UNIVERSIDAD NACIONAL DE CHIMBORAZO
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

Análisis multidimensional y cuadro de mando integral de los datos meteorológicos y de la calidad del aire de Quito

Trabajo de Titulación para optar al título de Ingeniero en Tecnologías de la Información

Autor:
Cueva Bonilla Kevin David

Tutor:
PhD. Fidel Vallejo Gallardo

Riobamba, Ecuador. 2025

DECLARATORIA DE AUTORÍA

Yo, Kevin David Cueva Bonilla, con cédula de ciudadanía 0650241177, autor del trabajo de investigación titulado: Análisis multidimensional y cuadro de mando integral de los datos meteorológicos y de la calidad del aire de Quito, certifico que la producción, ideas, opiniones, criterios, contenidos y conclusiones expuestas son de mí exclusiva responsabilidad.

Asimismo, cedo 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 autor de la obra referida, será de mi entera responsabilidad; librando a la Universidad Nacional de Chimborazo de posibles obligaciones.

En Riobamba, 20 de octubre de 2024.

_____

Kevin David Cueva Bonilla

C.I: 0650241177

ACTA FAVORABLE - INFORME FINAL DEL TRABAJO DE INVESTIGACIÓN

En la Ciudad de Riobamba, a los 29 días del mes de octubre de 2024, luego de haber revisado el Informe Final del Trabajo de Investigación presentado por el estudiante **KEVIN DAVID CUEVA BONILLA** con CC: **0650241177**, de la carrera **INGENIERÍA DE TECNOLOGÍAS DE LA INFORMACIÓN**, y dando cumplimiento a los criterios metodológicos exigidos, se emite el **ACTA FAVORABLE DEL INFORME FINAL DEL TRABAJO DE INVESTIGACIÓN** titulado **"ANÁLISIS MULTIDIMENSIONAL Y CUADRO DE MANDO INTEGRAL DE LOS DATOS METEOROLÓGICOS Y DE LA CALIDAD DEL AIRE DE QUITO"**, por lo tanto se autoriza la presentación del mismo para los trámites pertinentes.

FIDEL ERNESTO
VALLEJO GALLARDO

Ing. Fidel Vallejo Gallardo, Ph.D.
TUTOR

CERTIFICADO DE LOS MIEMBROS DEL TRIBUNAL

Quienes suscribimos, catedráticos designados Miembros del Tribunal de Grado para la evaluación del trabajo de investigación Análisis multidimensional y cuadro de mando integral de los datos meteorológicos y de la calidad del aire de Quito, presentado por Kevin David Cueva Bonilla con cédula de identidad número 0650241177, bajo la tutoría de Ing. Fidel Vallejo Gallardo, PhD; certificamos que recomendamos la APROBACIÓN de este con fines de titulación. Previamente se ha evaluado el trabajo de investigación y escuchada la sustentación por parte de su autor; no teniendo más nada que observar.

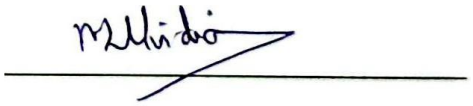
De conformidad a la normativa aplicable firmamos, en Riobamba a la fecha de su presentación.

PhD. Miryan Narváez

PRESIDENTE DEL TRIBUNAL DE GRADO

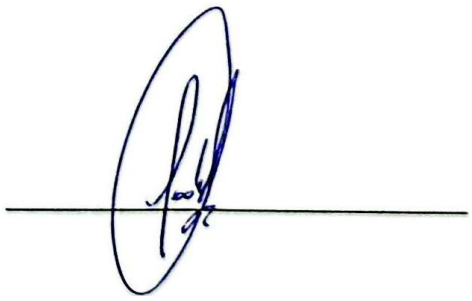
Mgs. María Isabel Uvidia

MIEMBRO DEL TRIBUNAL DE GRADO



Mgs. Lady Espinoza

MIEMBRO DEL TRIBUNAL DE GRADO



CERTIFICACIÓN

Que, **CUEVA BONILLA KEVIN DAVID** con CC: **0650241177**, estudiante de la Carrera **INGENIERÍA DE TECNOLOGÍAS DE LA INFORMACIÓN**, Facultad de **INGENIERÍA**; ha trabajado bajo mi tutoría el trabajo de investigación titulado "**ANÁLISIS MULTIDIMENSIONAL Y CUADRO DE MANDO INTEGRAL DE LOS DATOS METEOROLÓGICOS Y DE LA CALIDAD DEL AIRE DE QUITO**", cumple con el 10%, de acuerdo al reporte del sistema Anti plagio **TURNITIN**, porcentaje aceptado de acuerdo a la reglamentación institucional, por consiguiente autorizo continuar con el proceso.

Riobamba, 03 de diciembre de 2024

Ing. Fidel Vallejo Gallardo, Ph.D.
TUTOR

DEDICATORIA

Dedico este proyecto de investigación a mi familia, quienes han sido mi apoyo constante toda mi vida y han creído en mi siempre, a mis padres Willman Torres y Paulina Bonilla por ser el fundamento principal de mis días y mi guía para ser un gran ser humano, a mis hermanos Alejandro, Nicole, Islam, María del mar, Gabrielito por ser un gran equipo en toda mi trascendencia como persona. Dedico a mi abuelito Acevedo quien vive en mi memoria, gracias por inculcar todos los valores y el conocimiento que me hacen ser quien soy ahora y lo mucho que agradezco que formó a un gran padre, quien me enseñó la frase: "Empeñémonos por siempre hacer las cosas bellas, porque en lo bello está lo bueno. En las cosas buenas está Dios, porque las cosas bellas y buenas perseveran el universo", dedico a mi madre querida que sin duda da la vida por mí, y nunca me hace faltar el alimento, dedico a mis mishis, a mis perritos, a todas esas personas que amé, y he querido, fueron una parte de mi etapa universitaria y de mi aprendizaje.

Dedico a todos los ingenieros: Ximena Quintana, Diego Reina, Ana Congacha, Jorge Delgado, Gonzalo Allauca, Estela Narvaez, María Isabel Uvidia, Hugo Paz, Pamela Buñay, que en una parte de mi formación conversaron conmigo, me dieron un consejo, nos reímos, o pensamos en proyectos a futuro, hicimos eventos, organizamos cosas, etc. Fueron la transición de mi visión por sacar adelante a la carrera y sin duda creyeron en mi o me felicitaron, son profesionales que me hicieron amar mi título y me motivaron a compartir el conocimiento a los demás. Dedico a todos mis amigos, o personas que, aunque ya no nos hablemos siempre los estimaré por haberme acompañado en estos años, por todas esas horas de desvelo, risas, enojos, peleas, más risas, momentos compartidos que nunca se olvidan, Roger, Cristina, Bryan, Estefanía, Wilson. También dedico a todas esas personas que pudimos ser alguna vez Microsoft D´Soft.net, también a todos los que integran Nexus Community, siempre estaré orgulloso de verlos como futuros colegas y grandes profesionales.

Dedico a todos mis amigos y conocidos de la universidad, especialmente a CIU, gracias por su apoyo y dedicación en la representación de los intereses estudiantiles, por ser personas de contar y por ver lo que soy.

Finalmente, dedico esto a todos aquellos que, de una u otra forma, han sido parte de este camino. Su presencia y apoyo han sido fundamentales para alcanzar esta meta.

AGRADECIMIENTO

Agradezco a Dios por permitirme alcanzar mis logros y metas que me he propuesto, a mi tutor por su apoyo y guía durante el desarrollo de este trabajo de investigación, a los docentes de la carrera de Ingeniería en Tecnologías de la Información por impartir su conocimiento y dedicación en mi formación como ingeniero, y a Himbher por su apoyo en mi tesis.

Agradezco a Dios y a mi familia, por su apoyo y las fuerzas para superar numerables veces las enfermedades, obstáculos debido a la depresión, bajas de defensas, pensamientos suicidas, desmotivación, malas personas, personas injustas, personas que quise y ya no son parte de mi vida, la competencia, etc. Fueron mi motivación a superarme y demostrar que puedo llegar hacer tantas cosas después de todo y ser un ejemplo a seguir, gracias por estar siempre a mi lado, guiándome y sosteniéndome en los momentos más difíciles, gracias por creer en mí y saber quién soy, gracias por hacerme ver que de todo soy alguien especial y excepcional.

Expreso también mi sincero agradecimiento al Proyecto Internacional "Interactions between environmental compartments and their impact on the Andean ecohydrosphere under conditions of climate change", inscrito en la Universidad Nacional de Chimborazo y desarrollado en colaboración con la Universidad Técnica Federico Santa María, la Universidad Católica de Temuco, y la Universidad de Santiago de Chile. Este proyecto ha sido clave para mi formación y me ha brindado la oportunidad de comprender el impacto del cambio climático en la eco hidrosfera andina.

Kevin David Cueva Bonilla

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

La presente investigación se centra en el desarrollo de un análisis multidimensional y un cuadro de mando integral aplicado a datos meteorológicos y de calidad del aire de Quito, Ecuador, con el objetivo de brindar una herramienta que permita visualizar y comprender la evolución de estos datos a lo largo del tiempo. Este estudio surge de la necesidad de analizar grandes volúmenes de datos ambientales, los cuales son fundamentales para la toma de decisiones en temas de salud pública y gestión ambiental en la ciudad.

Para lograr este objetivo, se emplearon herramientas como R Studio, Power BI y Sistemas de Información Geográfica (SIG), facilitando tanto el procesamiento como la visualización de datos de calidad del aire y meteorológicos. La metodología utilizada fue CRISP-DM (Cross Industry Standard Process for Data Mining), que en sus seis fases permitió estructurar el proceso de análisis de datos y asegurar la calidad de la información. Para la evaluación de la calidad de los datos se utilizaron los criterios de completitud, credibilidad y precisión, siguiendo los lineamientos de la norma ISO/IEC 25012, utilizando la herramienta Talend Data Quality.

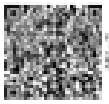
En cuanto a los resultados obtenidos, el análisis de calidad de datos mostró un 100% de completitud, un 100% de credibilidad y un 100% de precisión en las tablas de datos evaluadas. Estos resultados confirman que los datos utilizados cumplen con los estándares de calidad exigidos, garantizando la confiabilidad de los análisis y conclusiones del estudio. El producto final de esta investigación es un modelo de análisis multidimensional y un cuadro de mando integral que permite a los usuarios acceder de manera interactiva y visual a información crucial sobre la calidad del aire y las condiciones meteorológicas de Quito, apoyando así la toma de decisiones fundamentada en datos de alta calidad.

Palabras claves: Sistema Multidimensional, Cuadro de Mando, Calidad de Datos, CRISP-DM, Quito, Calidad del aire, Meteorología, Gestión ambiental.

ABSTRACT

This research focuses on developing a multidimensional analysis and a balanced scorecard applied to meteorological and air quality data from Quito, Ecuador, to provide a tool that allows one to visualize and understand the evolution of these data over time. This study arises from the need to analyze large volumes of environmental data, which are essential for decision-making on public health and environmental management issues in the city. To achieve this objective, tools such as R Studio, Power BI, and Geographic Information Systems (GIS) were used, facilitating both the processing and visualization of air quality and meteorological data. The methodology used was CRISP-DM (Cross Industry Standard Process for Data Mining), which has six phases that allow for structuring the data analysis process and ensuring the quality of the information. Evaluating the quality of the data, the criteria of completeness, credibility, and precision were used, following the guidelines of the ISO/IEC 25012 standard, using the Talend Data Quality tool. Regarding the results obtained, the data quality analysis showed 100% completeness, 100% credibility, and 100% accuracy in the data tables evaluated. These results confirm that the data used meet the required quality standards, guaranteeing the reliability of the analysis and conclusions of the study. The final product of this research is a multidimensional analysis model and a balanced scorecard that allows users to interactively and visually access crucial information on air quality and meteorological conditions in Quito, thus supporting decision-making based on high-quality data.

Keywords: Multidimensional System, Scorecard, Data Quality, CRISP-DM, Quito, Air Quality, Meteorology, Environmental Management.



Reviewed by:
Mg. Javier Andrés Saltos Chacán
ENGLISH TEACHER
c.c. 0202481438

CAPÍTULO I. INTRODUCCIÓN

La ciudad de Quito, epicentro de la región andina y capital de Ecuador, ha experimentado un crecimiento urbano acelerado en las últimas décadas, con un incremento en la actividad industrial, el tráfico vehicular y la demanda de energía [1]. Estos factores, aunque impulsan el desarrollo, también plantean desafíos significativos para la calidad del aire, un aspecto crítico que afecta directamente la salud pública y la sostenibilidad ambiental [2].

En este contexto, se desarrolla un análisis multidimensional y cuadro de mando integral de los datos meteorológicos y la calidad del aire de Quito durante el período 2005-2022, con el propósito de analizar cómo las variaciones climáticas influyen en la calidad del aire. Este proyecto aplica herramientas avanzadas de Tecnologías de la Información (TI) para abordar aspectos relevantes de la gestión ambiental y urbana en Quito, demostrando cómo la carrera de TI contribuye al desarrollo de sistemas de análisis de datos complejos y visualización interactiva que facilitan la toma de decisiones basada en evidencia.

Patrones estacionales y geográficos: ¿Existen patrones estacionales o geográficos discernibles en los datos de calidad del aire y meteorología en Quito a lo largo de los años?

Tendencias de contaminación: ¿Se ha registrado un aumento significativo en los niveles de contaminación durante el período de estudio?

Comparación con estándares internacionales: ¿Cómo se comparan los datos recopilados con las normativas actuales y las recomendaciones de la Organización Mundial de la Salud (OMS) en términos de calidad del aire?

Este estudio adoptó un enfoque metodológico riguroso que involucra la recopilación, limpieza, ordenamiento, análisis y visualización de datos. A través de herramientas como R (R Studio [3], paquetes TimeVariation [4], OpenAir [5]), Shiny [6], Power BI [7] y SIG (Sistemas de Información Geográfica) [8], se examinarán variables meteorológicas y contaminantes primarios y secundarios, como $PM_{2.5}$, NO2, SO2, O3, PM10, entre otras.

Esta investigación representa una contribución esencial para analizar, desde un enfoque de análisis de datos, el complejo problema ambiental de una ciudad andina. La comprensión de las interacciones entre el crecimiento urbano y la calidad del aire es fundamental para informar políticas y prácticas que promuevan un entorno urbano más saludable y sostenible en Quito.

1.1 Planteamiento del Problema

A medida que la urbe ha experimentado un crecimiento demográfico y una mayor actividad industrial y vehicular, se ha generado un aumento en las emisiones de contaminantes atmosféricos, incluyendo partículas finas (PM2.5), dióxido de nitrógeno (NO2), ozono (O3), entre otros. Estos contaminantes son conocidos por su influencia negativa en la salud humana y el medio ambiente.

La carencia de herramientas adecuadas para recopilar, procesar y visualizar datos meteorológicos y de contaminación atmosférica de manera eficiente y precisa ha dificultado el análisis integral de la calidad del aire en Quito. Esta fragmentación en la recolección de datos impide una comprensión exhaustiva de la evolución temporal y los patrones de contaminación atmosférica, así como la comparación con estándares nacionales e internacionales.

Esta investigación se planteó como un paso significativo para abordar estos desafíos, al introducir un enfoque innovador centrado en la aplicación de análisis multidimensional y la implementación de un cuadro de mando integral. Estos métodos avanzados permitirán la integración y visualización eficiente de los datos meteorológicos y de la calidad del aire de la ciudad de Quito. Al proporcionar una perspectiva más completa y accesible, se busca no solo comprender la evolución temporal y patrones estacionales de la contaminación atmosférica, sino también facilitar la toma de decisiones en la gestión ambiental en siguientes trabajos. Este enfoque se alinea con la naturaleza tecnológica de la Ingeniería en Tecnologías de la Información, ofreciendo herramientas prácticas y aplicables para mejorar la calidad de vida en la ciudad y contribuir a un entorno más saludable y sostenible.

1.2 Justificación

La implementación de un análisis multidimensional y de un cuadro de mando integral se justifica por la necesidad de abordar los desafíos actuales relacionados con la calidad del aire en la ciudad de Quito. El crecimiento demográfico, la actividad industrial y vehicular han generado un aumento en las emisiones de contaminantes atmosféricos, afectando tanto la salud humana como el medio ambiente. Sin embargo, la falta de una evaluación exhaustiva y actualizada de la calidad del aire ha llevado a un insuficiente análisis en el tiempo y una comparativa con las normativas nacionales e internacionales. Por lo tanto, es imperativo introducir métodos avanzados que permitan la integración y visualización eficiente de los datos meteorológicos y de la calidad del aire, con el fin de comprender mejor la evolución temporal y patrones estacionales de la contaminación atmosférica. Este enfoque no solo facilitará la toma de decisiones en la gestión ambiental, sino que también se alinea con la naturaleza tecnológica de la Ingeniería en Tecnologías de la Información, ofreciendo herramientas prácticas y aplicables para mejorar la calidad de vida en la ciudad y contribuir a un entorno más saludable y sostenible.

1.3 Formulación del Problema

En este trabajo de investigación se propuso abordar la siguiente pregunta: ¿Cómo el sistema multidimensional y el cuadro de mando integral permitirán mejorar la completitud, credibilidad y precisión de los datos meteorológicos y de la calidad del aire de Quito? Este enfoque se posiciona como un paso significativo para enfrentar los desafíos actuales, al ofrecer herramientas prácticas y aplicables para visualización y manejo de los datos meteorológicos y de la calidad del aire de Quito.

1.4 Objetivos

Objetivo General

Desarrollar un análisis multidimensional y cuadro de mando integral de los datos meteorológicos y de la calidad del aire de Quito.

Objetivos Específicos

- Analizar sistemas multidimensionales y cuadro de mando integral.
- Diseñar e implementar un modelo de análisis multidimensional y un cuadro de mando integral utilizando las herramientas R, R Studio, paquetes OpenAir, Power Bi y Sistemas de Información Geográfica (SIG) para la visualización y comprensión de los datos meteorológicos y de calidad del aire en Quito.
- Evaluar la completitud, credibilidad y precisión de los datos meteorológicos y de la calidad del aire de la ciudad de Quito, siguiendo estándares de la norma ISO/IEC-25012.

CAPÍTULO II. MARCO TEÓRICO

2.1 Data Warehouse

Conocido en español como almacén de datos, es un repositorio centralizado de datos que se utiliza para almacenar y gestionar información de diversas fuentes dentro de una organización. Está diseñado para facilitar el análisis y la toma de decisiones mediante el acceso a datos consolidados y consistentes [9].

2.1.1 Beneficios del Data Warehouse

El empleo de un Data Warehouse conlleva una serie de beneficios clave para las organizaciones [10], entre los que se destacan:

- **Integración de datos:** Centraliza datos de diversas fuentes para una visión unificada.
- **Análisis avanzado:** Facilita la identificación de patrones y tendencias clave.
- **Toma de decisiones:** Proporciona información precisa para decisiones estratégicas.
- **Eficiencia:** Mejora el rendimiento y la velocidad de acceso a los datos.
- **Historización:** Permite análisis retrospectivos y seguimiento de tendencias.

2.1.2 Características de un Data Warehouse

Según Noblejas [10], un Data Warehouse se caracteriza por las siguientes cualidades distintivas:

- **Orientación temática**: Los datos se organizan en torno a temas específicos de interés empresarial, facilita su acceso y comprensión.
- **Integración:** Combina datos de diversas fuentes y sistemas en un único repositorio centralizado.
- **Persistencia:** Los datos almacenados en el Data Warehouse son estables y no volátiles, esta garantiza su consistencia y fiabilidad.
- **Histórico**: Permite mantener un registro histórico de datos para análisis retrospectivos y seguimiento de tendencias.
- **Soporte para consultas complejas**: Diseñado para admitir consultas complejas y análisis de datos mediante herramientas de BI.
- **Estructura desnormalizada:** Los datos se desnormalizan para mejorar el rendimiento de las consultas y los informes.
- **No volátil:** Los datos almacenados en el Data Warehouse no se modifican ni se eliminan.

2.1.3 Requerimientos fundamentales

Según Kimball [11], los requerimientos fundamentales de un DataWarehouse, son:

- Integración de datos

- Escalabilidad
- Rendimiento
- Seguridad
- Usabilidad

Según diversos expertos en gestión de datos [12], las empresas enfrentan el desafío de manejar una gran cantidad de información no estructurada, que puede representar hasta el 80% de sus datos y se presenta en diversos formatos como texto o archivos xls. Para abordar este desafío, los Data Warehouses se convierten en herramientas fundamentales al proporcionar un entorno estructurado y centralizado para almacenar y gestionar estos datos, basándose en sistemas multidimensionales y Data Marts para organizar la información en unidades más manejables.

La integración de Data Marts dentro de un Data Warehouse es crucial para una gestión eficiente de los datos empresariales, permitiendo una estructuración clara de la información y cumpliendo con objetivos tanto empresariales como técnicos. Por otro lado, la naturaleza analítica de los Data Warehouses requiere un enfoque de procesamiento diferente, implicando la implementación de un modelo de base de datos conocido como sistema multidimensional, que facilita el acceso, la navegación y la recuperación de información para análisis detallados [12].

2.2 Sistema Multidimensional

Es un modelo de base de datos diseñado para representar y analizar datos de manera eficiente, especialmente para aplicaciones analíticas y de generación de informes. Permite organizar datos en una estructura multidimensional compuesta por tablas de hechos y dimensiones. Esta estructura facilita la visualización y el análisis de datos desde diferentes perspectivas, ayuda a los usuarios a comprender mejor las relaciones entre los diferentes elementos de los datos [13].

2.2.1 Tabla de hechos

Es una parte fundamental del modelo multidimensional y representa los eventos o medidas que se están analizando. Contiene datos numéricos o métricas que se pueden analizar y comparar. Algunos puntos clave para definir una tabla de hechos incluyen:

- Representación de medidas cuantificables.
- Contiene datos detallados y desagregados.
- Suele estar en el centro del esquema multidimensional y está conectada a las dimensiones a través de claves externas.

2.2.2. Dimensiones

Representan las categorías por las cuales se analizan los datos en la tabla de hechos. Proporcionan contextos para las medidas en la tabla de hechos y permiten segmentar y filtrar los datos según diferentes criterios. Algunos puntos para definir las dimensiones incluyen:

- Representación de aspectos cualitativos o descriptivos de los datos.
- Suelen tener una estructura jerárquica.
- Proporcionan la capacidad de agrupar y organizar datos para análisis.

2.2.3. Dimensión Tiempo

Es una dimensión especial que se utiliza para analizar datos en función del tiempo. Permite realizar análisis temporales y seguimiento de tendencias a lo largo del tiempo. Algunos puntos para definir la dimensión de tiempo incluyen:

- Organización de datos en unidades de tiempo como días, semanas, meses o años.
- Facilita la visualización de datos temporales mediante gráficos de series temporales.
- Permite realizar análisis comparativos y detectar patrones estacionales o tendencias a lo largo del tiempo.

