis padres Jaime Rodrigo Moina Guaman y Jenny Edilma Pinzón Moreno por que han sido siempre mis pilares de enseñanza como son los valores, siendo así el motor que impulsa mis sueños y esperanzas, quienes estuvieron siempre a mi lado en los días y noches más difíciles durante mis horas de estudio. Siempre han sido mis mejores guías de vida.

Clinton Jamil Moina Pinzón

AGRADECIMIENTO

Agradezco a Dios por permitirme lograr todos mis éxitos y metas que me he propuesto, a mi tutor por apoyarme y guiarme en todo el proceso de desarrollo del trabajo de investigación, a los docentes de la carrera de ingeniería en Sistemas y Computación por sus conocimientos impartidos para que como estudiante logre mis objetivos y culmine mi etapa universitaria con éxito.

Lidia Esperanza Enríquez Chamba - Clinton Jamil Moina Pinzón

ÍNDICE DE GENERAL

CAPÍTULO I. INTRODUCCIÓN.....	14
1.1 PLANTEAMIENTO Y JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	15
1.2 OBJETIVOS	16
1.2.1 Objetivo general	16
1.2.2 Objetivos específicos	16
CAPÍTULO II. MARCO TEÓRICO.....	17
2.1 Data Warehouse.....	17
2.2 Sistema Multidimensional	18
2.3 ETL.....	20
2.4 Pentaho.....	21
2.5 Cuadro de Mando Integral.....	21
2.6 Lenguaje de programación R	21
2.6.1 R Studio	22
2.6.2 Paquete Shiny	23
2.7 Metodología Crisp-dm.....	24
2.8 Norma ISO/IEC 25012	26
2.9 Talend Data Quality	28
CAPÍTULO III. METODOLOGIA.....	29
3.1 Tipo y Diseño de investigación	29
3.2 Técnicas de análisis e interpretación	29
3.3 Población y muestra	29
3.4 Operacionalización de las variables	31
3.5 Método de análisis y procedimiento de datos	32
3.6 Método de análisis y procedimiento de datos	32
3.6.1 Fase I. Comprensión de los datos	32
3.6.2 Fase II. Comprensión de los datos	33
3.6.3 FASE III. Preparación de los Datos.	38
3.6.4 FASE IV. Modelado de los datos	44

3.6.5	FASE V. Evaluación.....	44
3.6.6	FASE VI. Despliegue	45
CAPÍTULO IV. RESULTADOS Y DISCUSIÓN		50
4.1	Resultados.....	50
4.2	Discusión	58
CAPÍTULO V. CONCLUSIONES y RECOMENDACIONES		59
5.2	Recomendaciones.....	60
BIBLIOGRAFÍA		61
ANEXOS.....		64

ÍNDICE DE TABLAS

Tabla 1: Cuadro Comparativo de paquetes de R	22
Tabla 2: Criterios para medir la Completitud de los datos	27
Tabla 3: Criterios para medir la Credibilidad de los datos	27
Tabla 4: Criterios para medir la Precisión de los datos	28
Tabla 5: Tabla de operacionalización de las variables	31
Tabla 6: Colorimetría del Nivel de Calidad de los Indicadores	34
Tabla 7: Porcentaje del Nivel de Calidad del Indicador 1 por trimestre en la provincia de Chimborazo	34
Tabla 8: Porcentaje del Nivel de Calidad del Indicador 1 por cuatrimestre en la provincia de Chimborazo	35
Tabla 9: Porcentaje del Nivel de Calidad del Indicador 1 por semestre en la provincia de Chimborazo	36
Tabla 10: Porcentaje del Nivel de Calidad del Indicador 1 por año en la provincia de Chimborazo	37
Tabla 11: Datos de la tabla Cantón.....	38
Tabla 12: Datos de la tabla Provincia	38
Tabla 13: Descripción de la tabla Establecimiento	38
Tabla 14: Descripción de la tabla Indicadores.....	39
Tabla 15: Descripción de la tabla Zona	39
Tabla 16: Descripción de la tabla Tiempo	40
Tabla 17: Credibilidad, Completitud y Precisión de los datos de Establecimientos de la Coordinación 3.....	50
Tabla 18: Credibilidad, Completitud y Precisión de los datos de Hechos	50
Tabla 19: Credibilidad, Completitud y Precisión de los datos de Indicador	51
Tabla 20: Credibilidad, Completitud y Precisión de los datos de Cantón de la Coordinación Zonal 3	51

Tabla 21: Credibilidad, Completitud y Precisión de los datos de Provincia de la Coordinación Zonal 3	52
Tabla 22: Credibilidad, Completitud y Precisión de los datos de Zona de la Coordinación Zonal 3	52
Tabla 23: Credibilidad, Completitud y Precisión de los datos de Tiempo.....	52
Tabla 25: Descripción de la tabla Cantón	65
Tabla 26: Descripción de la tabla Provincia.....	65
Tabla 27: Descripción de la tabla Establecimiento	66
Tabla 28: Descripción de la tabla Indicadores.....	66
Tabla 29: Descripción de la tabla Zona	68
Tabla 30: Descripción de la tabla Tiempo	68
Tabla 31: Colorimetría del Nivel de Calidad para el indicador 10.....	68
Tabla 32: Colorimetría del Nivel de Calidad para el indicador 11.....	68
Tabla 33: Colorimetría del Nivel de Calidad para el indicador 12.....	68
Tabla 34: Colorimetría del Nivel de Calidad para el indicador 13.....	69
Tabla 35: Análisis de la calidad de los datos.....	69

ÍNDICE DE FIGURAS

Figura 1: Esquema tipo estrella	19
Figura 2: Esquema copo de nieve.....	20
Figura 3: Etapas de ciclo de vida de la metodología Crisp-dm.....	24
Figura 4: Características de la calidad de los datos según la norma IS/IEC 25012.....	26
Figura 5: Datos del indicador 1 por trimestre de la provincia de Chimborazo	35
Figura 6: Datos del Indicador 1 por cuatrimestre de la provincia de Chimborazo	36
Fuente: Lidia Enríquez, Clinton Moina	36
Figura 7: Datos del indicador 1 por semestre de la provincia de Chimborazo.....	37
Figura 8: Datos del indicador 1 por año de la provincia de Chimborazo	38
Fuente: Lidia Enriquez, Clinton Moina	38
Figura 9: Diseño de Sistema Multidimensional aplicando procesos de ETL.....	41
Figura 10: Filtro empleados en las tablas.....	42
Figura 11: Diagrama del Sistema Multidimensional Tipo estrella.....	43
Figura 12: Proceso para el diseño del Sistema Multidimensional y Cuadro de Mando Integral.	45
Figura 13: Librería Instaladas en R Studio para el desarrollo del Cuadro de Mando Integral.	46
Figura 14: Conexión a la base de datos.....	46
Figura 15: Consultas a las tablas de la base de datos	47
Figura 16: Login en R Studio.....	47
Figura 17: Configuración del archivo App.R.....	48
Figura 18: Código de gráficas y filtros aplicados.	48
Figura 19: Cuadro de mando integral a nivel de zona.....	49
Figura 20: Página principal, reporte trimestral y anual	53
Figura 21: Reporte por provincia trimestral y anual	54
Figura 22: Reporte por distrito trimestral y anual.....	55

Figura 23: Reporte por tipología trimestral y anual.....	56
Figura 24: Reporte por establecimiento mensual, trimestral, cuatrimestral, semestral y anual	57
Figura 25: Data Original: Matriz de reporte distrital del monitoreo de la calidad de la Atención Materno Neonatal.....	64
Figura 26: Estructurar la información en formato tabla	64
Figura 27: Estructuración de la información	65
Figura 28: Completitud, Credibilidad y Precisión de los datos de la tabla Cantón	71
Figura 30: Completitud, Credibilidad y Precisión de los datos de la tabla Establecimiento	72
Figura 31: Completitud, Credibilidad y Precisión de los datos de la tabla Indicador.....	73
Figura 32: Completitud, Credibilidad y Precisión de los datos de la tabla Provincia.....	73
Figura 33: Completitud, Credibilidad y Precisión de los datos de la tabla Tiempo	74
Figura 34: Completitud, Credibilidad y Precisión de los datos de la tabla Zona	75
Figura 35: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Cantón	76
Figura 36: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Establecimiento.....	76
Figura 37: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Hechos.....	77
Figura 38: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Indicador.....	77
Figura 39: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Provincia.....	78
Figura 40: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Tiempo	78
Figura 41: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Tiempo	79

RESUMEN

En la Coordinación Zonal 3 del Ministerio de Salud Pública existen grandes cantidades de datos que son escasamente analizados, actualmente existen reportes de 278 establecimientos en una matriz de reporte Distrital del Monitoreo de la Calidad de Atención Materno Neonatal, privándose de obtener información útil que les permita tomar decisiones a nivel organizacional y brindar atención de calidad.

Respecto a la necesidad, el presente trabajo de investigación se enfoca en el diseño de un sistema multidimensional y cuadro de mando integral, para lo cual se utilizó herramientas libres como Pentaho para realizar procesos de ETL, R Studio para la programación del Cuadro de Mando integral y la librería Shiny como Servidor para visualizar el cuadro de Mando Integral, se desarrolló bajo la metodología Crisp-dm la cual mediante seis fases ayudó a obtener calidad en los datos con criterios: completos, creíbles y precisos según establece la Norma ISO/IEC 25012, mediante la herramienta Talend Data Quality.

El proceso de análisis de datos consideró 6 tablas: Provincia, Cantón, Zona, Indicador, Hechos, Tiempo y Establecimiento, se obtuvo resultados de 100% de Completitud, 100% de credibilidad y 100% de Precisión de los datos en cada una de las tablas. Los resultados obtenidos son datos de calidad según lo establece las métricas que proporciona la norma ISO/IEC 25012. El resultado final de la tesis es un sistema multidimensional y cuadro de mando integral que permite a los usuarios acceder a información importante y valiosa sobre el sistema Materno Neonatal del Ministerio de Salud Pública.

Palabras Clave: Sistema Multidimensional, Cuadro de mando, Calidad de datos, Crisp-dm.

ABSTRACT

In the Zonal Coordination 3 of the Ministry of Public Health, there are large amounts of data that are scarcely analyzed. Currently, there are reports from 278 facilities in a District Report Matrix of Maternal Neonatal Care Monitoring, depriving them of obtaining useful information that would allow them to make organizational decisions and provide quality care.

Regarding this need, this research focuses on the design of a multidimensional system and an integral dashboard. Free tools such as Pentaho were used to perform ETL processes, R Studio for programming the integral dashboard, and the Shiny library as a server to visualize the integral dashboard. The research was developed under the Crisp-DM methodology, which, through six phases, helped obtain quality data with criteria: complete, credible, and accurate as established by ISO/IEC 25012 standard, through the Talend Data Quality tool.

The data analysis process considered six tables: Province, Canton, Zone, Indicator, Facts, Time, and Establishment. Results of 100% completeness, 100% credibility, and 100% accuracy of the data were obtained in each of the tables. The results obtained are of high-quality data according to the metrics provided by the ISO/IEC 25012 standard. The final result of the thesis is a multidimensional system and an integral dashboard that allows users to access important and valuable information about the Maternal Neonatal System of the Ministry of Public Health.

Keywords: Multidimensional System, Dashboard, Data Quality, Crisp-DM.



Reviewed by:
Mgs. Hugo Romero
ENGLISH PROFESSOR
C.C. 0603156258

CAPÍTULO I. INTRODUCCIÓN

En una organización o empresa el éxito depende de las decisiones acertadas que tomen los administrativos, para ello la inteligencia de negocios es una solución ya que ayuda a encontrar información exacta, oportuna y consistente (Alonso & Rico, 2010), además de mejorar la habilidad para la toma de decisiones, haciendo uso de aplicaciones, metodologías y herramientas tecnológicas las cuales permiten recopilar, transformar, analizar y cargar datos, para agilizar el procesamiento de la información y aumento de beneficios a la organización (Ranjan, 2009).

La inteligencia de negocios trabaja con el proceso de ETL (extracción, transformación y carga), donde se extrae datos alojados en fuentes heterogéneas, se los procesa con la intención de obtener información útil, que les permita tomar decisiones de manera precisa, rápida y oportuna.

La ciencia de los datos está enfocada en encontrar información procesable a partir de datos brutos y estructurados, se enfoca en descubrir respuestas a las cosas que no sabemos (Muller, 2014). En las últimas décadas las organizaciones y empresas generan diariamente grandes cantidades de datos, estos datos no están siendo usados correctamente o no están aportando ningún valor representativo para la toma de decisiones, por lo que es conveniente transformar dichos datos en información útil y está a su vez genere conocimiento.

Dentro de la ciencia de los datos exigen los sistemas multidimensionales con los cuales se puede lograr un rendimiento rápido al momento de realizar consultas empresariales, y ayuda a realizar un diseño lógico para modelar las particularidades de los procesos que ocurren dentro de una organización (Ranjan, 2009). Así como permite acceder a grandes cantidades de información y obtener análisis de las relaciones de distintos tipos de elementos relevantes para el negocio.

El cuadro de mando integral se ha convertido en una de las herramientas más usada ya que proporciona a los directivos los instrumentos necesarios para construir un camino hacia el éxito competitivo, basados en métodos y estrategias que se alineen a los objetivos de la empresa y supervisar su evolución (Alonso & Rico, 2010).

El Ministerio de Salud pública maneja gran cantidad de data en el área de Calidad Materno Neonatal, la cual está registrada en un archivo de Excel llamada matriz de reporte Distrital del Monitoreo de la Calidad de la Atención Materno Neonatal, sin embargo, necesita de procesos de ETL para la generación de información, Pentaho nos permitió generar un esquema multidimensional con información útil para la organización, información que fue cargada en un Data Warehouse y explotada con cuadros de mando integral realizados en lenguaje R con la librería Shiny.

La presente investigación se desarrolla en los siguientes capítulos: Capítulo I: Definición del problema de investigación, Capítulo II: Definición del marco teórico, donde se presenta el análisis e investigación para el desarrollo del sistema multidimensional y cuadro de mando integral, Capítulo III: Presenta los procesos de la metodología Crisp-dm, en la cual se detallan 6 etapas para la construcción de sistemas multidimensionales y cuadro de mando integral, Capítulo IV: Se presentan el análisis de los resultados obtenidos de completitud, credibilidad y precisión haciendo uso de la herramienta Talend de acuerdo con los objetivos planteados, y también se presenta la discusión de estos.

1.1 PLANTEAMIENTO Y JUSTIFICACIÓN DE LA INVESTIGACIÓN

Se identificó que en el Ministerio de salud pública en la Coordinación Zonal 3 - área materno neonatal tienen procesos con limitaciones para la generación de información que permita tomar decisiones en base al análisis de datos. En la actualidad la Coordinación Zonal 3 recopila datos de 278 establecimientos de salud en archivos Excel, sin embargo, la data es inconsistente porque existe errores de digitación.

Otro inconveniente tecnológico que se detectó es el uso de software no especializado para el procesamiento de información (Excel- matriz de reporte de indicadores de calidad materno neonatal). El proceso de recopilación de archivos de Excel de los establecimientos es manual, generando trabajos logísticos para la consolidación, integración y depuración de la información; una vez recopilada dicha información encontramos inconsistencia, data cruda, no precisa, etc.

Adicional a lo mencionado se identificó que la información registrada en los archivos de Excel llega incompleta, a destiempo y no es recopilada en tiempo real.

Ante esta problemática identificada, en proyecto de titulación se aplicó buenas prácticas que permitieron establecer un proceso de inteligencia de negocios, generar información de calidad, integrar procesos de ETL y mostrar cuadros de mando integral que permitan visualizar de manera clara el estado actual de la información de la organización.

Se ha considerado la siguiente pregunta de investigación: ¿Cómo el sistema multidimensional y cuadro de mando integral permitirá mejorar la completitud, credibilidad y precisión de las fuentes de información del sistema Materno Neonatal del Ministerio de Salud Pública?

1.2 OBJETIVOS

1.2.1 Objetivo general

Desarrollar un sistema multidimensional y cuadro de mando integral con fuentes de información del sistema Materno Neonatal del Ministerio de Salud Pública.

1.2.2 Objetivos específicos

- Analizar CRISP-DM para sistemas multidimensionales y cuadros de mando integral.
- Diseñar el sistema multidimensional y cuadro de mando integral utilizando la librería Shiny de R Studio con los datos del sistema Materno Neonatal del Ministerio de Salud Pública.
- Evaluar la completitud, credibilidad y precisión del sistema multidimensional según la norma ISO/IEC-25012.

CAPÍTULO II. MARCO TEÓRICO

2.1 Data Warehouse

Las organizaciones en la actualidad buscan elevar sus utilidades minimizando costes en su producción, para ello realizan procesos estratégicos y un análisis potente con el fin de obtener un mayor alcance y cubrir todas las necesidades que requiere, para realizar estos procedimientos hacen uso de tecnologías como el Data Warehouse, el cual es un almacén de datos variables en el tiempo, lógicos, no volátiles que se obtienen de múltiples fuentes y dan apoyo a la toma de decisiones (Inmon, 2002).

Beneficios del Data Warehouse

El data Warehouse provee varios beneficios a las empresas que lo implementan, entre los que se destacan:

- Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis.
- Facilita la integración de sistemas de aplicación no integrados.
- Organiza y almacena los datos que se necesitan para el procesamiento analítico informático sobre una amplia perspectiva de tiempo.
- Los datos que ingresan en el data Warehouse provienen del ambiente operacional en la mayoría de los casos.

Para obtener un tratamiento adecuado de los datos, según (Lanang, 2020), un data Warehouse presenta las siguientes características y requerimientos funcionales:

Características de un Data Warehouse

- Orientado hacia la información relevante de la organización
- Datos integrados
- Variable en el tiempo
- No volátil

Requerimientos fundamentales

- Acceso universal a los datos
- Implantación de un sistema abierto
- Separación lógica y física de la base de datos de almacén de todos
- Creación de herramientas de consulta para el usuario final.

Según (Inmon et al., 2008), las empresas tienen en su 80% la información relevante de forma no estructurada ya sea en formato texto o formatos xls, lo cual convertir esta información a forma estructurada es un verdadero desafío, para ello los Data Warehouse se apoyan en los sistemas multidimensionales y Data Mart para estructurar la información en pequeñas cantidades ya sea por áreas o departamentos.

Los data Warehouse tienen la capacidad de estar compuestos por Data Mart, que son una parte de los datos con las mismas propiedades de los data Warehouse pero orientándose a los diferentes departamentos que tiene una organización, ayudando a sí a establecer e integrar el sistema de manera más clara, el cual debe cumplir con objetivos que estén dirigidos al ámbito empresarial y técnico, un Data Warehouse debe ser un repositorio integrado y orientado al entorno empresarial y también debe ser un repositorio base que presente análisis y reportaría (Abelló, 2019).

Debido a que el Data Warehouse está orientado a la analítica, debe cumplir con un procesamiento diferente al de los sistemas operacionales, por lo que requiere un modelo de base de datos que se acerque más a la visión del usuario final, y debe ser de fácil acceso, navegación y recuperación de información, este modelo se lo conoce como sistema multidimensional (Cedeño, 2006).

