ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE SISTEMAS

UNIDAD DE TITULACIÓN

DISEÑO E IMPLEMENTACIÓN DEL DATA WAREHOUSE DW Y TABLEROS DE INTELIGENCIA DE NEGOCIO BI, PARA ANÁLISIS DE INDICADORES DE RENDIMIENTO EN EL ÁREA DE DESARROLLO Y VENTAS DE UNA EMPRESA QUE BRINDA SERVICIOS TECNOLÓGICOS DE SOFTWARE PARA OPERADORES DE TURISMO

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL GRADO DE MAGISTER EN SISTEMAS DE INFORMACIÓN MENCIÓN INTELIGENCIA DE NEGOCIOS Y ANALÍTICA DE DATOS MASIVOS

FREDDY MANUEL PANCHI GUAMANGALLO
freddy.panchi@epn.edu.ec

Director: SANG GUUN YOO, PHD
sang.yoo@epn.edu.ec

2023

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación “DISEÑO E IMPLEMENTACIÓN DEL DATA WAREHOUSE DW Y TABLEROS DE INTELIGENCIA DE NEGOCIO BI, PARA ANÁLISIS DE INDICADORES DE RENDIMIENTO EN EL ÁREA DE DESARROLLO Y VENTAS DE UNA EMPRESA QUE BRINDA SERVICIOS TECNOLÓGICOS DE SOFTWARE PARA OPERADORES DE TURISMO” desarrollado por Freddy Manuel Panchi Guamangallo, estudiante de la Maestría en Sistemas de Información, Mención en INTELIGENCIA DE NEGOCIOS Y ANALÍTICA DE DATOS MASIVOS, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.

Sang Guun Yoo, PhD

DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, Freddy Manuel Panchi Guamangallo, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Freddy Manuel Panchi Guamangallo

DEDICATORIA

El esfuerzo y trabajo que implicó la realización de este trabajo de titulación se lo dedico a Dios, a mi esposa y a mi familia por ser la fuente de mi alegría, motivación y fortaleza.

Freddy Panchi

AGRADECIMIENTO

A Dios por el regalo de la vida, la salud y la fuerza para salir adelante.

A la Escuela Politécnica Nacional y a sus excelentes profesores que con sus amplios conocimientos me han ayudado a crecer profesionalmente y como persona.

A TourOpp Go por brindarme la oportunidad de realizar este proyecto.

A Sang Guun Yoo, PhD por su ayuda, guía, apoyo y retroalimentación para poder culminar satisfactoriamente este proyecto.

A mi esposa que es mi inspiración y fortaleza en cada paso de mi vida.

A mi familia que siempre me apoya y motiva para alcanzar nuevas metas.

Freddy Panchi

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

RESUMEN

Este proyecto de titulación tiene como objetivo aplicar inteligencia de negocios en una empresa tecnológica que brinda servicios de software a operadores de turismo. El proyecto se enfoca en responder a las preguntas más importantes que tienen las áreas de desarrollo de software, ventas y marketing, así como analizar también a los usuarios, clientes finales y el chatbot de la empresa. Para esto se abordan todos los puntos en la implementación de herramientas de inteligencia de negocios como son: la selección, el análisis e implementación de la metodología más adecuada, el análisis de las herramientas de software Open Source más convenientes para llevar a cabo el proyecto, la creación de los Data Mart para cada área de estudio, la aplicación de modelos de inteligencia artificial para generar conocimiento a partir de los datos de la empresa y por su puesto la visualización de la información en un tableros que buscan responder a las diferentes preguntas del negocio. Finalmente, los resultados de este proyecto demostraron que se generó una herramienta ventajosa para la empresa y que aporta mucho valor al momento de tomar decisiones inteligentes.

Palabras clave: BI, Hefestos, Inteligencia Artificial, Dashboard, Visualizacion, Turismo.

ABSTRACT

This degree project aims to apply business intelligence in a technology company that provides software services to tour operators. The project focuses on answering the most important questions that software development, sales and marketing areas have, as well as analyzing users, end customers and the company's chatbot. For this, all the points in the implementation of tools for business intelligence are addressed, such as: the selection, analysis and implementation of the most appropriate methodology, the analysis of the most convenient Open Source software tools to carry out the project, the creation of the Data Marts for each area of study, the application of artificial intelligence models to generate knowledge from the company's data and, of course, the visualization of the information in a dashboard that seeks to answer the different questions of the business. Finally, the results of this project showed that an advantageous tool for the company was generated and adds a lot of value when making intelligent decisions.

Keywords: BI, Hefestos, Artificial Intelligence, Dashboard, Visualization, Tourism.

1. INTRODUCCIÓN

La toma de decisiones es un componente central que afecta directamente a las organizaciones ya que decisiones acertadas ayudan a que el negocio crezca rápidamente, mientras que decisiones equivocadas lo llevan a derrumbarse. La incorporación de tecnologías de información y comunicación ha promovido que los negocios recopilen mejor su información desde diferentes ejes de acción como son el desarrollo de software, ventas, marketing, entre otros. Las herramientas instaladas actualmente en la mayoría de las empresas proveen de información básica y tienen restricciones para analizar el comportamiento del negocio, por lo tanto, no brindan la seguridad necesaria para la toma de decisiones por eso es de suma importancia que las herramientas de Business Intelligence (BI) que se utilicen tengan un alto nivel de rendimiento al momento de medir y monitorear el crecimiento del negocio para poder definir si éste se encuentra en el camino correcto.

En este contexto, las empresas que desean prevalecer sobre las demás deben implementar herramientas de análisis de datos, inteligencia empresarial, minería de datos y técnicas de visualización de datos que en conjunto proporcionan la solución ideal para analizar tendencias comerciales, el crecimiento del negocio, la cantidad de ganancias, el desempeño de los empleados, la satisfacción del cliente, los puntos débiles y las oportunidades de mejora [1]. Cada día las empresas generan una gran cantidad de datos que son la base para la toma de decisiones estratégicas y ante este hecho los indicadores de rendimiento clave Key Performance Indicators (KPI) pueden cuantificar el desempeño de los procesos de gestión con la finalidad de recomendar acciones futuras [2]. Esto significa que establecer KPI’s ayuda a medir el éxito de una empresa, permitiendo conocer si se han cumplido o no los objetivos empresariales.

Las empresas que brindan servicios o productos de software poseen una naturaleza diferente en comparación a empresas tradicionales como, son las compañías de ventas de productos de consumo masivo. Las empresas de software utilizan a todo el personal en el proceso de ventas, es decir, requiere de una sinergia óptima entre el equipo de desarrollo de software y el equipo de ventas para conseguir el éxito comercial, siendo la inteligencia de negocios una herramienta necesaria para lograr dicha sinergia [3]. En este sentido, en este trabajo, se realizará un análisis de una empresa del área de software llamada TourOpp la cual se dedica a brindar servicios a operadores de turismo, todo esto, con la finalidad de crear una solución integral de inteligencia de negocios partiendo desde la selección de la

herramienta de software que mejor se ajuste a los requerimientos empresariales, el análisis y procesamiento de datos, definición de KPI's y por último, la implementación de un dashboard que permita visualizar la información más importantes.

1.1 Objetivo general

Diseñar e implementar un Data Warehouse y tableros de inteligencia de negocio, para análisis de indicadores de rendimiento en el área de desarrollo y ventas de una empresa que brinda servicios tecnológicos de software.

1.2 Objetivos específicos

- Comprender el estado del arte de las herramientas de dashboard interactivo para visualización de KPI´s.
- Definir los indicadores de rendimiento claves de la empresa TourOpp.
- Implementar un Data Warehouse y tableros de inteligencia de negocios.

1.3 Alcance

El presente trabajo de titulación pretende abarcar las diferentes etapas del diseño e implementación de un Data Warehouse y tableros de inteligencia de negocio para la empresa TourOpp.

En la primera etapa, se realizará el análisis de los requerimientos empresariales, las preguntas de negocio que se pretenden responder, las fuentes de datos que posee la empresa, y el estudio de las herramientas de BI que más se ajustan a su realidad empresarial.

En la segunda etapa se implementa la metodología Hefesto y los pasos que sugiere la misma para la construcción del Data Warehouse pasando por los diferentes procesos de extracción, transformación y carga (ETL por sus siglas en ingles), limpieza de datos y modelado de los mismos.

Como etapa final, se realizará la visualización de los indicadores más importantes con la ayuda de herramientas de visualización de datos.

1.4 Marco Teórico

En la actualidad, es muy común que las empresas almacenen y administren la información que son generados por sus procesos de desarrollo, marketing, ventas, etc. Ahora, los datos generados inicialmente son muy rústicos y requiere de los procesos de transformación para poder ser utilizados en los procesos de toma de decisiones. En este contexto, la inteligencia de negocios permite que la toma de decisiones esté sustentada sobre el conocimiento, lo cual permite minimizar riesgos e incertidumbres. Por otra parte, la inteligencia de negocios también genera el espacio para la construcción de indicadores claves diseñados para brindar información específica sobre preguntas del negocio que permiten conocer qué está sucediendo y cómo está sucediendo, además posibilitan la construcción de modelos que permiten predecir eventos futuros. La aplicación de inteligencia de negocios está destinada para quienes deseen tomar decisiones a través del análisis de sus datos sin importar la naturaleza de la empresa en cuestión, por lo cual, la inteligencia de negocios puede responder a las necesidades de diferentes tipos de empresas [4].