2.2.4. Modelado de un sistema multidimensional

Para realizar la representación de un sistema multidimensional, existen diferentes tipos de esquemas.

- **Esquema de Estrella (Star Scheme)**

  Es la estructura más comúnmente empleada en las bases de datos multidimensionales. Consiste en una tabla central o Fact Table donde se almacenan los datos clave y no redundantes. A esta tabla central se vinculan una serie de tablas de dimensiones, cada una representando un aspecto cualitativo o descriptivo de los datos. Estas tablas de dimensiones están formadas por conjuntos de atributos que pueden organizarse jerárquicamente o de manera parcial [14].

  En la **Figura 1** se representa un esquema de estrella:

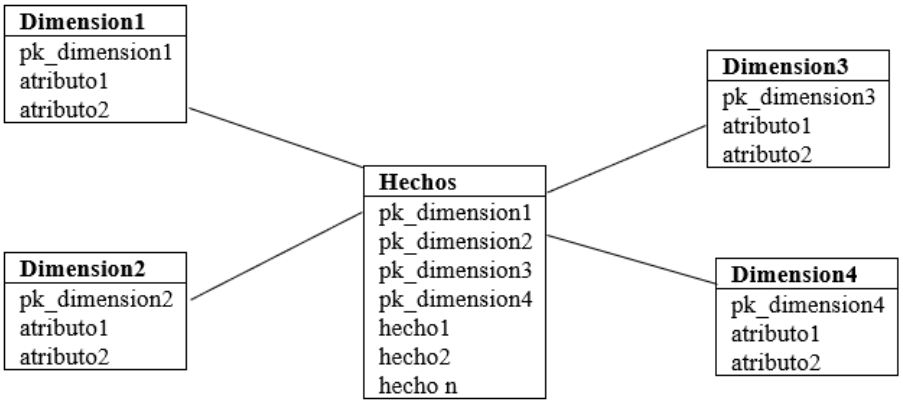


Figura 1: Esquema de estrella

- **Esquema de Copo de Nieve (Snowflake Scheme)**

Similar al esquema de estrella, pero con la diferencia de que algunas de las tablas que componen la base de datos están normalizadas. La normalización permite generar tablas adicionales, da lugar a una estructura que se asemeja a un copo de nieve. Esta normalización reduce la redundancia y ahorra espacio de almacenamiento en comparación con el esquema de estrella [14].

En la **Figura 2** se representa un esquema de Copo de Nieve:

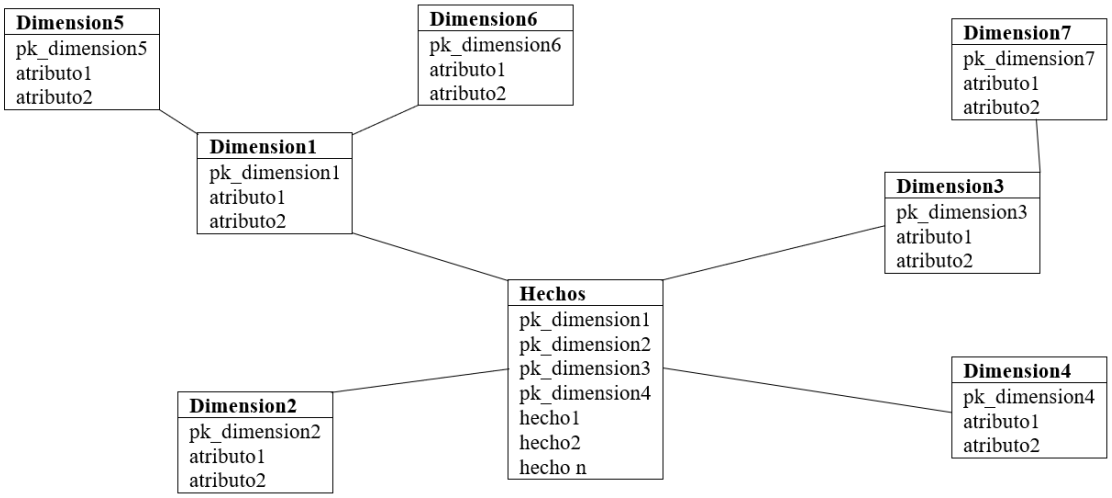


Figura 2: Esquema Copo de Nieve

- **Esquema de constelación de hechos**

Consiste en una estructura compuesta por varios esquemas de estrella. Cada esquema de estrella representa una tabla de hechos que se relaciona con otras tablas de dimensiones. Esta configuración permite modelar relaciones más complejas entre los datos y es útil en situaciones donde se necesitan analizar múltiples conjuntos de datos interrelacionados [14].

En la **Figura 3** se representa un esquema de Copo de Nieve:

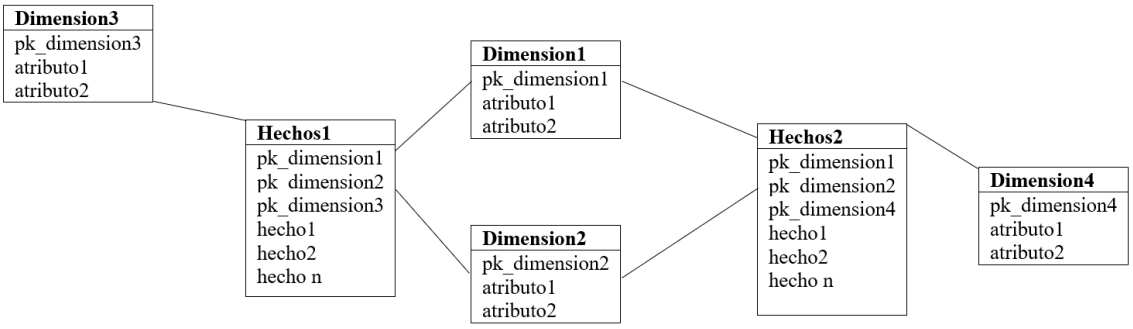


Figura 3: Esquema Constelación de Hechos.

2.3 ETL

El proceso de Extracción, Transformación y Carga es esencial para la gestión y análisis de datos. Consiste en consolidar información dispersa de diversas fuentes en un repositorio centralizado llamado almacenamiento de datos. Mediante reglas comerciales definidas, el ETL limpia, organiza y estructura los datos en bruto para su posterior almacenamiento, análisis y aplicación en inteligencia empresarial y machine learning (ML). Esto permite abordar necesidades específicas de inteligencia empresarial, como la generación de informes y la optimización de operaciones [15].

ETL es un proceso fundamental en la gestión de datos que consiste en tres etapas principales:

- **Extract (Extracción):** En esta etapa, los datos se extraen de diversas fuentes de origen, que pueden incluir bases de datos, archivos planos, sistemas en tiempo real, APIs, entre otros. La extracción puede implicar la recuperación de grandes volúmenes de datos de múltiples fuentes para su posterior procesamiento [16].
- **Transform (Transformación):** Durante la etapa de transformación, los datos extraídos se someten a diversas operaciones para limpiar, estructurar y prepararlos para su posterior análisis. Esto puede incluir la eliminación de datos duplicados o irrelevantes, la conversión de formatos, la normalización de datos, la agregación y la aplicación de reglas de negocio [16].
- **Load (Carga):** La etapa final del proceso ETL implica cargar los datos transformados en un almacén de datos o en una base de datos adecuada para su almacenamiento y análisis posteriores. La carga puede realizarse de forma incremental o completa, dependiendo de los requisitos del sistema y la frecuencia de actualización de los datos [16].

El proceso ETL es fundamental para garantizar la integridad, la calidad y la disponibilidad de los datos para su posterior análisis y uso en la toma de decisiones empresariales. Un ETL eficiente ayuda a garantizar la consistencia de los datos, mejora la eficiencia operativa y facilita la generación de informes y análisis significativos [17].

2.4 Pentaho

Es una plataforma de código abierto que ofrece una suite integral de herramientas para la integración de datos, la preparación de datos, el análisis de datos y la generación de informes. Utiliza un enfoque modular que permite a los usuarios seleccionar y combinar las herramientas según sus necesidades específicas. Pentaho ofrece capacidades para la extracción, transformación y carga (ETL) de datos, así como para la creación de paneles interactivos, informes y visualizaciones de datos. Es ampliamente utilizado en entornos empresariales debido a su flexibilidad, escalabilidad y facilidad de uso [18].

2.5 Cuadro de mando integral (CMI)

Es una metodología de gestión estratégica que se enfoca en traducir la visión y la estrategia de una organización en un conjunto coherente de indicadores de desempeño [19].

El CMI proporciona una visión equilibrada del rendimiento de la organización, permitiendo a los líderes tomar decisiones informadas y alinear las actividades diarias con los objetivos a largo plazo.

El CMI no solo se utiliza como una herramienta de medición retrospectiva del desempeño, sino también como un sistema de gestión proactiva que impulsa la mejora continua y la innovación dentro de la organización [20].

2.6 Lenguaje de programación R

Según estudios realizados por Vargas [3], se define a R como un lenguaje de programación de código abierto y multiplataforma diseñado para el análisis estadístico y la presentación gráfica de datos. Su popularidad ha crecido considerablemente debido a su facilidad de aprendizaje y su versatilidad, convirtiéndose en la herramienta más utilizada en campos como el aprendizaje automático, la minería de datos, la investigación biomédica, la bioinformática y la financiera.

Algunas de las características destacadas de R, según lo indicado por Santaela [21], incluyen:

- Capacidad integral para el manejo de datos.
- Una amplia colección de herramientas para el análisis estadístico.
- Funcionalidades gráficas para representar visualmente los resultados del análisis.
- La capacidad de combinar múltiples funciones estadísticas en un solo programa.
- La integración con otros lenguajes de programación como C, C++ o Fortran para realizar tareas de computación intensiva en datos.

2.6.1 R Studio

Es un entorno de desarrollo integrado (IDE) para R, que proporciona herramientas para la escritura y ejecución de código, visualización de datos, depuración y gestión de proyectos.

Es utilizado por su interfaz intuitiva y sus características que facilitan el análisis de datos y la generación de informes [22].

2.6.2 Paquetes R

Son extensiones de código desarrolladas por la comunidad de usuarios de R para realizar tareas específicas, como análisis estadístico, visualización de datos, modelado predictivo y más. Estos paquetes amplían las funcionalidades de R y permiten a los usuarios acceder a una amplia gama de herramientas especializadas para sus necesidades de análisis de datos [23].

Se analiza los siguientes paquetes: OpenAir, timevariation, lubridate, dplyr, shiny, ggplot2. La **Tabla 1** se compara los paquetes mencionados.

Tabla 1: Cuadro comparativo paquetes R

Nombre	Características	Ventajas	Desventajas	Configuración
OpenAir	Análisis de calidad del aire. Funciones específicas para visualizar y modelar datos ambientales.	Especializado en calidad del aire. Amplia gama de funciones.	Limitado a análisis de calidad del aire. Requiere conocimiento previo sobre el tema.	Configuración detallada requerida.
Lubridate	Paquete de R para manipulación de fechas y horas.	Facilita el trabajo con fechas en R (creación, extracción y modificación de elementos de fechas).	Tiene limitaciones con algunas zonas horarias muy específicas.	Aplicar funciones según la operación deseada.
Dplyr	Manejo y manipulación de datos en R, incluyendo filtrado, arreglos y resúmenes.	Muy eficiente para la manipulación de datos. Sencillo y intuitivo de usar con "verbs" claros.	Puede ser menos intuitivo para operaciones complejas.	Configuración simple, uso de verbos para manipulación de datos.
Shiny	Desarrollo de aplicaciones web interactivas directamente en R.	Permite crear aplicaciones interactivas fácilmente.	Requiere conocimientos básicos de desarrollo web y de R.	Configuración inicial para servidor y UI.
Ggplot2	Sistema de gráficos basado en gramática de gráficos. Permite crear gráficos complejos y personalizados de forma sencilla.	Altamente flexible y personalizable. Produce gráficos de alta calidad estéticamente atractivos.	Curva de aprendizaje empinada para usuarios nuevos. Algunas funciones pueden requerir conocimientos avanzados de sintaxis.	Requiere ajustes específicos para cada tipo de gráfico.

Fuente: [23]

2.7 Sistemas de información geográfica (SIG)

ESRI es una empresa líder en el desarrollo de software para Sistemas de Información Geográfica (SIG). Su plataforma principal, ArcGIS, es una de las herramientas más

utilizadas a nivel mundial en este campo. ArcGIS ofrece una amplia gama de aplicaciones y herramientas para capturar, almacenar, manipular, analizar y presentar datos geoespaciales. Los Sistemas de Información Geográfica (SIG) son herramientas que permiten capturar, almacenar, manipular, analizar y presentar datos geoespaciales. Estos sistemas integran datos geográficos con atributos específicos para crear mapas interactivos y realizar análisis espaciales. Son ampliamente utilizados en disciplinas como la geografía, la cartografía, la planificación urbana, la gestión de recursos naturales, la agricultura, la geomática y muchas otras áreas donde la ubicación geográfica es relevante. Los SIG ofrecen una variedad de funcionalidades, incluyendo la visualización de datos en mapas, la superposición de capas de información, el análisis de relaciones espaciales, la creación de modelos predictivos y la toma de decisiones basada en datos espaciales [8].

2.8 Power BI

Es una suite de herramientas de análisis de negocios desarrollada por Microsoft que permite transformar datos sin procesar en información significativa mediante visualizaciones interactivas y paneles [7].

Componentes de Power BI: Según Vásquez [24] Power BI consta de varios componentes que trabajan en conjunto:

- **Power BI Desktop**: Una aplicación de escritorio para diseñar, construir y publicar reportes.
- **Power BI Service**: Un servicio basado en la nube para compartir y colaborar en reportes y dashboards.
- **Power BI Mobile**: Aplicaciones móviles que permiten acceder a los dashboards y reportes desde dispositivos móviles.
- **Power BI Report Server**: Una solución on-premise para almacenar y gestionar reportes.

Características Principales: Según Vásquez [24] las características de Power BI son:

- **Integración de Datos**: Power BI puede conectarse a una amplia variedad de fuentes de datos, incluyendo bases de datos, servicios en la nube, archivos Excel, entre otros.
- **Transformación de Datos**: Utiliza Power Query para limpiar, transformar y preparar los datos para el análisis.
- **Visualización de Datos**: Ofrece una amplia gama de visualizaciones personalizables como gráficos, mapas y tablas.
- **DAX (Data Analysis Expressions)**: Un lenguaje de fórmulas que permite realizar cálculos avanzados y crear medidas y columnas calculadas.
- **Modelado de Datos**: Permite crear relaciones entre diferentes tablas de datos para construir un modelo de datos coherente.

2.9 Metodología CRISP-DM

Conocido como (Cross-Industry Standard Process for Data Mining) es un enfoque estándar utilizado para guiar proyectos de minería de datos. Consiste en un proceso iterativo y estructurado que abarca seis fases principales: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Cada fase tiene sus propios objetivos y actividades específicas que deben completarse antes de pasar a la siguiente etapa. La Metodología CRISP-DM proporciona un marco flexible y escalable que permite a los equipos de proyecto adaptarse a las necesidades y desafíos específicos de cada proyecto de minería de datos. Es ampliamente reconocida y utilizada en la industria debido a su enfoque práctico y su capacidad para gestionar de manera efectiva proyectos complejos de análisis de datos [25].

Para la creación del sistema multidimensional y el cuadro de mando integral, se aplicó la metodología CRISP-DM que emplea los procedimientos detallados en la **Figura 4**.

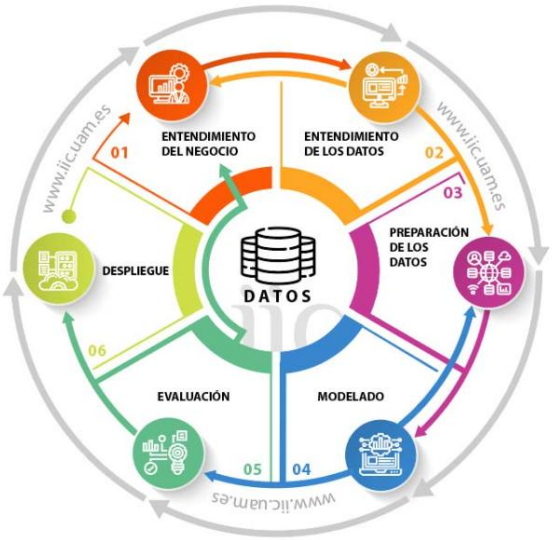


Figura 4: Metodología CRISP-DM
Fuente: [25]

Fase 1: Entendimiento del negocio (Business Understanding): En esta fase, se busca comprender los objetivos del negocio y cómo la minería de datos puede contribuir a alcanzarlos. Se establecen los criterios de éxito del proyecto y se definen los requerimientos desde una perspectiva empresarial. Es fundamental involucrar a los interesados clave y entender sus necesidades para orientar adecuadamente el proyecto de minería de datos [26]. En esta fase se debe alinear a los siguientes puntos:

- **Objetivos del Negocio**: Definir claramente los objetivos y necesidades del negocio.
- **Situación Actual**: Entender el contexto y las limitaciones del negocio [25].

Fase 2: Entendimiento de los datos (Data Understanding): En esta etapa, se recopilan los datos necesarios para el proyecto y se realiza una exploración inicial para comprender su

naturaleza y calidad. Se identifican posibles problemas o limitaciones de los datos y se determina su idoneidad para el análisis [26], esta se lleva a cabo tomando en cuenta lo siguiente:

- **Recolección de Datos Inicial:** Recopilar los datos iniciales necesarios para el proyecto.
- **Descripción de los Datos:** Explorar los datos recopilados y describir sus características.
- **Exploración de los Datos:** Realizar un análisis exploratorio para identificar patrones y relaciones.
- **Verificación de la Calidad de los Datos:** Evaluar la calidad de los datos y determinar si se requieren acciones adicionales de limpieza [25].

Fase 3: Preparación de los datos (Data Preparation): Aquí se llevan a cabo tareas de limpieza, integración y transformación de los datos para prepararlos adecuadamente para el modelado. Se eliminan valores atípicos, se resuelven los datos faltantes y se ajusta la estructura de los datos según sea necesario. Esta fase es crucial para garantizar la calidad y consistencia de los datos utilizados en el análisis [26].

Fase 4: Modelado (Modeling): En esta fase, se seleccionan y aplican técnicas de modelado para construir modelos basados en los datos preparados. Se exploran diferentes algoritmos y métodos de modelado para encontrar la mejor solución que satisfaga los objetivos del proyecto. Los modelos se evalúan y refinan iterativamente hasta alcanzar un nivel de rendimiento satisfactorio [26].

Fase 5: Evaluación (Evaluation): Aquí se evalúan los modelos generados en la fase anterior para determinar su eficacia y robustez. Se emplean métricas de desempeño y métodos de validación cruzada para evaluar la exactitud y la capacidad de generalización de los modelos. Se comparan diferentes modelos entre sí y con los criterios de éxito establecidos en la fase de comprensión del negocio [26].

Fase 6: Despliegue (Deployment): En la última fase, se implementan los resultados del proyecto en el entorno operativo del negocio. Se integran los modelos desarrollados en los sistemas existentes y se establecen procedimientos para su monitoreo y mantenimiento continuo. Se elabora un plan de acción para asegurar que los beneficios del proyecto se materialicen y se proporciona soporte para su adopción y uso efectivo en la organización [26].

Cada fase es importante y contribuye al éxito general del proyecto de minería de datos, proporcionando un marco estructurado para guiar el proceso desde la comprensión inicial del negocio hasta la implementación de soluciones prácticas y orientadas al valor.

2.9 Norma ISO/IEC 25012

También conocida como la Norma de Calidad de Datos de Software, establece un marco para evaluar la calidad de los datos en software y sistemas de información. Esta norma

proporciona un conjunto de requisitos y recomendaciones para la gestión y evaluación de la calidad de los datos, incluyendo aspectos como precisión, integridad, consistencia y accesibilidad. Su objetivo es garantizar que los datos utilizados en sistemas informáticos sean fiables, precisos y adecuados para su uso previsto [27].

Se observa la **Figura 5** donde se establece los parámetros para evaluar la calidad de los datos, abordando 15 características que pueden ser inherentes o dependientes del sistema.

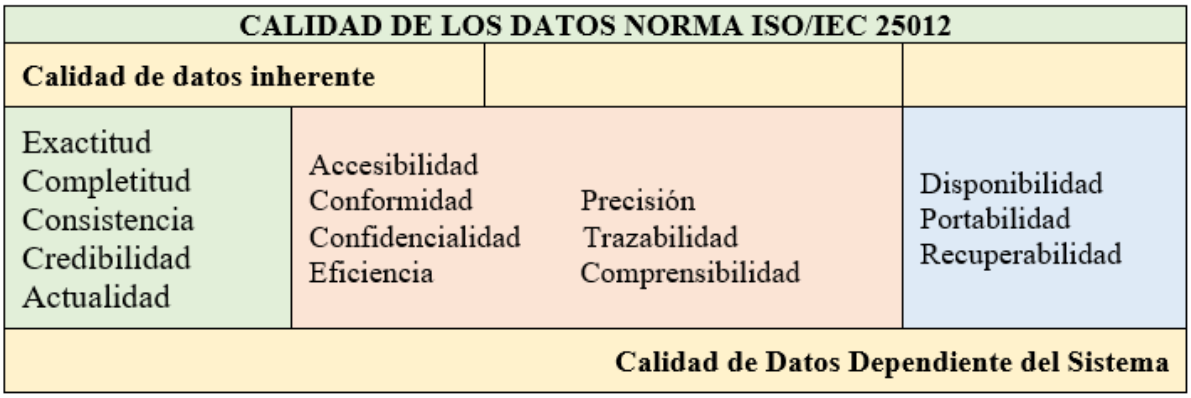


Figura 5: Características de la calidad de los datos norma ISO/IEC 25012

De los criterios establecidos, se enfocarán en la investigación los siguientes aspectos: completitud, credibilidad y precisión.

- **Completitud**

  Se refiere a la totalidad y exhaustividad de los datos disponibles en relación con el contexto específico de la información que se necesita. Una medida de completitud adecuada implica que no falten datos relevantes y que estos estén disponibles en su totalidad [27].

  En la **Tabla 2** se indica los parámetros para medir la completitud de los datos basado en la norma ISO/IEC 25012.