2.2 Sistema Multidimensional

Un sistema multidimensional permite a los analistas resumir, acceder y organizar los datos, dando la facilidad al usuario final de gestionar y visualizar la información de forma simple de todos los procesos que ocurren dentro de una organización. La razón principal para desarrollar, diseñar o crear un sistema multidimensional es lograr un rendimiento rápido al momento de realizar consultas de datos empresariales y tomar decisiones (Cedeño, 2006). Un sistema multidimensional se compone de hechos, dimensiones y debe constar una dimensión tiempo.

Tabla de hechos:

- Contiene las claves primarias de todas las dimensiones relacionadas.
- Almacena las medidas numéricas que facilitan múltiple cantidad de registros que se involucran en una consulta y dar una respuesta rápida cuando se ejecute una consulta que requiera gran cantidad de información.

Dimensiones:

- Contienen una clave y los atributos que representan la dimensión.
- La tabla de dimensión alimenta a la tabla de hechos.
- Los atributos de la tabla de dimensiones sirven de restricciones para las consultas realizadas por el usuario.
- Debe tener una estructura correcta ya que recae la responsabilidad de describir el pasado con precisión, además se puede sobrescribir un nuevo dato ingresado y borrar el que ya estaba.

Dimensión Tiempo:

- Es la dimensión con mayor importancia dentro de un sistema multidimensional ya que en ella se registran en qué momento ocurrió el hecho.

- Se registra el almacenamiento histórico de la información.
- Debido a que es de tipo tiempo debe asegurarse un orden implícito la inserción de los datos.

Modelado de un sistema multidimensional

Para hacer la representación de un sistema multidimensional existen dos tipos: esquema estrella y copo de nieve.

- **Esquema estrella**

Este esquema es el más usado en su estructura cuenta con dimensiones en las cuales se encuentra la información que sirve como criterio de búsqueda cuando el usuario realiza una consulta, debe existir una dimensión tiempo, una tabla adicional por cada dimensión y una tabla de hechos (Kimball, 2005).

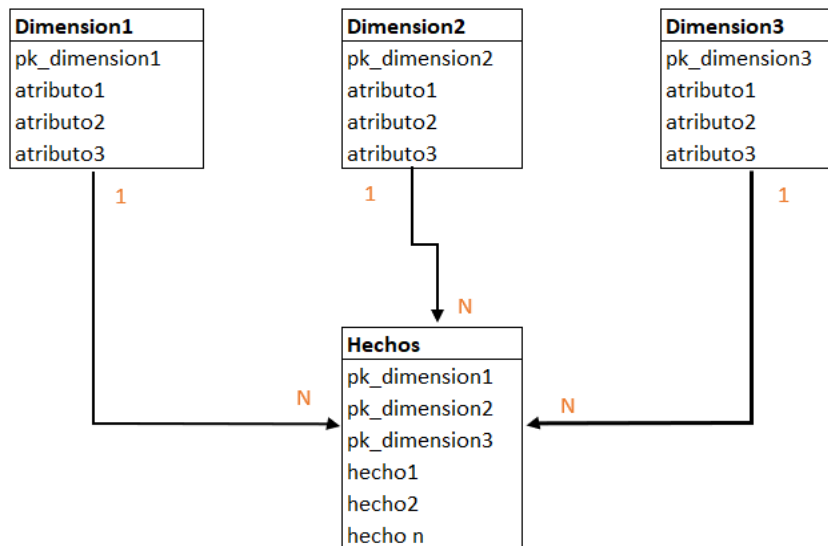


Figura 1: Esquema tipo estrella

Fuente: Lidia Enríquez, Clinton Moina

- **Esquema copo de nieve**

En su estructura cuenta con dimensiones, pero estas dimensiones cuentan con ramificaciones, y también con la tabla de hechos, la ventaja que presenta este esquema es que ayuda a reducir el espacio de almacenamiento y la desventaja es que aumenta el número de dimensiones por lo que dificulta que el usuario pueda realizar la consulta de manera optimizada. (Kimball, 2005).

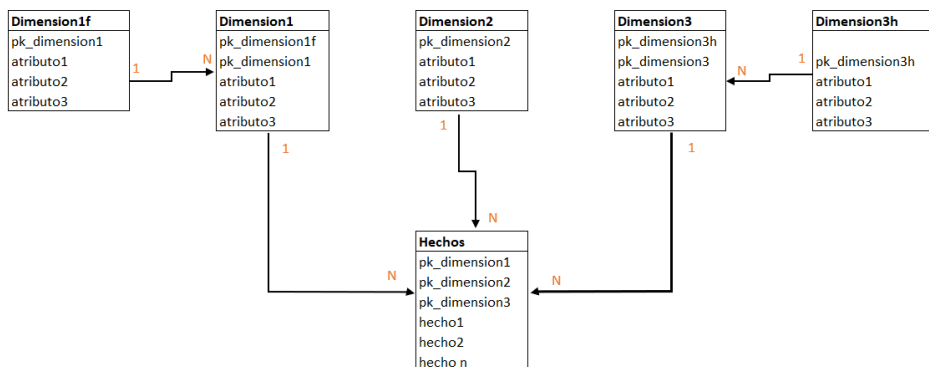


Figura 2: Esquema copo de nieve

Fuente: Lidia Enríquez, Clinton Moina

2.3 ETL

En la actualidad el incremento de soluciones de negocio hace que los procesos de ETL sean una necesidad para evaluar la calidad de los datos.

ETL es el proceso responsable de extraer los datos de diferentes fuentes y colocar en un almacén de datos proporcionando una generación automática de esquemas de datos conceptuales asignados directamente a la fuente de datos (Bargui et al., 2016).

Extracción. Extrae datos de diferentes sistemas, fuente de datos, archivos de texto, archivos xls de origen, realiza la operación de selección y procesa una salida dando como resultado datos cargados en memoria (Datos Crudos).

Transformación. El proceso de extracción toma los datos crudos, aplica las reglas de negocio y esta a su vez calcula las nuevas dimensiones y medidas.

- Limpia los datos
- Filtrado de los datos
- Combinación de diferentes datos por columnas.

Y en salida muestra datos resumidos de acuerdo con las necesidades del negocio, estructurados, con el formato correspondiente a cada dato, todo este proceso aún permanece en memoria. (Bustamante et al., 2013).

Carga. Toma los datos que se hizo en el proceso de transformación, realiza la operación de inserción y muestra como resultado final los datos con su formato respectivo, estructurados y de forma resumida en el sistema multidimensional, con este proceso finalizado los datos están listos para realizar informes. (Bustamante et al., 2013).

La utilización del ETL ayuda a simplificar la gestión de los datos y reducir el esfuerzo durante un análisis como también el tiempo y dinero eliminando la necesidad del desarrollo de un almacén de datos (Mali & Sachin, 2015).

2.4 Pentaho

Pentaho es una herramienta flexible, de código abierto y potente para hacer inteligencia de negocios, permite integrar los datos, servicios OLAP, informes, paneles de información, minería de datos y capacidad para extraer, transformar y cargar datos de diferentes almacenes para la inteligencia de negocios (Bouman, 2009).

2.5 Cuadro de Mando Integral

En la actualidad las empresas tienen la necesidad de contar con herramientas que les ayuden a medir y evaluar los resultados obtenidos, en este mismo contexto una de las herramientas novedosas es el cuadro de mando integral, que permite realizar un análisis de todos los procesos y principales áreas de la organización.

El CMI - Cuadro de Mando Integral mediante indicadores de gestión ofrece una definición práctica para llevar a cabo los contenidos estratégicos. Traduce la misión de la organización en un aglomerado de métricas que permiten a los directivos actuar y tomar decisiones en base a los resultados.

El CMI trabaja dentro de 4 perspectivas específicas: financiera, del cliente, de procesos internos y de aprendizaje y crecimiento (Ghiglione, 2021).

2.6 Lenguaje de programación R

Según (Contreras et al., 2014) R es un lenguaje de programación libre y multiplataforma, enfocado a realizar análisis y presentación de información (gráficos) estadística e investigación, incluye manuales y librerías con las que se puede realizar distintos análisis estadísticos y presentar de manera gráfica la información obtenida, en la actualidad ningún programa maneja todas las condiciones, recursos y manejabilidad que R presenta, es por eso que ha ganado popularidad porque la curva de aprendizaje es relativamente sencilla siendo el más utilizado para el aprendizaje automático, minería de datos, investigación biomédica, bioinformática financiera (RStudio, 2022) tiene varias características:

- Manejo completo de los datos.
- Almacena una gran colección de herramientas para análisis de datos.
- Por medio de graficas presenta el análisis realizado.
- Permite combinar en un solo programa múltiples funciones estadísticas.
- Integra varios lenguajes como es el C, C++ o Fortran para tareas de datos computacionales intensivas (José et al., 2014).

2.6.1 R Studio

Es un entorno de desarrollo integrado, multiplataforma, simplifica la creación y trabajo de proyectos de R de distintos tipos, además se puede implementar como servidor para los sistemas que se ejecutan en remoto y tengan acceso web (Vernazi, 2011).

Paquetes de R

La comunidad de R ha desarrollado varios paquetes que son un conjunto de funciones pueden encontrarse en repositorios línea de forma libre, ayudan a mantener un trabajo organizado, compartir con otros usuarios e incluye documentación,

Se analiza los siguientes paquetes: Shiny, viridis y ggplot2.

Cuadro comparativo de paquetes R

Tabla 1: Cuadro Comparativo de paquetes de R

	Características	Ventajas	Desventajas	Configuración
Shiny	Crea interfaces. Construcción de aplicaciones Desarrollado bajo la programación reactiva Incorporan componentes activos	Funciona en cualquier interfaz de R. Comunicación entre navegador. Fácil accesibilidad Automatización para crear proyectos	No ofrece tantas funcionalidades como las aplicaciones de escritorio.	Alta
Viridis	Fácil visualización Colores y apariencia similar Crea gráficos llamativos	Los gráficos son fáciles de comprender. Fáciles de leer para personas con daltonismo	Poca documentación	Baja
Ggplot2	Es un paquete para graficar Facilita crear gráficas complejas.	Estructura consistente de gráficos Alto nivel de abstracción	No se puede realizar gráficos tridimensionales No se puede Ver paquetes de r	Baja

Fuente: (RStudio, 2022)

Luego de haber analizado las librerías de R estudio, se decidió que el más adecuado para construir el cuadro de mando integral es el paquete Shiny.

2.6.2 Paquete Shiny

Shiny es un paquete de R creado en el año 2012 que permite la construcción de cuadro de mando o aplicaciones web con interacción entre el servidor y el usuario con el uso de R, (Jaramillo, 2019). Se puede obtener de forma rápida el análisis y visualización de datos ingresados, incorporando componentes y funcionalidades flexibles e interesantes (Gómez et al., 2016).

2.6.2.1 Desarrollo de aplicaciones web en Shiny

Para el desarrollo de aplicaciones web en Shiny se establecen los siguientes parámetros:

Funcionamiento de Shiny

Shiny nos permite desarrollar aplicaciones web dinámicas con interacción entre cliente y servidor mediante programación, que son aquellas que cambian su información o estado que se visualiza en pantalla cuando un usuario realiza una acción, las cuales hoy en día son muy utilizadas por su facilidad de entender (Chaure, 2022), una de las ventajas que presenta Shiny es que no se necesita tener conocimientos en desarrollo web, sin embargo se puede personalizar las aplicaciones con HTML Y CSS. Para desarrollar aplicaciones en Shiny, se basa en dos componentes que existen dentro del framework que son los reactantes y reactivos:

Para entender el concepto de estos dos componentes se ejemplifica de la siguiente manera, considere una página web que tenga dos elementos, un campo de texto y un gráfico de barras, pastel o el que se considere necesario para representar los datos. En el campo de texto el usuario podrá ingresar un número y por ende el gráfico de barras cambiará el número de barras, basándonos en este ejemplo, el componente reactante es el campo de texto, porque esté causa la reacción en la página web, y el componente reactivo es el gráfico de barras, ya que este reacciona a los cambios que le envía el componente reactivo, esto depende de cómo se haya configurado.

Estructura

Según (Ramirez, 2017), la tecnología Shiny para el desarrollo de las aplicaciones consta de los siguientes componentes:

- App.R: archivo que contiene los elementos de la interfaz y del servidor.
- Server.R: archivo que se desarrolla toda la lógica del servidor y el contenido dinámico para la interacción en las pantallas.
- Ui. R: archivo que se especifica los elementos de interfaz.

Crear archivos

En Shiny para poder visualizar algunos componentes web en la aplicación, el directorio que contiene los archivos y código se encuentra en app/views.

Para mostrar las reacciones, al momento de presionar un botón, cuando se ingresa o se cambie un valor dentro de un campo de texto, y además se puede subir un archivo, esto se hace en el directorio app/controllers.

En Shiny no se puede tener una aplicación que este en diferentes rutas (Chaure, 2022).

Publicación de la aplicación en el servidor web

Una vez que se desarrolla la parte visual y lógica de la aplicación web, y se tiene toda la información totalmente limpia, es decir, ya paso el proceso de ETL, esta queda lista para pasar a periodo de prueba, la mejor forma es que está información sea entendible para el usuario final, por lo tanto, se procede hacer la publicación en un servidor web que permita el acceso desde cualquier parte, en el servidor debe tener instalado Shiny como servicio (Chaure, 2022).

2.7 Metodología Crisp-dm

La metodología CRISP-DM (Cross Industry Estándar Process for Data Mining), es una metodología empleada para el desarrollo de proyectos de minería de datos y análisis de datos, que incluye fases bidireccionales, creada por las empresas SPSS, Daimenr Chrysler y NCR en el año 2000 (Moine et al., 2011).

Para el desarrollo del sistema multidimensional y cuadro de mando integral se hizo uso de la metodología Crisp-dm que utiliza los procesos, que se muestran en la figura:



Figura 3: Etapas de ciclo de vida de la metodología Crisp-dm

Fuente: (Haya, 2021)

Fase 1: Compresión del negocio: En esta fase el objetivo es alinear los objetivos del negocio con los objetivos del proyecto, con la finalidad de evitar crear un sistema multidimensional que no produzca un efecto real y fiable en la organización, es considerada la fase más fundamental de la metodología (Espinoza, 2020). En esta fase se debe alinear a los siguientes puntos:

- Determinación de objetivos: Proponer una solución donde los datos sean de calidad y que haga referencia a las metas puntualizadas.
- Evaluación de la situación actual: Antes de iniciar la implementación de la solución se debe comprender de forma puntual la situación en términos de negocio y así se logrará obtener éxito en proyecto (Galán, 2015).

Fase 2: Compresión de los datos: En esta fase se establece un primer contacto con el problema, familiarizarse con los datos que ayudarán a obtener la solución, verificar la calidad (Galán, 2015), esto se llevará a cabo tomando en cuenta lo siguiente.

- Recolección de datos: Recolectar los datos iniciales y prepararlos para realizar el procesamiento.
- Descripción de datos: En este punto se especifica el significado de cada campo, el formato inicial y el volumen de los datos.
- Exploración de datos: Realizados los pasos anteriores, se procede a encontrar una estructura para todos los datos, se crea tablas y gráficos de distribución.
- Verificar la calidad de los datos: Una vez hecha la exploración de los datos se puede afirmar que están completos, y ayudarán a cumplir con los objetivos propuestos del proyecto (Galán, 2015).

Fase 3: Preparación de los datos: Está es una de las fases que requiere mayor tiempo de ejecución, ya que los puntos que se mencionan a continuación son críticos para obtener un sistema multidimensional de calidad:

- Limpieza de datos: Aplicar diferentes técnicas, la cual incluye mayor esfuerzo y tiempo por las técnicas que existen.
- Creación de indicadores: Crear indicadores que complementen la capacidad predictiva de los datos a partir de los datos que ya existen y con esto entender o detectar ciertos comportamientos para modelar.
- Transformación de datos: Los datos deben conservar la integridad, lo que se cambia es el formato o estructura, con la finalidad de aplicar técnicas en la siguiente etapa (Galán, 2015).

Fase 4: Modelado de los datos: Depende de las propiedad y características de los datos. (Galán, 2015).

- Selección de técnica de modelado: Tomar en cuenta el tipo de problemática que se quiere resolver, considerando el objetivo principal y la relación con técnicas que ya existen.
- Generar el plan de prueba: Segmentar los datos de pruebas y validación

- Obtención del modelo: Crear el mejor modelo a través de un proceso iterativo de modificación de parámetros (Espinoza, 2020).

Fase 5: Evaluación

Según (Espinoza, 2020), En esta fase ya se realiza la evaluación del modelo, tomando en cuenta el cumplimiento de los criterios e indicadores de éxito de la problemática en cuestión, con los datos que se tomó para realizar el análisis.

Fase 6: Despliegue:

Está es la última fase de la metodología Crisp-dm, se obtienen los resultados que cumplen con el objetivo del modelo creado y se transformas la información en conocimiento que es de vital importancia para la organización (Zuñiga & Galán, 2020).

2.8 Norma ISO/IEC 25012

El estándar ISO/IEC 25012 define la calidad de los datos, contiene 15 características que se relacionan a calidad de datos: inherente y que dependen del sistema.

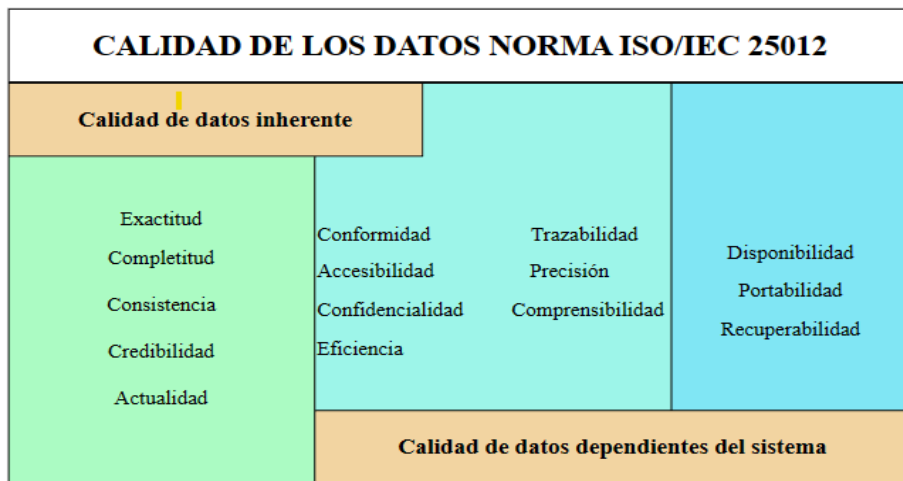


Figura 4: Características de la calidad de los datos según la norma IS/IEC 25012

Fuente: Lidia Enríquez, Clinton Moina

De todos los criterios anteriormente descritos los que se relacionan directamente con la investigación son: completitud, credibilidad y precisión.

Compleitud: Los datos requeridos para un uso específico están presentes de forma coherente y lógica en el conjunto de datos.

En la tabla 2 se indica los parámetros para medir la completitud de los datos basados en la norma ISO/IEC 25012.