Este proyecto, se encuentra enfocado particularmente en un desarrollo de inteligencia de negocios pragmático y ágil para obtener resultados sobre indicadores de rendimiento en una empresa de software. Y para ello, se requiere de una metodología para seguir un proceso formal y efectivo, y de herramientas tecnológicas para su implementación. A continuación, se realizará un análisis de las metodologías y herramientas más importantes para BI y poder seleccionar las más adecuadas para el presente caso de estudio.

1.4.1 Análisis de metodologías de BI

Según [5], las metodologías más destacadas que se han desarrollado para implementar BI son: Ralph Kimball, Hefestos y SAS Rapid Data Warehouse. Para elegir la metodología más adecuada es necesario considerar parámetros como: el cumplimiento de los objetivos, la capacidad de mejora y el grado de satisfacción de directivos y empleados.

A continuación, se detallan las características más importantes de estas metodologías.

1.4.1.1 Metodología Ralph Kimball

Esta metodología también es conocida como Modelo Dimensional y se basa en el ciclo de vida dimensional del negocio. Este modelo tiene como propósito mejorar la toma de decisiones con la ayuda de consultas a bases de datos relacionales ligadas a las mediciones de los resultados de los procesos del negocio. Se basa en 4 principios [6]:

1. Centrarse en la identificación de los requerimientos del negocio para comprender las relaciones, agudizar el análisis y la competencia consultiva.

2. El diseño de la base de datos debe ser único, integrado, fácil de usar y de alto rendimiento.

3. Las entregas deben realizarse en incrementos muy grandes en plazos de 6 a 12 meses, similar a las metodologías ágiles.

4. Debe ser una solución completa para los usuarios que incluya herramientas de consulta, reportes, análisis avanzado, soporte, sitio web y documentación.

1.4.1.2 SAS Rapid Data Warehouse

La metodología asegura un enfoque disciplinado e iterativo en la gestión e implementación de datos. Esta metodología considera 5 fases para permitir el éxito comercial y técnico en la elaboración de un Data Warehouse (DW), la cual se detalla a continuación [7].

1. Fase de evaluación. En esta fase se determinan las necesidades u oportunidades realistas para realizar un almacén de datos exitosos.

2. Definición de requerimientos. Se identifican las fuentes de datos para el DW, se diseña un modelo lógico, la transformación de datos y entrega de información deben ser documentadas. Se identifican las unidades de negocio, la infraestructura de la extracción OLTP, la transformación de los datos, estrategias de actualización de datos y la línea de tiempo para la construcción del DW.

3. Implementación. Se implementa el modelo lógico que fue diseñado en el paso anterior, esto se realiza en tres etapas que son: gestión, organización y explotación.

4. Fase de entrenamiento. Consiste en dos actividades principales, la primera considera la creación de un documento de alto nivel sobre el almacén de datos y la

forma de explotarlo; la segunda actividad consiste en la capacitación a usuarios y administradores del almacén.

5. Revisión. Una vez que la fase de entrenamiento esta completa y el sistema ha sido desplegado en producción es necesaria una evaluación sobre el éxito o fracaso de éste para cuantificar el impacto sobre la organización. Las recomendaciones que se encuentren deben ser documentadas para futuras expansiones del proyecto.

1.4.1.3 Hefestos

La metodología Hefestos presenta la ventaja de ser rápida y sencilla al momento de obtener resultados, es una metodología que permite la construcción de Data Warehouse que está ampliamente fundamentada en la investigación y evolución continua. Esta metodología plantea que es necesario conocer cada paso que se va a desarrollar para no caer en el problema de no saber qué se está desarrollando y por qué. La construcción de un Data Warehouse puede implementarse en cualquier etapa del ciclo de vida del desarrollo de software y es muy importante que la etapa de reunión de requerimientos sea muy concisa, el desarrollo no debe tomar demasiado tiempo ni despliegues muy largos y, por último, debe proporcionar implementaciones rápidas que demuestren las ventajas del Data Warehouse y motivar a los usuarios [4].

Algunas características principales de la metodología son las siguientes:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tomen decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.

- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.

Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente. En la Figura 1, se detallan los pasos de la metodología Hefestos

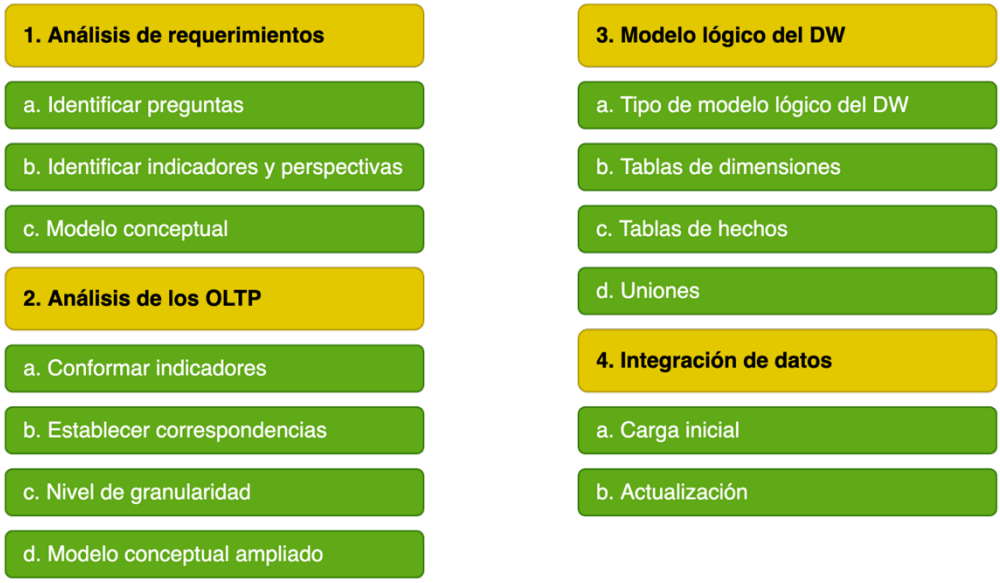


Figura 1 – Pasos de la metodología Hefestos.

1.4.2 Elección de la metodología más adecuada

Los indicadores más importantes que deben ser considerados para realizar la comparación y elección entre las metodologías Kimball, Hefestos y SAS se dividen en 6 fases [5]:

1. Requerimientos: se determinan las características y los requerimientos de la organización, así mismo se estudian a los usuarios.

2. Estrategia del proyecto: en esta etapa de cuantifica el tiempo y se definen las ventajas y las desventajas.

3. Planificación del proyecto: Se define el cronograma de actividades, se analiza el riesgo y se definen las responsabilidades y grupos de trabajo.

4. Selección de la tecnología: Se definen las tecnologías en función del entorno tecnológico de la organización.

5. Diseño del sistema de información: Se determina el modelo de información y la interfaz de usuario.

6. Elaboración del sistema de información: Comprende las etapas de análisis de indicadores, modelo conceptual, OLTP, modelo conceptual ampliado, modelo de DW, tablas de dimensiones, tablas de hechos y procesos ETL.

Para poder identificar la metodología que permita alcanzar las metas planteadas es necesario un estudio descriptivo y explicativo de las fases según los 6 criterios anteriores. Como parámetro de valoración se utiliza la escala de Likert que se muestra en la Tabla 1.

Tabla 1 - Escala de Likert

Escala	Valoración
Baja	1
Media	2
Alta	3

Según [5], la comparación entre las metodologías según los diferentes criterios y utilizando la escala de Likert que se detalla en la Tabla 2.

Tabla 2 - Comparación entre las metodologías: Kimball, Hefestos y SAS

Fase	Indicador	Kimball	Hefestos	SAS
Requerimientos	Definición de requisitos	2	3	2
	Características de la organización	2	3	2
	Análisis de usuarios	3	1	1
Estrategias del proyecto	Cuantificar el tiempo	2	3	2
	Definición de ventajas y desventajas	1	1	1
Planificación	Cronograma	3	1	3
	Riesgo	1	3	3
	Responsabilidades	3	3	1
Tecnología	Entorno actual	3	1	2
	Definición de tecnología	2	2	1
Diseño del sistema de información	Modelo de información	2	3	1
	Interfaz de usuario	1	1	1
Sistema de información	Indicadores	3	3	1
	Modelo conceptual	3	3	3
	Análisis OLTP	3	2	1
	Conformar indicadores	2	3	1

	Nivel de granularidad	3	3	3
	Modelo conceptual ampliado	1	3	1
	Modelo lógico DW	3	3	3
	Tablas de dimensiones	3	3	3
	Tablas de hechos	3	3	3
	Proceso ETL	3	3	3
		52	54	42

Como se puede observar en los resultados de la Tabla 2, la metodología Hefestos cumple con la mayoría de los indicadores expuestos en cada una de las 6 fases, por lo cual, se la considera como la metodología más adecuada. Considerando que TourOpp cambia constantemente para ajustarse a las necesidades del mercado y que necesita implementaciones ágiles que se adapten a los cambios generados en la empresa, se refuerza la decisión de optar por la metodología Hefestos por sobre Kimball ya que esta plantea un ciclo de vida demasiado amplio.