Tabla 2: Criterios para medir la Completitud de los datos

Tipo	**Inherente**
Documentación previa requerida	Para cada atributo a evaluar: indicar si es obligatorio.
Método para la medición	Para cada atributo, verificar que no haya datos vacíos o en blanco.
Variables	X= Porcentaje de datos del atributo que se encuentran completos.
Fórmula	Valor = x/100%
Escala	Casi ningún dato está completo (X<= 10%).
	Algunos datos están completos, pero hay muchos incompletos (X>10% y X <= 45%).

	Muchos datos están completos y hay algunos incompletos (X > 45% y X <= 85%).
	La mayoría de los datos están completos (X>85%).

Fuente: [28]

- **Credibilidad**

Se refiere a la confiabilidad y la fiabilidad de los datos, es decir, la medida en que los datos pueden ser considerados como verdaderos y precisos. Los datos creíbles son aquellos que provienen de fuentes confiables y están libres de errores o sesgos significativos [27]. Según Enríquez [29] la norma ISO/IEC para medir la credibilidad existen diferentes técnicas, como las siguientes:

- **Verificación**: se logra mediante una revisión manual de los datos o mediante el uso de herramientas especializadas.
- **Validación:** implica verificar la estructura de los datos, así como la consistencia y la integridad de los mismos.

Tabla 3. Criterios para medir la Credibilidad de los datos

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

Fuente: [27]

- **Pruebas de calidad:** se llevan a cabo pruebas de integración, aceptación o pruebas de sistema.
- **Cuestionarios o encuestas:** se utilizan estas técnicas con usuarios especializados en el dominio y manejo del sistema [27].

- **Precisión**

Se refiere a la exactitud y la consistencia de los datos en relación con el valor verdadero o esperado. Los datos precisos son aquellos que están libres de errores significativos y se acercan lo más posible a la realidad o a la verdad, permite una interpretación confiable y precisa de la información [27].

Existen diversos métodos y técnicas para obtener métricas, tales como:

- **Validación de los datos**: Incluye la verificación de la estructura y la integridad de los datos, distribuyendo los pesos de medición de la siguiente manera:

Tabla 4. Criterios para medir la Precisión de los datos

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

Fuente: [27]

- **Comparación con fuentes externas:** Se compara con fuentes externas que ofrezcan calidad y precisión para identificar posibles discrepancias.
- **Análisis estadísticos:** Se pueden calcular la media y la desviación estándar para determinar la precisión de los datos.
- **Revisión manual:** Esta técnica se puede aplicar cuando hay un usuario experto en el tema que maneja el sistema de información [27].

2.10 Talend Data Quality

Es una solución de software que ofrece una amplia gama de herramientas y funcionalidades para garantizar la calidad de los datos en toda la organización. Esta plataforma permite realizar diversas tareas relacionadas con la limpieza, enriquecimiento y estandarización de datos, así como la detección y corrección de errores. Talend Data Quality ofrece capacidades avanzadas de perfilado de datos y gestión de metadatos, facilita la identificación de problemas de calidad y la implementación de medidas correctivas. Esta herramienta es ampliamente utilizada en entornos empresariales para mejorar la precisión, integridad y fiabilidad de los datos, a su vez contribuye a la toma de decisiones más informadas y eficaces.

Es desafiante obtener datos precisos y completos de múltiples fuentes con la rapidez requerida por las empresas en la actualidad. Talend ofrece una solución integrada que abarca la integración, transformación y mapeo de datos, simplifica este proceso, según lo mencionado por Talend [30]. Se optó por esta herramienta para evaluar la completitud, credibilidad y precisión de los datos, utilizando las métricas establecidas por la norma ISO/IEC 25012, mencionadas anteriormente.

2.11 Estado del Arte

La calidad del aire y los datos meteorológicos son fundamentales para la gestión ambiental y el desarrollo urbano en Quito. Sin embargo, la falta de herramientas adecuadas para la recopilación, procesamiento y análisis de estos datos ha limitado la capacidad para realizar un análisis integral. A continuación, se presenta un estado del arte que subraya la relevancia de los sistemas multidimensionales y cuadros de mando integral.

Tabla 5. Estado del arte

Cita	Título	Resumen	Aporte
[31]	Análisis Multidimensional de Imágenes Digitales	Este artículo explora el uso de técnicas de análisis multidimensional aplicadas a imágenes digitales, destacando las metodologías y herramientas utilizadas para el procesamiento y análisis de datos visuales.	Proporciona una base metodológica para el análisis multidimensional, útil para la estructura del sistema de la investigación.
[32]	Data mining para determinar patrones del comportamiento de datos meteorológicos	El estudio aplica técnicas de minería de datos para identificar patrones en datos meteorológicos. Se enfoca en la detección de tendencias y anomalías en los registros históricos de clima.	Ofrece técnicas y métodos específicos para el análisis de datos meteorológicos, directamente aplicables a la investigación.
[33]	Análisis multidimensional y escalar del desarrollo territorial en Brasil	Analiza el desarrollo territorial en Brasil mediante enfoques multidimensionales y escalables. Utiliza diversas métricas para evaluar el progreso y los desafíos en diferentes regiones del país.	Aporta una perspectiva sobre el análisis multidimensional a nivel geográfico, relevante para el estudio de patrones en Quito.
[29]	Sistema multidimensional y cuadro de mando integral con fuentes de información del sistema materno neonatal del Ministerio	Desarrolla un sistema multidimensional y un cuadro de mando integral para gestionar y analizar datos del sistema materno neonatal, mejorando la toma de decisiones y la gestión de la información.	Proporciona un ejemplo práctico de implementación de sistemas multidimensionales y cuadros de mando integral.
[34]	Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA	Examina el uso de software libre para la explotación y análisis de datos provenientes de radares meteorológicos, destacando la eficiencia y accesibilidad de estas herramientas.	Destaca la utilización de herramientas libres, puede ser beneficioso para la accesibilidad y costo-efectividad al proyecto de investigación.
[35]	Análisis multidimensional de la segregación socioespacial en Tandil (Argentina) aplicando SIG	Investiga la segregación socioespacial en Tandil utilizando Sistemas de Información Geográfica (SIG) y técnicas de análisis multidimensional para visualizar y analizar datos espaciales.	Proporciona conocimientos sobre el uso de SIG en análisis multidimensional, útil para el aspecto geográfico del proyecto de investigación.
[36]	Prototipo móvil IoT para la predicción de la calidad del aire a través de Machine Learning	Desarrolla un prototipo de dispositivo IoT para predecir la calidad del aire utilizando algoritmos de Machine Learning, destacando la precisión y aplicabilidad en tiempo real de las predicciones.	Aporta metodologías de machine learning para la predicción de calidad del aire, directamente aplicables al estudio.
[37]	Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario	Aplica la metodología CRISP-DM para la minería de datos en un contexto universitario, detallando cada fase del proceso y los resultados obtenidos en la mejora de la gestión de datos.	Ofrece un marco claro y detallado sobre la aplicación de CRISP-DM, útil para estructurar la metodología de la investigación.

Los estudios mencionados destacan la importancia del análisis multidimensional y las técnicas de minería de datos como herramientas esenciales para el tratamiento y comprensión de grandes volúmenes de datos. La aplicación de estas metodologías no solo facilita la detección de patrones y anomalías, sino que también permite una visualización clara de la información, es fundamental en el contexto de la calidad del aire y las condiciones meteorológicas en Quito.

CAPÍTULO III. METODOLOGÍA

3.1 Metodología de investigación

En este proyecto de investigación, se adoptó un enfoque mixto que integró métodos cuantitativos y cualitativos para comprender y gestionar los datos meteorológicos y de calidad del aire en Quito. Se empleó la metodología CRISP-DM para guiar las seis etapas del proceso: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Se utilizaron herramientas como R Studio y paquetes OpenAir para realizar análisis estadísticos avanzados y generar visualizaciones significativas. En la evaluación se consideró los estándares establecidos en la norma ISO/IEC-25012 para asegurar la completitud, credibilidad y precisión de los datos utilizados.

3.2 Tipo de investigación

En esta investigación se utilizó un enfoque mixto, este enfoque fue seleccionado para obtener una comprensión más profunda del fenómeno estudiado y para enriquecer el análisis con diferentes perspectivas. En el ámbito cuantitativo, se aplicaron análisis estadísticos avanzados para visualizar los datos, analizarlos y hacer el proceso multidimensional. Por otro lado, en el enfoque cualitativo, se llevó a cabo una exhaustiva revisión de literatura que incluyó el análisis de normativas internacionales sobre calidad del aire, esta permitió realizar comparaciones entre los estándares globales y la situación local. Este enfoque dual no solo facilitó una evaluación integral de los datos, sino que también permitió contextualizar los hallazgos dentro de un marco normativo más amplio.

3.3 Diseño de la investigación

El diseño de la investigación siguió un enfoque exploratorio y descriptivo para analizar y comprender la relación entre variables meteorológicas y de calidad del aire en la ciudad de Quito.

3.4 Población y muestra

La población objetivo de este estudio estaba constituida por datos meteorológicos y de calidad del aire recopilados en las 13 estaciones de monitoreo de Quito, Ecuador, durante el período de estudio seleccionado, que abarcó desde 2005 hasta 2022. Las estaciones incluidas en el análisis fueron: Belisario Quevedo, Carapungo, Centro Histórico, Chillogallo, El Condado, Cotocollao, El Camal, Guamaní, Jipijapa, El Valle de los Chillos, San Antonio de Pichincha, Tumbaco y Turubamba. Dado que se trató de un análisis exhaustivo de los datos disponibles, la muestra incluyó todos los registros disponibles en las bases de datos de estas estaciones de monitoreo, totalizando aproximadamente 20.195.328 registros de datos a lo largo del período especificado.

3.5 Técnicas de recolección de datos

En esta investigación, se utilizó técnicas de recolección de datos que garantizaron la obtención de información relevante y confiable. La principal fuente de datos fue las

estaciones de monitoreo meteorológico y de calidad del aire ubicadas en la ciudad de Quito. Estas estaciones proporcionaron mediciones continuas de variables como temperatura, humedad, velocidad del viento y concentraciones de contaminantes atmosféricos. Los datos recopilados fueron obtenidos a través de la plataforma REMMAQ (Red de Monitoreo de la Calidad del Aire de Quito) [38], disponible en la web del municipio, asegurando la consistencia y la integridad de la información a través de protocolos estandarizados de recolección. Además, se realizó una revisión exhaustiva de la literatura científica y técnica relacionada con el tema de estudio para complementar y contextualizar los datos obtenidos de las estaciones de monitoreo.

3.6 Identificación de variables

3.6.1 Variable dependiente

Completitud, credibilidad y precisión del sistema multidimensional.

3.6.2 Variable independiente

Sistema multidimensional y cuadro de mando integral.

3.7 Operacionalización de variables

En la Tabla 2 se puede visualizar la operacionalización de variables:

Tabla 6. Operacionalización de Variables

Problema	Tema	Objetivos	Variables	Conceptualización	Dimensión	Indicadores
¿Cómo el sistema multidimen - sional y el cuadro de mando integral permitirá mejorar la completitud, credibilidad y precisión de los datos meteorológi-cos y de la calidad del aire de Quito?	Análisis multidi-mensional y cuadro de mando integral de los datos meteorológicos y de la calidad del aire de Quito.	**Objetivo General** Desarrollar un análisis Multidimensional y cuadro de mando integral de los datos meteorológicos y de la calidad del aire de Quito.	**Independiente** Sistema Multidimensio-nal y cuadro de mando integral	El sistema multidimensional y el cuadro de mando integral son herramientas de gestión que permiten analizar datos desde múltiples perspectivas y presentar información clave de manera visual y concisa. Utilizan indicadores estratégicos para evaluar el rendimiento y tomar decisiones informadas en áreas como negocios, salud, medio ambiente, entre otras. Su enfoque multidimensional facilita una comprensión profunda y una gestión eficaz de los recursos y procesos.	Análisis multidimen-sional y cuadro de mando integral.	**Independiente:** • Número módulos • Número de requisitos funcionales • Grado de integración de los datos meteorológicos y de calidad del aire en el sistema. • Facilidad de uso y accesibilidad del sistema para los usuarios.
		Objetivos específicos • Analizar sistemas multidimensionales y cuadro de mando integral. • Diseñar e implementar un modelo de análisis multidimensional y un cuadro de mando integral utilizando las herramientas R, R Studio, paquetes TimeVariation, OpenAir, Power BI y Sistemas de Información Geográfica (SIG) para la visualización y comprensión de los datos meteorológicos y de calidad del aire en Quito. • Evaluar la completitud, credibilidad y precisión de los datos meteorológicos y de la calidad del aire de Quito, siguiendo estándares de la norma ISO/IEC-25012.	**Dependiente** Evaluar la completitud, credibilidad y precisión de los datos meteorológicos y de la calidad del aire de Quito, siguiendo estándares de la norma ISO/IEC-25012.	• Completitud se refiere al grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto de uso específico. • Credibilidad se refiere al grado en el que los datos tienen atributos que se consideran ciertos y creíbles en un contexto de uso específico. • Precisión se refiere al grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico. • El análisis multidimensional y el cuadro de mando integral constituyen herramientas que permiten comprender los patrones y tendencias de la contaminación atmosférica en la ciudad.	Completitud, credibilidad y precisión	**Dependiente**: ISO/IEC 25012 • **Completitud:** 20 % de valores blancos y nulos. • **Credibilidad:** Cantidad de atributos vs cantidad de necesidades del negocio. • **Precisión:** Cantidad de datos vs cantidad de fuentes de referencia.

3.8 Método de análisis y procesamiento de datos

a) Utilización de la metodología CRISP-DM, que abarca las siguientes fases:

1. Tabular la data en un solo archivo, ordenado por fecha e identificando con claridad la estación y nombre de la variable (columnas) y hora/día (filas).
2. Identificar datos atípicos, outliers o debidos a errores de medición.
3. Establecer la cantidad de datos perdidos (NA) en la base de datos.

b) Utilización de la metodología CRISP-DM, que abarca las siguientes fases:

1. Entendimiento del negocio
2. Entendimiento de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Despliegue

c) Configuración del Entorno

Instalar R, Rstudio y los paquetes necesarios, incluyendo OpenAir.

d) Análisis de Datos

1. Realizar un análisis estadístico básico de las variables.
2. Crear gráficos utilizando las funciones de OpenAir, como SummaryPlot, CalendarPlot, PollutionRose, TrendLevel, entre otros.

e) Análisis Avanzado

1. Realizar un análisis de tendencias mediante comparación en TimeVariation.
2. Realizar un análisis básico de correlación entre variables meteorológicas (viento, temperatura) y contaminantes.

f) Visualización y Presentación

1. Utilizar tecnología SIG para ubicar geográficamente las estaciones y permitir la visualización de tendencias.
2. Desplegar los datos: resultados en el cuadro de mando integral diseñado para presentar los resultados de manera clara y accesible.

3.9 Metodología de desarrollo

En este punto se presentan las fases aplicadas en el desarrollo del sistema multidimensional y el cuadro de mando integral, siguiendo la metodología CRISP-DM. A continuación, se detallan las seis fases:

a. 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 históricos de Quito, correspondientes al periodo 2005 - 2022 mediante la creación de un sistema multidimensional y cuadro de mando integral que permitan un análisis exhaustivo y visualización de la información recopilada.

- **Observación de la situación actual**

Se dispone datos meteorológicos y de la calidad del aire de varias zonas de Quito, misma que se encuentran en formato .xlsx correspondiente al periodo 2005 – 2022, en la cual se almacena data como: Fecha, datos monóxido carbono (CO), datos dióxido de nitrógeno (NO2), datos ozono (O3), datos partículas menores a 2.5 micrómetros (PM2.5), datos partículas menores a 10 micrómetros (PM10), datos dióxido de azufre (SO2), datos dirección del viento (DIR), datos humedad relativa (HUM), datos radiación ultravioleta (IUV), datos precipitación (LLU), datos presión barométrica (PRE), datos radiación solar (RS), datos temperatura media (TMP), datos velocidad del viento (VEL) correspondientes de las siguientes zonas de Quito: El Camal, Belisario, Carapungo, Centro, Chillogallo, Condado, Cotocollao, Guamaní, Jipijapa, Los Chillos, San Antonio, Tumbaco, Turubamba, considerando que cada data depende de la zona y no todas tienen los mismos datos.

Fuente de datos: La fuente de datos es un archivo de excel .xlsx, como se puede observar en el **Anexo 1**, este archivo contiene data del periodo de 2004 – 2024, contienen 176065 filas de información en las hojas de Excel (Fecha, CO, NO2, O3, PM2.5, PM10, SO2, DIR, HUM, IUV, LLU, PRE, RS, TMP, VEL, contiene 14 filas de información en la hoja (Variables) y contiene 13 filas de información en la hoja (Parroquias). Se puede observar que la fuente de datos dependió de varias bases de datos, divididas en 14 archivos xlsx de cada variable como se puede observar en la **Figura 6,** esta misma que fue organizada en un solo archivo.

Figura 6: Bases de Datos

- **Definición de los objetivos del sistema multidimensional**

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

- Presentar un reporte por fecha.
- Presentar un reporte por parroquias.
- Presentar un reporte por variables.
- Presentar un reporte por datos meteorológicos y de la calidad del aire de Quito.

- **Realizar el plan del proyecto**

Se ha establecido un plan de proyecto que consta de seis etapas. Las que se detallan a continuación:

1. **Análisis inicial y definición de objetivos:** En esta fase, se llevó a cabo un análisis exhaustivo del negocio y se establecieron los objetivos del sistema multidimensional. Esta etapa se completó en una semana.
2. **Exploración y análisis de datos:** Durante dos semanas, se realizó una comprensión profunda y un análisis detallado de los datos relacionados con el área daos meteorológicos y de la calidad de aire de Quito.
3. **Preparación de los datos con ETL:** Se dedicaron tres semanas a la extracción, transformación y carga de datos utilizando las herramientas para preparar los datos.
4. **Diseño e implementación del modelo:** Durante otras tres semanas, se diseñó y ejecutó el modelo multidimensional.
5. **Evaluación de resultados:** Se asignaron tres semanas adicionales para evaluar los resultados obtenidos en la fase anterior y realizar ajustes según fuera necesario.

b. Fase II: Entendimiento de los datos

En esta fase, se llevó a cabo una recopilación de los datos iniciales, conocer su calidad, y entrar en contexto con la problemática que presenta la calidad del aire de Quito.

- **Obtención de los datos iniciales**

Los datos utilizados en este estudio corresponden a información histórica referentes a datos de varias zonas de Quito, que incluyen información por zonas: El Camal, Belisario, Carapungo, Centro, Chillogallo, Condado, Cotocollao, Guamaní, Jipijapa, Los Chillos, San Antonio, Tumbaco, Turubamba, cada zona contiene variables meteorológicas y de calidad del aire, dicha información fue extraída de la red de monitoreo REMMAQ de datos históricos de Quito, que

estaban almacenados en archivos xlsx, data que está relacionada directamente con el problema en cuestión, ver anexo 1.

Adicional se hizo la estructuración de la información en formato tabla para posterior hacer la exploración de los datos y el ETL, ver anexo 2.

Esta estructuración resultó en un archivo de Excel con 13 hojas, cada uno contiene 157776 datos de información, en la **Tabla 7** se observa según corresponde.

Tabla 7. Variables meteorológicas por zonas

Estación	Fecha	CO	DIR	HUM	LLU	NO2	O3	PM2.5	PM10	PRE	RS	SO2	TMP	VEL	IUV
1. Belisario Quevedo	×	×	×	×	×	×	×	×		×	×	×	×	×	
2. Carapungo	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
3. Centro Histórico	×	×	×	×	×	×	×	×		×	×	×	×	×	×
4. Chillogallo	×	×				×	×	×				×			
5. El Condado	×	×				×	×	×				×			
6. Cotocollao	×	×	×	×	×	×	×	×		×	×	×	×	×	×
7. El Camal	×	×	×	×	×	×	×	×		×	×	×	×	×	
8. Guamaní	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
9. Jipijapa	×														×
10. El Valle de los Chillos	×	×	×	×	×	×	×	×		×	×	×	×	×	
11. San Antonio de Pichincha	×	×	×	×	×		×	×	×	×	×	×	×	×	
12. Tumbaco	×		×	×	×	×	×	×	×	×			×	×	
13. Turubamba	×	×				×	×	×				×			

Nota:

X indica que la variable existe en la estación correspondiente.

Las celdas vacías indican que la variable no existe para esa estación.

- **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 fuente de datos extraída de los datos meteorológicos y de la calidad del aire de Quito.

- **Descripción General de los datos**

  1. **Dim_Fecha:** Contiene 157776 registros con fechas desde 2005 hasta 2022.
  2. **Dim_Variables:** Contiene 14 registros, cada uno representando una variable ambiental o meteorológica con su descripción, abreviatura y unidad de medida.
  3. **Dim_Parroquias:** Contiene 13 registros con información sobre las parroquias de Quito, su zona y sus coordenadas.
  4. **Dim_DatMC:** Tabla de hechos con 20195328 (Veinte millones ciento noventa y cinco mil trescientos veinte y ocho) registros que integra las dimensiones de fecha, variable y parroquia, junto con los valores medidos.

- **Análisis Descriptivo Inicial**

  El objetivo principal de esta fase fue realizar un análisis descriptivo inicial para comprender la distribución y las características básicas de los datos. Para analizar la **Distribución Temporal**, se crean gráficos de series temporales para cada variable con el fin de observar tendencias y estacionalidades. Para la **Distribución Geográfica**, se examina cómo se distribuyen los datos en las diferentes parroquias de Quito, que permiten visualizar la distribución espacial de cada variable. Además, se calculan **Estadísticas Descriptivas** (media, mediana, valores nulos, mínimo y máximo) para cada variable, generando tablas que resumen estos estadísticos descriptivos.

- **Detección de Anomalías**

  El objetivo aquí es identificar datos atípicos o anómalos que podrían afectar el análisis posterior. Para la detección de valores atípicos, se utilizan métodos estadísticos implementados a través de boxplots para identificar outliers en los datos. En cuanto a los datos faltantes, se analiza su presencia en las diferentes variables creando una matriz de valores faltantes y calculando el porcentaje de datos faltantes para cada variable.

- **Correlación de Variables**

  Esta fase tiene como objetivo evaluar las relaciones entre las diferentes variables para entender las posibles interacciones y dependencias. Se calcula una Matriz de Correlación

entre todas las variables, creando una visualización que muestra las relaciones entre ellas, facilitando la identificación de posibles correlaciones significativas.

- **Análisis Específico de Variables Clave**

El objetivo de este análisis detallado fue centrarse en las variables clave como CO, NO2, O3, PM2.5, PM10 y TMP. Para la Distribución de Variables Clave, se crean histogramas y gráficos de densidad para cada una de estas variables, permitiendo un análisis detallado de su distribución. En el Análisis de Tendencias y Estacionalidades, se evalúan las tendencias y patrones estacionales de estas variables clave, realizando análisis de series temporales que descomponen las series en componentes de tendencia, estacionalidad y ruido.