Tabla 2: Criterios para medir la Completitud de los datos

Tipo	Inherente
Documentación previa requerida	Por cada atributo a evaluar: indicar si es obligatorio
Método para la medición	Por cada atributo, verificar que no haya datos vacíos/en blanco
Variables	X= Porcentaje de datos del atributo que se encuentran completos
Fórmula	Valor = $x/100\%$
Escala	Casi ningún dato se encuentra completo ($X \leq 10\%$). Algunos datos se encuentran completos, pero se observan muchos incompletos ($X > 10\% \& X \leq 45\%$). Muchos datos se encuentran completos y se observan algunos incompletos ($X > 45\% \& X \leq 85\%$). La gran totalidad de los datos se encuentran completos ($X > 85\%$).

Fuente: (Calabrese et al., 2019)

Credibilidad: Capacidad de los datos para que el usuario los considere creíbles o verdaderos, según la norma ISO/IEC para medir la credibilidad existen múltiples técnicas, como las siguientes:

- Verificación: se puede obtener mediante una revisión manual de los datos y también usando herramientas específicas.
- Validación: incluye la verificación de la estructura de los datos, junto con la consistencia y la integridad de los datos.

Tabla 3: Criterios para medir la Credibilidad de los datos

Peso	Valor	Valor %
Bajo	0,00 – 0,33	0% - 33%
Medio	0,34 – 0,75	34% - 75%
Alto	0,76 – 1,00	76% - 100%

Fuente: (González & González, 2013)

- Pruebas de calidad: Se puede realizar pruebas de integración, aceptación o pruebas de sistema.
- Cuestionario o encuesta: Se realiza esta técnica a los usuarios especializados en el dominio y manejo del sistema (ISO/IEC, 2008).

Precisión: Capacidad de los datos para ser exactos y permitir discriminar adecuadamente, existen diferentes métodos y técnicas para obtener la métrica, como son las siguientes:

- Validación de los datos: Incluye verificar la estructura y la integridad de los datos, siendo repartidos los pesos de medición de la siguiente manera:

Tabla 4: Criterios para medir la Precisión de los datos

Peso	Valor	Valor %
Bajo	0,00 – 0,33	0% - 33%
Medio	0,34 – 0,75	34% - 75%
Alto	0,76 – 1,00	76% - 100%

Fuente: (González & González, 2013)

- Comparación con fuentes externas: Se realiza la comparación con fuentes externas que proporcionen calidad y precisión para determinar si existen discrepancias.
- Análisis estadísticos: Se puede calcular la media, desviación estándar para obtener la precisión de los datos.
- Revisión manual: Está técnica se la puede aplicar cuando existe un usuario experto en el tema y domina el sistema de información (ISO/IEC, 2008).

2.9 Talend Data Quality

La herramienta Talend ayuda a las organizaciones a obtener datos saludables luego de haber realizado todo el proceso de extracción, transformación y carga o a su vez haber aplicado las reglas de negocio necesarias. Talend proporciona a los usuarios el beneficio de dedicarse el 30% de su tiempo a recopilar la data y el 70% a realizar el análisis y tomar decisiones certeras (Talend, 2023).

Es un tanto complicado recopilar datos que sean correctos de diversas fuentes y mayormente a la velocidad que los negocios exigen hoy en día, Talend proporciona un enfoque unificado donde combina la integración, transformación y mapeo de datos (Talend, 2023). Se eligió esta herramienta para obtener los resultados de completitud, credibilidad y precisión según las métricas que proporciona la norma ISO/IEC 25012 que se mencionaron anteriormente.

CAPÍTULO III. METODOLOGIA

3.1 Tipo y Diseño de investigación

En esta investigación se usa un enfoque mixto debido a que involucra la recolección y análisis de datos cuantitativos y cualitativos, así como integración y discusión para lograr un mayor entendimiento del objeto de estudio (Fernández & Baptista, 2014). Se aplicó la metodología Crisp-dm, que permitió llevar a cabo el proceso de análisis de datos, para el desarrollo de la investigación se aplicó los pasos que se describen a continuación:

- **Recolección y análisis de datos:** Los datos se obtuvieron de la matriz de reporte Distrital del Monitoreo de la Calidad de Atención Materno Neonatal de la Coordinación Zonal 3 del Ministerio de Salud Pública que se encuentra en formato Excel y contiene información de zona, provincia, cantón, distrito, establecimiento, tiempo e indicadores.

Se utilizó el método analítico, debido a que mediante el análisis de la información obtenida se obtendrá los datos necesarios para el desarrollo del proyecto.

- **Comprensión de los datos:** Se hizo una descripción general de los datos mediante el uso de gráficos y tablas, y se efectuó un análisis de los datos obteniendo el número de valores nulos, blancos y no válidos, este proceso se lo realizó mediante la investigación descriptiva y exploratoria.
- **Preparación de los datos:** Se aplicó el método deductivo ya que se seleccionaron los datos finales para el estudio aplicando procesos de ETL.

También se aplicó diseño documental el cual se basa en un proceso sistemático de recolección, análisis e interpretación de datos o información de un determinado tema, porque analiza y ofrece de una manera organizada toda la información disponible sobre un tema específico (Morales, 2013). Además, se desarrolla en cualquier ámbito donde exista un registro de información, misma que la participación de personal aporte comparaciones, producción de nuevos conocimientos y propuestas.

3.2 Técnicas de análisis e interpretación

Para el presente proyecto de investigación se utilizó la herramienta Talend Data Quality para obtener la completitud, credibilidad y precisión de los datos según la Norma ISO/IEC 25012, se aplicó reglas de ETL (extracción, transformación y carga) y se aplicó la metodología Crisp-dm para el desarrollo del sistema Multidimensional y Cuadro de Mando Integral.

3.3 Población y muestra

Para la presente investigación se establece como población de estudio la data de la matriz de reporte distrital del monitoreo de Calidad de la atención Materno Neonatal la misma que contiene 278 establecimientos de la Coordinación Zonal 3 del Ministerio de Salud Pública.

Para cada establecimiento se registran en una matriz 22 indicadores de la estrategia calidad materno neonatal.

Para el proyecto de investigación se considera toda la población de los 278 establecimientos, no se aplica muestra porque la población es procesable por las herramientas de inteligencia de negocio que existen en la actualidad.

3.4 Operacionalización de las variables

Tabla 5: Tabla de operacionalización de las variables

Problema	Tema	Objetivos	Variables	Indicadores
<p>Se ha logrado identificar que en Ministerio de salud pública en la Coordinación Zonal 3 - área materno neonatal tienen procesos con limitaciones para la generación de información que les permita tomar decisiones en base al análisis de datos, en la actualidad la coordinación zonal 3 recopila datos de 278 establecimientos de salud en archivos de Excel, sin embargo, existe errores de digitación, la data es inconsistente, por lo tanto, requieren aplicar ETL.</p> <p>Otro inconveniente tecnológico detectado es el uso de software no especializado para el procesamiento de información (Excel- matriz de reporte de indicadores de calidad materno neonatal). El proceso de recopilación de archivos en Excel de los 278 establecimientos es manual generando trabajos logísticos para la consolidación, integración y depuración de la información; una vez recopilada dicha información encontramos inconsistencia, data cruda, no precisa, etc.</p> <p>Adicional a lo mencionado se ha identificado que la información registrada en los archivos de Excel llega incompleta, a destiempo y no es recopilada en tiempo real.</p> <p>Ante esta problemática identificada, el proyecto de titulación pretende desarrollar una solución informática y la aplicación de buenas prácticas que permita establezca un proceso de inteligencia de negocios, genere información de calidad, integre procesos de ETL y muestre cuadros de mando integral que permitan visualizar de manera clara el estado actual de la información de la organización.</p> <p>Se ha considera la siguiente pregunta de investigación: ¿Cómo el sistema multidimensional y cuadro de mando integral en SHINY permitirá mejorar la completitud, credibilidad y precisión de las fuentes de información del sistema Materno Neonatal del Ministerio de Salud Pública?</p>	<p>Sistema multidimensional y cuadro de mando integral en SHINY con fuentes de información del sistema Materno Neonatal Ministerio de Salud Pública</p>	<p>General Desarrollar un sistema multidimensional y cuadro de mando integral en Shiny con los datos del sistema Materno Neonatal del Ministerio de Salud Pública.</p> <p>Específicos: Analizar sistemas multidimensionales y cuadro de mando integral. Diseñar el sistema multidimensional y cuadro de mando integral utilizando la librería Shiny para el sistema Materno Neonatal del Ministerio de Salud Pública. Evaluar la completitud, credibilidad y precisión del sistema multidimensional según la norma ISO/IEC-25012.</p>	<p>Dependiente: Fuentes de información del sistema Materno Neonatal Ministerio de Salud Pública</p> <p>Independiente: Sistema multidimensional y cuadro de mando integral.</p>	<p>Dependiente: Calidad de los datos ISO/IEC 25012 Completitud: 0 % de valores blancos y nulos.</p> <p>Credibilidad: Cantidad de atributos vs cantidad de necesidades del negocio. Precisión: Cantidad de datos vs cantidad de fuentes de referencia.</p> <p>Independiente: Número de requisitos funcionales</p>

Fuente: Lidia Enríquez, Clinton Moina

3.5 Método de análisis y procedimiento de datos

- a. Análisis exploratorio de la data – Perfilamiento de la data
- b. Uso de la metodología Crisp-dm, que considera las siguientes fases
 1. Comprensión del negocio.
 2. Comprensión de los datos.
 3. Preparación de los datos: Estructurar los datos de la matriz de Excel proporcionado.
 4. Modelado de los datos: Aplicar ETL a la data.
 5. Evaluación de los datos: Análisis de los datos obtenidos luego de aplicar la metodología Crisp-dm y verificando los atributos de calidad de los datos: completitud, credibilidad y precisión.
 6. Despliegue de los datos: resultados en el cuadro de mando integral diseñado.

3.6 Método de análisis y procedimiento de datos

En este punto se presenta cada una de las fases que se aplicó en el desarrollo del sistema multidimensional y cuadro de mando integral, las mismas que conforman la metodología CRIPS-DM, a continuación, se detalla las 6 fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado de los datos, Evaluación y Despliegue.

3.6.1 Fase I. Comprensión de los datos

Determinación de objetivos

El objetivo del trabajo de investigación es obtener la completitud, credibilidad y precisión de los datos de la Matriz de Reporte Distrital del Monitoreo de la Calidad de la Atención Materno Neonatal que pertenece a la coordinación zonal 3 del Ministerio de Salud Pública mediante la creación de un sistema multidimensional y cuadro de mando integral que apoye a la toma de decisiones y a la gestión continua de la coordinación zonal 3.

Evaluación de la situación actual

Se dispone de la data del área de calidad materno neonatal entregada por la coordinadora de la coordinación zonal 3, misma que se encuentra en formato xmls correspondiente al año 2021, 2022 en la cual se almacena data de provincias, cantones, zona, tiempo, establecimientos e indicadores correspondientes a la coordinación zonal 3 del Ministerio de Salud Pública.

Fuente de datos: La fuente de datos es un archivo de Excel xmls, como se puede observar en el Anexo1, este archivo contiene data desde el año 2021 y 2022, y contiene 1667 registros de provincias, 1667 registros de cantones, 1667 registros de establecimientos, 56174 registros de indicadores, 56173 registros de tiempo, 141 registros de zona.

Definición de los objetivos del Sistema Multidimensional

Para este proyecto de investigación se han especificado los siguientes objetivos:

- Presentar un reporte por zona

- Presentar un reporte por distrito
- Presentar un reporte por tipología
- Presentar un reporte por establecimientos
- Presentar un reporte por provincia

Realizar el plan del proyecto

Se ha definido 6 etapas para el desarrollo del proyecto, las que se detallan a continuación:

1. Comprensión del negocio y planteamiento de los objetivos del sistema multidimensional en 1 semana.
2. Comprensión y análisis de la data del área de calidad de materno neonatal en 2 semanas.
3. Preparación de los datos hacer el ETL (Extracción, transformación y carga) en pentaho en 3 semanas.
4. Diseño y ejecución del modelo en 3 semanas.
5. Evaluación de los resultados del punto 4 en 3 semanas.

3.6.2 Fase II. Comprensión de los datos

En la metodología CRISP-DM en su segunda fase se lleva a cabo la recopilación de datos iniciales, conocer su calidad, y entrar en contexto con la problemática que presenta el Ministerios de Salud pública en la estrategia de Calidad Materno Neonatal de la Coordinación Zonal 3.

Recolectar los datos iniciales

Los datos recolectados son referentes al área de calidad Materno Neonatal del MSP, los mismo que incluyen información de provincia, cantón, establecimiento, indicadores y tiempo, zona, dicha información fue tomada de la matriz proporcionada por la coordinadora de la Coordinación de zonal 3 que están almacenados en un archivo en formato xmls, data que está relacionada directamente con el problema en cuestión, ver anexo 1.

Adicional se hizo estructuración de la información en formato tabla para posterior sea leída en la herramienta Pentaho, ver anexo 2.

Descripción de los datos

Se ha detallado los datos recopilados en tablas, ver anexo 3.

Exploración de los datos

En este punto se realizó la exploración de la data proporcionada por la coordinación zonal 3, se basa en el “INSTRUCTIVO PARA EL USO DE LA BASE DE DATOS EXCEL MONITOREO Y REPORTE DE INDICADORES CONSOLIDADO DE LOS ESTABLECIMIENTOS DE SALUD DEL DISTRITO” proporcionado por la Coordinación Zonal 3. En la tabla 5 se puede observar los colores que determinan el nivel de calidad de los Indicadores 1 al 9C de acuerdo con el rango establecido, para los indicadores 10, 11, 12 y 13, ver anexo 4.

Tabla 6: Colorimetría del Nivel de Calidad de los Indicadores 1 al 9C

Color	Nivel de Calidad	Rango
Rojo	No Aceptable	0 a 70,99%
Amarillo	Alerta	71 a 90,99%
Verde	Aceptable	91 a 100%

Fuente: Coordinación Zonal 3 del Ministerio de Salud Pública

De acuerdo con lo establecido en la tabla 6, se hizo la exploración de los datos del indicador 1 (% Insumos esenciales atención Materna y RN) por trimestre, cuatrimestre, semestre y año de la provincia de Chimborazo.

La tabla 7 muestra el nivel de calidad en porcentaje **del indicador 1** (% Insumos esenciales atención Materna y RN), obtenido en la provincia de Chimborazo en los 4 trimestres del año 2022.

Tabla 7: Porcentaje del Nivel de Calidad del Indicador 1 por trimestre en la provincia de Chimborazo

Trimestre	Porcentaje	Nivel de Calidad
1	83,4494%	Alerta
2	84,7556%	Alerta
3	83,4688%	Alerta
4	0	No Aceptable

Fuente: Lidia Enríquez, Clinton Moina

La figura 5 muestra el porcentaje por trimestre del indicador 1 (% Insumos esenciales atención Materna y RN), de acuerdo con los colores en los 3 primeros trimestres tienen un nivel de calidad de alerta y en el último trimestre el nivel de calidad es no aceptable.

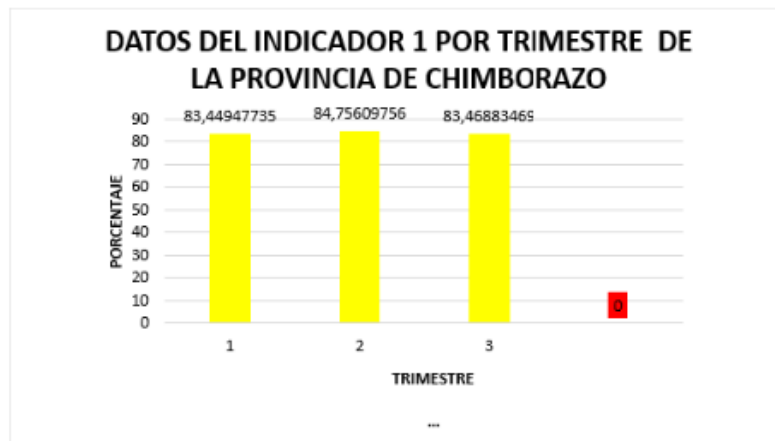


Figura 5: Datos del indicador 1 por trimestre de la provincia de Chimborazo

Fuente: Lidia Enríquez, Clinton Moina

La tabla 8 muestra el nivel de calidad en porcentaje que se encuentra el Indicador 1 (% Insumos esenciales atención Materna y RN), por cuatrimestre en la provincia de Chimborazo.

Tabla 8: Porcentaje del Nivel de Calidad del Indicador 1 por cuatrimestre en la provincia de Chimborazo

Trimestre	Porcentaje	Nivel de Calidad
1	84,24589348	Alerta
2	84,14634146	Alerta
3	83,46883469	Alerta

Fuente: Lidia Enríquez, Clinton Moina

La figura 6 muestra la distribución del porcentaje por cuatrimestre del indicador 1 (% Insumos esenciales atención Materna y RN), de acuerdo con los colores en los 3 cuatrimestres tiene un nivel de calidad de alerta.

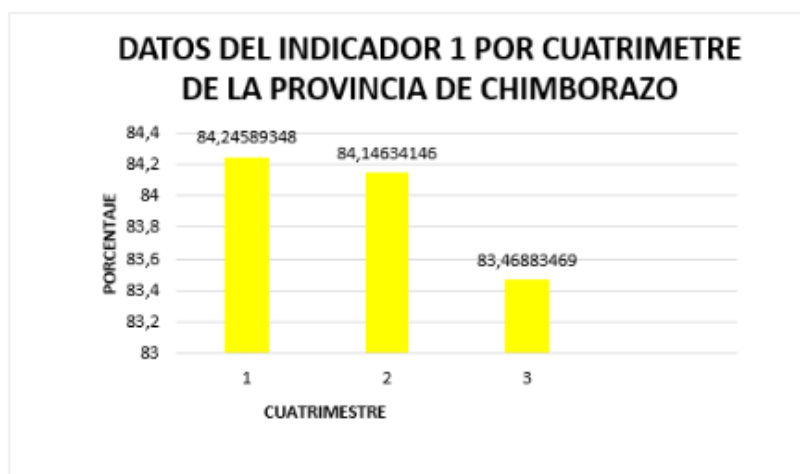


Figura 6: Datos del Indicador 1 por cuatrimestre de la provincia de Chimborazo

Fuente: Lidia Enríquez, Clinton Moina

En la tabla 9 muestra el nivel de calidad en porcentaje que se encuentra el Indicador 1 (% Insumos esenciales atención Materna y RN), por semestre en la provincia de Chimborazo.

Tabla 9: Porcentaje del Nivel de Calidad del Indicador 1 por semestre en la provincia de Chimborazo

Trimestre	Porcentaje	Nivel de Calidad
1	84,2459	Alerta
2	83,4688	Alerta

Fuente: Lidia Enríquez, Clinton Moina

La figura 7 muestra la distribución del porcentaje por semestre del indicador 1 (% Insumos esenciales atención Materna y RN), de acuerdo con los colores en los 2 semestres el nivel de calidad es de alerta.