Es por este motivo que, implementar inteligencia de negocios a través de la metodología Hefesto es una opción viable para desarrollar este proyecto acompañado también de una adecuada visualización de datos.

1.4.3 Selección de las herramientas

A continuación, se detallan algunas de las restricciones planteadas por la gerencia de la empresa de estudio que deben cumplir las herramientas utilizadas en este proyecto.

- **Open source:** Dado el carácter tecnológico de TourOpp donde desarrolla sus propios servicios de software, es necesario que las herramientas o lenguajes de programación sean de código abierto y que se puedan modificar y personalizar según los requerimientos de la empresa.

- **Compatibilidad con el entorno de Amazon web Services (AWS):** TourOpp hace uso de la infraestructura de AWS para desplegar sus servicios, por lo cual es necesario que las herramientas sean compatibles con dicho entorno para evitar realizar integraciones innecesarias con otros proveedores de servicios en la nube.

- **Actualización asincrónica:** Los servicios de TourOpp se realizan durante largos períodos de tiempo, por lo cual no es necesario que los procesos de tratamiento de datos se realicen en tiempo real, la información en el DW puede actualizarse en intervalos de tiempo largos.

- **Evitar la intrusión en el sistema actual:** La nueva implementación de software para BI debe ser paralelo al sistema actual, es decir que no puede conectarse a las bases de datos de producción, no puede aumentar carga a los servicios para evitar que el rendimiento del sistema actual se vea afectado.

Tomando en cuenta lo anterior se realiza el análisis correspondiente de cara a infraestructura, despliegue y servicios para definir cuáles son las herramientas más apropiadas para realizar la implementación de herramientas de inteligencia de negocios en la empresa.

1.4.3.1 Elección de los lenguajes de programación

En TourOpp, el equipo de desarrollo de software utiliza principalmente los lenguajes de programación JavaScript en el entorno de ejecución de Node.js y el lenguaje de programación Python. JavaScript es conocido por su aplicabilidad en desarrollo web mientras que Python es ampliamente reconocido por su aplicación en tareas relacionadas a inteligencia artificial, el análisis de datos, big data, entre otros. En la Tabla 3, se plantea la comparación entre JavaScript y Python, en cuanto a términos de escalabilidad, casos de uso, procesos de uso de memoria intensivo y el rendimiento.

Tabla 3 - Comparación entre Python y Node.js [8]

Python	**Node.js**
Tiene un buen soporte asíncrono.	Posee un potente entorno de tiempo de ejecución para solicitudes asíncronas.
Usado para aplicaciones de negocios.	Usado para frontend y backend.
Permite escribir código altamente legible.	Menos legible.
Se adapta a proyectos grandes.	Adecuado para procesos con uso intensivo de memoria.
Intérprete PyPy.	Interprete JavaScript.
No es la mejor opción para aplicaciones en tiempo real.	La mejor opción para aplicaciones en tiempo real.
Ideal para cálculos científicos y numéricos.	No apto para computación científica compleja.

Ambos lenguajes de programación tienen sus puntos fuertes y se complementan muy bien para las actividades que se deben realizar en este proyecto por lo que no se considera necesario añadir otros lenguajes de programación. A continuación, en la Tabla 4, se

muestra un resumen de las actividades que requiere el proyecto y el lenguaje de programación adecuado para la misma.

Tabla 4 - Definición de las tareas a realizar con cada lenguaje de programación.

Tarea	Lenguaje	Razón de uso de lenguaje
Definición de modelos de tablas y operaciones en las bases de datos	Js	Actualmente la empresa mantiene los modelos de bases de datos con un ORM (Object relational mapping) en JavaScript.
Definición de los nuevos modelos de tablas del DW y operaciones en las bases de datos	Js	Para mantener una estructura similar a la que ya existe.
Conexión con fuentes de datos en el entorno empresarial	Js	Ya existen esquemas de conexión.
Aplicación de modelos de inteligencia artificial	Python	Este lenguaje posee una variedad muy grande de librerías para entrenar y ejecutar modelos de inteligencia artificial
Limpieza y tratamiento de datos	Python	Python ofrece varias librerías que están optimizadas para estas tareas.

Las tareas planteadas para este proyecto se las puede trabajar de una manera muy eficiente con una combinación de las mejores características de estos dos lenguajes. En el mercado existen otros lenguajes de programación que podrían utilizarse, pero con el análisis anterior determinamos que JavaScript y Python son suficientes para este proyecto.

1.4.3.2 Microservicios.

El proyecto se compone de varias subtareas específicas que trabajan de forma independiente. Además, cada subtarea necesita de un entorno diferente para ser ejecutado, dependencias, versiones de librerías y entornos de ejecución. Por lo tanto, la implementación de microservicios es un enfoque muy apropiado para trabajar en este proyecto.

Por lo anterior, se considera el uso de Docker para desplegar los servicios que se crean en este proyecto. Docker es un proyecto de código abierto que permite crear, probar y ejecutar aplicaciones en unidades de software llamadas contenedores. Estos contenedores están

dotados de bibliotecas, sistema operativo, código y tiempo de ejecución para tener la certeza de que el código se ejecute correctamente [9].

Como este proyecto se compone de varios servicios, se considera el uso de Docker Compose, el cual, es una herramienta diseñada para poder definir, conectar y ejecutar varios servicios con un simple comando. Docker Compose se puede utilizar en entornos de desarrollo, prueba y producción [10].

1.4.3.3 Automatización

Realizar una implementación de herramientas BI requiere de la ejecución de varias tareas secundarias para ello es necesaria una herramienta de automatización. Como antecedente hay que mencionar que la empresa utiliza ampliamente Jenkins, el cual es un servidor de automatización open source [11].

En el caso puntual de este proyecto, dado que esta implementación no tiene permitido realizar una conexión directa a la base de datos de producción es necesaria una tarea automática que realice una copia de dicha base de datos y la ponga a disposición de los microservicios en este proyecto y como paso final iniciar el proceso de carga o actualización del Data Warehouse.

La tarea automática que debe realizar Jenkins debe seguir los siguientes pasos:

1. Crear un snapshot de la base de datos de producción.
2. Copiar el snapshot anterior a la zona de disponibilidad donde se encuentran los microservicios del proyecto
3. Eliminar la base de datos de desarrollo.
4. Levantar la base de datos desde el snapshot creado de producción.
5. Enviar un evento de inicio para el proceso de actualización de datos.

1.4.3.4 Servidor

Las características del servidor EC2 en el cual se van a desplegar todos los microservicios que componen el proyecto será determinado más adelante en base al consumo de memoria RAM, CPU y almacenamiento de todo el proyecto ejecutándose y realizando los

procesos necesarios para cargar o actualizar el DW, esto se analiza con más detalle en la sección 2.11.

1.4.3.5 Base de datos

El servicio de bases de datos relacionales RDS es un servicio de AWS que permite crear, configurar, manejar y escalar bases de datos de tipo relacional de una manera muy sencilla. Al ser un servicio administrado libera al operador de bases de datos de tareas como configuraciones, parches, copias de seguridad, etc. [12]

Para realizar la implementación del DW se aprovisiona una nueva base de datos de tipo MySQL en el servicio RDS de AWS con la finalidad de no interferir en las bases de datos que actualmente soportan los servicios de TourOpp.

1.4.3.6 Tablero de visualización

La visualización de datos es una parte muy importante de este proyecto para responder gráficamente la respuesta a las preguntas de negocio más importantes. Las restricciones que indica la empresa que se deben tener en cuenta para elegir la herramienta de visualización de datos son las siguientes: La herramienta debe ser totalmente gratuita, debe permitir la conexión con la fuente de datos, debe permitir crear tableros interactivos, compartir reportes y aplicar variedad de gráficas.

Para esta tarea existen herramientas en el mercado como PowerBI, Tableu, QuickSight o Pentaho PDI, que tienen características muy potentes para el análisis de datos y son ampliamente utilizadas, algunas de estas tienen versiones gratuitas pero algunas funcionalidades como compartir reportes no son de uso gratuito lo cual va en contra de las restricciones planteadas por la empresa. Por otro lado, la herramienta de visualización de datos de Google llamada Data Studio ofrece todas sus funcionalidades de forma gratuita.

Google Data Studio es una herramient que se ejecuta en línea y forma parte de Google Analytics 360. Las principales ventajas de ésta son las siguientes según [13].

- Permite compartir los informes con otros usuarios.
- Permite crear informes de forma sencilla y rápida.
- Permite interactuar con los informes.
- Permite la conexión con varias fuentes de datos.

- Permite visualizar los datos en gráficas y tablas.

Las 3 herramientas principales que proporciona Google Data Studio son: Fuentes de datos, gráficos y controles.

1. **Fuentes de datos.**

Como paso inicial para poder crear reportes es necesario incluir las fuentes de datos disponibles y para ello realizan los siguientes pasos:

- Seleccionar la opción “Añadir datos” que se muestra en la Figura 2.

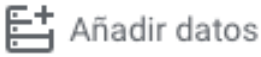


Figura 2. Añadir datos en Google Data Studio

- Seleccionar el conector de MySQL que se muestra en la Figura 3.