De acuerdo con la Organización Mundial de la Salud (OMS) en términos de calidad del aire se tiene la siguiente clasificación de variables meteorológicas como contaminantes ambientales, se visualiza en la **Tabla 8** y **9**:

- **Contaminantes**

Tabla 8. Clasificación de contaminantes ambientales

Contaminantes	Parámetro	Rango normal	Clasificación	Clasificación detallada
Contaminantes Gaseosos				
Monóxido de Carbono (CO)	mg/m³	0 - 10	Calidad del aire	Buena: 0-1, Moderada: 1-2, Mala: 2-10, Peligrosa: >10
Dióxido de Nitrógeno (NO2)	µg/m³	0 - 40	Calidad del aire	Buena: 0-20, Moderada: 20-40, Mala: 40-200, Peligrosa: >200
Ozono (O3)	µg/m³	0 - 100	Calidad del aire	Buena: 0-50, Moderada: 50-100, Mala: 100-240, Peligrosa: >240
Dióxido de Azufre (SO2)	µg/m³	0 - 20 (24 horas) / 0 - 500 (10 min)	Calidad del aire	Buena: 0-20, Moderada: 20-50, Mala: 50-500, Peligrosa: >500
Material Particulado				
PM2.5 (Material particulado menor a 2.5 micrómetros)	µg/m³	0 - 10 (anual) / 0 - 25 (24 horas)	Calidad del aire	Buena: 0-10, Moderada: 10-25, Mala: 25-75, Peligrosa: >75
PM10 (Material particulado menor a 10 micrómetros)	µg/m³	0 - 20 (anual) / 0 - 50 (24 horas)	Calidad del aire	Buena: 0-20, Moderada: 20-50, Mala: 50-150, Peligrosa: >150

Fuente: [39]

- **Variables Meteorológicas**

Tabla 9. Clasificación de Variables Meteorológicas.

Variable	Parámetro	Rango Normal	Clasificación	Clasificación Detallada
Condiciones Meteorológicas				
Dirección del Viento (DIR)	°	0 - 360	Meteorología	-
Humedad Relativa (HUM)	%	0 - 100	Meteorología	Baja: 0-30, Moderada: 30-60, Alta: 60-100
Radiación Ultravioleta (IUV)	IUV	0 - 11+	Meteorología	Baja: 0-2, Moderada: 3-5, Alta: 6-7, Muy Alta: 8-10, Extrema: >11
Precipitación (LLU)	mm	0 - 50+ (varía por región y clima)	Meteorología	Baja: 0-10, Moderada: 10-20, Alta: 20-50, Muy Alta: >50
Presión Barométrica (PRE)	mb	870 - 1080	Meteorología	Baja: <980, Normal: 980-1030, Alta: >1030
Radiación Solar (RS)	W/m²	0 - 1361	Meteorología	Baja: 0-200, Moderada: 200-600, Alta: 600-1000, Muy Alta: >1000
Temperatura Media (TMP)	°C	-50 - 50	Meteorología	Muy Fría: <-10, Fría: -10-0, Templada: 0-20, Caliente: 20-30, Muy Caliente: >30
Velocidad del Viento (VEL)	m/s	0 - 15	Meteorología	Baja: 0-1.5, Moderada: 1.5-3.3, Alta: 3.3-10.8, Muy Alta: >10.8

Fuente: [39]

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 4.

c. Fase III: Preparación de los datos

La preparación de los datos fue una etapa fundamental para garantizar la calidad y coherencia de la información utilizada en el análisis. En esta sección se describen las tareas de limpieza, transformación e integración de datos para corregir errores, eliminar valores atípicos y asegurar la coherencia entre las diferentes fuentes de datos. Además, se aplicaron técnicas de ingeniería de características para crear variables relevantes y significativas para el análisis.

Se puede observar en la **Figura 7**, la preparación de los datos.

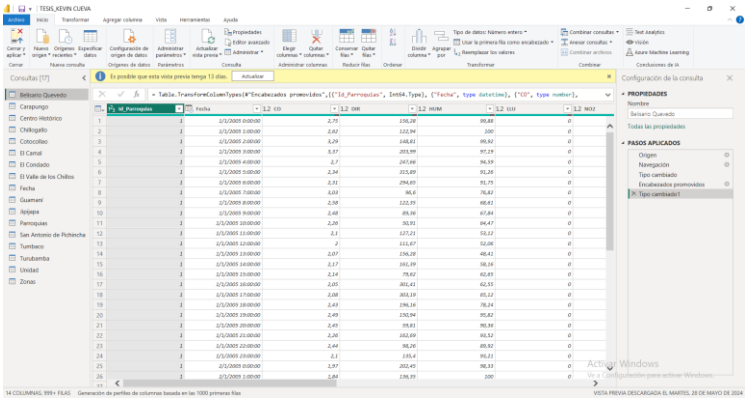


Figura 7. Preparación de los datos

En base a la data en bruto como se puede observar en la **Figura 8**, la organización de la data de manera estructurada.

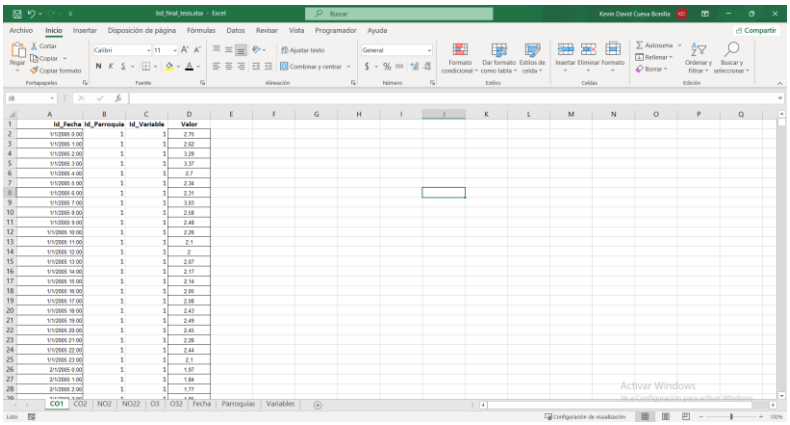


Figura 8. Data organizada

- **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 10** hasta la **Tabla 13** 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 10: Descripción de la tabla variables

Atributo	Tipo	Naturaleza
Id_Variable	Código	Original
Descripción	Independiente	Original
Abreviatura	Independiente	Original
Unidad	Independiente	Original

Tabla 11: Descripción de la tabla parroquias

Atributo	Tipo	Naturaleza
Id_Parroquia	Código	Original
ParroquiaUrbana	Independiente	Original
ZonaMetropolitana	Independiente	Original
Latitud	Independiente	Original
Longitud	Independiente	Original

Tabla 12: Descripción de la tabla fecha

Atributo	Tipo	Naturaleza
Id_Fecha	Fecha	Original

Tabla 13: Descripción de la tabla de datmc

Atributo	Tipo	Naturaleza
Id_Fecha	Código	Original
Id_Parroquia	Código	Original
Id_Variable	Código	Original
Valor	Independiente	Original

- **Limpieza de los datos**

En el proceso de limpieza de datos se realizaron varias tareas para asegurar que la información fuera precisa y libre de errores antes del análisis. Esto incluyó la eliminación de registros duplicados, la verificación de rangos y valores anómalos, y la eliminación de valores fuera del período 2005-2022 para mantener la coherencia temporal. Adicionalmente, se eliminaron valores nulos o incompletos utilizando Pentaho, asegurando la completitud de los registros en la tabla consolidada de hechos Dim_DatMC. Estas acciones permitieron mantener la integridad y precisión de los datos, cumpliendo con los criterios de completitud, credibilidad y precisión definidos por la norma ISO/IEC 25012.

Extracción

- **Recolección de datos originales**: Obtención de datos meteorológicos y de calidad del aire de varias zonas de Quito en formato XLSX correspondiente al período 2005 – 2022.

Transformación

- **Filtrado temporal**: Eliminación de registros fuera del período 2005-2022 para enfocarse en el rango de tiempo relevante.

- **Organización de datos por parroquias**: División de los datos por parroquias en 13 hojas de Excel, asegurando que cada hoja contiene datos específicos de una parroquia.
- **Creación de dimensiones**:
    - **Dim_Fecha**: 157776 registros con identificadores de fecha.
    - **Dim_Variables**: 14 registros con información sobre las variables (descripción, abreviatura, unidad).
    - **Dim_Parroquias**: 13 registros con información sobre las parroquias (nombre, zona metropolitana, latitud, longitud).
- **Consolidación de Datos**: Combinación de los datos de las 13 hojas en una sola tabla de hechos (Dim_DatMC) con 20195328 registros, identificando claramente las dimensiones de fecha, variable y parroquia, y asociando el valor correspondiente.

  En esta dimensión se procedió a usar Pentaho como se observa en la **Figura 9** para observar la data sin valores nulos.

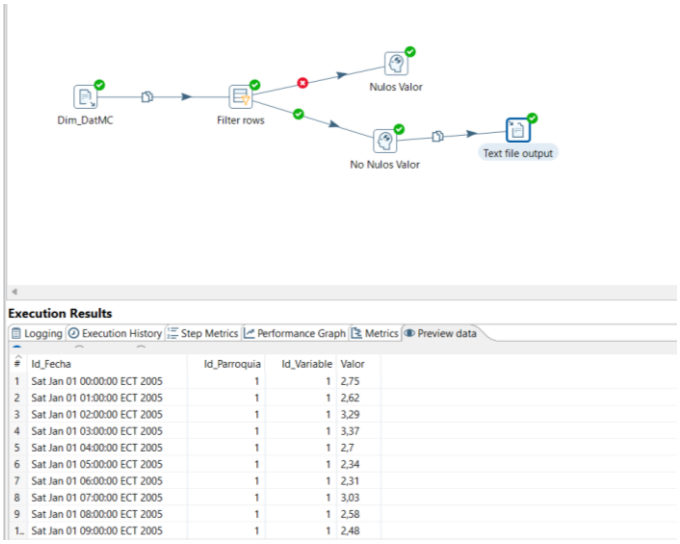


Figura 9. Diseño de limpieza de datos de la dimensión Dim_DatMCz

En la **Figura 10** se puede observar los parámetros que se usaron en el filtro.

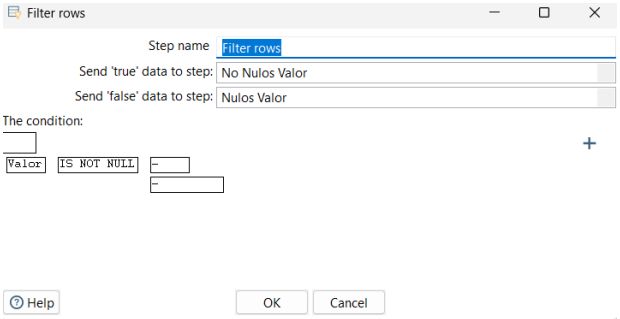


Figura 10. Filtro Dim_DatMC

Carga

- **Carga de datos transformados**: Almacenamiento de los datos transformados en una estructura adecuada para su posterior análisis y uso en el cuadro de mando integral.
- **Atributos derivados:** Para crear el sistema multidimensional tipo estrella, se agregó la tabla de hechos (Dim_DatMC, contiene todos los identificadores de las tablas correspondientes a fecha, zona geográfica y variables (Dim_Fecha, Dim_Variables, Dim_Parroquias)
- **Integrar los datos**: Al construir el modelo de datos en formato estrella, se realizó varias integraciones clave, se agregó las tablas de cada dimensión y la tabla de hechos, se observa en la **Figura 11**. Estas integraciones permiten una estructura eficiente para organizar los datos, que facilitan el análisis multidimensional y la visualización de la información. A través de este enfoque, se vinculan las distintas tablas de dimensiones con la tabla de hechos, optimizando el acceso y la consulta de los datos relacionados con las variables meteorológicas y de calidad del aire.

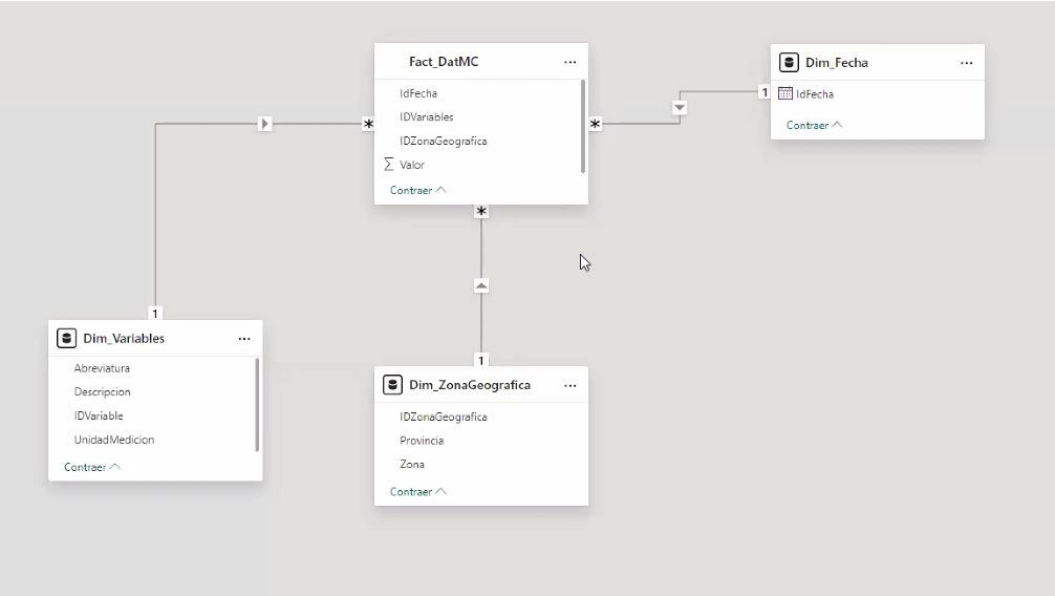


Figura 11. Modelado de datos tipo estrella

- Se relacionó la tabla Dim_Fecha que contiene la información de la fecha y hora de cada variable meteorológica con registros del período del 2005 al 2022, solo contiene su id que es tipo fecha.
- Se relacionó la tabla Dim_Variables que contiene la información de las 14 variables meteorológicas y de la calidad del aire, esta contiene el identificador, la descripción, la unidad de medida y la abreviatura de la variable meteorológica.
- Se relacionó la tabla Dim_Parroquias que contiene el identificador de las 13 parroquias, también contiene el nombre de la parroquia, la zona metropolitana que pertenece y las coordenadas en latitud y longitud para georreferenciación.

- Se relacionó la tabla de hechos Dim_DatMC que tiene relación con las tablas Dim_Fecha, Dim_Variables y Dim_Parroquias, también contiene el Valor que corresponde a las variables meteorológicas y de la calidad del aire.

Formateo de los datos: No fue necesario realizar el formateo de los datos, ya que desde su origen tenían la estructura requerida para crear el sistema multidimensional.

d. Fase IV: Modelado

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 varias herramientas como Pentaho, PowerBi, se creó el sistema multidimensional que incluyó las tablas: fecha, parroquias, variables y la tabla DatosMC (Datos meteorológicos y de la calidad del aire). Resultado de esto se obtuvo un esquema de datos tipo estrella con las siguientes tablas: las dimensiones fecha, parroquias, variables y la tabla de hechos DatosMC, estas 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 basará en el 100% de los datos. Después de realizar el proceso de Extracción, Transformación y Carga (ETL), se evaluará la credibilidad, completitud y precisión de los datos utilizando la herramienta Talend Data Quality, esta proporciona los resultados de manera automática.

Construcción del modelo

Se ejecutaron las 3 dimensiones que conforman el sistema multidimensional y la tabla de hechos utilizando la herramienta Talend Data Quality, como se muestra en el anexo 5. Se evaluó la calidad de los datos por tabla: fecha, variables, parroquias y hechos. Con estos datos se construyeron 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 herramienta Talend Data Quality, ver Anexo 5.

e. 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 Anexo 5, 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 esta fase fue el siguiente:

- Se comenzó con una matriz de Excel que contenía los datos meteorológicos y de calidad del aire para varias zonas de Quito, correspondientes al período 2005–2022. Se realizó el proceso de estructuración de la información en filas y columnas de manera tabular, dividiendo los datos en hojas específicas para cada zona y creando tablas para fechas, parroquias y variables.
- El proceso de ETL se llevó a cabo durante la organización de los datos. Se usaron herramientas como Pentaho para realizar la limpieza inicial de datos, incluyendo la identificación de valores nulos y datos inválidos. La transformación de los datos incluyó la conversión de valores nulos a 0, y la carga final se realizó en una base de datos estructurada para el análisis.
- A pesar de que Pentaho no se utilizó para eliminar valores nulos, se implementaron controles rigurosos para asegurar la integridad de la información.
- Se utilizó la herramienta Talend para verificar la calidad de los datos, evaluando aspectos como la completitud, credibilidad y precisión. Esto permitió asegurar que los datos fueran fiables y adecuados para el análisis posterior.
- Con los datos verificados, se procedió a crear dos cuadros de mando integral utilizando herramientas de visualización. Se desarrolló un reporte en PowerBI para una visualización interactiva y otro en R Studio utilizando la librería Shiny para una exploración más detallada de la información. Ambos cuadros de mando integral permitieron una representación efectiva de los datos meteorológicos y de calidad del aire, facilitando el análisis y la toma de decisiones.

Todo el proceso ejecutado hasta este punto se realizó sin inconvenientes, obteniendo resultados

f. Fase VI: Despliegue

En la fase final de la metodología CRISP-DM, se integra y consolida la información obtenida de las fases anteriores para transformarla en conocimientos útiles. Esta fase se centra en la programación y visualización de los cuadros de mando integral.

Para el despliegue en PowerBI, se siguieron los siguientes pasos:

1. **Preparación de Datos**: Los datos procesados y validados durante las fases previas fueron cargados en PowerBI. Se importaron desde las fuentes previamente estructuradas, garantizando la integridad y consistencia de la información.
2. **Diseño de Informes**: Se creó un informe interactivo que permite la visualización detallada de los datos meteorológicos y de calidad del aire. El informe incluye gráficos dinámicos, tablas y mapas que facilitan el análisis de las variables clave.
3. **Configuración de Dashboards**: Se configuró el dashboard en PowerBI para ofrecer una vista integral y consolidada de los datos, ver anexo 8.

4. **Publicación y Acceso**: Una vez diseñado y configurado el dashboard, se publicaron en PowerBI Service, permitiendo el acceso a los usuarios autorizados. La plataforma PowerBI facilita la compartición de informes y la colaboración en tiempo real. En la Figura 12, se puede observar el despliegue de los datos en la herramienta de PowerBI.

Figura 12. Dashboard desplegado en Power BI

Link de dashboard desplegado:

https://app.powerbi.com/groups/me/reports/aacdae5d-9e8e-47fe-8dca-881416946294?ctid=3d285e75-2402-401a-aa82-b00278f48a41&pbi_source=linkShare

Para el despliegue del segundo cuadro de mando integral se utilizó R Studio, un conjunto de librerías para generar el análisis y los gráficos correspondientes, y la librería Shiny como servidor para el despliegue y visualización de los gráficos en el cuadro de mando integral, para visualizar la programación ver **anexo 6**, a continuación se detalla cada paso realizado de manera generalizada.

En la **Figura 13** se muestra las librerías instaladas que se usaron en el desarrollo.

```
#PROYECTO DE INVESTIGACION
#Por: Kevin Cueva
#Tutor: Ing. Fidel Vallejo Gallardo, Ph.D.

#Cargar librerías
library(shiny)
library(shinydashboard)
library(readxl)
library(ggplot2)
library(dplyr)
library(lubridate)
library(openair)
library(leaflet)
library(gridExtra)
library(corrplot)
library(shinymanager)
```

Figura 13. Librerías instaladas en R Studio para el desarrollo del Cuadro de Mando Integral

En esta sección, se cargan las librerías necesarias para el desarrollo del cuadro de mando integral:

- **shiny** y **shinydashboard** para construir la aplicación web y su interfaz de usuario.
- **readxl** para leer archivos Excel.

- **ggplot2** y **dplyr** para análisis de datos y visualización.
- **lubridate** para manejar fechas.
- **openair** para análisis de datos ambientales.
- **leaflet** para visualización de mapas.
- **gridExtra** para combinar múltiples gráficos.
- **corrplot** para visualizar matrices de correlación.
- **shinymanager** para la autenticación de usuarios.

Luego se realiza la definición de usuarios y la carda de datos, como se muestra en la **Figura 14**.

```
# Definir usuarios para la autenticación
credentials <- data.frame(
  user = c("admin", "user"),
  password = c("admin", "user123"),
  stringsAsFactors = FALSE
)

# Cargar los datos
file_path <- "Data_Quito_Zonas.xlsx"
sheet_names <- excel_sheets(file_path)

data_list <- lapply(sheet_names, function(sheet) {
  df <- read_excel(file_path, sheet = sheet) %>%
    filter_all(any_vars(!is.na(.))) # Eliminar nulos
  df$date <- as.POSIXct(df$Fecha, format="%Y-%m-%d %H:%M:%S")
  df$year <- year(df$date)
  df$parroquia <- sheet
  return(df)
})

names(data_list) <- sheet_names
```

Figura 14. Definición de usuarios y carga de datos

En esta sección, se define una tabla de credenciales para la autenticación de usuarios. Luego, se carga la data y se procesa cada hoja en una lista de datos (data_list), convirtiendo las fechas al formato POSIXct y añadiendo información adicional como el año y la parroquia**.**

Posteriormente se definen las coordenadas de las parroquias y las clasificaciones para variables de calidad del aire (según los estándares de la OMS) y variables meteorológicas, como se observa en la **Figura 15**.

```
# Lista de coordenadas
coords <- data.frame(
  parroquia = sheet_names,
  lat = c(-0.21018967079232298, -0.0941631913233585, -0.21832583248105075,
          -0.28562187961221214, -0.10540754721213547, -0.12362342615905705,
          -0.2498508781175223, -0.3276388278339068, -0.15889892779898834,
          -0.2777388604427264, -0.29186521267391957, -0.21088541467023406,
          -0.28024998690802544),
  lon = c(-78.4428128135124, -78.4508210751852, -78.51334865638042,
          -78.56433896894426, -78.51307533949115, -78.50452321914054,
          -78.51951562015202, -78.56759102092217, -78.46782686636986,
          -78.53815966645821, -78.5668932410136, -78.39665643866752, -78.54093890391562)
)

# Estándares de la OMS
classification_oms <- data.frame(
  variable = c("CO", "NO2", "O3", "SO2", "PM2.5", "PM10"),
  good = c(1, 20, 50, 20, 10, 20),
  moderate = c(2, 40, 100, 50, 25, 50),
  bad = c(10, 200, 240, 500, 75, 150),
  dangerous = c(Inf, Inf, Inf, Inf, Inf, Inf),
  category = c("Buena", "Moderada", "Mala", "Peligrosa", "Buena", "Moderada")
)
```

Figura 15. Definición de coordenadas y clasificaciones

Posteriormente se realiza el diseño de interfaz de usuario como se observa en la **Figura 16**.

```
# Interfaz de usuario
ui <- secure_app(
    dashboardPage(
  dashboardHeader(title = "CMDAQ (Cuadro de Mando de Datos Ambientales de Quito)"),

  dashboardSidebar(
    sidebarMenu(
      menuItem("Inicio", tabName = "home", icon = icon("home")),
      menuItem("Datos", tabName = "data", icon = icon("database")),
      menuItem("Análisis", tabName = "analysis", icon = icon("chart-bar")),
      menuItem("Mapa", tabName = "map", icon = icon("map")),
      menuItem("Resumen", tabName = "summary", icon = icon("info-circle"))
    )
  ),

  dashboardBody(
    tags$head(
      tags$style(HTML("
        .content-wrapper {
          background-color: #f4f6f9;
        }
        .box-header {
          background-color: #3c8dbc;
          color: white;
        }
        .box {
          border-radius: 10px;
          margin-bottom: 20px;
        }
        .box-body {
          padding: 15px;
        }
        .form-control {
          border-radius: 5px;
        }
      "))
    ),
    tabItems(
```

Figura 16. Diseño de interfaz de usuario

En la interfaz de usuario del cuadro de mando integral se realiza:

- **dashboardHeader** define el encabezado del panel.
- **dashboardSidebar** crea un menú de navegación con secciones para "Inicio", "Datos", "Análisis", "Mapa" y "Resumen".
- **dashboardBody** contiene el contenido de cada sección en pestañas, que incluyen una bienvenida, selección de datos, análisis, un mapa interactivo y un resumen.