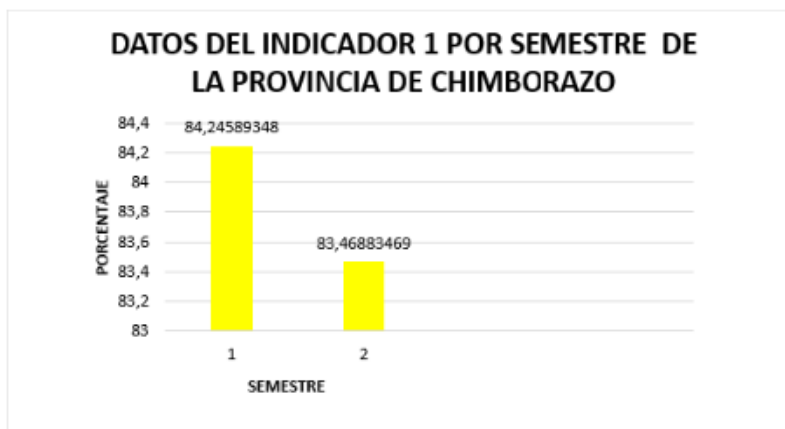


Figura 7: Datos del indicador 1 por semestre de la provincia de Chimborazo

Fuente: Lidia Enríquez, Clinton Moina

En la tabla 10 muestra el nivel de calidad en porcentaje que se encuentra el Indicador 1 (% Insumos esenciales atención Materna y RN), por año en la provincia de Chimborazo.

Tabla 10: Porcentaje del Nivel de Calidad del Indicador 1 por año en la provincia de Chimborazo

Trimestre	Porcentaje	Nivel de Calidad
1	83,9285%	Alerta

Fuente: Lidia Enríquez, Clinton Moina

La figura 8 muestra la distribución del porcentaje por año del indicador 1 (% Insumos esenciales atención Materna y RN), de acuerdo con los colores el nivel de calidad es de alerta

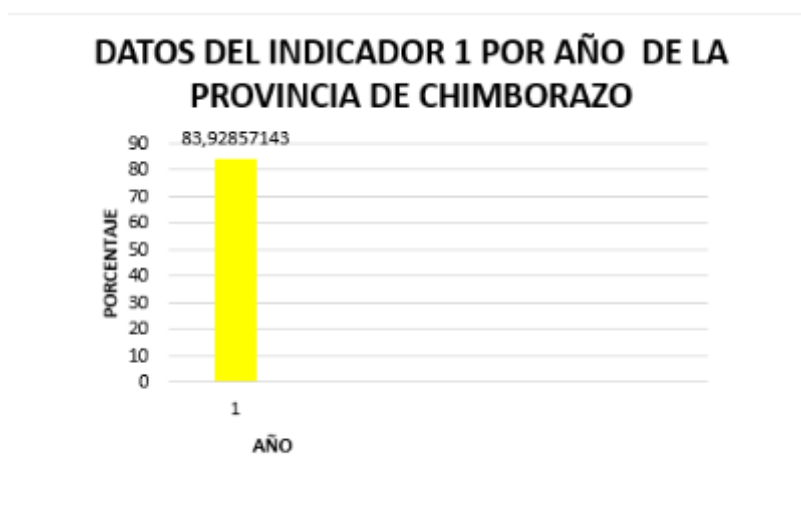


Figura 8: Datos del indicador 1 por año de la provincia de Chimborazo

Fuente: Lidia Enriquez, Clinton Moina

Verificar la calidad de los datos

Luego de la exploración de los datos, se verificó la calidad de los datos, obteniendo valores nulos, blancos, no válidos, ver anexo 5.

3.6.3 FASE III. Preparación de los Datos.

Seleccionar los datos

El proceso de selección de datos estuvo enfocado en la data que permita cumplir con los objetivos que planteados en la fase I de la metodología CRISP-DM, y determinar los campos a usar en este punto.

En la tabla 11 hasta la tabla 16 se presenta los campos seleccionados que se utilizará en el desarrollo del proyecto, se describen los campos de tipo código que hacen referencia al id de cada tabla y los atributos independientes, es decir que no tienen relación entre ellas, esto se debe al modelo tipo estrella que se generó, como se observa en la figura 11.

Tabla 11: Datos de la tabla Cantón

Atributo	Tipo	Naturaleza
Id_canton	Código	Original
Distrito	Independiente	Original
cantón	Independiente	Original
Fecha_ingreso	Independiente	Original

Fuente: Lidia Enríquez, Clinton Moina

Tabla 12: Datos de la tabla Provincia

Atributo	Tipo	Naturaleza
Id_provincia	Código	Original
Provincia	Independiente	Original
Fecha_ingreso	Independiente	Original

Fuente: Lidia Enríquez, Clinton Moina

Tabla 13: Descripción de la tabla Establecimiento

Atributos	Tipo	Naturaleza
Id	Código	Original
Id establecimiento	Independiente	Original
Establecimiento	Independiente	Original

Parroquia	Independiente	Original
Tipología	Independiente	Original

Fuente: Lidia Enríquez, Clinton Moina

Tabla 14: Descripción de la tabla Indicadores

Atributos	Tipo	Descripción
Id	Código	Original
Indicador	Independiente	Original
Nombre	Independiente	Original
Universo	Independiente	Original
Numerador	Independiente	Original
Denominador	Independiente	Original
Porcentaje	Independiente	Original
Fecha_ingreso	Independiente	Original
Grupom	Independiente	Original
Trimestre	Independiente	Original
Numeradort	Independiente	Original
Denominadort	Independiente	Original
Procentajet	Independiente	Original
Grupot	Independiente	Original
Cuatrimestre	Independiente	Original
Numeradorc	Independiente	Original
Denominadorc	Independiente	Original
Procentajec	Independiente	Original
Grupoc	Independiente	Original
Semestre	Independiente	Original
Numeradors	Independiente	Original
Denominadors	Independiente	Original
Porcentajes	Independiente	Original
Grupos	Independiente	Original
Anual	Independiente	Original
Numeradora	Independiente	Original
Denominadora	Independiente	Original
Porcentajea	Independiente	Original
Grupoa	Independiente	Original

Fuente: Lidia Enríquez, Clinton Moina

Tabla 15: Descripción de la tabla Zona

Atributos	Tipo	Descripción
------------------	-------------	--------------------

Id	Código	Original
Zona	Independiente	Original
Fecha_ingreso	Independiente	Original

Fuente: Lidia Enríquez, Clinton Moina

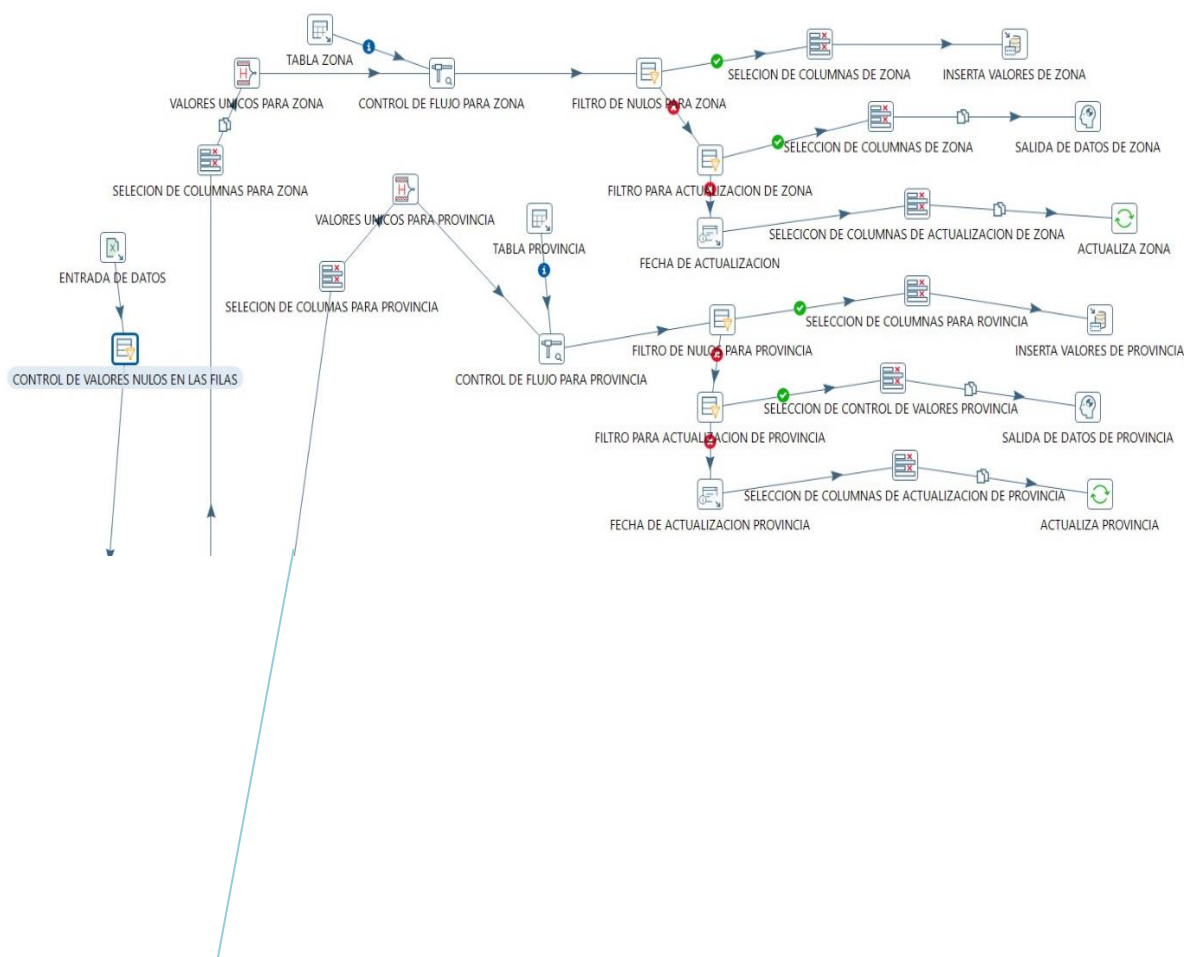
Tabla 16: Descripción de la tabla Tiempo

Atributos	Tipo	Descripción
Id	Código	Original
Indicador	Independiente	Original
Meses	Independiente	Original

Fuente: Lidia Enríquez, Clinton Moina

Limpieza de los datos

Para realizar la limpieza de los datos se lo realizo con la herramienta Pentaho, en la cual se inició importando el archivo de Excel, selección de filas y selección de valores distintos de nulos, blancos y datos reales. En la figura 9 se puede ver el sistema multidimensional diseñado.



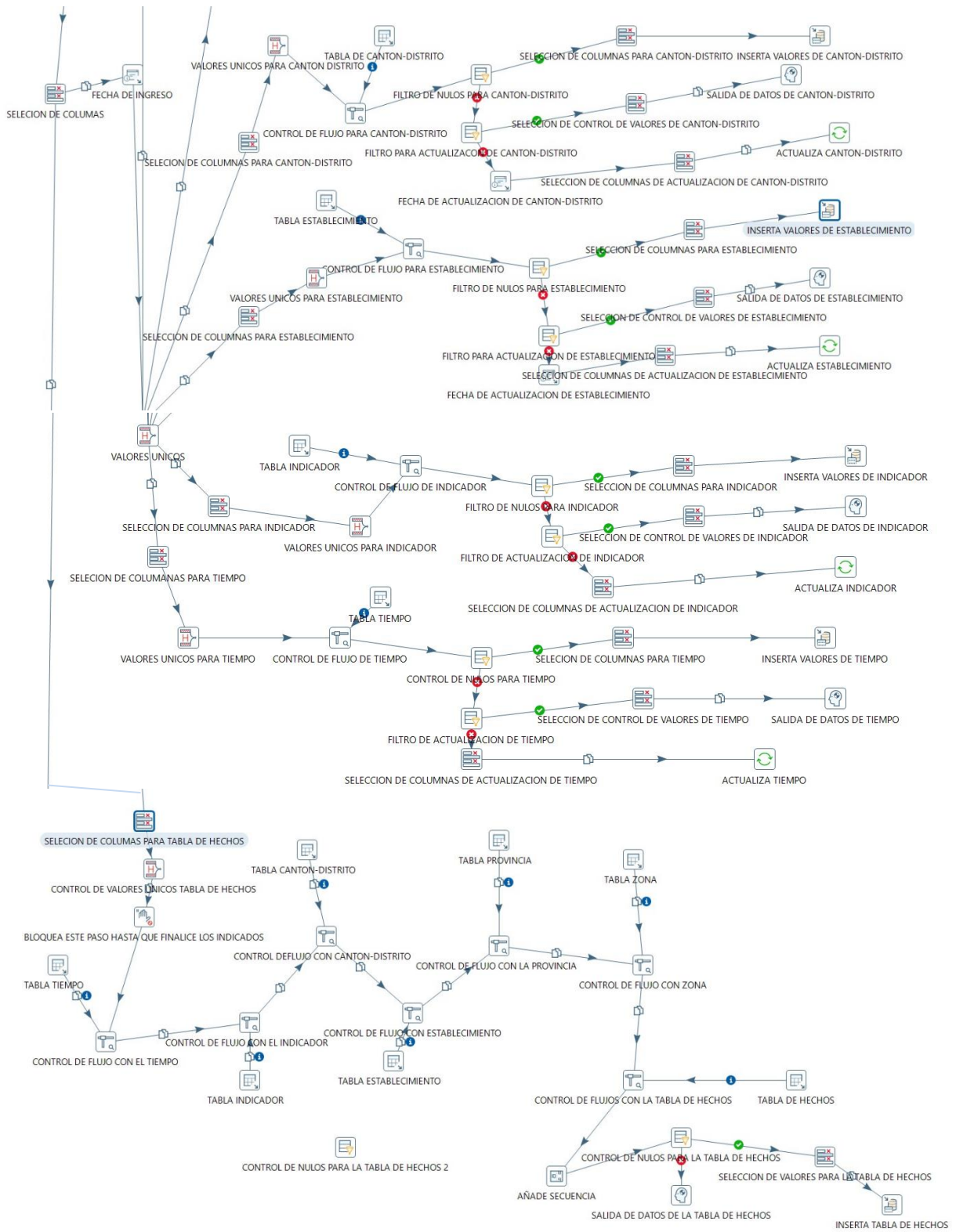


Figura 9: Diseño de Sistema Multidimensional aplicando procesos de ETL

Fuente: Lidia Enríquez, Clinton Moina

En la figura 10 podemos observar los parámetros que se han usado en el filtro.

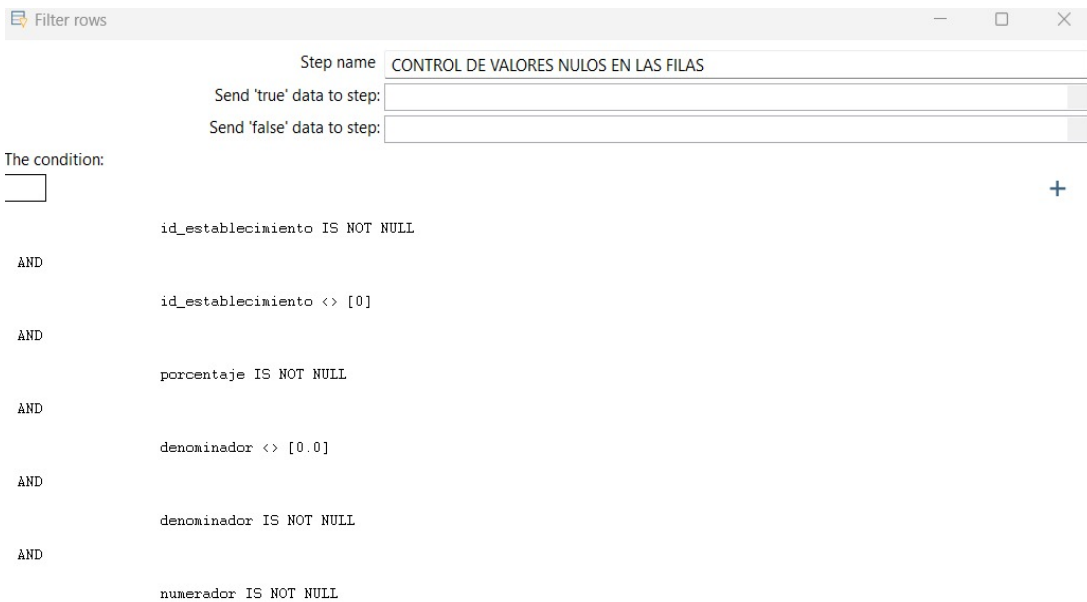


Figura 10: Filtro empleados en las tablas

Fuente: Lidia Enríquez, Clinton Moina

Creación de indicadores:

Para la creación de indicadores se toman en cuenta los siguientes puntos:

Atributos derivados

Para formar el sistema multidimensional tipo estrella se añadió la tabla Hechos la cual consta de todos los id de las tablas cantón, provincia, establecimiento, indicadores, tiempo, zona

Integrar los datos

Al construir un modelo de datos tipo estrella, las integraciones que se hicieron son las siguientes:

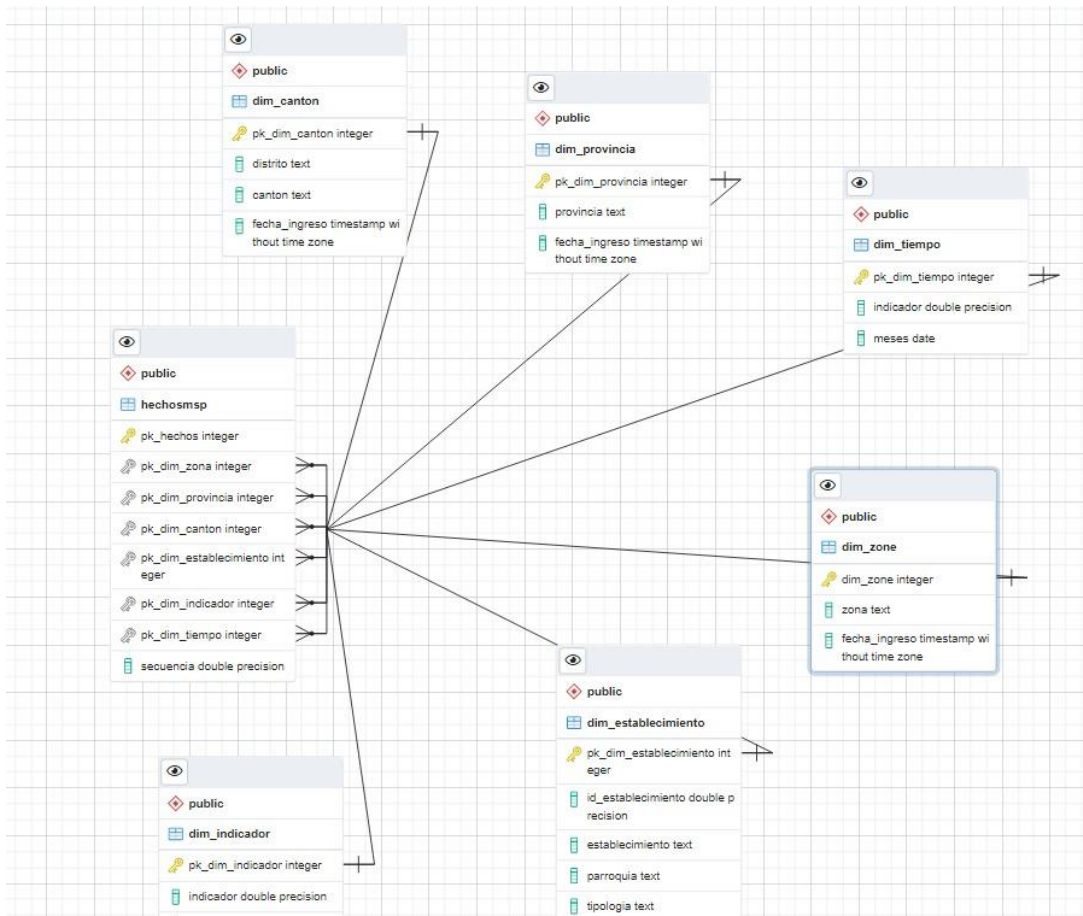


Figura 11: Diagrama Tipo estrella

Fuente: Lidia Enríquez, Clinton Moina

- Se relacionó la tabla provincia que contiene la información de las 4 provincias que pertenecen a la coordinación zonal 3 con la tabla de hechos que contiene todos los id principal de todas las tablas.
- Se relacionó la tabla indicadores que contiene la información del tiempo en trimestres, cuatrimestres, anuales el grupo, numerador e indicador de las historias clínicas registradas en la matriz con la tabla de hechos que contiene todos los id principal de todas las tablas.
- Se relacionó la tabla tiempo que contiene la información de la fecha cuando se ingresan los datos al sistema multidimensional con la tabla de hechos que contiene todos los id principal de todas las tablas.
- Se relacionó la tabla establecimiento que contiene la información de los 278 establecimientos que pertenecen a la coordinación zonal 3 con la tabla de hechos que contiene todos los id principal de todas las tablas.
- Se relacionó la tabla cantón que contiene la información de los 26 cantones que pertenecen a la coordinación zonal 3 con la tabla de hechos que contiene todos los id principal de todas las tablas.