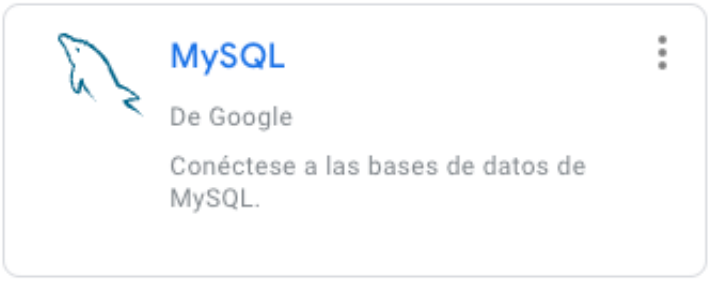


Figura 3 – Conector de Google Data Studio con MySQL.

- Completar la información para la conexión con la base de datos que se muestra en la Figura 4.

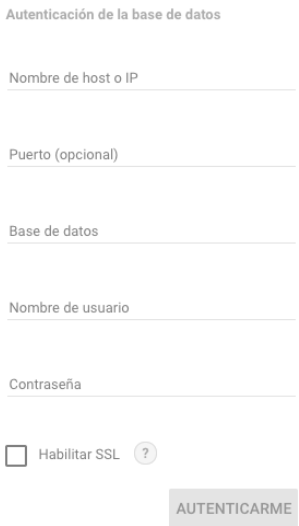


Figura 4 – Información de autenticación con la base de datos MySQL.

Una vez realizado el proceso de conexión se puede escoger cada tabla que se generó en el desarrollo de los Data Mart y se debe repetir este proceso para cada una de las tablas que componen los Data Mart.

2. Gráficos

Una vez que se han incluido todas las fuentes de datos es momento de empezar a crear visualizaciones sobre los mismos. Para esto, Google Data Studio provee de una variedad de gráficos como se muestra en la Figura 5.

Figura 5 – Tipos de gráficos que provee Google Data Studio.

3. Controles.

Los controles son elementos que permiten modificar las gráficas previamente implementadas, los controles actúan como filtros que permiten adaptar las gráficas a la visualización que se espera obtener. En la Figura 6 se muestran los controles que permite utilizar Google Data Studio.

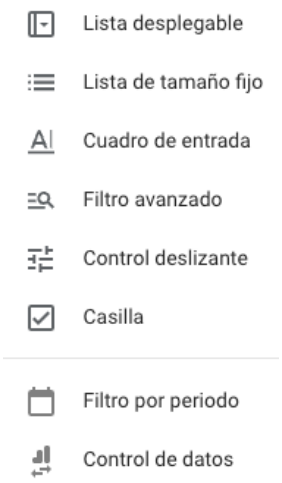


Figura 6 – Controles que provee Google Data Studio.

Técnicas para escoger el gráfico adecuado.

Escoger el gráfico adecuado permite que el usuario entienda la información que se desea compartir de una manera más rápida, para ello hay que considerar que la mayoría de las visualizaciones pueden clasificarse en 5 tipos [14]

- Comparación

Los gráficos de comparación responden a como se comparan los diferentes valores o atributos dentro del conjunto de datos, pueden mostrar fácilmente valores altos y bajos y se utilizan los siguientes tipos de gráficos: Columna, Barras, Bala, Torta, etc.

- Tendencia

Si se desea conocer información sobre el desempeño de un conjunto de datos a lo largo de un periodo de tiempo, los siguientes gráficos son los más adecuados: Línea, Línea de doble eje y barras.

- Composición

Se utiliza cuando se desea visualizar el cómo las partes individuales conforman un todo. Para este propósito se utilizan los siguientes tipos de gráficos: Torta, Barra Apilada, Zona, Cascada.

- Distribución.

Estos gráficos son útiles para conocer la tendencia normal y el rango de los datos, además de los valores atípicos. Los gráficos utilizados para este propósito son: Gráfico de dispersión, Mekko, Línea y Columna

- Relación.

Se emplean estos gráficos para conocer como una variable se relaciona con otra o con otras diferentes, por ejemplo, para demostrar como algo afecta positivamente, no afecta o afecta negativamente a otra variable. Para esto se tienen los gráficos: Gráfico de dispersión, Burbuja y Línea.

1.4.3.7 Entrenamiento de modelos de inteligencia artificial.

Al trabajar con modelos de inteligencia artificial una de las etapas más importantes y que consume mayor cantidad de recursos computacionales es el entrenamiento. Google Colab es un servicio de Google que permite ejecutar código de Python en la nube. Asimismo, es una herramienta muy útil para entrenar modelos de inteligencia artificial y realizar ciencia de datos ya que permite utilizar la potencia de los servidores de Google de manera gratuita. También, el servicio de Google permite utilizar GPUs y TPUs para entrenar modelos de manera más rápida [15]

De manera resumida, en la Tabla 5, se detallan las herramientas seleccionadas para llevar a cabo el proyecto.

Tabla 5 - Resumen de las herramientas a utilizar

Ítem	Herramienta
Lenguaje de programación	Javascript + Python
Microservicios	Docker + Docker Compose
Automatización	Jenkins
Servidor	Instancia EC2
Entrenamiento de modelos IA	Google Colab
Visualización	Google Data Studio

2 METODOLOGÍA

En este capítulo, se desarrollan los pasos que propone la metodología Hefestos para la construcción de los Data Mart en cada área de estudio, el análisis de los modelos de inteligencia artificial a utilizar y se aborda también la visualización de los datos.

Para empezar, se realiza una recopilación de las preguntas más importantes que la gerencia de la empresa necesita resolver para poder conocer el estado actual de la misma, descubrir información importante y tomar decisiones acertadas que le ayuden a la empresa a seguir el camino correcto.

2.1 Descripción de la organización del caso de estudio

A nivel mundial la pandemia del COVID 19 ha tenido un impacto muy negativo en el sector turístico dadas las restricciones de viajes internacionales para contener la propagación de brotes del virus lo cual paralizó en gran medida la industria de turismo. Según [16] cuanto más grande es el sector turístico mayor es el estímulo económico de los gobiernos para mitigar el impacto económico y estabilizar las economías tambaleantes, esto refleja la importancia que posee el sector turístico.

Asimismo, según [17], está comprobado que las operadoras tradicionales de turismo con oficinas offline serán desplazadas por aquellas operadoras que adopten la digitalización y automatización en sus procesos cotidianos. En consecuencia, las operadoras que logren brindar la mejor personalización a sus clientes en sus servicios obtendrán una mayor ventaja competitiva.

En este contexto, TourOpp es una empresa que reconoce la importancia del sector turístico a nivel mundial e identifica la oportunidad de ofrecer servicios de software a operadoras de turismo con el fin de ayudarles a automatizar procesos, mejorar la relación con los clientes, mejorar su posicionamiento en el mercado e incrementar sus ventas. Para lograr estos objetivos, los servicios se enfocan en crear un vínculo más cercano entre el operador de turismo y sus clientes.

Como primer paso, TourOpp, entabla una conversación con los clientes del operador haciendo uso de mensajes de texto personalizados y a través de estos envía recordatorios y responde a las preguntas que pudiera tener con la ayuda de un chatbot que hace uso de procesamiento de lenguaje natural. De esta manera, TourOpp logra automatizar procesos repetitivos para los operadores y que de otra manera le consumirían demasiado tiempo.

Una vez que existe la interacción entre los clientes finales y el servicio de TourOpp, éste solicita una calificación sobre la actividad que realizó, si la calificación es positiva, se incentiva al cliente a que registre una reseña en las plataformas de Google y TripAdvisor con lo cual el Operador mejora su presencia en las redes. En el caso de que la calificación fuere negativa, se le notifica al operador para que actúe al respecto, evitando así las reseñas negativas en las redes. De esta manera, TourOpp logra que los operadores de turismo mejoren la relación con los clientes y su posicionamiento en el mercado.

Por último, TourOpp envía enlaces de recomendación de actividades similares para incentivar a los clientes de los operadores de turismo que compren una actividad nueva haciendo uso de incentivos como cupones de descuento. De esta manera, TourOpp genera venta cruzada para los operadores.

Mediante el uso de los servicios de TourOpp, los operadores de turismo logran un retorno de inversión (ROI) en base a aspectos como el ahorro de tiempo que toma responder a preguntas frecuentes, el ahorro en publicidad y la venta cruzada. Esta información es de alta relevancia para los operadores de turismo y TourOpp les facilita la misma a través de reportes mensuales.

2.1.1 Definición de términos

A continuación, se realiza la definición de términos más importantes que se usan en este documento.

- **Usuario:** Entidad que se suscribe a los servicios de TourOpp, en este caso puntual los operadores de turismo.
- **Cliente final:** Personas con las que TourOpp interactúa directamente a través de mensajería personalizada, enviando recordatorios y resolviendo dudas.
- **Actividad:** Hace referencia a los tours que ofrece un usuario hacia los clientes finales.
- **Sistema de reservas:** El software de terceros que usan los usuarios principalmente para manejar sus clientes finales, actividades y reservas. TourOpp se conecta directamente al sistema de reservas del usuario para recibir las transacciones.
- **Transacciones:** Es un arreglo de información que TourOpp recibe desde el sistema de reserva cuando un cliente final compra una actividad de un Usuario.

- **Comandos de chatbot:** cuando un cliente final interactúa con el chatbot le realiza preguntas sobre diferentes temas como el lugar de recogida, la hora a la que empieza el tour, las atracciones principales, etc. Estos diferentes temas se conocen como comandos de chatbot.