Seguido de la creación de la interfaz de usuario se define el servidor como se observa en la **Figura 17**.

```
# Servidor
server <- function(input, output, session) {
  res_auth <- secure_server(check_credentials = check_credentials(credentials))
  observe({
    if (!is.null(input$parroquia) && !is.null(data_list[[input$parroquia]])) {
      df <- data_list[[input$parroquia]]
      if (nrow(df) > 0) {
        start_date <- min(df$date)
        end_date <- max(df$date)
        updateDateRangeInput(session, "dateRange", start = start_date, end = end_date)
      }
    }
  })

  output$variableSelector <- renderUI({
    variables <- names(data_list[[input$parroquia]])
    variables <- variables[!(variables %in% c("date", "year", "parroquia", "Fecha"))]  # Excluir c
    selectInput("variables", "Seleccionar Variables", choices = variables, multiple = TRUE)
  })

  selected_data <- reactive({
    data_list[[input$parroquia]] %>%
      filter(date >= input$dateRange[1] & date <= input$dateRange[2])
  })

  output$timeSeriesPlot <- renderPlot({
    input$plotButton
    isolate({
      data <- selected_data()
      pollutants <- input$variables
      p <- list()
      for (pollutant in pollutants) {
        p[[pollutant]] <- ggplot(data, aes_string(x = "date", y = pollutant)) +
          geom_line() +
```

Figura 17. Definición del servidor

En el servidor:

- **secure_server** maneja la autenticación.
- **observe** actualiza el menú desplegable de variables basado en la selección de parroquia.
- **filtered_data** filtra los datos según el rango de fechas seleccionado.
- **output$variableSelector** actualiza el menú desplegable de variables basado en los datos filtrados.
- **output$timeSeriesPlot, output$distributionPlot y output$correlationPlot** generan gráficos basados en la variable seleccionada.
- **output$mapPlot** crea un mapa interactivo con marcadores para cada parroquia.
- **output$summaryText** muestra un resumen estadístico de los datos filtrados.

Finalmente, el código para ejecutar la aplicación Shiny, como se observa en la **Figura 18**.

```
# Ejecutar la aplicación
shinyApp(ui = ui, server = server)
```

Figura 18. Código para ejecutar la aplicación Shiny

Una vez completado el desarrollo del Cuadro de Mando Integral para el análisis de datos ambientales y meteorológicos en Quito, se procede a desplegar la aplicación utilizando la librería Shiny en R. En el **anexo 7**, se presenta el cuadro de mando integral diseñado, que incluye varias secciones interactivas para facilitar la exploración y el análisis de los datos.

En el **panel izquierdo** de la interfaz, se encuentra el menú de navegación principal, que permite acceder a diferentes secciones del cuadro de mando: **Inicio**, **Datos**, **Análisis**, **Mapa** y **Resumen**. Cada sección está diseñada para ofrecer funcionalidades específicas:

- **Inicio**: Proporciona una visión general y bienvenida a los usuarios, explicando el propósito y las funcionalidades del cuadro de mando integral.
- **Datos**: Permite la selección de la parroquia, el rango de fechas y la variable de interés.
- **Análisis**: Presenta varias pestañas para análisis detallados, incluyendo gráficos de **Series Temporales**, **Distribución** y **Correlaciones**. Estos gráficos permiten a los usuarios realizar un análisis exhaustivo de las tendencias y relaciones entre las variables.
- **Mapa**: Ofrece una visualización interactiva en un mapa utilizando Leaflet. Los usuarios pueden explorar las coordenadas y datos de cada parroquia en un formato geográfico, facilitando la identificación de patrones espaciales.
- **Resumen**: Proporciona un resumen estadístico de los datos filtrados, ofreciendo una visión general rápida y útil para la toma de decisiones.

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: Dim_Fecha, Dim_Parroquias, Dim_Variables, Dim_DatMC, estos criterios se observan en el **Anexo 5**, y a continuación se detalla los resultados obtenidos:

Dim_Fecha: En la data original recolectada referente a datos de varias zonas de Quito; la tabla Dim_Fecha contenía 176064 registros correspondientes al período 2004 - 2024, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros disminuyo a 157776 registros del período 2005-2022 y esta se utilizó en los campos: Id_Fecha.

La información cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión, se puede observar a continuación en la Tabla 14.

Tabla 14. Credibilidad, Completitud y Precisión de los datos de la tabla Dim_Fecha.

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

Dim_Parroquias: En la data original recolectada referente a datos de varias zonas de Quito; no contenía una tabla especificando exactamente las parroquias pero si se dividía en 14 hojas donde cada una tenía la data y sus zonas correspondientes, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros se transformó a 13 registros, esos registros se utilizó en los campos: Id_Parroquia, ParroquiaUrbana, ZonaMetropolitana, Latitud, Longitud.

La información cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión, se puede observar a continuación en la Tabla 15.

Tabla 15. Credibilidad, Completitud y Precisión de los datos de la tabla Dim_Parroquias.

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

Dim_Variables: En la data original recolectada referente a datos de varias zonas de Quito; la data contenía 176064 registros correspondientes al período 2004 – 2024, esta no contenía una tabla especificando exactamente las variables pero si se dividía en 14 hojas donde cada una tenía la data, las zonas correspondientes y cada una se identificaba por una variable meteorológica y de la calidad del aire, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros se transformó a 14 registros, se utilizó los campos: Id_Variable, Descripción, Abreviatura, Unidad.

La información cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión, se puede observar a continuación en la Tabla 16.

Tabla 16. Credibilidad, Completitud y Precisión de los datos de la tabla Dim_Variables.

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

Dim_DatMC: En la data original recolectada referente a datos de varias zonas de Quito; la data contenía 176064 registros correspondientes al período 2004 – 2024, estas tenían cada variable con sus datos respectivas divididos en 14 hojas donde cada una tenía la información de los datos meteorológicos y de la calidad del aire de Quito, juntando toda esa información se tenía 2.464.896 datos, al aplicar las reglas de ETL respectivas: eliminación de campos nulos, blancos, datos inválidos, la cantidad de registros se transformó a 20.195.328 registros, la data juntada mejora la comprensión de los datos y facilita la toma de decisiones de la información respectiva, se utilizó los campos: Id_Fecha, Id_Parroquia, Id_Variable, Valor.

Obteniendo así información que cumple con los criterios de la norma ISO/IEC 25012: completitud, credibilidad y precisión, se puede observar a continuación en la Tabla 17.

Tabla 17. Credibilidad, Completitud y Precisión de los datos de la tabla Dim_DatMC.

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

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, ver Anexo 5, los cuales son aptos para generar conocimiento para Quito, a continuación, se presenta los resultados de los cuadros de mando integral.

En la **Figura 19**, de manera generalizada para visualizar lo realizado en Power BI, donde se observa el menú general del cuadro de mando integral.

Figura 19. Cuadro de mando integral

En R Studio se obtiene lo siguiente:

En la **Figura 20** se puede observar la autenticación para ingresar al cuadro de mando integral.

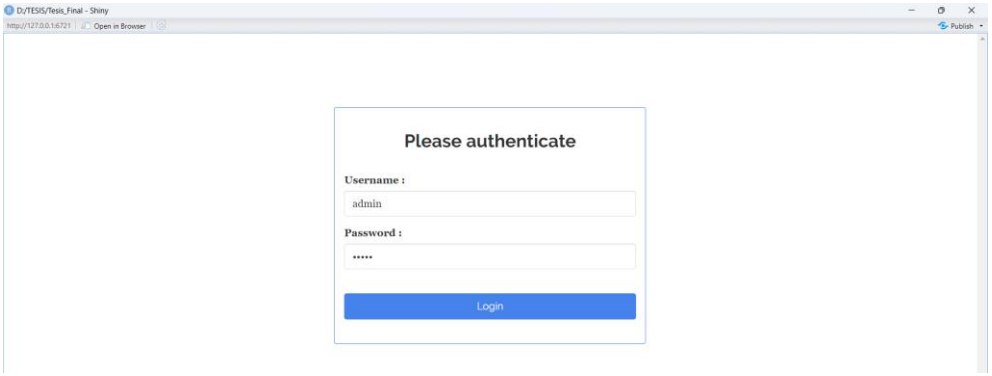


Figura 20. Autenticación para ingresar al Cuadro de mando integral

Una vez autenticado se presenta el inicio donde se puede ver un mensaje al usuario de bienvenida y las indicaciones respectivas, como se puede observar en la **Figura 21**.

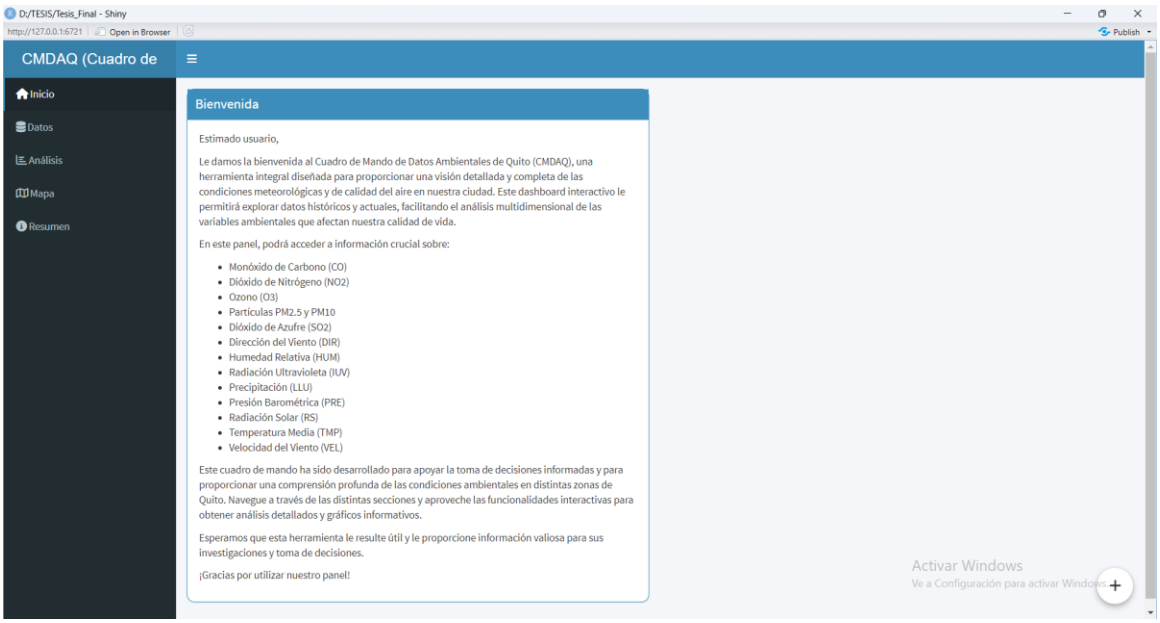


Figura 21. Inicio del cuadro de mando integral en R Studio.

Como ejemplo, se presenta los resultados del análisis realizado en R Studio sobre los datos de calidad del aire para la parroquia Belisario Quevedo. El análisis se centra en las variables CO (monóxido de carbono) y PM2.5 (partículas de menos de 2.5 micrómetros) para el período comprendido entre el 1 de enero de 2005 hasta el 31 de diciembre de 2006. A continuación se detallan los principales resultados y visualizaciones obtenidas.

1. Selección de variables y filtro de fechas

Como ejemplo de visualización, se realizó un filtrado de datos para seleccionar las variables CO y PM2.5, y se aplicó un rango de fechas que abarca desde el 1 de enero de 2005 hasta el 31 de diciembre de 2006. Este filtrado se aplicó a los datos de la parroquia Belisario

Quevedo para enfocar el análisis en el período y las variables de interés, se observa en la **Figura 22**.

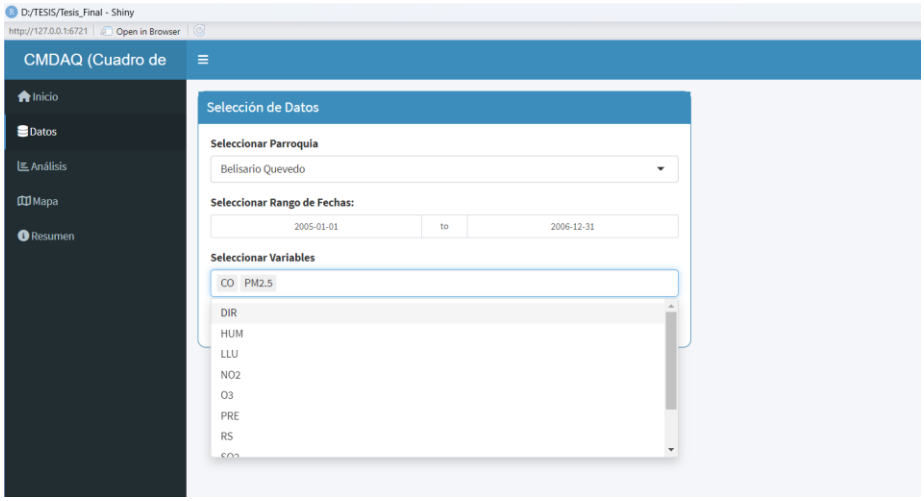


Figura 22. Sección de datos de cuadro de mando integral en R Studio

2. Visualización de datos

Las siguientes visualizaciones se generaron para ilustrar las concentraciones de CO y PM2.5 en la sección de análisis.

Tabla 18. Resultados R Studio

Gráfico de Series Temporales	Gráfico de Distribución	Gráfico de Correlación
Descripción: Se creó un gráfico de series temporales para mostrar cómo las concentraciones de CO y PM2.5 cambiaron a lo largo del tiempo. Este gráfico permitió identificar variaciones, tendencias y posibles patrones estacionales en la evolución de los contaminantes durante el periodo analizado, ver **Figura 23**.	**Descripción:** Se generó un histograma para las concentraciones de CO y PM2.5 con el propósito de observar la distribución de estos contaminantes a lo largo del periodo estudiado. El objetivo de este gráfico fue visualizar los niveles más comunes de concentración y entender la variabilidad de los datos, ver **Figura 24**.	**Descripción:** Se creó un diagrama de dispersión para analizar la relación entre las concentraciones de CO y PM2.5. El propósito de este gráfico fue observar si existía una correlación entre ambos contaminantes y explorar su comportamiento conjunto a lo largo del tiempo, ver **Figura 25**.

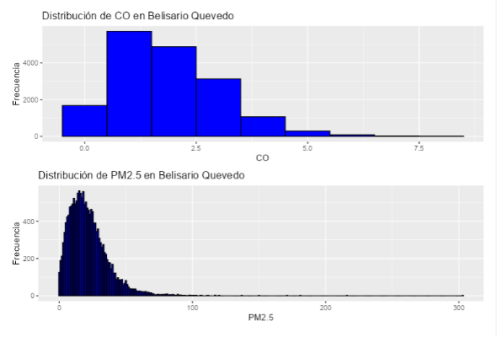


Figura 23. Gráfica de frecuencias de las variables CO y PM2.5

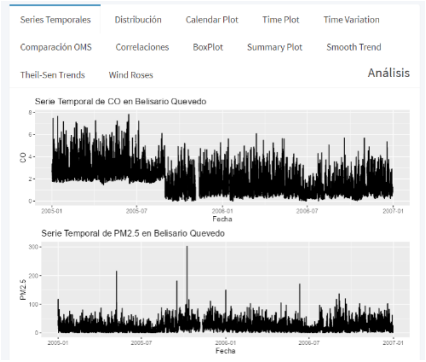


Figura 24. Gráfica de series temporales de las variables CO y PM2.5

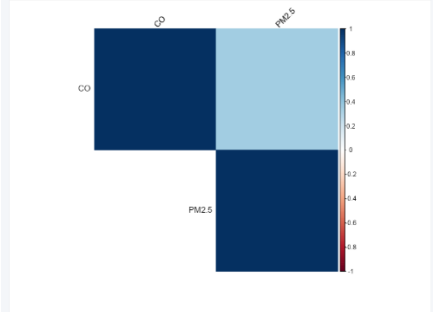


Figura 25. Gráfica de correlación de las variables CO y PM2.5

Tabla 19. Resultados R Studio

CalendarPlot	TimePlot	TimeVariation
Descripción: El CalendarPlot proporciona una vista gráfica de los datos a lo largo del tiempo en un formato de calendario. Cada día del año está representado por un cuadro, con el color indicando la magnitud de la concentración de CO y PM2.5, ver **Figura 26**.	**Descripción:** El TimePlot muestra las concentraciones de CO y PM2.5 a lo largo del tiempo en un gráfico de líneas. Cada línea representa la variación en la concentración de un contaminante específico, ver **Figura 27**.	**Descripción:** El gráfico TimeVariation muestra la variación de las concentraciones de CO y PM2.5 a lo largo del tiempo con un enfoque en la variabilidad y las tendencias estacionales, ver **Figura 28**.

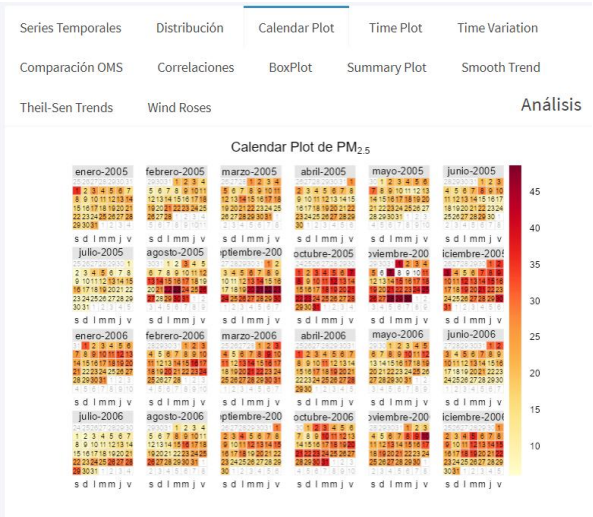


Figura 26. Gráfico CalendarPlot de las variables CO y PM2.5

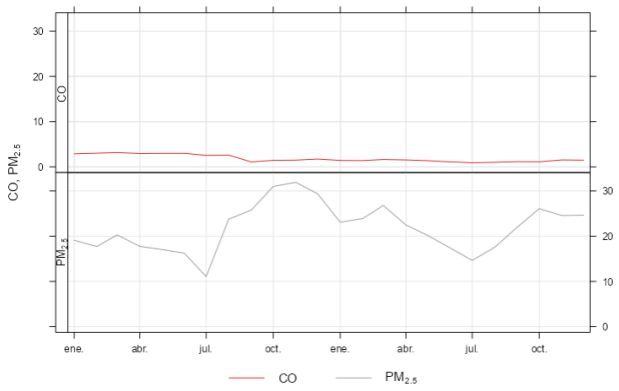


Figura 27. Gráfico TimePlot de las variables CO y PM2.5

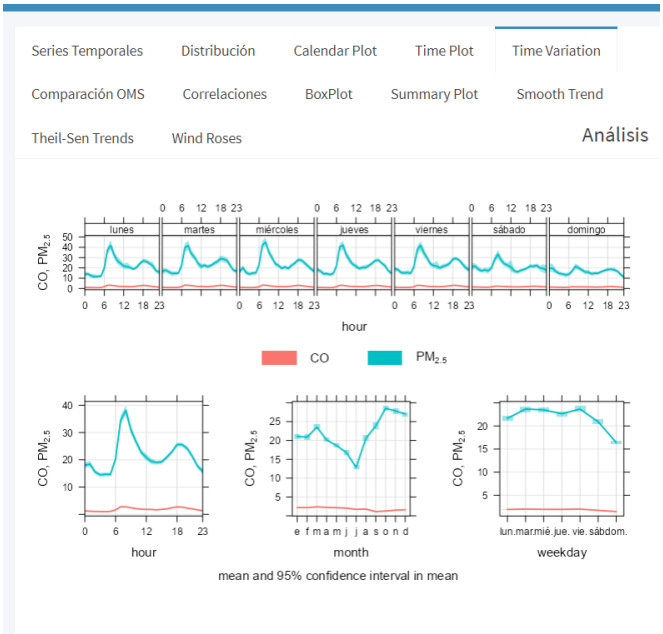


Figura 28. Gráfico TimeVariation de las variables CO y PM2.5

Tabla 20. Resultados R Studio

Comparación OMS	BoxPlot	Smooth Trend
Descripción: El gráfico Comparación OMS compara las concentraciones de CO y PM2.5 con los límites de calidad del aire establecidos por la Organización Mundial de la Salud (OMS), ver **Figura 29**.	**Descripción:** El BoxPlot presenta la distribución estadística de las concentraciones de CO y PM2.5 mediante cajas y bigotes, mostrando los cuartiles, la mediana y los valores atípicos, ver **Figura 30**.	**Descripción**: El gráfico Smooth Trend muestra la tendencia suavizada de las concentraciones de CO y PM2.5 mediante una línea de ajuste suave, ver **Figura 31**.