- La tabla hechos tiene relación con las tablas cantón, provincia, establecimiento, zona, indicador, tiempo.

Formateo de los datos.

No ha sido necesario realizar el formateo de los datos ya que desde el origen tuvieron la estructura que se necesitó para crear el sistema multidimensional.

3.6.4 FASE IV. Modelado de los datos

Selección de la técnica de modelado

Para el presente proyecto, no es necesario implementar técnicas de modelado de datos, ya que mediante la metodología Crisp-dm se logró obtener datos de calidad, se aplicó procesos de integración, transformación y limpieza de datos (ETL) haciendo uso de la herramienta Pentaho, se creó el Sistema Multidimensional que incluyó las tablas: zona, provincia, cantón, indicadores, establecimientos y tiempo. Resultado de esto se obtuvo un esquema de datos tipo estrella con las siguientes tablas: zona, provincia, cantón, indicadores, tiempo, establecimiento y hechos, las cuales ya contienen datos de calidad. Una vez aplicado el proceso se obtuvo información que proporcionó conocimiento mediante el desarrollo del Cuadro de Mando Integral.

Generar plan de prueba

El plan de prueba se obtendrá del 100% de los datos que, una vez realizado el proceso de Extracción, transformación y carga en Pentaho, se medirá la credibilidad, completitud y precisión de los datos en la herramienta Talend Data Quality la cual nos arroja los resultados automáticamente.

Construir el modelo

Se ejecuto las 7 dimensiones que conforman el sistema multidimensional sobre la herramienta Talend Data Quality, ver anexo 6.

Se obtiene la calidad de los datos por tabla: cantón, provincia, establecimiento, zona, indicador, tiempo, hechos, con los cuales se construyó las visualizaciones de los reportes en el Cuadro de Mando Integral.

Evaluar el modelo

Se presenta los resultados obtenidos de credibilidad, completitud y precisión obtenidos de la

777

herramienta Talend Data Quality, los mismo que presentamos en los Resultados y Discusión en el capítulo IV.

3.6.5 FASE V. Evaluación

Evaluar el resultado

Una vez realizado el proceso de evaluación de la calidad de los datos de acuerdo con las características propuestas por la Norma ISO/IEC 25012: completitud, credibilidad y precisión, los cuales se detallan en el capítulo IV de Resultados y Discusión se procede a crear el cuadro de mando integral como se indica en la fase VI. Despliegue.

Revisar el proceso

El proceso para llegar hasta este punto fue el siguiente:

Figura 12: Proceso para el diseño del Sistema Multidimensional y Cuadro de Mando Integral.

Fuente: Lidia Enríquez, Clinton Moina

1. Se inicio con una matriz de Excel proporcionada por la coordinadora de la coordinación Zonal 3 “Matriz de reporte distrital del monitoreo de la calidad de la atención Materno Neonatal”, se realizó el proceso de estructura la información en filas y columnas a manera de tabla.
2. Se cargo la información en Pentaho (Spoon), se aplicó las reglas de ETL respectivas: limpieza de datos, eliminar campos nulos, eliminar datos inválidos.
3. La data se envió a la base de datos PostgreSQL con los respectivos controles que se aplicó en el paso 2.
4. Se uso la herramienta Talend para verificar la calidad de los datos: completitud, credibilidad y precisión.
5. Finalmente, con la información de calidad se procedió a crear los cuadros de mando usando R Studio y la librería Shiny para la visualización de la información.

Todo el proceso ejecutado hasta este apartado no ha presentado ningún inconveniente, obteniendo los resultados óptimos para el presente trabajo de investigación.

3.6.6 FASE VI. Despliegue

En la última fase de la metodología Crisp-dm, se consolida las 5 fases anteriores y se transforma la información en conocimiento, se completa el proceso que se muestra en la figura 8: la programación y visualización del cuadro de mando integral.

Se utilizó R Studio y la librería Shiny como servidor para el despliegue y visualización de los gráficos en el cuadro de mando integral.

En la figura 13 se muestran las librerías instaladas que se usaron en el desarrollo.

```

#Librerías
library(shiny)
library(shinydashboard)
library(DT)
library(shinyjs)
library(sodium)
library(tidyverse)
library(DBI)
library(RPostgreSQL)
library(DT)
library(pool)
library(quantmod)
library(reactable)
library(dplyr)
library(devtools)
library(ggplot2)
library(ggthemes)
library(lubridate)
library(rsconnect)
library(tidyverse)
library(urltools)
library(rsconnect)

```

Figura 13: Librería Instaladas en R Studio para el desarrollo del Cuadro de Mando Integral.

Fuente: Lidia Enríquez, Clinton Moina

Luego se realizó la conexión a la base de datos, como se muestra en la figura 14.

```

dvr <- RPostgres::Postgres()
db <- 'DATAMART'
host_db <- 'localhost'
db_port <- '5433'
db_user <- 'postgres'
db_password <- '1234'

con <- dbPool(dvr,
  dbname = db,
  host=host_db,
  port=db_port,
  user=db_user,
  password=db_password)
#permite abrir y cerrar conexiones sql
onStop(function(){
  poolClose(con)
})

```

Figura 14: Conexión a la base de datos

Fuente: Lidia Enríquez, Clinton Moina

Una vez realizada la conexión a la base de datos, se realizaron las consultas a las tablas: provincia, cantón, indicador, establecimiento, tiempo, zona, como se observa en la figura 15.

```

dim_provincia<- tbl(con,"dim_provincia")
dim_canton<- tbl(con,"dim_canton")
dim_indicador<-tbl(con,"dim_indicador")
dim_establecimiento<- tbl(con,"dim_establecimiento")
dim_tiempo<- tbl(con,"dim_tiempo")
porcentaje <-dim_indicador%>% select(porcentaje) %>%pull()
meses<-dim_indicador%>%select(fecha_ingreso)%>%pull()
dim_tiempo<-dim_tiempo%>%select(meses)%>%pull()

```

Figura 15: Consultas a las tablas de la base de datos

Fuente: Lidia Enríquez, Clinton Moina

Siguiendo la estructura para la programación de aplicaciones en R Studio con Shiny, en el archivo Ui.R se realizó el login, como se observa en la figura 16.

```

source("global.R") #Utilizamos el archivo global.r para la utilización de las librerías.

# Main login screen
loginpage <- div(id = "loginpage", style = "width: 500px; max-width: 100%; margin: 0 auto;
padding: 20px;",
  wellPanel(
    tags$h2("LOGIN MSP", class = "text-center", style = "padding-top: 0;color:#333; font-
weight:600;"),
    textInput("userName", placeholder="Username", label = tagList(icon("user"),
"Username")),
    passwordInput("passwd", placeholder="Password", label = tagList(icon("unlock-alt"),
"Password")),
    br(),
    div(
      style = "text-align: center;",
      actionButton("login", "SIGN IN", style = "color: white; background-color:#3c8dbc;
padding: 10px 15px; width: 150px; cursor: pointer;
font-size: 18px; font-weight: 600;"),
      shinyjs::hidden(
        div(id = "nomatch",
          tags$p("Oops! Usuario o contrasenia incorrecta!",
            style = "color: red; font-weight: 600;
padding-top: 5px;font-size:16px;",
            class = "text-center"))),

```

Figura 16: Login en R Studio

Fuente: Lidia Enríquez, Clinton Moina

En la gráfica 17 se muestra la configuración del archivo App.R donde se realizó la programación de los gráficos y se aplicaron los filtros respectivos por cada tabla.

```

tabitems(
  tabitem(tabName="Rzonal",
    fluidRow(style = "font-size: 75%;cursor: pointer;margin:0 ;
      text-align: center;",
      box(
        solidHeader = T,
        collapsible = T,
        width=4,
        height=130,
        selectInput(
          "indicadorz",
          "INDICADOR",
          rprovincia %>% select(nombre)%>% pull(),
          selected = "All",
        ),
      ),
      box(
        solidHeader = T,
        collapsible = T,
        width=3,
        height=130,
        selectInput(
          "trimestrez",
          "TRIMESTRE",

```

Figura 17: Configuración del archivo App.R

Fuente: Lidia Enríquez Clinton Moina

En la figura 15 se muestra el código para obtener las gráficas con los filtros respectivos.

```

output$grafica<-renderPlot({
  ggplot(data=df()),mapping=aes( x=meses, y=porcentaje, color=grupom, labels =porcentaje ))+
  geom_point(size=4)+
  scale_color_manual(breaks=c('1.0','2.0','3.0'),
    values = c("#FC0C05", "#FBFE3E", "#05FC1F"))+
  #scale_color_discrete("grupom",labels = c("1.0", "2.0", "3.0"))+
  labs(#title="Reporte mensual",
    caption="Reporte: de indicadores en escala del 0 a 100%",
    x="Meses del año",
    y="Porcentaje por mes",
    color="Grupo" ) +
  geom_line(colour='black')+theme_minimal()+

  theme(axis.text.x=element_text(angle=45))+
  geom_text(aes(label = sprintf("%.2f%%",porcentaje))),

    position = position_dodge(0.9),
    vjust = -0.2,colour='black')
})

```

Figura 18: Código de gráficas y filtros aplicados.

Fuente: Lidia Enriquez, Clinton Moina

Finalmente, terminado el desarrollo del cuadro integral se procede a desplegarlo haciendo uso de la librería Shiny como servidor, en la figura 19 se visualiza el cuadro de mando integral construido, en el panel izquierdo se muestra el menú de reportes por zona, provincia,

distrito, tipología y establecimientos, en la parte superior filtro por indicador, trimestres y año al que pertenece la información y las gráficas de reportes trimestral y anual. El mismo procedimiento de visualización se realiza para todos los reportes.

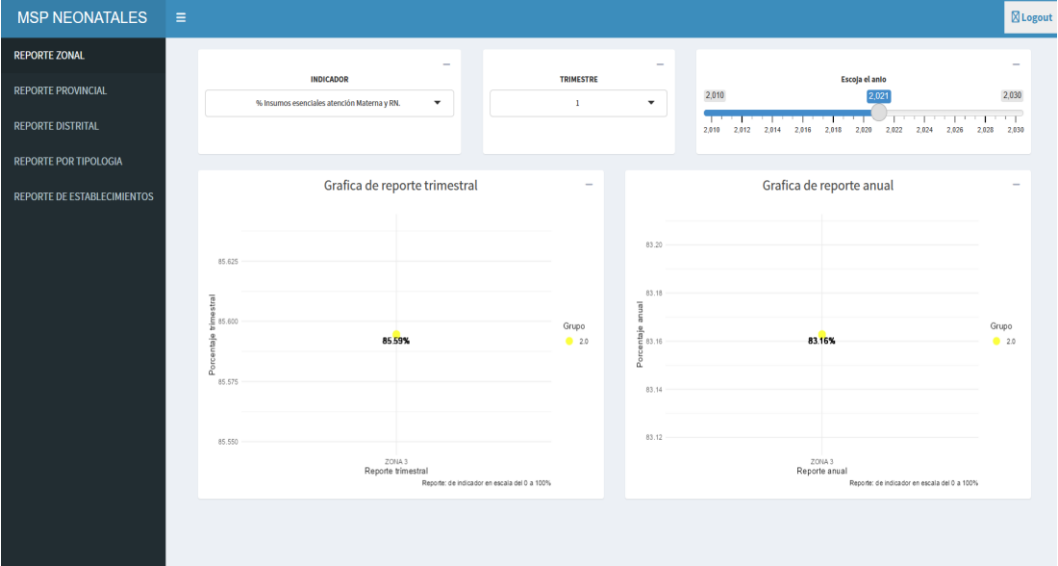


Figura 19: Cuadro de mando integral a nivel de zona.

Fuente: Lidia Enríquez, Clinton Moina

CAPÍTULO IV. RESULTADOS Y DISCUSIÓN

4.1 Resultados

Una vez diseñado el Sistema Multidimensional para obtener información de calidad, los datos obtenidos fueron procesados en la herramienta Talend Data Quality donde se obtuvo los criterios de calidad: completitud, credibilidad y precisión. Para obtener los resultados se tomó los datos de las tablas: establecimiento, cantón, hechosmsp, provincia, indicador, tiempo y zona, como se detalla a continuación:

Establecimiento: En la data original provista por la coordinación 3 del MSP; la tabla establecimiento contenía 1667 registros correspondientes al año 2021 y 2022, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyó a 129 registros, se utilizó los campos: id_establecimiento, establecimiento, parroquia y tipología.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 17: Credibilidad, Completitud y Precisión de los datos de Establecimientos de la Coordinación 3.

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%
Precisión	100%		0%

Fuente: Lidia Enríquez, Clinton Moina

Hecho: En la data original esta tabla no existía, para crear el sistema multidimensional tipo estrella se requiere crearla y consta de 4596 registros, la cual contiene los siguientes campos: id_hecho, indicador, nombre, universo, numerador, denominador, porcentaje, Fecha_ingreso, Grupom, trimestre, Numeradort, denominadort, porcentaje, Grupot.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 18: Credibilidad, Completitud y Precisión de los datos de Hechos

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%
Precisión	100%		0%

Fuente: Lidia Enríquez, Clinton Moina

Indicador: En la data original provista por la coordinación 3 del MSP; la tabla Indicador contenía 56174 registros correspondientes al año 2021 y 2022, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyó a 4596 registros, se utilizó los campos: Id, Indicador, Nombre, Universo, Numerador, Denominador, Porcentaje, Fecha_ingreso, Grupom, Trimestre, Numeradort, Denominadort, Procentajet, Grupot, Cuatrimestre, Numeradorc, Denominadorc, Procentajec, Grupoc, Semestre, Numeradors, Denominadors, Porcentajes, Grupos, Anual, Numeradora, Denominadora, Porcentajea, Grupoa.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 19: Credibilidad, Completitud y Precisión de los datos de Indicador

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%
Precisión	100%		0%

Fuente: Lidia Enríquez, Clinton Moina

Cantón: En la data original provista por la coordinación 3 del MSP; la tabla Cantón contenía 1667 registros correspondientes al año 2021 y 2022, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyó a 129 registros, se utilizó los campos: id, cantón.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 20: Credibilidad, Completitud y Precisión de los datos de Cantón de la Coordinación Zonal 3

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%
Precisión	100%		0%

Fuente: Lidia Enríquez, Clinton Moina

Provincia: En la data original provista por la coordinación 3 del MSP; la tabla Provincia contenía 1667 registros correspondientes al año 2021 y 2022, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyó a 129 registros, se utilizó los campos: Id, provincia, fecha de ingreso

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 21: Credibilidad, Completitud y Precisión de los datos de Provincia de la Coordinación Zonal 3

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%
Precisión	100%		0%

Fuente: Lidia Enríquez, Clinton Moina

Zona: En la data original provista por la coordinación 3 del MSP; la tabla Zona contenía 141 registros correspondientes al año 2021 y 2022, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyo a 129 registros, se utilizó los campos: id_zona, zona, Fecha_ingreso.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 22: Credibilidad, Completitud y Precisión de los datos de Zona de la Coordinación Zonal 3

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%
Precisión	100%		0%

Fuente: Lidia Enríquez, Clinton Moina

Tiempo: En la data original provista por la coordinación 3 del MSP; la tabla Zona contenía 56174 registros correspondientes al año 2021 y 2022, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyo a 4596 registros, se utilizó los campos: id_tiempo, indicador, meses.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión.

Tabla 23: Credibilidad, Completitud y Precisión de los datos de Tiempo

Características	Calidad de los datos	X/100%	Error
Credibilidad	100%		0%
Completitud	100%		0%

Precisión	100%	0%
-----------	------	----

Fuente: Lidia Enríquez, Clinton Moina

Una vez obtenidos los resultados de calidad de los datos: completitud credibilidad y precisión de los datos según los establece la norma ISO/IEC 25012, los cuales son aptos para generar conocimiento para la coordinación zonal 3, a continuación, se presenta los resultados del cuadro de mando integral.

El análisis de calidad fue realizado tomando en cuenta la tabla 6: Colorimetría del Nivel de Calidad de los Indicadores 1 al 9C.

En la figura 20 se observa, el panel izquierdo muestra reportes por zona, provincia, distrito, tipología y establecimientos, en la parte superior derecha los filtros por indicador, trimestre y también el año en el que se desea obtener la información.

En la gráfica de reporte trimestral por zona se interpreta que el indicador 1 (% Insumos esenciales atención Materna y RN) en el primer trimestre del año 2022, el nivel de calidad está en alerta con un porcentaje de 85,59% representado de color amarillo, como se establece en la tabla 6.

En la gráfica de reporte anual por zona se interpreta que el indicador 1: % Insumos esenciales atención Materna y RN en el año 2022, el nivel de calidad está en alerta con un porcentaje de 83,16% representado por color amarillo.