- **Review:** Es un mensaje tipo reseña, que escribe el cliente final en la página de la actividad que realizó indicando su experiencia sobre el mismo.

- **Rating:** Es la calificación que otorga el cliente final sobre la actividad que realizó.

- **Recordatorios:** Son mensajes programados que el usuario configura de acuerdo con la actividad que realiza el cliente final.

- **Fallback:** Se produce cuando el chatbot no entiende alguna pregunta del cliente final. En otras palabras, es un tema para el cual el chatbot no está entrenado para entenderlo y emitir una respuesta.

- **Data Mart:** Es un Data Warehouse enfocado en un área específica de la empresa.

2.1.2 Uso de BI en TourOpp

Hacer uso de inteligencia de negocios permitirá a TourOpp tomar decisiones inteligentes en base al análisis del amplio conjunto de datos que posee sobre su negocio. Para esto es necesario que TourOpp transforme los objetivos empresariales en indicadores que puedan ser analizados desde diferentes perspectivas y comprender los eventos que sucedieron en el pasado, que están sucediendo en el presente y poder así predecir los eventos que van a suceder en el futuro.

TourOpp, al ser una empresa que brinda servicios de tecnología, posee la infraestructura donde recolecta la información de sus servicios principales que servirán de base para la realización de este proyecto.

Con la implementación de este proyecto, TourOpp busca conocer a sus usuarios y clientes finales, asimismo busca resolver interrogantes de cómo está creciendo, cómo mejorar sus servicios y así poder escalar a un mercado cada vez más amplio.

2.2 Análisis de los requerimientos empresariales

Como paso inicial, se desarrolló el proceso de recopilación de preguntas de negocio más importantes que se desean responder y para ello, se desarrolló un diálogo con los directivos de la empresa, con los empleados del equipo de ventas y del equipo de desarrollo de software donde se detectaron los requerimientos en función de los objetivos de la empresa como se muestra en el ANEXO A. Luego se filtraron considerando la relevancia de estos para el negocio y si existe o no la información para poder responder dichas preguntas.

Se detectaron 6 ámbitos principales a estudiar que son:

- **Ventas:** El equipo de ventas posee una lista de preguntas sobre el crecimiento de suscripciones, los clientes perdidos y sobre los ingresos generados.

- **Desarrollo:** El actuar del equipo de ventas tiene una relación directa con el equipo de desarrollo, ya que éste debe estar preparado para hacer que los servicios de TourOpp cumplan con la oferta de valor de la empresa y también que soporten al aumento progresivo en el número de usuarios y clientes finales.

- **Marketing:** En el equipo de marketing se desea conocer si los blogs y post publicados están teniendo el impacto esperado y si están generando o no intención de suscripción en nuevos usuarios potenciales.

- **Análisis de clientes finales:** Se desea conocer como están interactuando los clientes finales con el servicio.

- **Análisis de usuarios:** La empresa desea conocer las características principales que diferencian a unos usuarios de otros para poder clasificarlos y descubrir cuales son los mejores usuarios.

- **Chatbot:** El chatbot es una parte muy importante de la empresa ya que interactúa directamente con los clientes finales, por lo cual es necesario conocer cómo se puede mejorar su rendimiento.

A continuación, se desarrollaron los Data Mart para las áreas de estudio.

2.3 Data Mart para ventas

2.3.1 Análisis de requerimientos

A continuación, se detallan los requerimientos que se recopilaron con ayuda del equipo de ventas y que servirán de base para desarrollar el Data Mart.

2.3.1.1 Identificar preguntas

A continuación, se muestran las preguntas de negocio que se recopilaron, las cuales se estructuraron de una forma más clara y se filtraron en función de los objetivos del negocio y de la disponibilidad de datos para poder responder a las mismas.

1. ¿Cómo están creciendo las suscripciones en el transcurso del año y cuál es la predicción de crecimiento hasta finalizar este año?

El área de ventas tiene un objetivo para el año en curso que es alcanzar un total de 188 suscripciones, considerando esto, se desea conocer el ritmo de adquisición de clientes por semana y la predicción del número de suscripciones hasta la finalización del año.

2. ¿Cuántos clientes están cancelando sus suscripciones y cuál es la predicción de cancelaciones hasta finalizar este año?

Para el área de ventas, el máximo aceptable de cancelaciones mensuales es de una, es decir 12 cancelaciones por año, por esto se desea conocer el número de cancelaciones por semana y la predicción de cancelaciones hasta la finalización del año.

3. ¿Cómo se comporta el promedio de ingresos a lo largo del tiempo?

El promedio de ingresos mensuales es un indicador del tamaño de los usuarios. Si este valor está por encima de la meta significa que existen más clientes grandes que pequeños, mientras que si está por debajo de la meta implica que la mayoría de los clientes son pequeños. Lo ideal es que este promedio se encuentre sobre un valor establecido.

2.3.1.2 Análisis de perspectivas e indicadores

1. ¿Cómo están creciendo las suscripciones en el transcurso del año y cuál es la predicción del número de suscripciones hasta finalizar este año?

    - **Perspectivas:** Tiempo.
    - **Indicadores:** Suscripciones.

2. ¿Cuántos clientes están cancelando sus suscripciones y cuál es la predicción del número de cancelaciones hasta finalizar este año?

    - **Perspectivas:** Tiempo.
    - **Indicadores:** Cancelaciones.

3. ¿Cómo se comporta el promedio de ingreso a lo largo del tiempo?

    - **Perspectivas:** Tiempo
    - **Indicador:** Promedio de ingresos.

2.3.1.3 Modelo conceptual

En la Figura 7 se muestra el modelo conceptual para el Data Mart en el área de ventas.

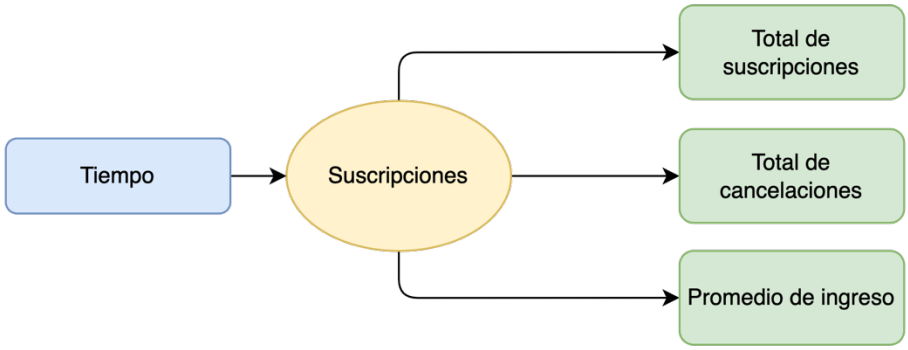


Figura 7 – Modelo conceptual para ventas

2.3.2 Análisis de OLTP

En este paso se analizan las fuentes de datos OLTP (OnLine Transaction Processing) para establecer el cálculo de los indicadores y la relación entre el modelo conceptual anteriormente definido y las fuentes de datos.

2.3.2.1 Conformar indicadores

A continuación, se explica la forma en la que se van a calcular los respectivos indicadores en función de los hechos y las funciones.

- **“Total de suscripciones”**
    - Hechos: Suscripciones
    - Función de sumarización: SUM

El indicador “Total de suscripciones” representa la sumatoria de las suscripciones que se han realizado por cada semana del año.

- **“Total de cancelaciones”**
  - Hechos: Cancelaciones
  - Función de sumarización: SUM

El indicador “Total de cancelaciones” representa la sumatoria de todos los usuarios que decidieron cancelar su suscripción por cada semana del año.

- **“Promedio de ingreso mensual”**
  - Hechos: (Ingreso mensual) / (Suscripciones)
  - Función de sumarización: AVG

El indicador “Promedio de ingreso” representa el promedio de los ingresos nuevos semanales.

2.3.2.2 Establecer correspondencias

El área de ventas utiliza la plataforma Chargebee para manejar las suscripciones de los usuarios. Chargebee es un software de facturación que impulsa la facturación recurrente, la gestión de suscripciones y la facturación automática de extremo a extremo [18].

Chargebee posibilita la integración a través de un API que puede ser consultado para recuperar información relevante sobre clientes, suscripciones, pedidos, pagos, etc. Touropp realiza una consulta una vez a la semana a este API, ordena la información y lo guarda en la tabla “subscriptions”. En la Figura 8, se muestran las correspondencias entre los OLTP y el modelo conceptual.