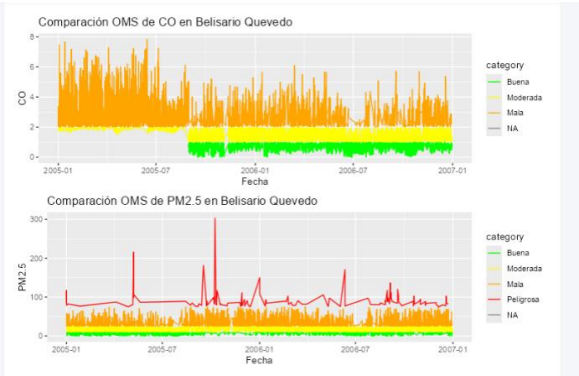


Figura 29. Gráfico de comparación OMS de las variables CO y PM2.5

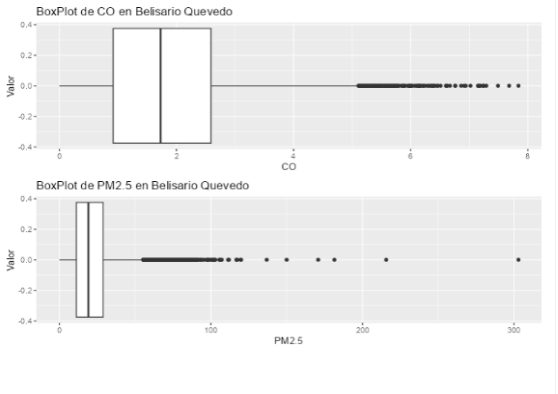


Figura 30. Gráfico BoxPlot de las variables CO y PM2.5

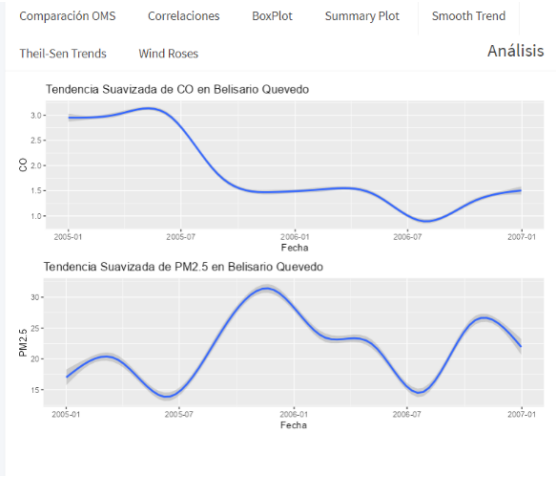


Figura 31. Gráfica de Smooth Trend de las variables CO y PM2.5

Tabla 21. Resultados R Studio

WindRose	SummaryPlot	Theil-Sen Trends
Descripción: El gráfico WindRose representa la distribución de las concentraciones de contaminantes en función de la dirección del viento, ver **Figura 32**.	**Descripción:** El SummaryPlot proporciona un resumen gráfico de las concentraciones de CO y PM2.5 en forma de gráficos combinados que pueden incluir histogramas, densidades y gráficos de dispersión, ver **Figura 33**.	**Descripción:** El gráfico Theil-Sen Trends muestra las tendencias en las concentraciones de CO y PM2.5 utilizando el estimador Theil-Sen, que es robusto frente a valores atípicos, ver **Figura 34**.

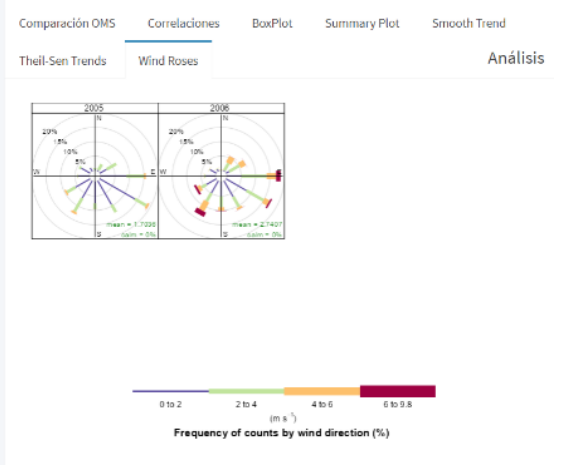


Figura 32. Gráfica WindRose de las variables CO y PM2.5

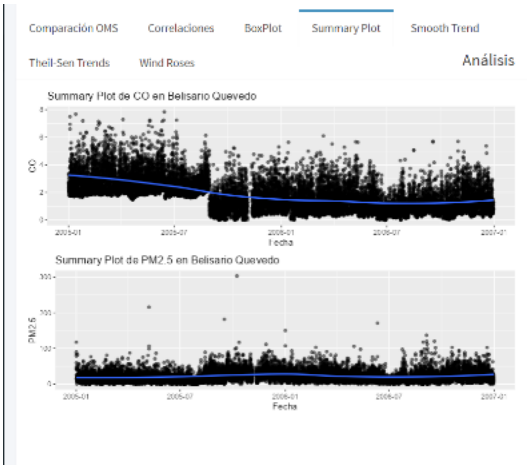


Figura 33. Gráfico Summary Plot de las variables CO y PM2.5

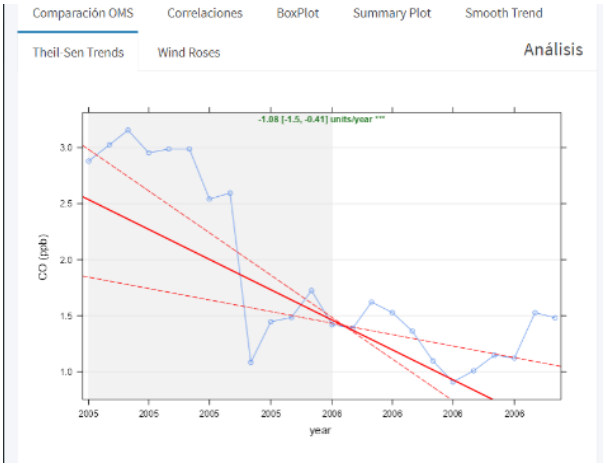


Figura 34. Gráfica Theil-Sen Tends de las variables CO y PM2.5

4.2 Discusión

El análisis de los datos meteorológicos y de calidad del aire en Quito (2005 – 2022), como ejemplo de las variables CO (Monóxido de Carbono) y PM2.5 (Partículas Menores a 2.5 micrómetro) en el periodo 2005 - 2006, reveló patrones importantes en la variabilidad de las concentraciones. Utilizando herramientas de R Studio, se generaron diversas visualizaciones, incluyendo TimePlot, TimeVariation, Comparación OMS, BoxPlot, Smooth Trend, Theil-Sen Trends, CalendarPlot, WindRose, y SummaryPlot. Estas visualizaciones mostraron variaciones estacionales claras, así como valores de contaminación que superan los límites establecidos por la OMS en ciertos periodos. La CalendarPlot y la WindRose fueron particularmente útiles para identificar días críticos de alta contaminación y patrones de dispersión de contaminantes por la dirección del viento.

La evaluación de la calidad de los datos fue realizada utilizando la herramienta Talend Data Quality, que aseguró la completitud, credibilidad y precisión de los datos conforme a la norma ISO/IEC 25012. Para el desarrollo del proyecto se utilizó la metodología CRISP-DM que según Haya [25], es la guía más utilizada en el desarrollo de proyectos de análisis de datos y minería de datos. La Smooth Trend y las Theil-Sen Trends proporcionaron una visión robusta de las tendencias a largo plazo, destacando la importancia de monitorear y analizar eventos excepcionales que podrían tener un impacto significativo en la salud pública y el medio ambiente.

El Cuadro de Mando Integral desarrollado en este estudio integra datos meteorológicos y de calidad del aire en Quito, similar al enfoque multidimensional propuesto por Gómez Díaz y Ramos [1], quienes destacaron los desafíos urbanos-industriales en Quito. Este sistema, apoyado en la metodología CRISP-DM [25], coincide con la estructura descrita por Pedrini [20], al permitir la toma de decisiones informadas y la evaluación de políticas públicas en contextos críticos.

En comparación con trabajos previos, como el de Vargas y Mesa-Fúquen [3], que exploran análisis con RStudio, y el uso de herramientas libres para meteorología como Bellini [34], este proyecto resalta la eficiencia de la implementación de software libre y de paquetes especializados como openair para realizar análisis exhaustivos. Las tendencias identificadas mediante Smooth Trend y Theil-Sen Trends se alinean con los enfoques de minería de datos propuestos por Peña [32], quienes analizaron patrones meteorológicos históricos. Sin embargo, nuestro estudio amplía este enfoque al incluir visualizaciones como CalendarPlot y WindRose, proporcionando una perspectiva única sobre patrones de contaminación y dinámica del viento en Quito.

Además, la comparación de datos espaciales en este trabajo se ve fortalecida por el uso de técnicas similares a las empleadas en [35] para analizar segregación socioespacial. En nuestro caso, estas técnicas permiten identificar no solo la distribución geográfica de los contaminantes, sino también correlaciones entre variables meteorológicas y de calidad del aire.

CAPÍTULO V. CONCLUSIONES y RECOMENDACIONES

5.1 Conclusiones

- A partir del análisis realizado sobre sistemas multidimensionales y cuadros de mando integral, se identificó que estas herramientas son fundamentales para la gestión estratégica y la toma de decisiones basadas en datos. En particular, se evidenció que, al combinar estas técnicas con herramientas de visualización como R Studio, Power BI y SIG, se logró una comprensión más profunda y detallada de fenómenos complejos como los datos meteorológicos y de calidad del aire. R Studio permitió realizar un análisis estadístico detallado, identificando patrones y correlaciones entre las variables, mientras que Power BI facilitó la creación de un dashboard interactivos que simplificaron la interpretación de grandes volúmenes de datos, permitiendo explorar la información de manera dinámica. SIG, por su parte, añadió una capa espacial a la visualización, proporcionando una comprensión geográfica precisa de cómo las condiciones meteorológicas y la calidad del aire varían en distintas zonas de Quito, lo que enriqueció la interpretación de los datos.
- Se diseñó el sistema multidimensional aplicando procesos ETL (Extracción, Transformación y Carga) a los datos provenientes de las tablas de tiempo, parroquias, variables y mediciones de calidad del aire. Para este propósito, se utilizaron herramientas como Power BI, que es especialmente efectiva para el manejo de grandes volúmenes de datos y la integración de diversas fuentes. Además, se empleó R Studio para realizar análisis estadísticos y modelado de datos, aprovechando su capacidad para manejar complejidades en los conjuntos de datos. Power BI y R Studio se utilizó para construir el cuadro de mando integral, permitiendo crear visualizaciones interactivas que facilitan la comprensión de los resultados.
- Se utilizó la herramienta Talend Data Quality para evaluar la calidad de los datos, teniendo en cuenta los criterios de completitud, credibilidad y precisión. Los resultados mostraron un 100% de datos completos, un 100% de datos creíbles y un 100% de datos precisos, lo que confirma que los datos son de alta calidad. Estos resultados están alineados con los estándares establecidos en la norma ISO/IEC-25012, que define la calidad de los datos como un conjunto de características esenciales para garantizar que la información sea útil y fiable. La aplicación de estos estándares contribuye a asegurar la fiabilidad del análisis y la toma de decisiones basadas en estos datos.

5.2 Recomendaciones

- Dado que la combinación de herramientas como Power BI y R Studio ha demostrado ser eficaz para el análisis y la visualización de datos meteorológicos y de calidad del aire, se recomienda la ampliación de la infraestructura de análisis de datos en organizaciones gubernamentales y de investigación mediante el uso de estas herramientas. Además, se sugiere explorar la integración de nuevas tecnologías como la inteligencia artificial (IA) y el aprendizaje automático (ML) para mejorar la predicción y el análisis en tiempo real de los fenómenos climáticos y ambientales.
- Como se evidenció en el diseño del sistema multidimensional, los procesos ETL (Extracción, Transformación y Carga) son clave para integrar datos de diferentes fuentes y generar información útil. Se recomienda fortalecer la automatización y la optimización de estos procesos, utilizando herramientas como Talend y otras soluciones de integración de datos, para garantizar la eficiencia y escalabilidad de los sistemas de información. Esto podría incluir la mejora de los tiempos de procesamiento de datos y la inclusión de más fuentes de datos relevantes.
- Se recomienda la implementación de procesos continuos de monitoreo y validación de calidad de datos, especialmente al incorporar nuevos datos o fuentes de información. Esto asegurará que la integridad de los análisis y las decisiones basadas en datos se mantenga a lo largo del tiempo, y se prevendrán problemas derivados de la calidad de los datos a medida que se expanda el uso del sistema multidimensional.

BIBLIOGRÁFIA

[1] M. G. Gómez Díaz y L. Ramos, «El crecimiento urbano-industrial en Quito: del neoliberalismo al socialismo del siglo XXI», *Estoa Rev. Fac. Arquit. Urban. Univ. Cuenca*, vol. 12, n.º 23, pp. 129-147, 2023.

[2] C. E. para A. L. y el Caribe, «Gran potencial para solucionar problemas ambientales». Accedido: 28 de octubre de 2024. [En línea]. Disponible en: https://www.cepal.org/es/comunicados/gran-potencial-solucionar-problemas-ambientales

[3] L. E. Vargas y E. Mesa-Fúquen, «Introducción al análisis de datos con RStudio», 2021, Accedido: 30 de abril de 2024. [En línea]. Disponible en: http://52.200.198.20/handle/123456789/141281

[4] Lopez, «timeVariation function - RDocumentation». Accedido: 8 de julio de 2024. [En línea]. Disponible en: https://www.rdocumentation.org/packages/openair/versions/0.3-9/topics/timeVariation

[5] D. C. Carslaw y J. Davison, «The openair book». Accedido: 13 de mayo de 2024. [En línea]. Disponible en: https://bookdown.org/david_carslaw/openair/

[6] RStudio, «Shiny», Shiny. Accedido: 22 de julio de 2024. [En línea]. Disponible en: https://shiny.posit.co/

[7] Microsoft, «Power BI: visualización de datos | Microsoft Power Platform». Accedido: 2 de julio de 2024. [En línea]. Disponible en: https://www.microsoft.com/es-es/power-platform/products/power-bi

[8] ESRI, «¿Qué es SIG? | SIGSA». Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://www.sigsa.info/es-mx/what-is-gis/overview

[9] Oracle, «¿Qué es un almacén de datos?» Accedido: 28 de abril de 2024. [En línea]. Disponible en: https://www.oracle.com/cl/database/what-is-a-data-warehouse/

[10] D. Noblejas, «Data Warehouse: Qué es, ventajas y objetivos», Nunsys. Accedido: 28 de abril de 2024. [En línea]. Disponible en: https://www.nunsys.com/data-warehouse/

[11] R. Kimball, «The Data Warehouse Toolkit». Accedido: 28 de abril de 2024. [En línea]. Disponible en: http://160592857366.free.fr/joe/ebooks/ShareData/The%20Data%20Warehouse%20Toolkit.pdf

[12] R. González, «El 80% de los datos que manejan las empresas son desestructurados», Big Data Magazine. Accedido: 28 de abril de 2024. [En línea]. Disponible en:

https://bigdatamagazine.es/el-80-de-los-datos-que-manejan-las-empresas-son-desestructurados

[13] J. Proaño, «Bases de datos multidimensionales. ¿Qué son? Ejemplos», Ayuda Ley Protección Datos. Accedido: 29 de abril de 2024. [En línea]. Disponible en: https://ayudaleyprotecciondatos.es/bases-de-datos/multidimensionales/

[14] D. F. Valbuena, «Esquemas en Data Warehousing», Data Management. Accedido: 29 de abril de 2024. [En línea]. Disponible en: https://datamanagement.es/2020/04/03/esquemas-data-warehousing/

[15] AWS, «¿Qué es ETL? - Explicación de extracción, transformación y carga (ETL) - AWS», Amazon Web Services, Inc. Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://aws.amazon.com/es/what-is/etl/

[16] IBM, «¿Qué es ETL (extracción, transformación, carga)? | IBM». Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://www.ibm.com/es-es/topics/etl

[17] J. Trujillo, «ETL process modeling conceptual for data warehouses: a systematic mapping study», *IEEE Lat. Am. Trans.*, vol. 9, n.º 3, pp. 358-363, 2022.

[18] S. Navarro, «¿Qué es Pentaho Data Integration? | KeepCoding Bootcamps». Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://keepcoding.io/blog/que-es-pentaho-data-integration/

[19] F. Cremona, «El CMI rueda: Un modelo alternativo desde las perspectivas de los Stakeholders», *Costos Gest.*, n.º 100, pp. 106-128, 2021.

[20] J. H. Pedrini, «Cuadro de Mando Integral (CMI): relevancia y perspectivas», *Doc. Trab. CECIN*, 2022, Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://sedici.unlp.edu.ar/handle/10915/139604

[21] J. Santaela, «¿Qué es la programación en R y cómo funciona? - Talently», Talently Blog. Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://talently.tech/blog/programacion-en-r/

[22] F. Kronthaler y S. Zöllner, *Data Analysis with RStudio: An Easygoing Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021. doi: 10.1007/978-3-662-62518-7.

[23] R. Sánchez, «Paquetes · ciencia-de-datos-con-r». Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://rsanchezs.gitbooks.io/ciencia-de-datos-con-r/content/paquetes/paquetes.html

[24] R. A. D. Vásquez, «Power bi como herramienta de apoyo a la toma de decisiones», *Univ. Soc.*, vol. 14, n.º S3, Art. n.º S3, jun. 2022.

[25] P. Haya, «La metodología CRISP-DM en ciencia de datos - IIC», Instituto de Ingeniería del Conocimiento. Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/

[26] C. Roberto, «Crisp-DM: las 6 etapas de la metodología del futuro», Blog MBA Esalq USP. Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://blog.mbauspesalq.com/es/2022/05/31/crisp-dm-las-6-etapas-de-la-metodologia-del-futuro/

[27] ISO/IEC, «ISO 25012». Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://iso25000.com/index.php/normas-iso-25000/iso-25012

[28] J. Calabrese, «Guía para evaluar calidad de datos basada en ISO/IEC 25012», en *XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019)*, 2021. Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://sedici.unlp.edu.ar/handle/10915/91086

[29] L. E. Enríquez Chamba y C. J. Moina Pinzón, «Sistema multidimensional y cuadro de mando integral con fuentes de información del sistema materno neonatal del Ministerio de Salud Pública», B.S. thesis, Riobamba, Universidad Nacional de Chimborazo, 2023. Accedido: 30 de abril de 2024. [En línea]. Disponible en: http://dspace.unach.edu.ec/handle/51000/10562

[30] Talend, «Talend Data Quality: Trusted Data for the Insights You Need», Talend - A Leader in Data Integration & Data Integrity. Accedido: 30 de abril de 2024. [En línea]. Disponible en: https://www.talend.com/products/data-quality/

[31] E. L. Portiansky, «Análisis multidimensional de imágenes digitales», *Portal Libr. Univ. Nac. Plata*, 2022, Accedido: 25 de junio de 2024. [En línea]. Disponible en: https://libros.unlp.edu.ar/index.php/unlp/catalog/book/437

[32] J. Huere-Peña, J. Gave-Chagua, R. Yaulilahua-Huacho, W. Salas-Contreras, T. Gonzales, y J. Ayuque-Rojas, *Data mining para determinar patrones del comportamiento de datos meteorológicos*. Instituto Universitario de Innovación Ciencia y Tecnología Inudi Perú, 2022. doi: 10.35622/inudi.b.066.

[33] S. [et al Schneider, «Análisis multidimensional y escalar del desarrollo territorial en Brasil», *Multidimensional and multiscalar analisis of territorial rural development in Brazil*, nov. 2010, Accedido: 25 de junio de 2024. [En línea]. Disponible en: http://repositorio.flacsoandes.edu.ec/handle/10469/2980

[34] Y. N. Bellini Saibene, M. Volpacchio, S. Banchero, y R. Mezher, *Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA*. 2014.

[35] S. Linares y D. Lan, «Análisis multidimensional de la segregación socioespacial en Tandil (Argentina) aplicando SIG», *Investig. Geográficas*, n.º 44, Art. n.º 44, dic. 2007, doi: 10.14198/INGEO2007.44.08.