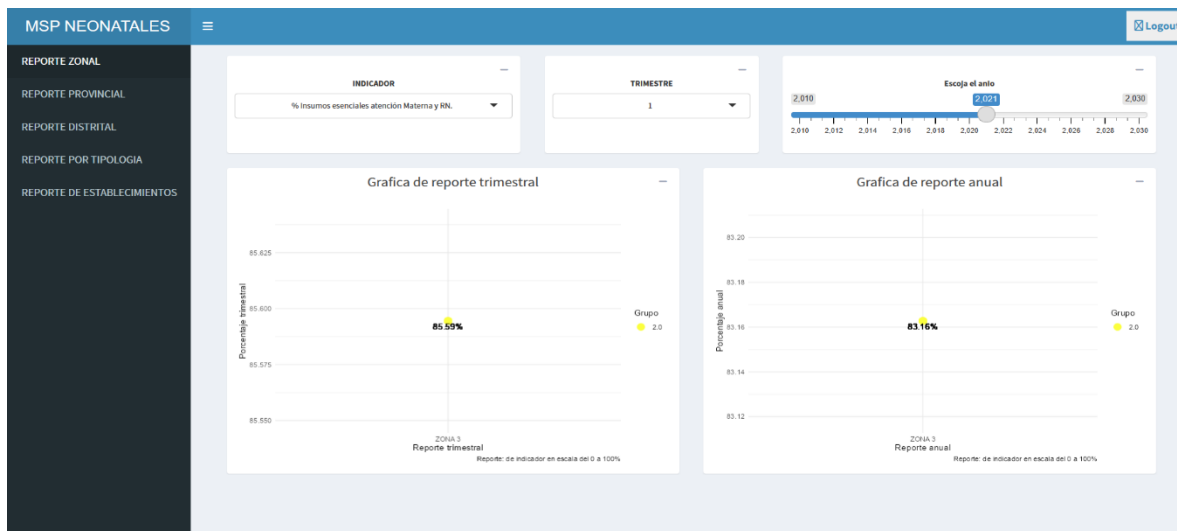


Figura 20: Página principal, reporte trimestral y anual

La figura 21, en la gráfica de reporte trimestral por provincia, se interpreta que el indicador 1: % Insumos esenciales atención Materna y RN en el año 2022, en la provincia Chimborazo el nivel de calidad es de alerta ya que se encuentra en 85,81% y se representa de color amarillo, en la provincia Cotopaxi el nivel de calidad es aceptable ya que se encuentra en 94,05% y se representa de color verde, en la provincia Pastaza el nivel de calidad es de alerta ya que se encuentra en 78,05% y se representa de color amarillo y en la provincia Tungurahua el nivel de calidad es de alerta ya que se encuentra en 87,85% y se representa de color amarillo.

En la gráfica de reporte anual por provincia: se interpreta que el indicador 1: % Insumos esenciales atención Materna y RN en el año 2022, en la provincia Chimborazo el nivel de calidad es de alerta ya que se encuentra en 85,86% y se representa de color amarillo, en la provincia Cotopaxi el nivel de calidad es de alerta ya que se encuentra en 90,06% y se representa de color amarillo, en la provincia Pastaza el nivel de calidad es de alerta ya que se encuentra en 78,05% y se representa de color amarillo y en la provincia Tungurahua el nivel de calidad es de alerta ya que se encuentra en 85,49% y se representa de color amarillo.

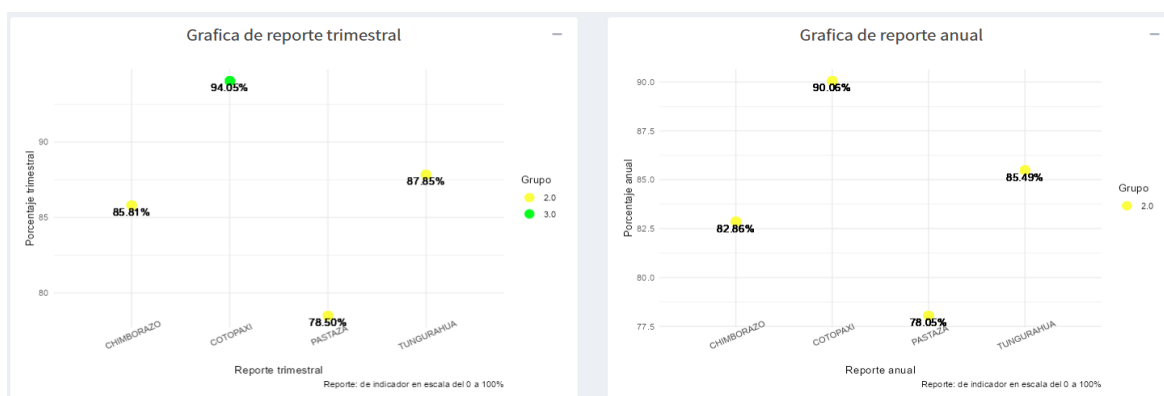


Figura 21: Reporte por provincia trimestral y anual

En la figura 22, en la gráfico de reporte trimestral por Distrito, se interpreta que el indicador 1: % Insumos esenciales atención Materna y RN en el año 2022, en el distrito 05D01 el nivel de calidad es aceptable ya que se encuentra en 92,12% y se representa de color verde, en el distrito 05D03 el nivel de calidad es aceptable ya que se encuentra en 94,39% y se representa de color verde, en el distrito 05D04 el nivel de calidad es de alerta ya que se encuentra en 80,49% y se representa de color amarillo, en el distrito 05D06 el nivel de calidad es aceptable ya que se encuentra en 97,71% y se representa de color verde, en el distrito 06D01 el nivel

de calidad es de alerta ya que se encuentra en 90,64% y se representa de color amarillo, en el distrito 06D02 el nivel de calidad es de alerta ya que se encuentra en 85,69% y se representa de color amarillo, en el distrito 06D03 el nivel de calidad de alerta ya que se encuentra en 81,47% y se representa de color amarillo, en el distrito 06D01 el nivel de calidad es de alerta ya que se encuentra en 78,50% y se representa de color verde, en el distrito 18D01 el nivel de calidad es de alerta ya que se encuentra en 86,82% y se representa de color amarillo, en el distrito 18D02 el nivel de calidad es de alerta ya que se encuentra en 89,57,12% y se representa de color amarillo, de igual forma se realiza la misma interpretación para la gráfica de reporte anual.

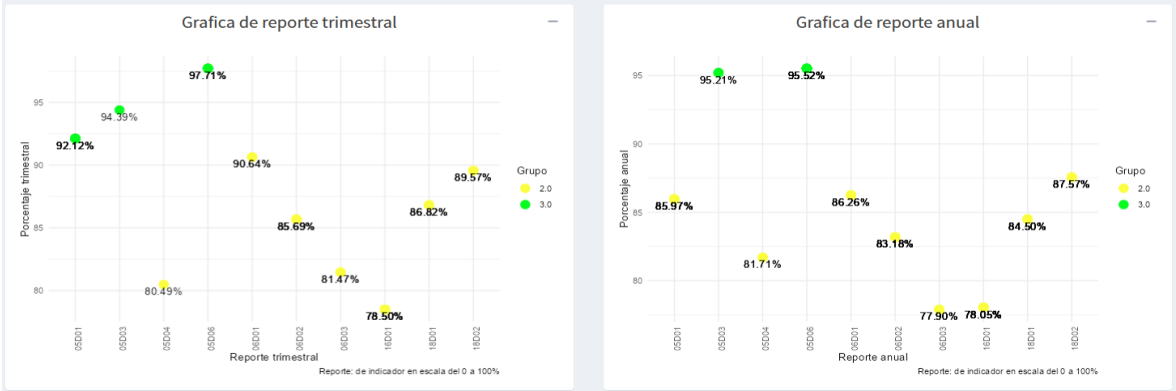


Figura 22: Reporte por distrito trimestral y anual

En la figura 23, en la gráfica de reporte trimestral por tipología: se interpreta que el indicador 1: % Insumos esenciales atención Materna y RN en el año 2022, en los centros de salud tipo A el nivel de calidad es de alerta ya que se encuentra en 85,51% y se representa de color amarillo, en los centros de salud tipo B el nivel de calidad es de alerta ya que se encuentra en 88,72% y se representa de color amarillo, en los centros de salud tipo C el nivel de calidad es de alerta ya que se encuentra en 90,01,% y se representa de color amarillo, en los Hospitales Básicos el nivel de calidad es aceptable ya que se encuentra en 96,50% y se representa de color verde, en los Hospital General el nivel de calidad es de alerta ya que se encuentra en 81,78% y se representa de color amarillo, en los Puestos de Salud el nivel de calidad es de alerta ya que se encuentra en 78,05% y se representa de color amarillo. Así como también se realiza la misma interpretación para la gráfica de reporte anual por tipología.

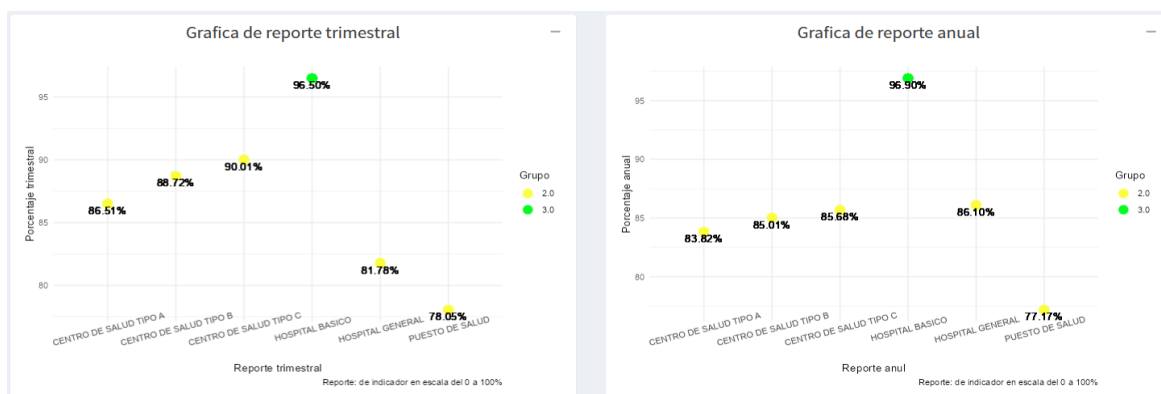


Figura 23: Reporte por tipología trimestral y anual

En la figura 24, en la gráfica de reporte mensual por establecimiento, se interpreta que el indicador 1: % Insumos esenciales atención Materna y RN en el año 2022, en la provincia de Cotopaxi, distrito 06D03, establecimiento el Corazón de la parroquia el Corazón el Hospital Básico se encuentra en con más de 90,99% en todos los meses obteniendo un nivel de calidad aceptable y se representa de color verde. Se puede observar que en la gráfica de reporte por trimestre, cuatrimestre, semestre y anual también cumple con más del 90,99% por lo tanto tienen un nivel de calidad aceptable y se representa de color verde.

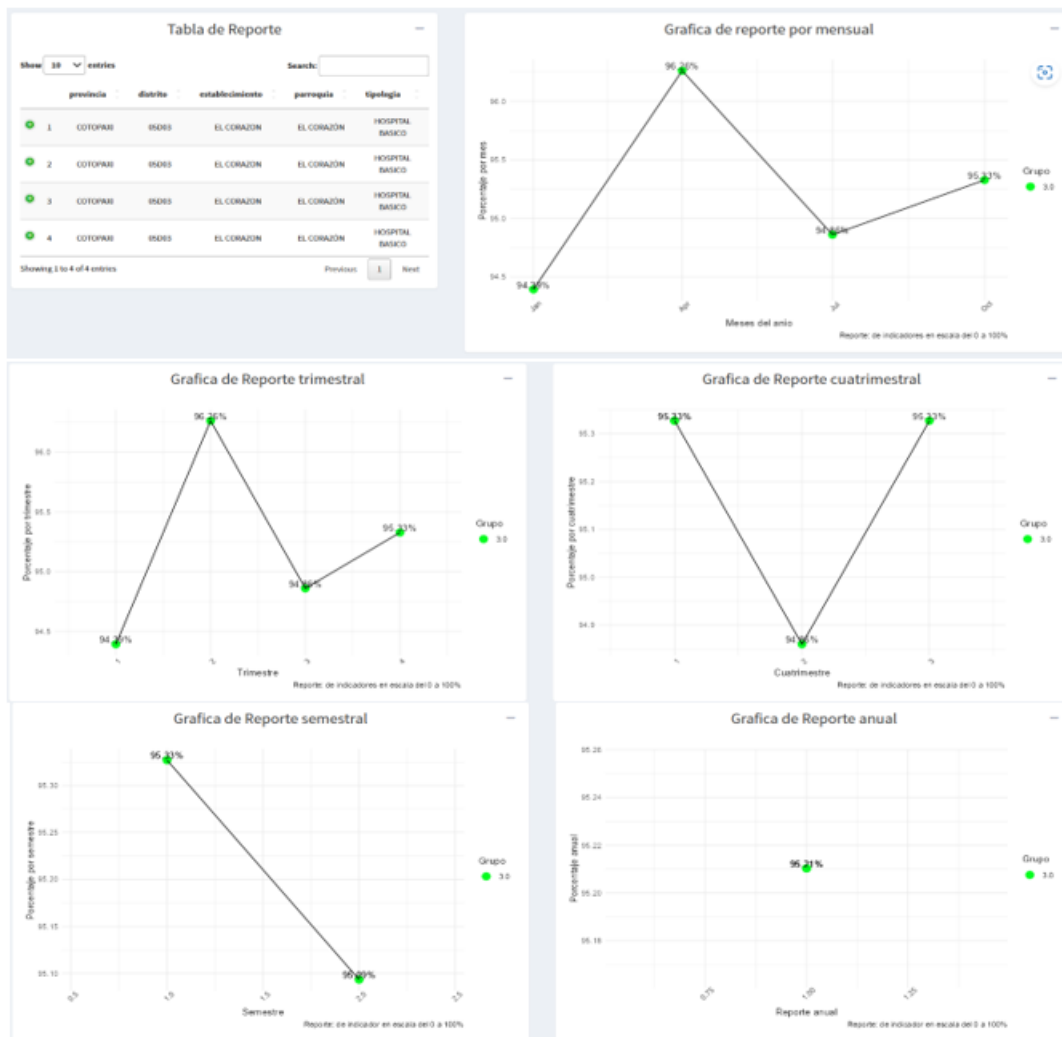


Figura 24: Reporte por establecimiento mensual, trimestral, cuatrimestral, semestral y anual

4.2 Discusión

La herramienta Talend Data Quality fue utilizada para evaluar los criterios de completitud, credibilidad y precisión de los datos que se obtuvo al diseñar el sistema multidimensional y cuadro de mando integral para el área de calidad materno neonatal de la Coordinación Zonal 3 del Ministerio de Salud Pública.

Los resultados obtenidos fueron comparados con las métricas que propone la norma ISO/IEC 25012 (ver Capítulo II, apartado 2.8); la completitud de los datos fue del 100% comparados con relación al 85% que establece la norma ISO/IEC 25012, la credibilidad de los datos fue del 100% comparados con relación al 75% que establece la norma ISO/IEC 25012, en cuanto a la precisión de los datos fue 100% comparado con el 75% que establece la norma ISO/IEC 25012.

Para obtener datos de calidad se debe considerar las siguientes buenas prácticas: Estructuras la información, selección de datos representativos, limpieza de datos, validación de datos, filtrado de datos.

Los valores de los criterios de credibilidad, completitud y precisión de conformidad que establece la norma ISO/IEC 25012 cumplen con los parámetros de calidad de la información. Para el desarrollo del proyecto se utilizó la metodología Crisp - dm que según (KDNuggets, 2014), en la actualidad es la guía más utilizada en el desarrollo de proyectos de análisis de datos y minería de datos, según encuestas aplicadas en los últimos años el porcentaje de uso es de 43% comparado con 6 metodologías más.

El cuadro de mando integral presenta reportes por zona, provincia, distrito, tipología y establecimientos en los cuales según el “Instructivo para el uso de colores de la base de datos Excel: Monitoreo y reporte de indicadores consolidado de los establecimientos de salud del distrito” establecido por el Ministerio de Salud Pública se presenta los porcentajes que hacen referencia a los niveles de calidad, dentro del cual el color rojo hace referencia que el nivel de calidad no es aceptable y considera un rango de 0 a 70,99%, el color amarillo el nivel de calidad está en alerta con un rango de 71 a 90,99% y el color verde el nivel de calidad es aceptable con un rango de 91 a 100%.

CAPÍTULO V. CONCLUSIONES y RECOMENDACIONES

5.1 Conclusiones

- La metodología Crisp-Dm fue aplicada en la construcción de Sistemas multidimensionales y cuadro de mando integral, mediante Crisp-dm se pudo realizar el trabajo siguiendo las 6 fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación, Despliegue; donde las dos fases más críticas son comprensión del negocio y Preparación de los datos, la comprensión del negocio consiste en alinear los objetivos del negocio con los del proyecto y en la fase de preparación de los datos se realiza limpieza de datos, creación de indicadores, transformación de los datos, esta tarea se realizó usando la herramienta Pentaho.
- Se diseñó el sistema multidimensional aplicando ETL a los datos de las tablas: tiempo, zona, provincia, cantón, establecimientos e indicadores; a continuación, se creó el cuadro de mando integral, en el cual se pudo visualizar los resultados para el cual fue creado el proyecto: reporte de zona, provincia, distrito, tipología y establecimiento, se transformó esa información en conocimiento que sirve de apoyo a la toma de decisiones.
- Para determinar la calidad de los datos se usó la herramienta Talend Data Quality considerando los criterios de completitud, credibilidad y precisión, obteniendo 100% de los datos completos, 100% de los datos creíbles y 100% de datos precisos, lo cual determina que los datos son de calidad

5.2 Recomendaciones

- Al seleccionar los datos para un sistema multidimensional, se debe elegir aquellos que presenten menos irregularidades (nulos, blancos, incompletos), para cumplir con los criterios de calidad según la norma ISO/IEC 25012: completitud, credibilidad y precisión.
- El proceso de ETL es importante para el análisis de los datos, se recomienda tomar en cuenta el tipo de entidad donde se aplicará el proyecto, al tratarse de una entidad pública se debe elegir herramientas libres, gratuitas, que tengan conectividad con diferentes fuentes de información y que se adapte a las necesidades actuales y futuras del negocio, por ejemplo: Pentaho para realizar procesos de ETL, R Studio para realizar la programación del cuadro de mando y Shiny como servidor para visualizar el cuadro de mando integral, y para la verificación de calidad de la data con Talend Data Quality.
- Establecer de forma clara los objetivos organizacionales que se desean obtener cuando se desarrolle el cuadro de mando integral, para obtener gráficos que representen conocimiento para la organización donde se esté aplicando.