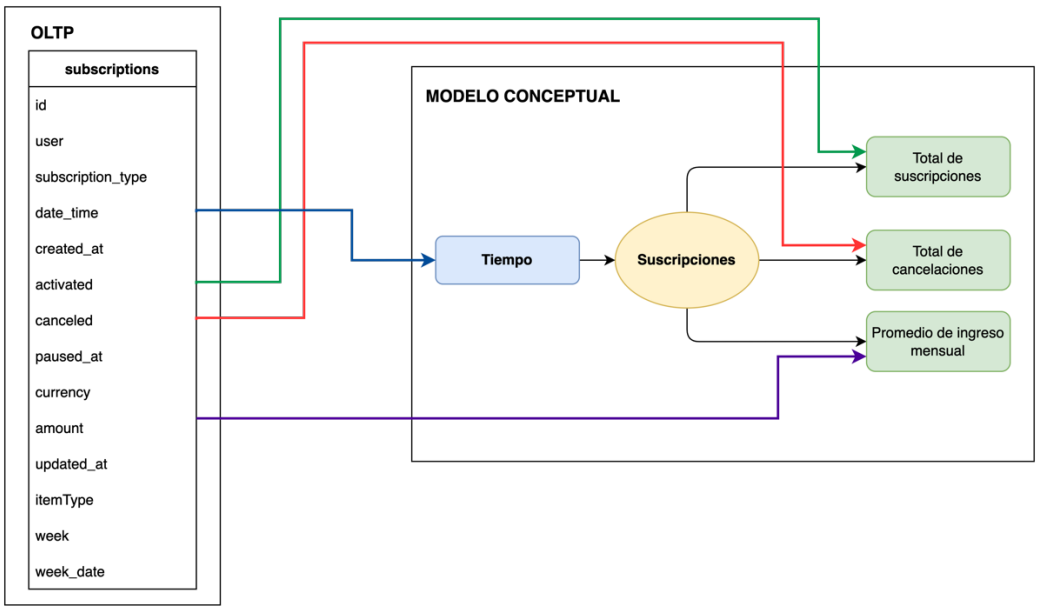


Figura 8 – Correspondencias entre OLTP y modelo conceptual de ventas

Las relaciones identificadas son las siguientes:

- El campo “date_time” de la tabla “subscriptions” se relaciona con la perspectiva “Tiempo” ya que indica la fecha en la que se realizó la operación.
- La columna “activated” se relaciona con el indicador “Total de suscripciones” ya que indica que el registro corresponde a una suscripción.
- La columna “canceled” se relaciona con el indicador “Total de cancelaciones” ya que indica que el registro corresponde a una cancelación.
- El campo “amount” se relaciona con el indicador “Promedio de ingreso”

2.3.2.3 Nivel de granularidad

A continuación, se van a seleccionar los campos que contendrá cada perspectiva.

- Perspectiva “Tiempo”: los datos que se van a emplear para el análisis son los siguientes:
  - Año
  - Trimestre
  - Mes
  - Semana

2.3.2.4 Modelo conceptual ampliado

Luego de analizar el nivel de granularidad, en la Figura 9 se presenta el modelo conceptual ampliado.

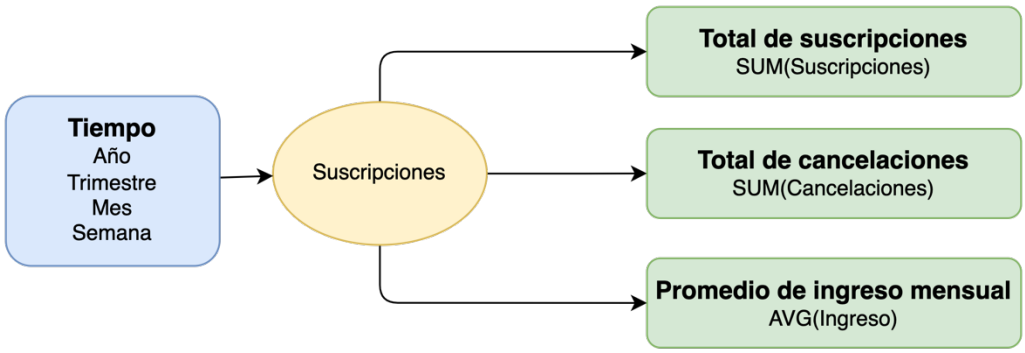


Figura 9 – Modelo conceptual ampliado para ventas

2.3.3 Modelo lógico

A continuación, se conforma el modelo lógico teniendo como base el modelo conceptual que ha sido creado.

2.3.3.1 Tipo de modelo lógico

Para este Data Mart de ventas se utilizará un esquema de tipo estrella dado que es un modelo que consta de una sola tabla de dimensiones y una de hechos.

2.3.3.2 Tablas de dimensiones

En este paso se diseñan las tablas de dimensiones que forman parte de la Data Mart.

- Perspectiva “Tiempo”:

La nueva tabla tendrá el nombre “FECHA”.

Tendrá una clave principal con el nombre “idFecha”.

Se mantendrán los campos “Año”, “Trimestre”, “Mes”, “Semana”.

La tabla de dimensión Fecha se muestra en la Figura 10.

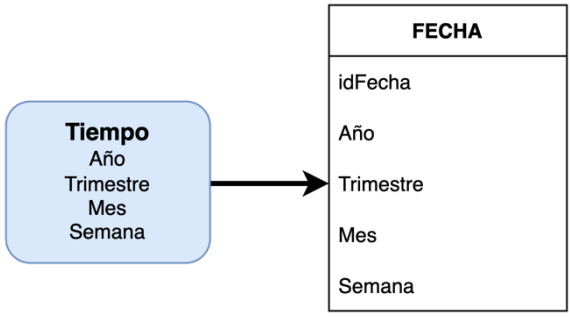


Figura 10- Tabla de dimensión Fecha

2.3.3.3 Tablas de hechos

En este paso se define la tabla de hecho, que contendrán los indicadores de estudio.

La tabla de hechos se llamará “SUSCRIPCIONES”.

Su clave principal “idFecha” tiene relación con la tabla de dimensión “FECHA”

Se crean los siguientes hechos: “Suscripciones”, “Cancelaciones” e “IngresoPromedio”. El diseño de la tabla de hechos “SUSCRIPCIONES” se muestra en la Figura 11.

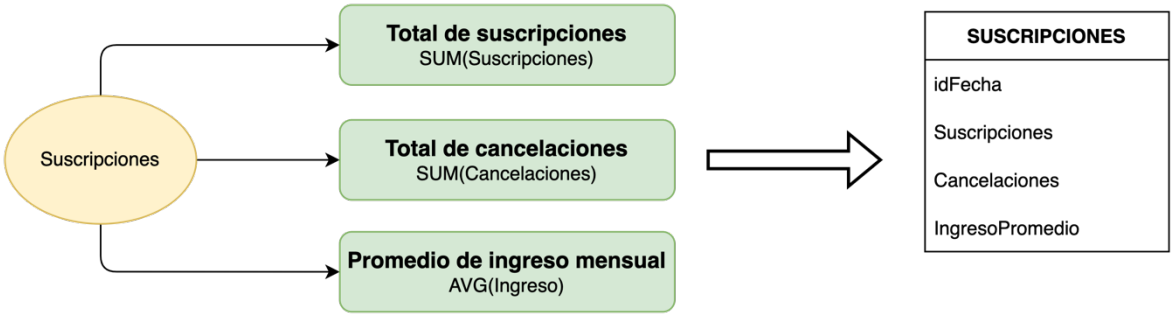


Figura 11 – Tabla de hechos suscripciones.

2.3.3.4 Uniones

En este paso, se realizan las uniones entre las tablas de dimensiones y las tablas de hechos como se muestra en la Figura 12.

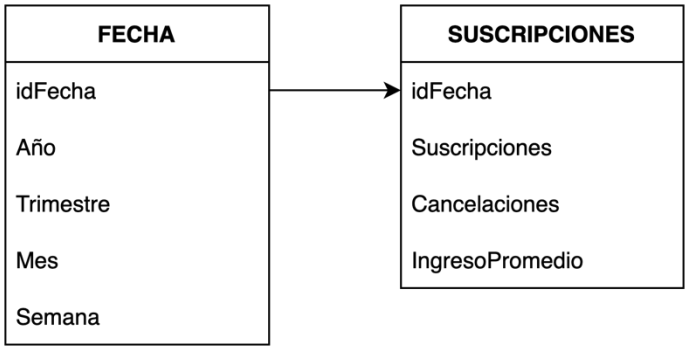


Figura 12- Uniones entre las tablas del Data Mart de ventas

2.3.4 Integración de datos

A continuación, se procede a probar el modelo creado con datos, utilizando técnicas de limpieza y calidad de datos, luego se define la política de actualización y los procesos que se van a llevar a cabo.

2.3.4.1 Carga inicial

El proceso de carga inicial se muestra en la Figura 13.

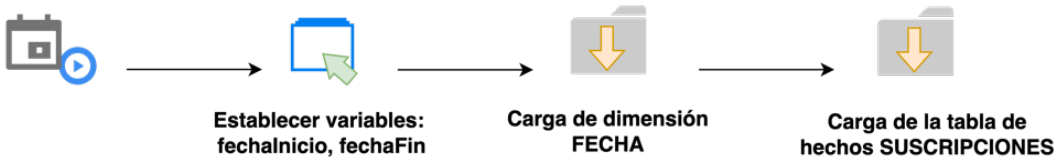


Figura 13 – Carga inicial para ventas

A continuación, se detallan los pasos a seguir en el proceso de carga inicial.

- Inicio: Inicia el proceso en el momento que se le indique.
- Establecer variables: Se establece la variable “fechaInicio” como la fecha de la primera suscripción y la variable “fechaFin” como la fecha actual.
- Carga dimensión FECHA: Para la carga de la dimensión fecha se creó un script que genera fechas en un rango desde la fecha de inicio hasta la fecha de finalización (fin del año en curso) con un rango semanal. El formato para la columna “idFecha” se compone de “año + semana”. Un ejemplo del resultado de este script, se muestra la Tabla 6.