[36] A. N. Santana Andrade, «Prototipo móvil IoT para la predicción de la calidad del aire a través de Machine Learning», masterThesis, 2022. Accedido: 25 de junio de 2024. [En línea]. Disponible en: http://dspace.ups.edu.ec/handle/123456789/23510

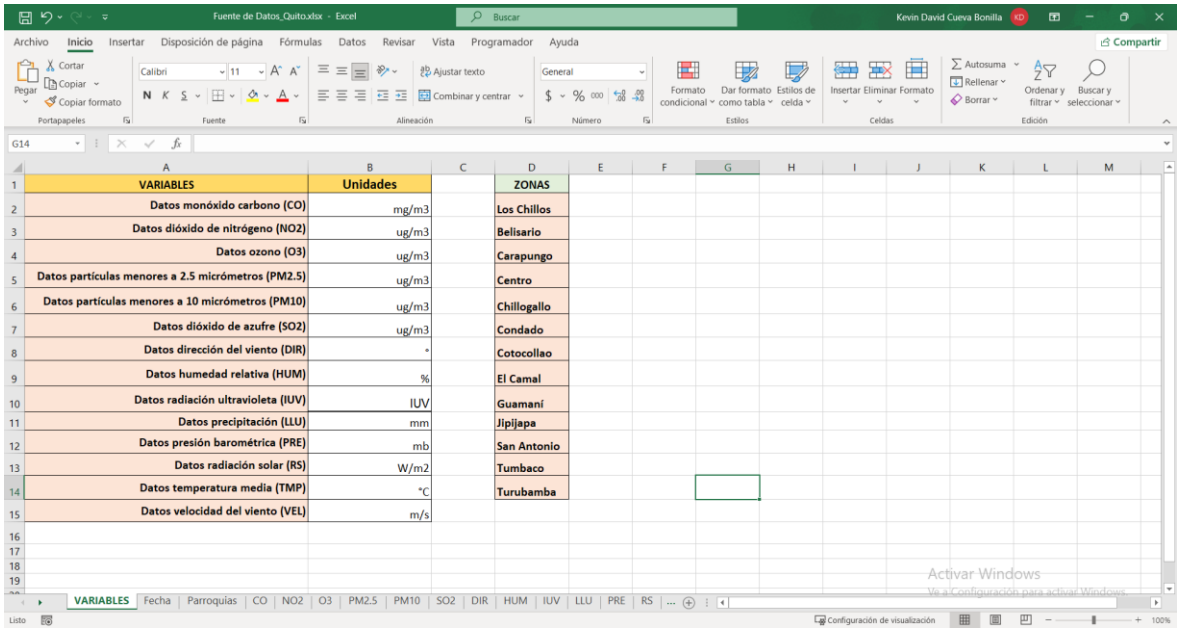
[37] V. Galán, «Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario». Accedido: 25 de junio de 2024. [En línea]. Disponible en: https://e-archivo.uc3m.es/rest/api/core/bitstreams/714c5452-962e-44cf-993f-ebb3088d4aa5/content

[38] «DATOS_HISTORICOS_REMMAQ». Accedido: 29 de octubre de 2024. [En línea]. Disponible en: https://datosambiente.quito.gob.ec/

[39] OPS/OMS, «Calidad del Aire Ambiente - OPS/OMS | Organización Panamericana de la Salud». Accedido: 23 de julio de 2024. [En línea]. Disponible en: https://www.paho.org/es/temas/calidad-aire/calidad-aire-ambiente

ANEXOS

Anexo 1. Origen de datos



VARIABLES	Unidades
Datos monóxido carbono (CO)	mg/m3
Datos dióxido de nitrógeno (NO2)	ug/m3
Datos ozono (O3)	ug/m3
Datos partículas menores a 2.5 micrómetros (PM2.5)	ug/m3
Datos partículas menores a 10 micrómetros (PM10)	ug/m3
Datos dióxido de azufre (SO2)	ug/m3
Datos dirección del viento (DIR)	°
Datos humedad relativa (HUM)	%
Datos radiación ultravioleta (IUV)	IUV
Datos precipitación (LLU)	mm
Datos presión barométrica (PRE)	mb
Datos radiación solar (RS)	W/m2
Datos temperatura media (TMP)	°C
Datos velocidad del viento (VEL)	m/s

ZONAS
Los Chillos
Belisario
Carapungo
Centro
Chillogallo
Condado
Cotocollao
El Camal
Guamaní
Jipijapa
San Antonio
Tumbaco
Turubamba

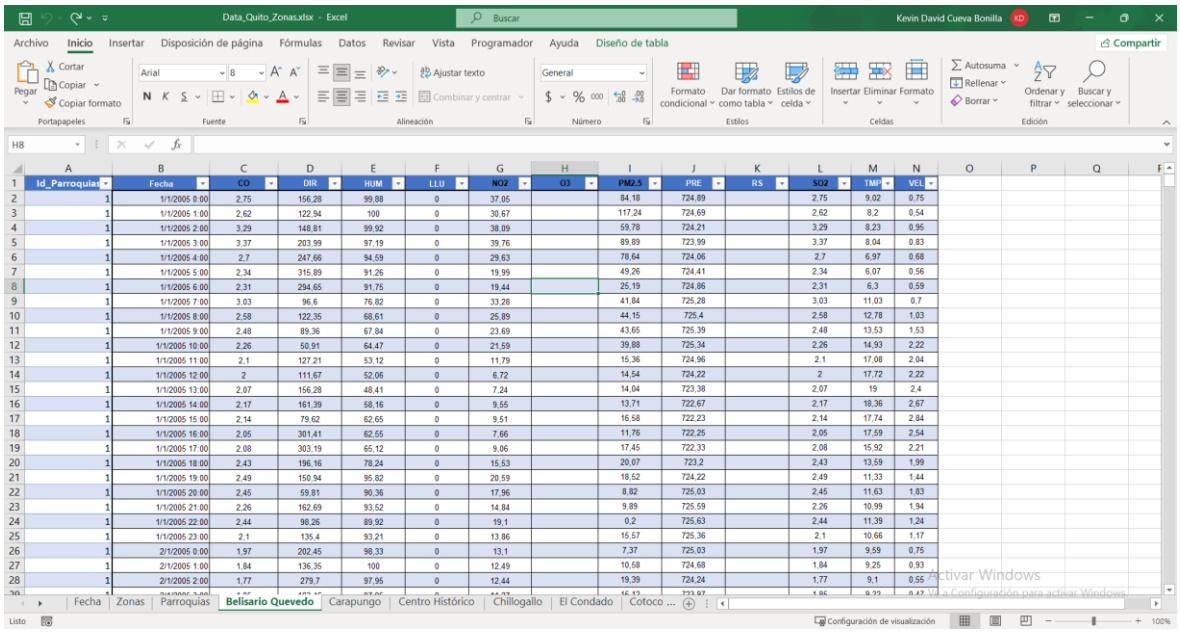
Figura 35: Data Original: Data organizada y unida de los datos meteorológicos y de la calidad del aire de Quito.

Fecha / Unidad	Belisario mg/m3	Carapungo mg/m3	Centro mg/m3	Cotocollao mg/m3	ElCamal mg/m3	Guamani mg/m3	LosChillos mg/m3	SanAntonio mg/m3	Condado mg/m3	Turubamba mg/m3	Chillogallo mg/m3
1/1/2004 0:00	7,42		4,33		3,96				3,49	4,3	
1/1/2004 1:00	7,96		4,45		3,49				3,73	4,01	
1/1/2004 2:00	8,42		3,53		1,78				5,36	3,75	
1/1/2004 3:00	9,06		3,36		1,29				4,42	3,23	
1/1/2004 4:00	6,57		2,99		0,9				3,46	3,07	
1/1/2004 5:00	5,92		2,86		0,9				2,67	3,04	
1/1/2004 6:00	5,88		2,89		0,71				2,66	3,03	
1/1/2004 7:00	5,88		2,86		0,67				2,5	3,01	
1/1/2004 8:00	5,9		2,81		0,73				2,35	2,99	
1/1/2004 9:00	5,96		2,81		0,63				2,38	2,98	
1/1/2004 10:00	6,06		2,82		0,65				2,5	2,94	
1/1/2004 11:00	6,19		2,81		0,69				2,66	2,89	
1/1/2004 12:00	6,24		2,84		0,7				2,73	2,9	

176045	31/01/2024 02:00	0,352	0,235	0,273	-0,086		0,236	0,005	0,303			
176046	31/01/2024 03:00	0,322	0,252	0,274	-0,095		0,218	0,051	0,349			
176047	31/01/2024 04:00	0,304	0,281	0,314	-0,053		0,304	0,03	0,356			
176048	31/01/2024 05:00	0,343	0,528	0,461	0,105		0,351	0,167	0,409			
176049	31/01/2024 06:00	0,817	0,93	1,339	0,624		0,582	0,527	0,787			
176050	31/01/2024 07:00	1,278	1,012	2,022	0,621		0,512	1,078	1,117			
176051	31/01/2024 08:00	1,347	0,666	1,039	0,69		0,306	0,585	0,728			
176052	31/01/2024 09:00	1,052	0,683	0,868	0,668		0,397	0,309	0,54			
176053	31/01/2024 10:00	0,983	0,601	0,733	0,316		0,437	0,167	0,43			
176054	31/01/2024 11:00	0,891	0,375	0,791	0,13		0,347	0,238	0,446			
176055	31/01/2024 12:00	0,709	0,379	0,562	0,058		0,543	0,258	0,4			
176056	31/01/2024 13:00	0,677	0,389	0,365	0,022		0,573	0,219	0,386			
176057	31/01/2024 14:00	0,625	0,342	0,416	0		0,546	0,128	0,353			
176058	31/01/2024 15:00	0,534	0,367	0,472	0,003		0,525	0,041	0,383			
176059	31/01/2024 16:00	0,51	0,361	0,547	0,032		0,583	0,132	0,373			
176060	31/01/2024 17:00	0,709	0,387	0,667	0,081		0,705	0,512	0,417			
176061	31/01/2024 18:00	0,821	0,477	0,666	0,139		0,758	0,534	0,4			
176062	31/01/2024 19:00	0,882	0,464	0,71	0,357		0,585	0,236	0,431			
176063	31/01/2024 20:00	0,805	0,436	0,736	0,373		0,599	0,165	0,354			
176064	31/01/2024 21:00	0,974	0,541	0,797	0,476		0,555	0,176	0,34			
176065	31/01/2024 22:00	0,669	0,295	0,62	0,071		0,512	0,151	0,323			
176066	31/01/2024 23:00	0,481	0,331	0,477	0,154		0,412	0,042	0,307			

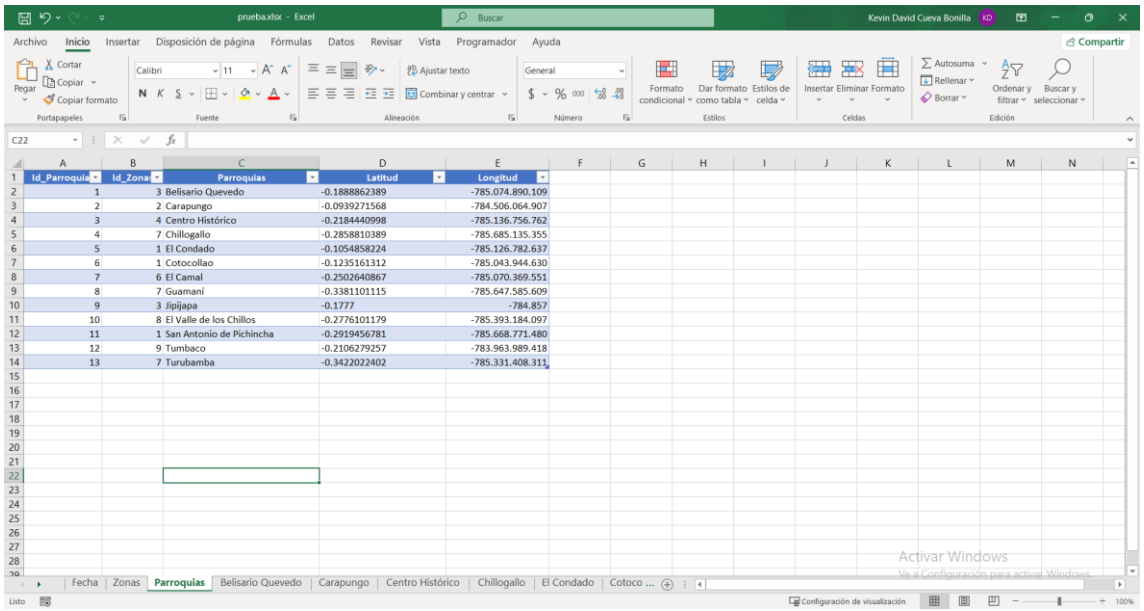
Figura 36: Data Original: Con información del período 2004 - 2024

Anexo 2. Estructuración de los datos



Id_Parroquias	Fecha	CO	DIR	HUM	LLU	NO2	O3	PM2.5	PRE	RS	SO2	TMP	VEL
1	1/1/2005 0:00	2,75	156,28	99,88	0	37,05		84,18	724,89		2,75	9,02	0,75
1	1/1/2005 1:00	2,62	122,94	100	0	30,67		117,24	724,69		2,62	8,2	0,54
1	1/1/2005 2:00	3,29	148,81	99,92	0	38,09		59,78	724,21		3,29	8,23	0,95
1	1/1/2005 3:00	3,37	203,99	97,19	0	39,76		89,89	723,99		3,37	8,04	0,83
1	1/1/2005 4:00	2,7	247,66	94,59	0	29,63		78,64	724,06		2,7	6,97	0,68
1	1/1/2005 5:00	2,34	315,89	91,26	0	19,99		49,26	724,41		2,34	6,07	0,56
1	1/1/2005 6:00	2,31	294,65	91,75	0	19,44		25,19	724,86		2,31	6,3	0,59
1	1/1/2005 7:00	3,03	96,6	76,82	0	33,28		41,84	725,28		3,03	11,03	0,7
1	1/1/2005 8:00	2,58	122,35	68,61	0	25,89		44,15	725,4		2,58	12,78	1,03
1	1/1/2005 9:00	2,48	89,36	67,84	0	23,69		43,65	725,39		2,48	13,53	1,53
1	1/1/2005 10:00	2,26	50,91	64,47	0	21,59		39,88	725,34		2,26	14,93	2,22
1	1/1/2005 11:00	2,1	127,21	53,12	0	11,79		15,36	724,96		2,1	17,08	2,04
1	1/1/2005 12:00	2	111,67	52,06	0	6,72		14,54	724,22		2	17,72	2,22
1	1/1/2005 13:00	2,07	156,28	48,41	0	7,24		14,04	723,38		2,07	19	2,4
1	1/1/2005 14:00	2,17	161,39	58,16	0	9,55		13,71	722,67		2,17	18,36	2,67
1	1/1/2005 15:00	2,14	79,62	62,65	0	9,51		16,58	722,23		2,14	17,74	2,84
1	1/1/2005 16:00	2,05	301,41	62,55	0	7,66		11,76	722,25		2,05	17,59	2,54
1	1/1/2005 17:00	2,08	303,19	65,12	0	9,06		17,45	722,33		2,08	15,92	2,21
1	1/1/2005 18:00	2,43	196,16	78,24	0	15,53		20,07	723,2		2,43	13,59	1,99
1	1/1/2005 19:00	2,49	150,94	95,82	0	20,59		18,52	724,22		2,49	11,33	1,44
1	1/1/2005 20:00	2,45	59,81	90,36	0	17,96		8,82	725,03		2,45	11,63	1,83
1	1/1/2005 21:00	2,26	162,69	93,52	0	14,84		9,89	725,59		2,26	10,99	1,94
1	1/1/2005 22:00	2,44	98,26	89,92	0	19,1		0,2	725,63		2,44	11,39	1,24
1	1/1/2005 23:00	2,1	135,4	93,21	0	13,86		15,57	725,36		2,1	10,66	1,17
1	2/1/2005 0:00	1,97	202,45	98,33	0	13,1		7,37	725,03		1,97	9,59	0,75
1	2/1/2005 1:00	1,84	136,35	100	0	12,49		10,58	724,68		1,84	9,25	0,93
1	2/1/2005 2:00	1,77	279,7	97,95	0	12,44		19,39	724,24		1,77	9,1	0,55

Figura 37: Estructuración de la información en formato tabla (Una parroquia).



Id_Parroquia	Id_Zona	Parroquias	Latitud	Longitud
1	3	Belisario Quevedo	-0.1888862389	-785.074.890.109
2	2	Carapungo	-0.0939271568	-784.506.064.907
3	4	Centro Histórico	-0.2184440998	-785.136.756.762
4	7	Chillogallo	-0.2858810389	-785.685.135.355
5	1	El Condado	-0.1054858224	-785.126.782.637
6	1	Cotocollao	-0.1235161312	-785.043.944.630
7	6	El Camal	-0.2502640867	-785.070.369.551
8	7	Guamaní	-0.3381101115	-785.647.585.609
9	3	Jipijapa	-0.1777	-784.857
10	8	El Valle de los Chillos	-0.2776101179	-785.393.184.097
11	1	San Antonio de Pichincha	-0.2919456781	-785.668.771.480
12	9	Tumbaco	-0.2106279257	-783.963.989.418
13	7	Turubamba	-0.3422022402	-785.331.408.311

Figura 38: Estructuración de la información en formato tabla (Parroquias)

Anexo 3. Descripción de los datos

Tabla 22. Descripción de la tabla Fecha.

Atributos	Tipo	Descripción
Fecha	Date	Identificador como fecha histórica del período 2005 - 2022.

Tabla 23. Descripción de la tabla Zonas

Atributos	Tipo	Descripción
Id_Zona	Integer	Identificador único del establecimiento
Zona Metropolitana	String	Nombre le la zona de Quito.

Tabla 24. Descripción de la tabla Parroquias

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Identificador único del establecimiento
Id_Zona	Integer	Llave Foránea del identificador único de la tabla Zonas.
Parroquias	String	Nombre de la parroquia de Quito.
Latitud	String	Coordenadas latitudinales de la parroquia.
Longitud	String	Coordenadas longitudinales de la parroquia.

Tabla 258. Descripción de la tabla Belisario Quevedo

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, LLU, NO2, O3, PM2.5, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 26. Descripción de la tabla Carapungo

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, LLU, NO2, O3, PM2.5, PM10, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 27. Descripción de la tabla Centro Histórico

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, IUV, LLU, NO2, O3, PM2.5, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 28. Descripción de la tabla Chillogallo

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, NO2, O3, PM2.5, SO2	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 29. Descripción de la tabla El Condado

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, NO2, O3, PM2.5, SO2	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 30. Descripción de la tabla Cotocollao

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, IUV, LLU, NO2, O3, PM2.5, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 31. Descripción de la tabla El Camal

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, LLU, NO2, O3, PM2.5, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 32. Descripción de la tabla Guamaní

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.

CO, DIR, HUM, IUV, LLU, NO2, O3, PM2.5, PM10, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 33. Descripción de la tabla Jipijapa

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
IUV	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 34. Descripción de la tabla El Valle de los Chillos

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, LLU, NO2, O3, PM2.5, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 35. Descripción de la tabla San Antonio de Pichincha

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, DIR, HUM, LLU, O3, PM2.5, PM10, PRE, RS, SO2, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 36. Descripción de la tabla Tumbaco

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
DIR, HUM, LLU, NO2, O3, PM2.5, PM10, PRE, RS, TMP, VEL	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Tabla 37. Descripción de la tabla Turubamba

Atributos	Tipo	Descripción
Id_Parroquia	Integer	Llave Foránea del identificador único de la tabla Zonas.
Fecha	Date	Llave Foránea como identificador de fecha histórica del período 2005 -2022.
CO, NO2, O3, PM2.5, SO2	Number	Valor numérico de variable meteorológica o de la calidad del aire de Quito.

Anexo 4. Verificar la calidad de los datos

Tabla 38. Análisis de calidad de los datos

Atributo	Campo	Valores Nulos	Valores blancos	Valores no válidos
Variables(14)	Id_Variable Descripción Abreviatura Unidad	0%	0%	0%
Parroquias(13)	Id_Parroquia ParroquiaUrbana ZonaMetropolitana Latitud Longitud	0%	0%	0%
Fecha(157777)	Id_Fecha	0%	0%	0%
Data(20195328)	Id_Fecha Id_Parroquia Id_Variable Valor	0%	0%	0%

Anexo 5. Modelos Generados

Dim_DatMC
From : **Local Connection** - Sample: Head

Head sample
2024-07-22 at **06:14:18** by **Kevin David Cueva**

Sample quality 0% 0% 100%

	Id_Fecha* Date	Id_Parroquia* Integer	Id_Variable* Integer	Valor* Decimal
1	1/1/2005 0:00	1	1	2,75
2	1/1/2005 1:00	1	1	2,62
3	1/1/2005 2:00	1	1	3,29
4	1/1/2005 3:00	1	1	3,37
5	1/1/2005 4:00	1	1	2,7
6	1/1/2005 5:00	1	1	2,34
7	1/1/2005 6:00	1	1	2,31
8	1/1/2005 7:00	1	1	3,03
9	1/1/2005 8:00	1	1	2,58
10	1/1/2005 9:00	1	1	2,48
11	1/1/2005 10:00	1	1	2,26
12	1/1/2005 11:00	1	1	2,1

Figura 39. Completitud, Credibilidad y Precisión de los datos de la tabla de hechos Dim_DatMC

Dim_Fecha
From : **Local Connection** - Sample: Head

Head sample
2024-07-22 at **06:40:02** by **Kevin David Cueva**

Sample quality 0% 0% 100%

	field0* Date	field1* Text
1	1/1/2005	0:00
2	1/1/2005	1:00
3	1/1/2005	2:00
4	1/1/2005	3:00
5	1/1/2005	4:00
6	1/1/2005	5:00
7	1/1/2005	6:00

Figura 40. Completitud, Credibilidad y Precisión de los datos de la tabla de Dim_Fecha

Dim_Parroquias
From : **Local Connection** - Sample: Head

Head sample
2024-07-22 at **12:00:54** by **Kevin David Cueva**

Sample quality 0% 0% 100%

	Id_Parroquia* Integer	ParroquiaUrbana* Text	ZonaMetropolitana* Text	Latitud* Text	Longitud* Text
1	1	Belisario Quevedo	Eugenio Espejo	-0.210189670792322…	-78.4428128135124
2	2	Carapungo	Calderón	-0.0941631913233585	-78.4508210751852
3	3	Centro Histórico	Manuela Sáenz	-0.218325832481050…	-78.51334865638042
4	4	Chillogallo	Quitumbe	-0.285621879612212…	-78.56433896894…
5	5	El Condado	La Delicia	-0.105407547212135…	-78.51307533949115
6	6	Cotocollao	La Delicia	-0.123623426159057…	-78.50452321914…
7	7	El Camal	Eloy Alfaro	-0.2498508781175223	-78.51951562015202
8	8	Guamani	Quitumbe	-0.3276388278339068	-78.56759102092217
9	9	Jipijapa	Eugenio Espejo	-0.158898927798988…	-78.46782686636986
10	10	El Valle de los Ch…	Los Chillos	-0.2777388604427264	-78.53815966645…
11	11	San Antonio de Pic…	La Delicia	-0.291865212673919…	-78.56689324101…
12	12	Tumbaco	Tumbaco	-0.210885414670234…	-78.39665643866752
13	13	Turubamba	Quitumbe	-0.280249986908025…	-78.54093890391562

Figura 41. Completitud, Credibilidad y Precisión de los datos de la tabla Dim_Parroquias

Dim_Variables
From : **Local Connection** - Sample: Head

Head sample
2024-07-22 at **12:00:52** by **Kevin David Cueva**

Sample quality 0% 0% 100%

	Id_Variable* Integer	Descripción* Text	Abreviatura* Text	Unidad* Text
1	1	Datos monóxido car…	CO	mg/m3
2	2	Datos dióxido de n…	NO2	ug/m3
3	3	Datos ozono	O3	ug/m3
4	4	Datos partículas m…	PM2.5	ug/m3
5	5	Datos partículas m…	PM10	ug/m3
6	6	Datos dióxido de a…	SO2	ug/m3
7	7	Datos dirección de…	DIR	°
8	8	Datos humedad rela…	HUM	%
9	9	Datos radiación ul…	IUV	IUV
10	10	Datos precipitación	LLU	mm
11	11	Datos presión baro…	PRE	mb
12	12	Datos radiación so…	RS	W/m2
13	13	Datos temperatura …	TMP	°C
14	14	Datos velocidad de…	VEL	m/s

Figura 42. Completitud, Credibilidad y Precisión de los datos de la tabla Dim_Variables