BIBLIOGRAFÍA

- Abelló, A. (2019). *Introducción al Data Warehouse*. Catalunya: Universidad Oberta de Catalunya.
- Alonso, G., & Rico, D. (2010). Inteligencia de negocios. *Grupo de Investigación en Teleinformática y Desarrollo de Software (GITYD)*, 6.
- Bargui, F., Ben-Abdallah, H., & Feki, J. (2016). A natural language-based approach for a semi-automatic data mart design and ETL generation. *Journal of Decision Systems*.
- Bouman, J. R. (2009). *Pentaho Solutions*. Canada: Wiley Publishing, Inc.
- Bustamante, A., Galvis, E., & Gómez, L. (2013). *Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI*. Pereira: Scientia Et Technica, vol. 18.
- Calabrese, J., Esponda, S., Pisini, A., & Boracchia, M. (2019). *Guía para evaluar calidad de los datos basada en ISO/IEC 25012*. Río Cuarto: Instituto de Investigación en Informática LIDI (III-LIDI).
- Cedeño, A. (2006). *Modelo Multidimensional*. Cuba: Sistema de Información Científica.
- Chauré, P. (2022). *APLICACIÓN WEB SHINY PARA EL ANÁLISIS Y EXPERIMENTACIÓN DE MODELOS DE PREDICCIÓN DE SERIES TEMPORALES*. Madrid: Facultad de Ciencias Económicas y Empresariales.
- Espinoza, J. (2020). Aplicación de la metodología CRISP-DM para segmentación geográfica de una base de datos pública. *SciELO Analytics*, 14.
- Fernández, C., & Baptista, P. (2014). *Metodología de la investigación 6ta Edición*. Mexico: Miembro de la Cámara Nacional de la Industria Editorial Mexicana.
- Galán, V. (2015). *Aplicación de la metodología Crisp-dm a un proyecto de minería de datos*. Madrid. Obtenido de https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf
- Ghiglione, F. (2021). EL cuadro de mando integral como herramienta de eficiencia en la gestión empresarial. *Consejo Nacional de Investigaciones Científicas y Técnicas Facultad de Agronomía*, 10.
- Gómez, D. S., Molina, M. D., Mulero, J., Nueda, M. J., & Pascual, A.-}. (2016). Aplicaciones diseñadas con Shiny: un recurso docente para. *Departamento de Matemáticas*.
- González, M., & González, J. (2013). *Aplicación del estándar ISO/IEC 9126-3 en el modelo de datos conceptual entidad-relación*. Obtenido de

- http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-11292013000200010&lng=en&tlng=es
- Haya, P. (2021). *Instituto de Ingeniería del Comercio*. Obtenido de <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>
- Inmon, W. H. (2002). *Building the Data Warehouse*. New York: Third Edition. Obtenido de <https://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf>
- Inmon, W., Strauss, D., & Neushloos, G. (2008). *The Architecture for the Next Generation of Data Warehousing*. London: Morgan Kaufmann Publishers.
- ISO/IEC. (2008). *ISO/IEC 25012:2008*. Obtenido de <https://www.iso.org/standard/41049.html>
- Jaramillo, M. (2019). *COLOQUIO DE ESTADISTICA*. Medellín.
- José, C., Molina, E., & Arteaga, P. (2014). *Introducción a la programación estadística con R para profesores*. SEJ2007-60110/EDUC (MEC-FEDER) y EDU2010-.
- KDNuggets. (2014). *KDNuggets*. Obtenido de www.kdnuggets.com
- Kimball. (2005). *Is ER Modeling Hazardous to DSS*. Obtenido de <http://www.dbmsmag.com/9510d05.html>
- Lanang, P. (2020). *Data Warehouse Design for the Bank X with Inmon Approach*. Udayana: International Journal of Engineering and Emerging Technology.
- Mali, N., & Sachin, B. (2015). A Survey of ETL Tools. *International Journal of Computer Techniques*.
- Moine, J., Gordillo, S., & Haedo, S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. *Grupo de Investigación en Minería de Datos*, 8.
- Morales, O. A. (2013). *FUNDAMENTOS DE LA INVESTIGACIÓN DOCUMENTAL*. Merida: Grupo de Estudios Odontológicos, Discursivos y Educativos.
- Muller, O. (2014). Systems Information Systems (Business Informatics) Business Intelligence. *Business & Information Systems Engineering*, 30.
- Ramirez, A. (21 de 08 de 2017). *Vision*. Obtenido de <https://synergy.vision/corpus/shiny/2017-08-15-shiny.html>
- Ranjan, J. (2009). Business intelligence: Concepts, components, techniques and benefits. *Journal of theoretical and applied information technology*, 12. Obtenido de <https://d1wqtxts1xzle7.cloudfront.net/33997956/9Vol19No1-with-cover-page->

v2.pdf?Expires=1646801823&Signature=GMO~aKVdhgnKpHGPOJrttJzCnhY7dg
xsoWw3L~ck5nGSCJNLHdatPIRojmqXmB0fKttggwzuyXRv~rzPPMjG-
q0xJe3S6TueOsa5TnZWacEXwga-25-l-
F6Xf3bCR7ePNDQggR5E1AlgNU4oLwU

RStudio. (2022). *R Sutdio*. Obtenido de <https://cran.r-project.org>

RStudio. (2022). *rstudio.com*. Obtenido de <https://www.rstudio.com/products/rpackages/>

RStudio. (2022). *The Comprehensive R Archive Network*. Obtenido de <https://cran.r-project.org>

Software, E. (2021). *Evaluando Software*. Obtenido de <https://www.evaluandosoftware.com/abc-del-data-warehouse/#:~:text=Un%20data%20warehouse%20o%20dep%C3%B3sito,de%20toma%20de%20decisiones%20gerenciales>.

Talend. (2023). *Talend*. Obtenido de <https://www.talend.com>

Vernazi, J. (2011). *Getting Started with RStudio*. United States of America: O'REILLY.

Zuñiga, J., & Galán, V. (Agosto de 2020). *Scielo.org.mx*. Obtenido de http://www.scielo.org.mx/scielo.php?pid=S1405-77432020000100008&script=sci_arttext

ANEXOS

Anexo 1. Origen de datos

MATRIZ DE REPORTE DISTRITAL DEL MONITOREO DE LA CALIDAD DE LA ATENCIÓN MATERNO NEONATAL

MINISTERIO DE SALUD PÚBLICA DEL ECUADOR
DIRECCIÓN NACIONAL DE CALIDAD DE LOS SERVICIOS DE SALUD

DISTRICTO: 06D04 Establecimiento de Salud: [Nombre]

Número de Embarazadas Asignadas al Establecimiento que un 15% de los embarazos esperados pueden llegar a tener complicaciones: 26
Complicaciones obstétricas esperadas mensual: 2,19

Proceso Estándar de entrada para la AM y RN	Estándar	Indicador	Marzo		Abril		Mayo		Junio		Julio		Agosto		Septiembre		Octubre		
			N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	
1	1	% Insumos esenciales atención Materna y RN	30	97	29	88	29	97	29	97	30	100							
2	2	Control prenatal atendido según norma	30	97	29	88	29	97	29	97	30	100							
3A	3A	% Partos con partograma graficado y con control																	
3B	3B	% Partos con decisiones frente a desviaciones de curva de partograma																	
4	4	% Partos con Datoconia (manejo activo del tercer periodo)																	
5	5	% de post partos con control según la norma																	
6	6	% de RN en los que se realizó actividades de la norma																	
7A	7A	% partos atendidos por médica u obstetra																	
7B	7B	% partos atendidos por médica u obstetra																	

Figura 25: Data Original: Matriz de reporte distrital del monitoreo de la calidad de la Atención Materno Neonatal

Anexo 2. Estructuración de los datos

D2 =FECHA[DERECHA[UNICARDENAS], REPORTE ESTABLECIMIENTO SALUD\ISA6:SH56],4),1,1)

tipología	id_establecimiento	indicador	id_mes	univero	numador	denominador	porcentaje	grupom	nombre	distrito	establecimiento	canton	parroquia	provincia
CENTRO DE SALUD TIPO A	316	1	2021-01-01	0	31.00	41.00	75.61	2	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-02-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-03-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-04-01	0	39.00	41.00	95.12	3	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-05-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-06-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-07-01	0	39.00	41.00	95.12	3	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-08-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-09-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-10-01	0	36.00	41.00	87.80	2	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-11-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	1	2021-12-01	0	0.00	0.00	0.00	1	% Insumos esenciales atención Materna y RN	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-01-01	31	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-02-01	26	26.00	26.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-03-01	30	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-04-01	26	26.00	26.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-05-01	26	26.00	26.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-06-01	30	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-07-01	27	27.00	27.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-08-01	42	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-09-01	48	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-10-01	56	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-11-01	58	30.00	30.00	100.00	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	2	2021-12-01	54	29.00	30.00	96.67	3	Control prenatal atendido según norma	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-01-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-02-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-03-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-04-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-05-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-06-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-07-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-08-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-09-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-10-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-11-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	3	2021-12-01	0	0.00	0.00	#DIV/0!	#DIV/0!	% Partos con partograma graficado y con control	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	4	2021-01-01	0	0.00	0.00	#DIV/0!	#DIV/0!	ton decisiones frente a desviaciones de curva de p	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	4	2021-02-01	0	0.00	0.00	#DIV/0!	#DIV/0!	ton decisiones frente a desviaciones de curva de p	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	4	2021-03-01	0	0.00	0.00	#DIV/0!	#DIV/0!	ton decisiones frente a desviaciones de curva de p	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	4	2021-04-01	0	0.00	0.00	#DIV/0!	#DIV/0!	ton decisiones frente a desviaciones de curva de p	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI
CENTRO DE SALUD TIPO A	316	4	2021-05-01	0	0.00	0.00	#DIV/0!	#DIV/0!	ton decisiones frente a desviaciones de curva de p	065005	SIGCHOS	MITON SIGCHI	SIGCHOS	COTOPAXI

Figura 26: Estructurar la información en formato tabla

D2 =FECHA(DERECHA(UNIRCADENAS(,"REPORTE ESTABLECIMIENTO SALUD"(SA6;SH56),4),1,1)

P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	
1	zona	trimestre	numerador	denominador	porcentaje	grupo	cuatrimestre	numerador	denominador	porcentaje	grupo	semestre	numerador	denominador	porcentaje	grupos	anual	numerador	denominador	porcentaje	grupo
2	ZONA 3	1	31.00	41.00	75.61	2.00	1	70.00	82.00	85.36585366	2	1	70.00	82.00	85.36585366	2	1	145.00	164.00	88.41463415	2
3	ZONA 3	1	31.00	41.00	75.61	2.00	1	70.00	82.00	85.36585366	2	1	70.00	82.00	85.36585366	2	1	145.00	164.00	88.41463415	2
4	ZONA 3	1	31.00	41.00	75.61	2.00	1	70.00	82.00	85.36585366	2	1	70.00	82.00	85.36585366	2	1	145.00	164.00	88.41463415	2
5	ZONA 3	2	39.00	41.00	95.12	3.00	1	70.00	82.00	85.36585366	2	1	70.00	82.00	85.36585366	2	1	145.00	164.00	88.41463415	2
6	ZONA 3	2	39.00	41.00	95.12	3.00	2	39.00	41.00	95.12195122	3	1	70.00	82.00	85.36585366	2	1	145.00	164.00	88.41463415	2
7	ZONA 3	2	39.00	41.00	95.12	3.00	2	39.00	41.00	95.12195122	3	1	70.00	82.00	85.36585366	2	1	145.00	164.00	88.41463415	2
8	ZONA 3	3	39.00	41.00	95.12	3.00	2	39.00	41.00	95.12195122	3	2	75.00	82.00	91.46341463	3	1	145.00	164.00	88.41463415	2
9	ZONA 3	3	39.00	41.00	95.12	3.00	2	39.00	41.00	95.12195122	3	2	75.00	82.00	91.46341463	3	1	145.00	164.00	88.41463415	2
10	ZONA 3	3	39.00	41.00	95.12	3.00	3	36.00	41.00	87.80487805	2	2	75.00	82.00	91.46341463	3	1	145.00	164.00	88.41463415	2
11	ZONA 3	4	36.00	41.00	87.80	2.00	3	36.00	41.00	87.80487805	2	2	75.00	82.00	91.46341463	3	1	145.00	164.00	88.41463415	2
12	ZONA 3	4	36.00	41.00	87.80	2.00	3	36.00	41.00	87.80487805	2	2	75.00	82.00	91.46341463	3	1	145.00	164.00	88.41463415	2
13	ZONA 3	4	36.00	41.00	87.80	2.00	3	36.00	41.00	87.80487805	2	2	75.00	82.00	91.46341463	3	1	145.00	164.00	88.41463415	2
14	ZONA 3	1	86.00	86.00	100.00	3.00	1	112.00	112.00	100	3	1	168.00	168.00	100	3	1	344.00	345.00	99.71014493	3
15	ZONA 3	1	86.00	86.00	100.00	3.00	1	112.00	112.00	100	3	1	168.00	168.00	100	3	1	344.00	345.00	99.71014493	3
16	ZONA 3	1	86.00	86.00	100.00	3.00	1	112.00	112.00	100	3	1	168.00	168.00	100	3	1	344.00	345.00	99.71014493	3
17	ZONA 3	2	82.00	82.00	100.00	3.00	1	112.00	112.00	100	3	1	168.00	168.00	100	3	1	344.00	345.00	99.71014493	3
18	ZONA 3	2	82.00	82.00	100.00	3.00	2	113.00	113.00	100	3	1	168.00	168.00	100	3	1	344.00	345.00	99.71014493	3
19	ZONA 3	2	82.00	82.00	100.00	3.00	2	113.00	113.00	100	3	1	168.00	168.00	100	3	1	344.00	345.00	99.71014493	3
20	ZONA 3	3	87.00	87.00	100.00	3.00	2	113.00	113.00	100	3	2	176.00	177.00	99.43502825	3	1	344.00	345.00	99.71014493	3
21	ZONA 3	3	87.00	87.00	100.00	3.00	2	113.00	113.00	100	3	2	176.00	177.00	99.43502825	3	1	344.00	345.00	99.71014493	3
22	ZONA 3	3	87.00	87.00	100.00	3.00	3	119.00	120.00	99.16666667	3	2	176.00	177.00	99.43502825	3	1	344.00	345.00	99.71014493	3
23	ZONA 3	4	89.00	90.00	98.89	3.00	3	119.00	120.00	99.16666667	3	2	176.00	177.00	99.43502825	3	1	344.00	345.00	99.71014493	3
24	ZONA 3	4	89.00	90.00	98.89	3.00	3	119.00	120.00	99.16666667	3	2	176.00	177.00	99.43502825	3	1	344.00	345.00	99.71014493	3
25	ZONA 3	4	89.00	90.00	98.89	3.00	3	119.00	120.00	99.16666667	3	2	176.00	177.00	99.43502825	3	1	344.00	345.00	99.71014493	3
26	ZONA 3	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
27	ZONA 3	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
28	ZONA 3	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
29	ZONA 3	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
30	ZONA 3	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
31	ZONA 3	2	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
32	ZONA 3	3	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
33	ZONA 3	3	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
34	ZONA 3	3	0.00	0.00	#DIV/0!	#DIV/0!	3	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
35	ZONA 3	4	0.00	0.00	#DIV/0!	#DIV/0!	3	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
36	ZONA 3	4	0.00	0.00	#DIV/0!	#DIV/0!	3	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
37	ZONA 3	4	0.00	0.00	#DIV/0!	#DIV/0!	3	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
38	ZONA 3	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
39	ZONA 3	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
40	ZONA 3	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
41	ZONA 3	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!
42	ZONA 3	2	0.00	0.00	#DIV/0!	#DIV/0!	2	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!	1	0.00	0.00	#DIV/0!	#DIV/0!

Figura 27: Estructuración de la información

Anexo 3. Descripción de los datos

Tabla 25: Descripción de la tabla Cantón

Atributos	Tipo	Descripción
Id	Integer	Identificador único del cantón
Cantón	String	Nombre del cantón
Distrito	String	Nombre del distrito que pertenece al cantón
Fecha_ingreso	Date	Fecha cuando ingresa los datos al sistema

Fuente: Lidia Enríquez, Clinton Moina

Tabla 26: Descripción de la tabla Provincia

Atributos	Tipo	Descripción
Id	Integer	Identificador único de provincia
Provincia	String	Nombre de la provincia
Fecha_ingreso	Timestamp	Fecha cuando ingresa los datos al sistema

Fuente: Lidia Enríquez, Clinton Moina

Tabla 27: Descripción de la tabla Establecimiento

Atributos	Tipo	Descripción
Id	Integer	Identificador único del establecimiento
Id_establecimiento	Number	Código único del establecimiento
Establecimiento	String	Nombre del establecimiento
Parroquia	String	Nombre de la parroquia que pertenece al establecimiento
Tipología	String	Nombre del tipo de establecimiento
Fecha_ingreso	Timestamp	Fecha cuando ingresa los datos al sistema

Fuente: Lidia Enríquez, Clinton Moina

Tabla 28: Descripción de la tabla Indicadores

Atributos	Tipo	Descripción
Id	Integer	Identificador único de indicador
Indicador	Number	Número del indicador
Nombre	String	Nombre del indicador
Universo	String	Total, de historias clínicas registradas
Numerador	String	Numero de historias clínicas registradas
Denominador	String	Numero de historias clínicas que deben registrar
Porcentaje	Number	Total, de dividir el numerador para el denominador
Fecha_ingreso	Date	Fecha cuando ingresan los datos del sistema
Grupom	String	Grupo en que pertenece el mes
Trimestre	Number	Numero de trimestre por año
Numeradort	String	Total, de historias clínicas registradas
Denominadort	String	Numero de historias clínicas que deben registrar

Procentajet	Number	Total, de dividir el numerador para el denominador
Grupot	String	Grupo en que pertenece el mes
Cuatrimestre	Number	Total, de dividir el numerador para el denominador
Numeradorc	String	Total, de historias clínicas registradas
Denominadorc	String	Numero de historias clínicas que deben registrar
Procentajec	Number	Total, de dividir el numerador para el denominador
Grupoc	String	Grupo en que pertenece el mes
Semestre	Number	Total, de dividir el numerador para el denominador
Numeradors	String	Total, de historias clínicas registradas
Denominadors	String	Numero de historias clínicas que deben registrar
Porcentajes	Number	Total, de dividir el numerador para el denominador
Grupos	String	Grupo en que pertenece el mes
Anual	Number	Total, de dividir el numerador para el denominador
Numeradora	String	Total, de historias clínicas registradas
Denominadora	String	Numero de historias clínicas que deben registrar
Porcentajea	Number	Total, de dividir el numerador para el denominador
Grupoa	String	Grupo en que pertenece el mes

Fuente: Lidia Enríquez, Clinton Moina

Tabla 29: Descripción de la tabla Zona

Atributos	Tipo	Descripción
Id	Integer	Identificador único de zona
Zona	String	Nombre de la zona
Fecha_ingreso	Timestamp	Fecha cuando los datos ingresan al sistema

Fuente: Lidia Enríquez, Clinton Moina

Tabla 50: Descripción de la tabla Tiempo

Atributos	Tipo	Descripción
Id	Integer	Identificador único de tiempo
Indicador	Number	Identificador del indicador
Meses	Date	Fecha que el indicador tiene datos

Fuente: Lidia Enríquez, Clinton Moina

Anexo 4. Exploración de los datos

Tabla 31: Colorimetría del Nivel de Calidad para el indicador 10

Color	Nivel de Calidad	Rango
Rojo	No aceptable	1,01 a 100%
Verde	Aceptable	0 a 1,00%

Fuente: Coordinación Zonal 3 del Ministerio de Salud Pública

Tabla 32: Colorimetría del Nivel de Calidad para el indicador 11

Color	Nivel de Calidad	Rango
Rojo	No Aceptable	1,01 en adelante
Verde	Aceptable	0 a 1,00%