Tabla 6 – Ejemplo de datos para la carga de la dimensión FECHA.

idFecha	Año	Trimestre	Mes	Semana
202201	2022	1	1	1
202202	2022	1	1	2
202203	2022	1	1	3

El resultado de este script se guarda en la tabla “FECHA”.

- Carga de la tabla de hechos SUSCRIPCIONES: La carga de esta tabla de hechos se compone de dos pasos como se indica en la Figura 14.

Figura 14 – Carga de la tabla de hechos Suscripciones

- **Obtener datos OLTP**

El primer paso consiste en cargar la información de las columnas: “Suscripciones” e “IngresoPromedio” desde la tabla “suscriptions” con la siguiente consulta SQL:
SELECT COUNT (*) as Suscripciones, AVG (amount) as IngresoPromedio, week WHERE activated = 1 GROUP BY week ORDER BY week DESC;
El tercer paso consiste en cargar la información de las columnas “Cancelaciones” desde la tabla “suscriptions” con la siguiente consulta SQL:
SELECT count(*) as Cancelaciones FROM suscriptions WHERE canceled = 1 GROUP BY week ORDER BY week DESC;

- **Cargar SUSCRIPCIONES**

En este paso se guarda en la tabla “SUSCRIPCIONES” los datos obtenidos en el paso anterior.

2.3.4.2 Actualización

Las políticas de actualización definidas con la ayuda de los usuarios son las siguientes:

- La información se actualizará todos los martes a las 12 de la noche.

- Para la tabla de hechos “SUSCRIPCIONES”, las columnas “Suscripciones”, “IngresoPromedio” y “Cancelaciones” se cargarán de manera incremental teniendo en cuenta la fecha de la última actualización.

El proceso ETL para la actualización de Data Mart se parece mucho al de la carga, pero tiene las siguientes diferencias:

- Inicio: el proceso inicia todos los martes a las 12 de la noche.
- Establecer variables: se establece “fechaInicio” como la fecha actual menos una semana y “fechaFin” como la fecha actual.
- Carga dimensión “FECHA”: no es necesario ya inicialmente la carga de la dimensión se realizó hasta la finalización del año en curso.
- Carga de la tabla de hechos “SUSCRIPCIONES”: a las consultas SQL se le añade la condición “WHERE date_time >= fechaInicio and date_time <= fechaFinal”.

2.3.5 Indicadores dinámicos.

El data mart de ventas está construido para responder a las preguntas sobre el comportamiento de las suscripciones y cancelaciones en el transcurso del tiempo mientras que para poder conocer el comportamiento futuro es necesario extender el data mart y tomar en cuenta indicadores dinámicos que permitan predecir las nuevas suscripciones, utilizando el modelo que se detalla en la sección 2.9.1, y predecir las nuevas cancelaciones, utilizando el modelo que se detalla en la sección 2.9.2. Para esto se plantea el modelo que se muestra en la Figura 15.

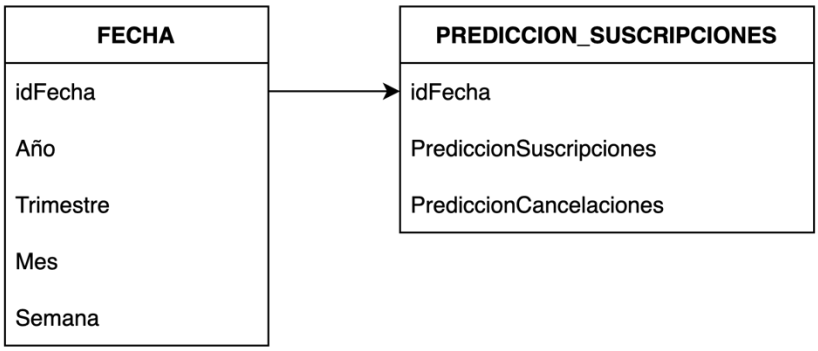


Figura 15 – Modelo para predicción de suscripciones.

Este modelo “PREDICCION_SUSCRIPCIONES” cuenta con las siguientes características:

- La llave primaria “idFecha” tiene relación con la tabla de dimensión “FECHA”.
- Se definen los hechos “PrediccionSuscripciones” y “PrediccionCancelaciones”.

El proceso de actualización de este modelo tiene los siguientes los pasos:

- Inicio: La actualización de la tabla “PREDICCION_SUSCRIPCIONES” se realiza inmediatamente después de la actualización de la tabla de hechos “SUSCRIPCIONES” ya que esta tabla es la fuente de datos para realizar las predicciones.
- Establecer variables: se establece la variable “fechaInicio” como la fecha actual y “fechaFin” como 31 de diciembre del año en curso.
- Predicción: La columna “Suscripciones” y la columna “Cancelaciones” de la tabla de hechos “SUSCRIPCIONES” sirven como fuente de datos para alimentar los modelos detallados en la sección 2.9.1 y 2.9.2. Estos modelos devuelven los valores de predicciones semanales hasta la finalización del año.
- Carga: Se limpian los registros existentes en la tabla “PREDICCION_SUSCRIPCIONES” para luego guardar los nuevos valores de predicciones.

2.4 Data Mart para desarrollo

2.4.1 Análisis de requerimientos

A continuación, se detallan los requerimientos que se recopilaron en conjunto con el equipo de desarrollo.

2.4.1.1 Identificar preguntas

1. ¿Cómo se ha comportado la disponibilidad del servicio en el transcurso del tiempo?

TourOpp es una empresa tecnológica que mantiene y despliega sus propios microservicios. Para conocer el estado de los microservicios, existe uno que constantemente realiza una consulta a los demás cada 5 min. De esta manera es posible conocer cuáles de estos se encuentran trabajando normalmente y cuales no para poder tomar acciones correctivas.

2. ¿Cuándo será necesario realizar un escalamiento horizontal de la infraestructura?

Para entender los límites de procesamiento que posee la infraestructura se realizaron pruebas de estrés al sistema completo donde se descubrió que si se superan las 10,000 transacciones al día el sistema empieza a ralentizarse por lo cual es necesario realizar un escalamiento en los microservicios.

3. ¿Cuándo será necesario escalar la base de datos?

La base de datos está montada en volúmenes no administrados, por lo cual es necesario conocer el espacio que está ocupando para poder predecir cuándo será necesario asignar un espacio más grande.

4. ¿Cuáles son las horas del día en las que el uso del sistema es más intensivo?

Este indicador sirve para conocer en que horarios es más recomendable realizar tareas de mantenimiento en la infraestructura de la empresa.

2.4.1.2 Análisis de perspectivas e indicadores

1. ¿Cómo se ha comportado la disponibilidad del servicio en el transcurso del tiempo?

- **Perspectivas:** Tiempo
- **Indicadores:** Porcentaje de disponibilidad

2. ¿Cuándo será necesario realizar un escalamiento horizontal de la infraestructura?

- **Perspectivas:** Tiempo
- **Indicadores:** Total de transacciones

3. ¿Cuándo será necesario escalar el volumen de las bases de datos de cada sistema de reservas?

- **Perspectivas:** Tiempo, Sistema de reservas
- **Indicadores:** Almacenamiento

4. ¿Cuáles son las horas del día en las que el uso del sistema es más intensivo?

- **Perspectivas:** Tiempo
- **Indicadores:** Total de transacciones

2.4.1.3 Modelo conceptual

A continuación, en la Figura 16 se muestra el modelo conceptual para el Data Mart en el área de desarrollo.

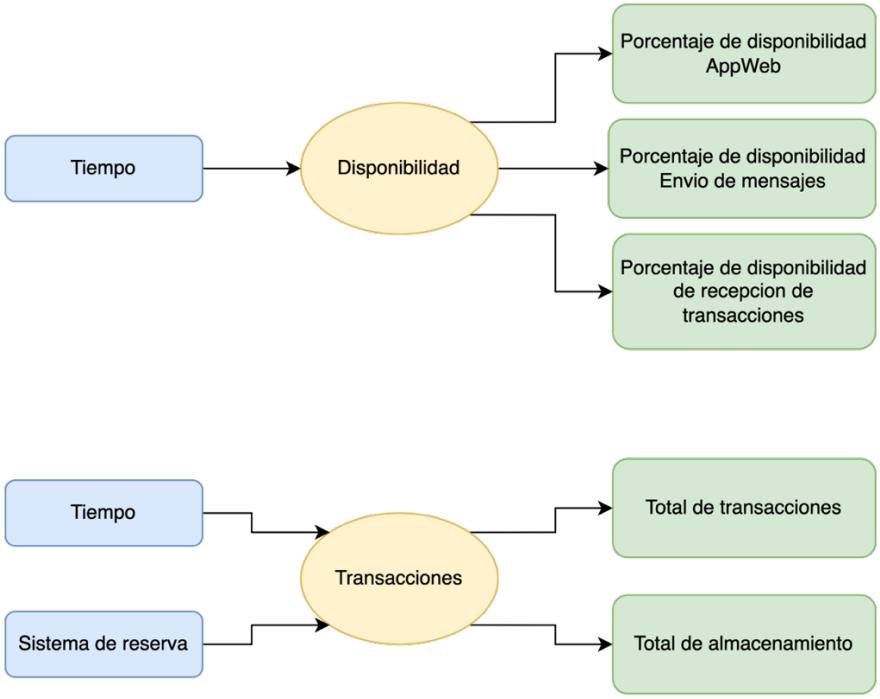


Figura 16 – Modelo conceptual para desarrollo.