- **Resultados Talend Trust Score**

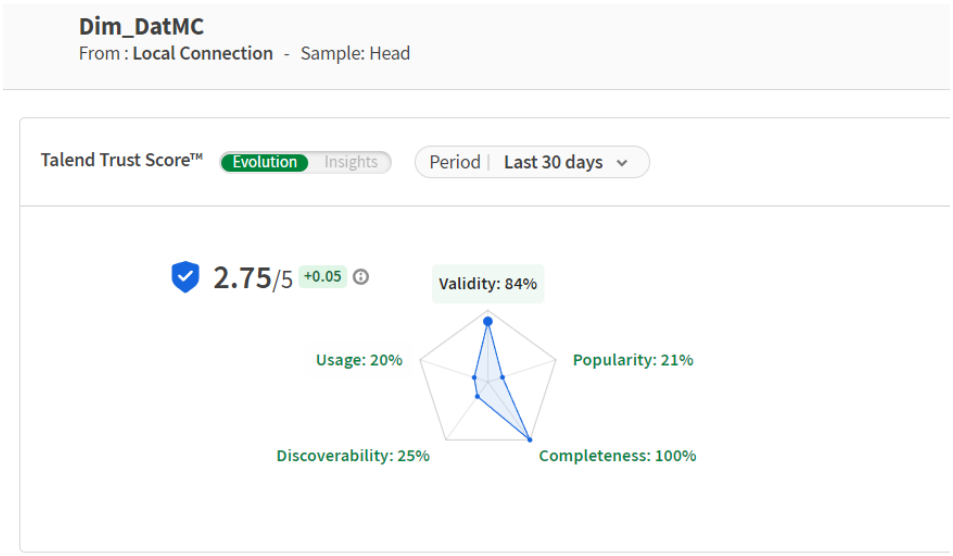


Figura 43. Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla de hechos Dim_DatMC

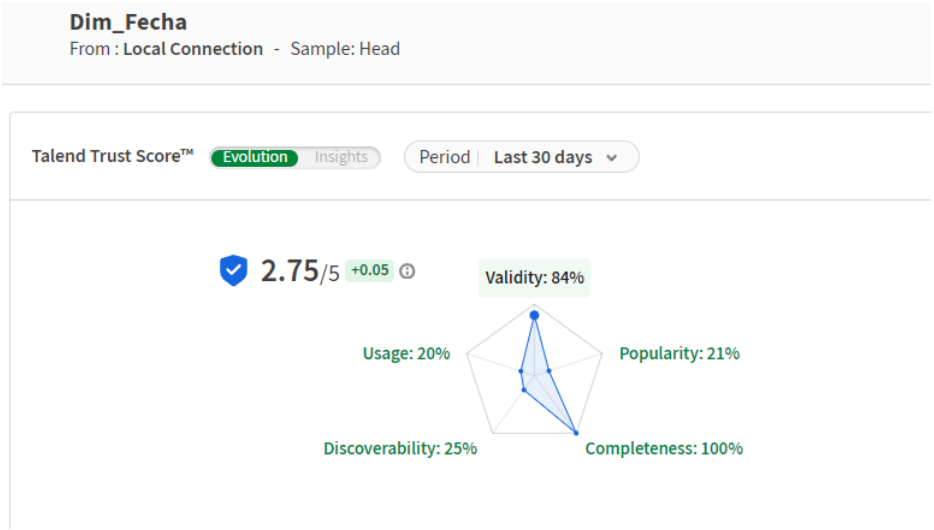


Figura 44. Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Dim_Fecha

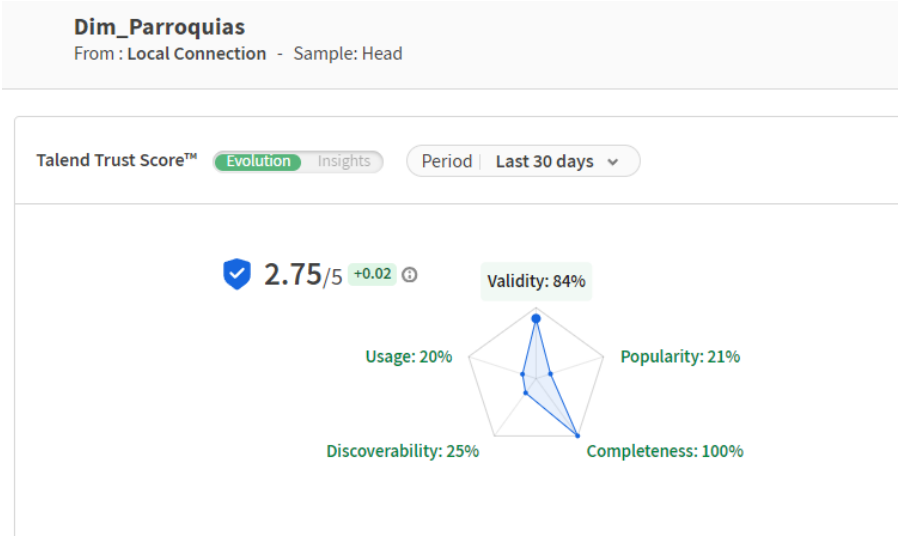


Figura 45 Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Dim_Parroquias

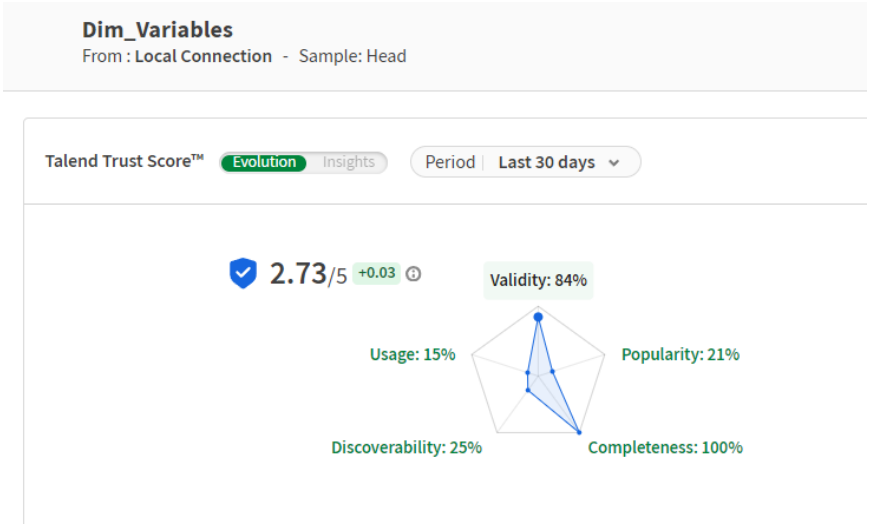


Figura 46. Resultados de completitud, Credibilidad y Precisión en la herramienta Talend de la tabla Dim_Variables

Anexo 6. Código en R Studio

```
#PROYECTO DE INVESTIGACIÓN
#Por: Kevin Cueva
#Tutor: Ing. Fidel Vallejo Gallardo, PhD.

#Cargar librerías
library(shiny)
library(shinydashboard)
library(readxl)
library(ggplot2)
```

```
library(dplyr)
library(lubridate)
library(openair)
library(leaflet)
library(gridExtra)
library(corrplot)
library(shinymanager)

# Definir usuarios para la autenticación
credentials <- data.frame(
  user = c("admin", "user"),
  password = c("admin", "user123"),
  stringsAsFactors = FALSE
)

# Cargar los datos
file_path <- "Data_Quito_Zonas.xlsx"
sheet_names <- excel_sheets(file_path)

data_list <- lapply(sheet_names, function(sheet) {
  df <- read_excel(file_path, sheet = sheet) %>%
    filter_all(any_vars(!is.na(.))) # Eliminar nulos
  df$date <- as.POSIXct(df$Fecha, format="%Y-%m-%d %H:%M:%S")
  df$year <- year(df$date)
  df$parroquia <- sheet
  return(df)
})

names(data_list) <- sheet_names

# Lista de coordenadas
```

```
coords <- data.frame(
  parroquia = sheet_names,
  lat = c(-0.21018967079232298, -0.0941631913233585, -0.21832583248105075, -0.28562187961221214, -0.10540754721213547, -0.12362342615905705, -0.2498508781175223, -0.3276388278339068, -0.15889892779898834, -0.2777388604427264, -0.29186521267391957, -0.21088541467023406, -0.28024998690802544),
  lon = c(-78.4428128135124, -78.4508210751852, -78.51334865638042, -78.56433896894426, -78.51307533949115, -78.50452321914054, -78.51951562015202, -78.56759102092217, -78.46782686636986, -78.53815966645821, -78.5668932410136, -78.39665643866752, -78.54093890391562)
)

# Estándares de la OMS
classification_oms <- data.frame(
  variable = c("CO", "NO2", "O3", "SO2", "PM2.5", "PM10"),
  good = c(1, 20, 50, 20, 10, 20),
  moderate = c(2, 40, 100, 50, 25, 50),
  bad = c(10, 200, 240, 500, 75, 150),
  dangerous = c(Inf, Inf, Inf, Inf, Inf, Inf),
  category = c("Buena", "Moderada", "Mala", "Peligrosa", "Buena", "Moderada")
)

# Rangos de Clasificación Meteorológica
classification_meteorology <- data.frame(
  variable = c("HUM", "IUV", "LLU", "PRE", "RS", "TMP", "VEL"),
  low = c(0, 0, 0, -Inf, 0, -Inf, 0),
  extreme = c(Inf, 11, Inf, Inf, Inf, 30, Inf),
  category = c("Baja", "Moderada", "Alta", "Normal", "Moderada", "Alta", "Baja")
)

# Interfaz de usuario
ui <- secure_app(
```

dashboardPage(

dashboardHeader(title = "CMDAQ (Cuadro de Mando de Datos Ambientales de Quito)")

……

Anexo 7. Cuadro de Mando integral en R Studio

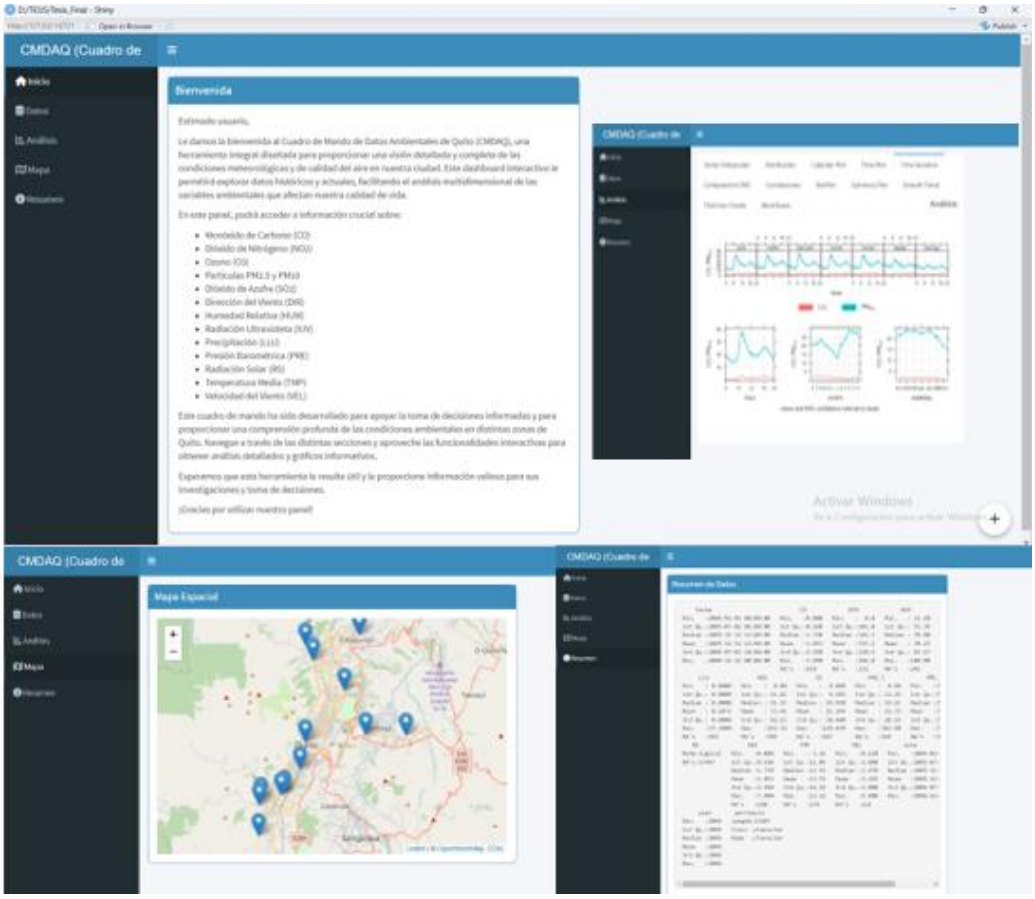


Figura 47. Cuadro de Mando Integral en R Studio

Anexo 8. Cuadro de Mando integral en Power BI

Figura 48. Cuadro de Mando Integral en Power BI – Pantalla 1

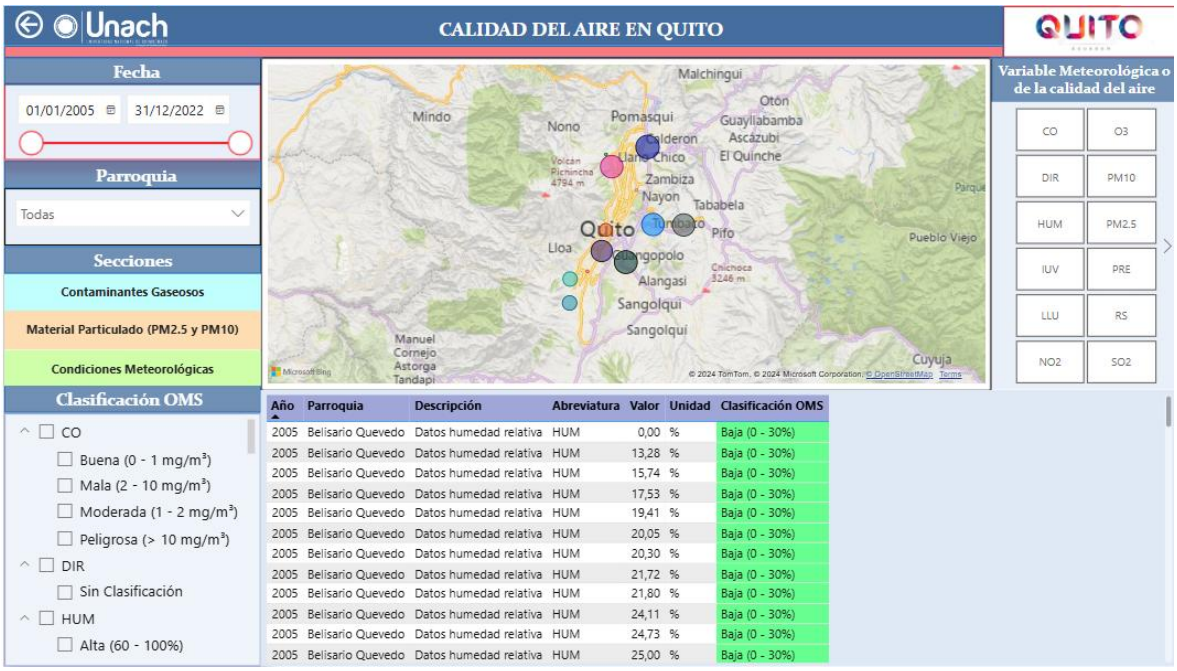


Figura 49. Cuadro de Mando Integral en Power BI – Pantalla 2

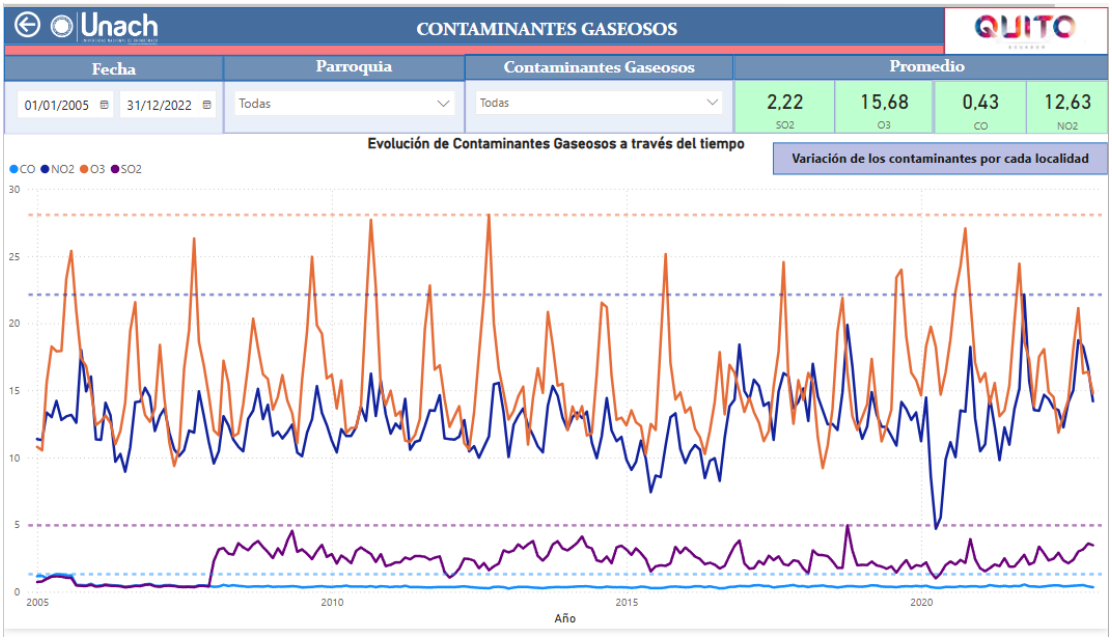


Figura 50. Cuadro de Mando Integral en Power BI – Pantalla 3

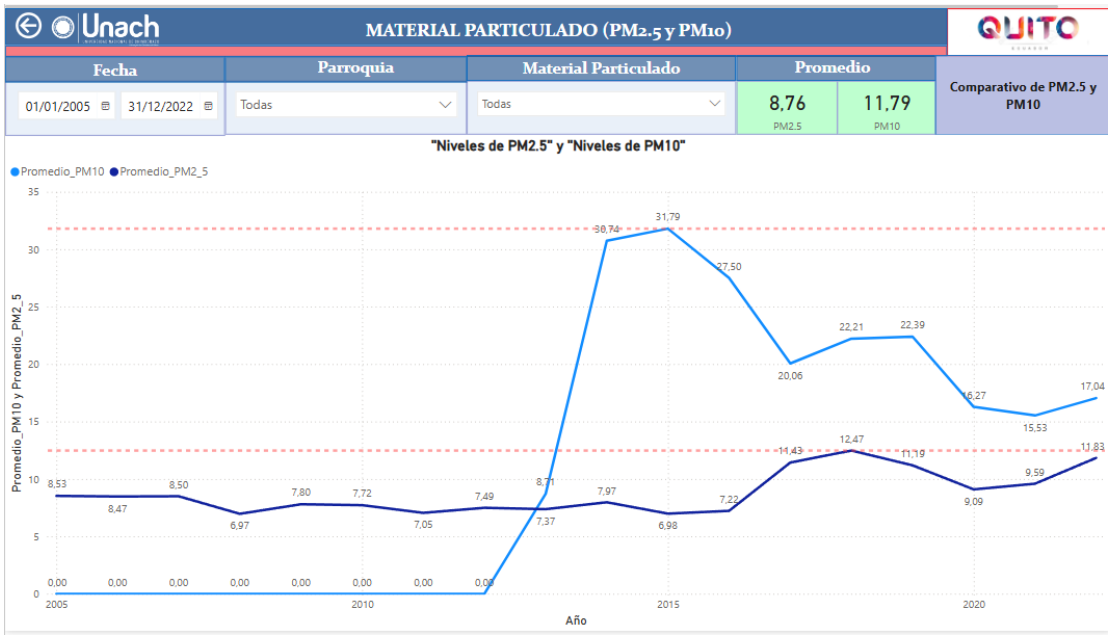


Figura 51. Cuadro de Mando Integral en Power BI – Pantalla 4

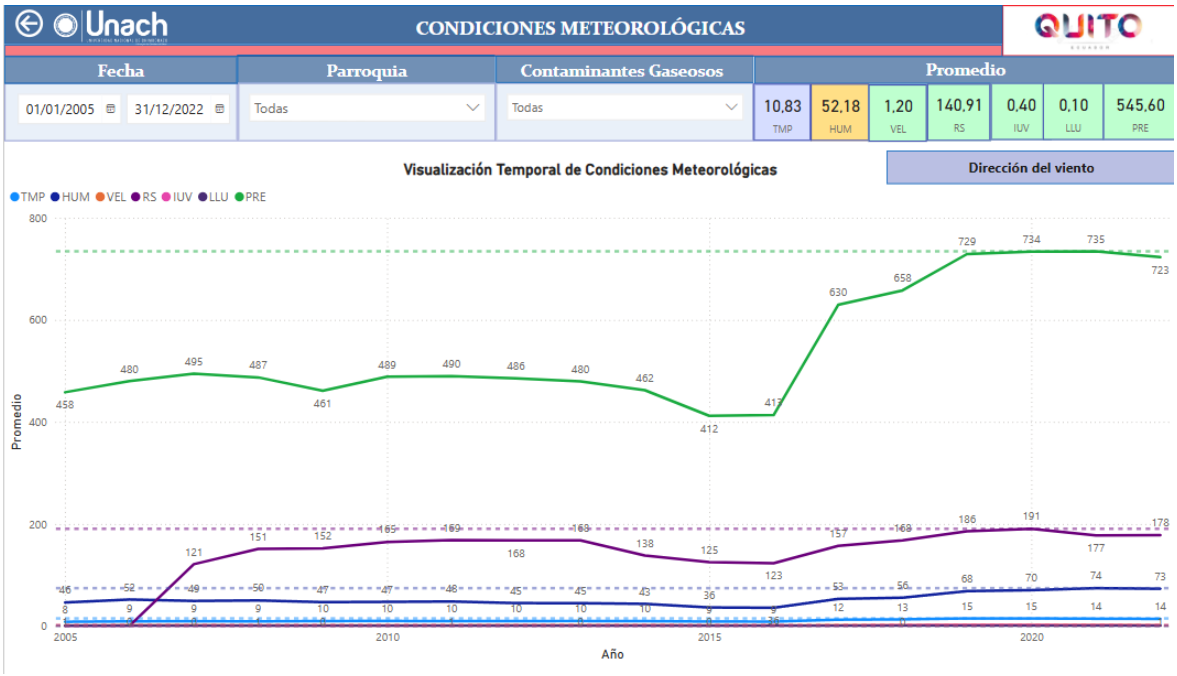


Figura 52. Cuadro de Mando Integral en Power BI – Pantalla 5

Link de grabación de uso del reporte publicado en Power BI: https://unachedu-my.sharepoint.com/:v:/g/personal/kevin_cueva_unach_edu_ec/EfVyQRC9D2NAiAfpExc03boBy1ftnEMj6TFAQQDvTB4u9A?nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOiJPbmVEcml2ZUZvckJ1c2luZXNzIiwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldlYiIsInJlZmVycmFsTW9kZSI6InZpZXciLCJyZWZlcnJhbFZpZXciOiJNeUZpbGVzTGlua0NvcHkifX0&e=Oh7st9