Fuente: Coordinación Zonal 3 del Ministerio de Salud Pública

Tabla 33: Colorimetría del Nivel de Calidad para el indicador 12

Color	Nivel de Calidad	Rango
Rojo	No Aceptable	0 a 70,99%
Verde	Aceptable	71 a 100%

Fuente: Coordinación Zonal 3 del Ministerio de Salud Pública

Tabla 34: Colorimetría del Nivel de Calidad para el indicador 13

Color	Nivel de Calidad	Rango
Rojo	No Aceptable	0 a 99,99%
Verde	Aceptable	100%

Fuente: Coordinación Zonal 3 del Ministerio de Salud Pública

Anexo 5. Verificar la calidad de los datos

Tabla 35: Análisis de la calidad de los datos

Tabla	Campo	Valores nulos	Valores blancos	Valores validos	no
Cantón (1667)	Id	0%	0%	0%	
	Cantón	0%	0%	0%	
	Distrito	0%	0%	0%	
	Fecha Ingreso	0%	0%	0%	
Provincia (1667)	Id	0%	0%	0%	
	Provincia	0%	0%	0%	
	Fecha_ingreso	0%	0%	0%	
Establecimiento (1667)	Id	0%	0%	0%	
	Id_establecimiento	0%	0%	0%	
	Establecimiento	0%	0%	0%	
	Parroquia	0%	0%	0%	
	Tipología	0%	0%	0%	
	Id	0%	0%	0%	
	Indicador	0%	0%	0%	
	Nombre	0%	0%	0%	
	Universo	0%	0%	0%	
	Numerador	0%	0%	0%	
	Denominador	0%	0%	0%	




	Porcentaje	0%	0%	0%
	Fecha_ingreso	0%	0%	0%
Indicador	Grupom	0%	0%	0%
(56174)	Trimestre	0%	0%	0%
	Numeradort	0%	0%	0%
	Denominadort	0%	0%	0%
	Procentajet	0%	0%	0%
	Grupot	0%	0%	0%
	Cuatrimestre	0%	0%	0%
	Numeradorc	0%	0%	0%
	Denominadorc	0%	0%	0%
	Procentajec	0%	0%	0%
	Grupoc	0%	0%	0%
	Semestre	0%	0%	0%
	Numeradors	0%	0%	0%
	Denominadors	0%	0%	0%
	Porcentajes	0%	0%	0%
	Grupos	0%	0%	0%
	Anual	0%	0%	0%
	Numeradora	0%	0%	0%
	Denominadora	0%	0%	0%
	Porcentajea	0%	0%	0%
	Grupoa	0%	0%	0%
	Id	0%	0%	0%
Zona (141)	Zona	0%	0%	0%
	Fecha_ingreso	0%	0%	0%
	Id	0%	0%	0%
Tiempo	Indicador	0%	0%	0%
(56174)	Meses	0%	0%	0%

Fuente: Lidia Enríquez, Clinton Moina

Anexo 6. Modelos generados

dim_canton
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:10:46 by clinton.moina@trial43074.us.talend.com

Sample quality
0%  0%  100% 

	pk_dim_canton* Integer (string)	distrito* Text (string)	canton* Text (string)	fecha_ingreso* Date (string)	id_establecimie... Integer (string)
1	568	05D01	LATACUNGA	2023-02-07 23:59:...	347
2	569	05D01	LATACUNGA	2023-02-07 23:59:...	320
3	570	05D01	LATACUNGA	2023-02-07 23:59:...	348
4	571	05D01	LATACUNGA	2023-02-07 23:59:...	349
5	572	05D01	LATACUNGA	2023-02-07 23:59:...	3986
6	573	05D01	LATACUNGA	2023-02-07 23:59:...	352
7	574	05D01	LATACUNGA	2023-02-07 23:59:...	318
8	575	05D01	LATACUNGA	2023-02-07 23:59:...	358
9	576	05D01	LATACUNGA	2023-02-07 23:59:...	351
10	577	05D01	LATACUNGA	2023-02-07 23:59:...	353
11	578	05D01	LATACUNGA	2023-02-07 23:59:...	322
12	579	05D01	LATACUNGA	2023-02-07 23:59:...	301
13	580	05D01	LATACUNGA	2023-02-07 23:59:...	354

Figura 28: Completitud, Credibilidad y Precisión de los datos de la tabla Cantón

dim_establecimiento
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:19:52 by clinton.moina@trial43074.us.talend.com

Sample quality
0% ————— 100%

	pk_dim_estable... Integer (string)	establecimiento* Text (string)	parroquia* Text (string)	tipologia* Text (string)	fecha_ingreso* Date (string)	id_establecimie... Integer (string)
1	568	LATACUNGA	LA MATRIZ	CENTRO DE SALUD T...	2023-02-07 23:59:...	347
2	569	LASSO	TANICUCHÍ	CENTRO DE SALUD T...	2023-02-07 23:59:...	320
3	570	PATUTAN	ELOY ALFARO	CENTRO DE SALUD T...	2023-02-07 23:59:...	348
4	571	SAN BUENAVENTURA	SAN BUENAVENTURA	CENTRO DE SALUD T...	2023-02-07 23:59:...	349
5	572	LOMA GRANDE	ELOY ALFARO	CENTRO DE SALUD T...	2023-02-07 23:59:...	3986
6	573	ALAQUEZ	ALAQUES (ALÁQUEZ)	CENTRO DE SALUD T...	2023-02-07 23:59:...	352
7	574	POALO	POALÓ	CENTRO DE SALUD T...	2023-02-07 23:59:...	318
8	575	PALOPO	IGNACIO FLORES	CENTRO DE SALUD T...	2023-02-07 23:59:...	358
9	576	BELISARIO QUEVEDO	BELISARIO QUEVEDO	CENTRO DE SALUD T...	2023-02-07 23:59:...	351
10	577	MULALO	MULALO	CENTRO DE SALUD T...	2023-02-07 23:59:...	353
11	578	TOACASO	TOACASO	CENTRO DE SALUD T...	2023-02-07 23:59:...	322
12	579	JOSEGUANGO ALTO	MULALO	CENTRO DE SALUD T...	2023-02-07 23:59:...	301
13	580	JOSEGUANGO BAJO	JOSEGUANGO BAJO	CENTRO DE SALUD T...	2023-02-07 23:59:...	354

Figura 29: Completitud, Credibilidad y Precisión de los datos de la tabla Hechos

dim_hechosmp
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:20:06 by clinton.moina@trial43074.us.talend.com

Sample quality
0% ————— 100%

2.63/5 No rating

	pk_hechos* Integer (string)	pk_dim_zona* Integer (string)	pk_dim_provin... Integer (string)	pk_dim_canton* Integer (string)	pk_dim_estable... Integer (string)	pk_dim_indicad... FR Postal Code (string)	pk_dim_tiempo* FR Postal Code (string)	id_establecimie... Integer (string)	secuencia* Integer (string)
1	131	397	409	568	568	49393	46538	347	1
2	183	397	409	568	568	49394	46531	347	2
3	192	397	409	568	568	49395	46532	347	3
4	200	397	409	568	568	49396	46533	347	4
5	210	397	409	568	568	49397	46534	347	5
6	216	397	409	568	568	49398	46535	347	6
7	225	397	409	568	568	49399	46536	347	7
8	235	397	409	568	568	49400	46537	347	8
9	245	397	409	568	568	49401	46538	347	9
10	252	397	409	568	568	49402	46539	347	10
11	255	397	409	568	568	49403	46540	347	11
12	262	397	409	568	568	49404	46541	347	12
13	267	397	409	568	568	49405	46542	347	13

Figura 30: Completitud, Credibilidad y Precisión de los datos de la tabla Establecimiento

dim_indicador
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:29:22 by clinton.moina@trial43074.us.talend.com

Sample quality 0% 100%

	pk_dim_indicador* PK Postal Code (string)	id_establecimie... Integer (string)	indicador* Integer (string)	nombre* Text (string)	univero* Decimal (string)	numerador* Decimal (string)	denominador* Decimal (string)	porcentaje* Decimal (string)	fecha_ingreso* Date (string)	trimestre* Integer (string)
1	49393	347	1	% Insumos esencia...	878.0	162.0	168.0	96.42857142857143	2021-01-01	1
2	49394	347	1	% Insumos esencia...	0.0	176.0	188.0	97.77777777777777	2021-04-01	2
3	49395	347	1	% Insumos esencia...	0.0	120.0	188.0	66.66666666666666	2021-10-01	4
4	49396	347	2	Control prenatal ...	127.0	30.0	30.0	100.0	2021-01-01	1
5	49397	347	2	Control prenatal ...	108.0	29.0	30.0	96.66666666666667	2021-02-01	1
6	49398	347	2	Control prenatal ...	157.0	30.0	30.0	100.0	2021-03-01	1
7	49399	347	2	Control prenatal ...	129.0	30.0	30.0	100.0	2021-04-01	2
8	49400	347	2	Control prenatal ...	98.0	29.0	30.0	96.66666666666667	2021-05-01	2
9	49401	347	2	Control prenatal ...	104.0	28.0	30.0	93.33333333333333	2021-06-01	2
10	49402	347	2	Control prenatal ...	146.0	29.0	30.0	96.66666666666667	2021-07-01	3
11	49403	347	2	Control prenatal ...	165.0	29.0	30.0	96.66666666666667	2021-08-01	3
12	49404	347	2	Control prenatal ...	98.0	30.0	30.0	100.0	2021-09-01	3
13	49405	347	2	Control prenatal ...	107.0	12.0	30.0	40.0	2021-10-01	4

Figura 31: Completitud, Credibilidad y Precisión de los datos de la tabla
Indicador

dim_provincia
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:20:28 by clinton.moina@trial43074.us.talend.com




Sample quality 0% 100%

	pk_dim_provin... Integer (string)	id_establecimie... Integer (string)	provincia* Text (string)	fecha_ingreso* Date (string)
1	409	347	COTOPAXI	2023-02-07 23:59:...
2	410	320	COTOPAXI	2023-02-07 23:59:...
3	411	348	COTOPAXI	2023-02-07 23:59:...
4	412	349	COTOPAXI	2023-02-07 23:59:...
5	413	3986	COTOPAXI	2023-02-07 23:59:...
6	414	352	COTOPAXI	2023-02-07 23:59:...
7	415	318	COTOPAXI	2023-02-07 23:59:...
8	416	358	COTOPAXI	2023-02-07 23:59:...
9	417	351	COTOPAXI	2023-02-07 23:59:...
10	418	353	COTOPAXI	2023-02-07 23:59:...
11	419	322	COTOPAXI	2023-02-07 23:59:...
12	420	301	COTOPAXI	2023-02-07 23:59:...

Figura 32: Completitud, Credibilidad y Precisión de los datos de la tabla
Provincia

dim_tiempo
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:20:42 by clinton.moina@trial43074.us.talend.com




Sample quality
0%  0%  100% 

	pk_dim_tiempo* FR Postal Code (string)	id_establecimie... Integer (string)	indicador* Integer (string)	meses* Date (string)
1	46530	347	1	2021-01-01
2	46531	347	1	2021-04-01
3	46532	347	1	2021-10-01
4	46533	347	2	2021-01-01
5	46534	347	2	2021-02-01
6	46535	347	2	2021-03-01
7	46536	347	2	2021-04-01
8	46537	347	2	2021-05-01
9	46538	347	2	2021-06-01
10	46539	347	2	2021-07-01
11	46540	347	2	2021-08-01
12	46541	347	2	2021-09-01
13	46542	347	2	2021-10-01

Figura 33: Completitud, Credibilidad y Precisión de los datos de la tabla Tiempo

dim_zone
From Local connection: Local Connection - Sample: Head

Head sample ▾
2023-02-17 at 16:20:48 by clinton.moina@trial43074.us.talend.com

Sample quality
0%  0%  100% 

	dim_zone* Integer (string)	id_establecimie... Integer (string)	zona* Text (string)	fecha_ingreso* Date (string)
1	397	347	ZONA 3	2023-02-07 23:59:...
2	398	320	ZONA 3	2023-02-07 23:59:...
3	399	348	ZONA 3	2023-02-07 23:59:...
4	400	349	ZONA 3	2023-02-07 23:59:...
5	401	3986	ZONA 3	2023-02-07 23:59:...
6	402	352	ZONA 3	2023-02-07 23:59:...
7	403	318	ZONA 3	2023-02-07 23:59:...
8	404	358	ZONA 3	2023-02-07 23:59:...
9	405	351	ZONA 3	2023-02-07 23:59:...
10	406	353	ZONA 3	2023-02-07 23:59:...
11	407	322	ZONA 3	2023-02-07 23:59:...
12	408	301	ZONA 3	2023-02-07 23:59:...
13	409	354	ZONA 3	2023-02-07 23:59:...

Figura 34: Completitud, Credibilidad y Precisión de los datos de la tabla Zona

Anexo 7. Resultados Herramienta Talend: Credibilidad, Completitud y Precisión.

dim_canton

From Local connection: [Local Connection](#) - Sample: Random

Talend Trust Score™

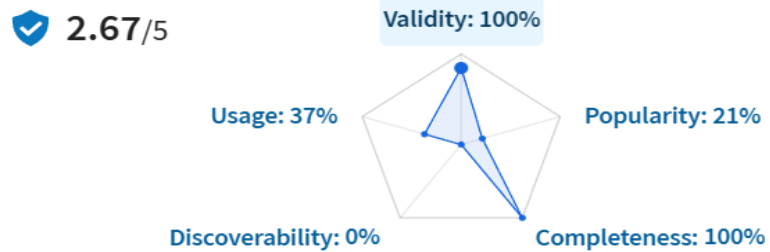


Figura 35: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Cantón

dim_establecimiento

From Local connection: [Local Connection](#) - Sample: Head

Talend Trust Score™

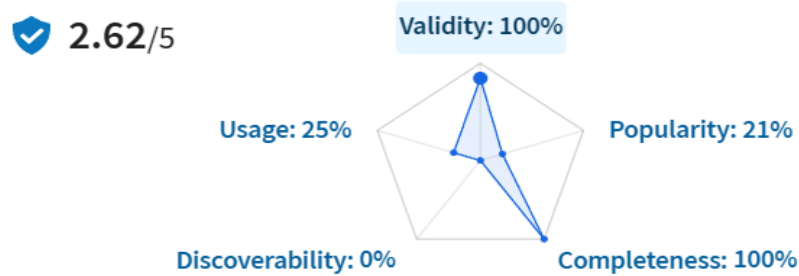


Figura 36: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Establecimiento

dim_hechosmsp

From Local connection: Local Connection - Sample: Head

Talend Trust Score™

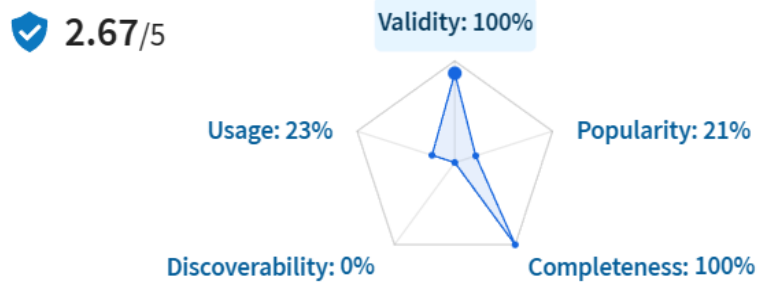


Figura 37: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Hechos

dim_indicador

From Local connection: Local Connection - Sample: Head

Talend Trust Score™

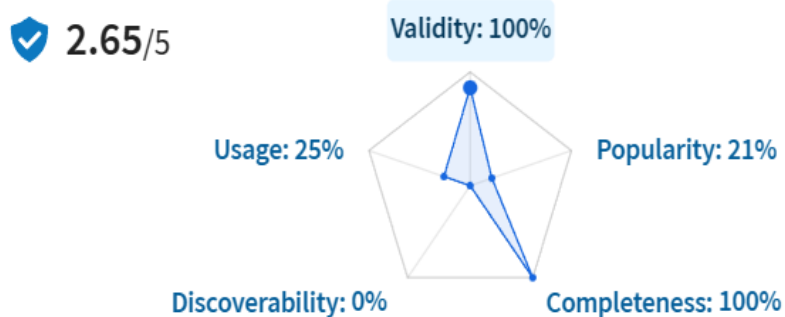


Figura 38: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Indicador

dim_provincia

From Local connection: Local Connection - Sample: Head

Talend Trust Score™

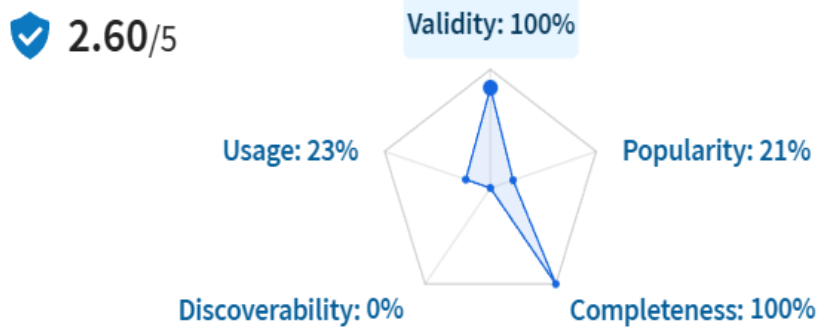


Figura 39: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Provincia

dim_tiempo

From Local connection: Local Connection - Sample: Head

Talend Trust Score™

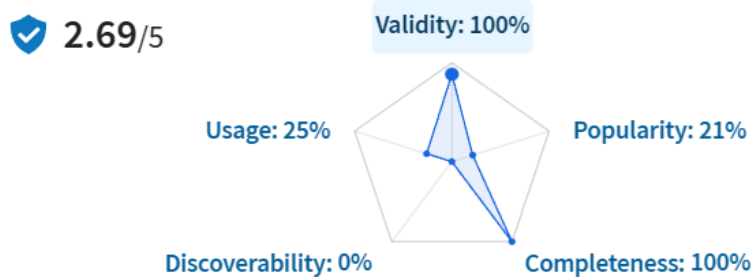


Figura 40: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Tiempo

dim_zone

From Local connection: [Local Connection](#) - Sample: Head

Talend Trust Score™

 2.60/5

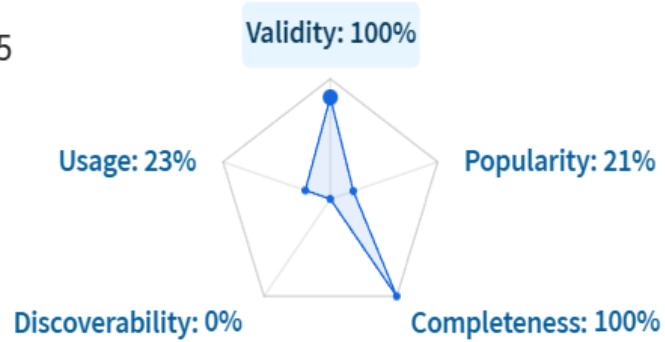


Figura 41: Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Tiempo