2.4.2 Análisis de OLTP

A continuación, se analizan las fuentes de datos para establecer el cálculo de los indicadores y la relación con el modelo conceptual anteriormente definido.

2.4.2.1 Conformar indicadores

La forma en la que se van a calcular los respectivos indicadores en función de los hechos y las funciones es la siguiente:

- **"Porcentaje de disponibilidad App Web"**
  - Hechos: ((Tiempo de disponibilidad de la aplicación en la semana) * 100) / (Tiempo total por semana)
  - Función de sumarización: PERCENT

  El indicador "Porcentaje de disponibilidad de la aplicación" representa el porcentaje de tiempo que estuvo disponible la aplicación en el transcurso de la semana.
- **"Porcentaje de disponibilidad del servicio de envió de mensajes"**
  - Hechos: ((Tiempo de disponibilidad del envío de mensajes en la semana) * 100) / (Tiempo total por semana)
  - Función de sumarización: PERCENT

  El indicador "Porcentaje de disponibilidad del servicio de envío de mensajes" representa el porcentaje de tiempo que estuvo disponible el servicio de envió de mensajes en el transcurso de la semana.
- **"Porcentaje de disponibilidad del servicio de recepción de transacciones"**
  - Hechos: ((Tiempo de disponibilidad del envío de mensajes en la semana) * 100) / (Tiempo total por semana)
  - Función de sumarización: PERCENT

  El indicador "Porcentaje de disponibilidad del servicio de envío de mensajes" representa el porcentaje de tiempo que estuvo disponible el servicio de envió de mensajes en el transcurso de la semana.
- **"Total de transacciones"**
  - Hechos: Transacciones
  - Función de sumarización: SUM
- **"Total de almacenamiento"**
  - Hechos: Almacenamiento
  - Función de sumarización: SUM

  El indicador "Total de almacenamiento" representa el volumen que ocupan las bases de datos.

2.4.2.2 Establecer correspondencias

En la Figura 17 se muestran las fuentes de datos OLTP y las relaciones que tiene con el modelo conceptual.

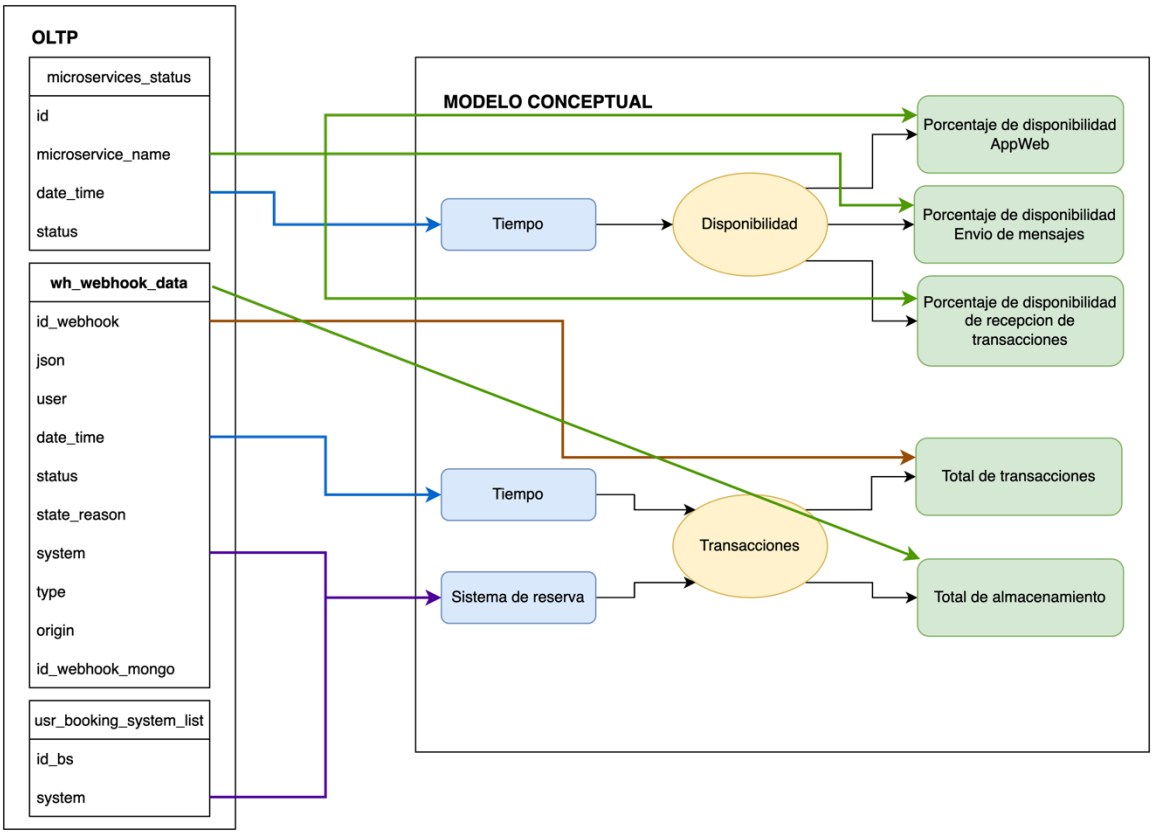


Figura 17 – Relaciones entre los OLTP y el modelo conceptual de desarrollo.

Las relaciones identificadas son las siguientes:

- La columna de “date_time” de la tabla “microservices_status” se relaciona con la perspectiva “Tiempo”.
- La columna “microservice_name” de la tabla “microservices_status” se relaciona con los indicadores “Porcentaje de disponibilidad app web”, “Porcentaje de disponibilidad Envío de mensajes” y “Porcentaje de disponibilidad recepción de transacciones”.
- La columna “id_webhook” de la tabla “wh_webhook_data” se relaciona con el indicador “Total de transacciones”.
- La columna “date_time” de la tabla “wh_webhook_data” se relaciona con una perspectiva “Tiempo”.
- La columna “system” de la tabla “wh_webhook_data” se relaciona con la perspectiva “Sistema de reserva”.
- La tabla “wh_webhook_data” se relaciona con el indicador total de almacenamiento.

2.4.2.3 Nivel de granularidad

A continuación, se van a seleccionar los campos que concentran cada perspectiva.

- Perspectiva “Tiempo” para disponibilidad: los datos que se van a emplear para el análisis son los siguientes:
  - Año
  - Trimestre
  - Mes
  - Semana
- Perspectiva “Tiempo” para transacciones: los datos que se van a emplear para el análisis son los siguientes:
  - Año
  - Mes
  - Día
  - Hora.
- Perspectiva “Sistema”: los datos que se van a emplear para el análisis son los siguientes:
  - Nombre

2.4.2.4 Modelo conceptual ampliado

Luego de analizar el nivel de granularidad, se presenta el modelo conceptual ampliado en la Figura 18.

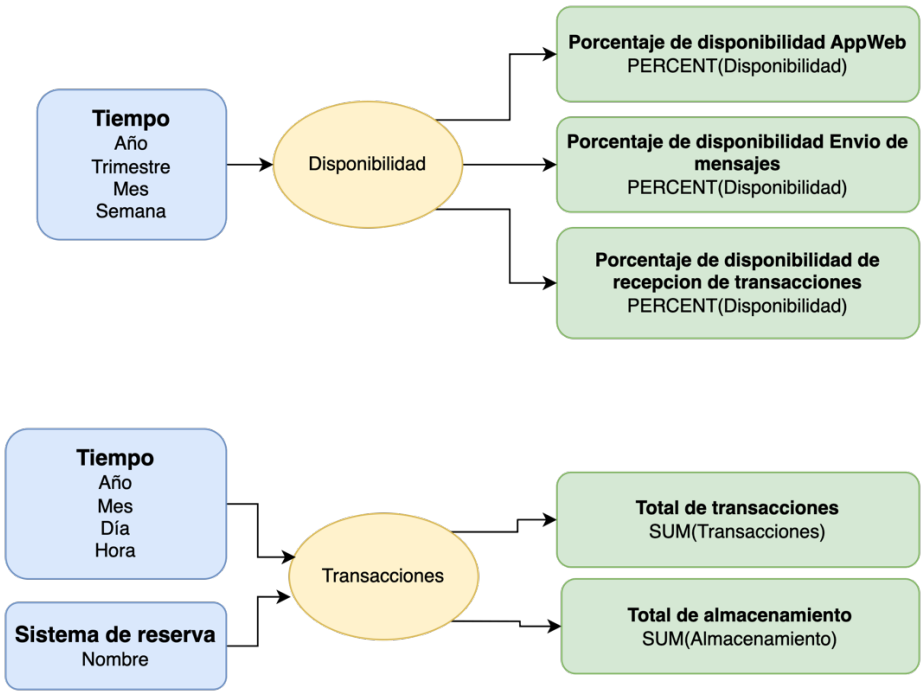


Figura 18 - Modelo conceptual ampliado de desarrollo

2.4.3 Modelo lógico

A continuación, se conforma el modelo lógico teniendo como base el modelo conceptual que ha sido creado.

2.4.3.1 Tipo de modelo lógico

Para este Data Mart de desarrollo se utilizará un esquema de tipo estrella dado que es un modelo simple.

2.4.3.2 Tabla de dimensiones

En este paso se diseñan las tablas de las dimensiones que conforman la Data Mart.

- Perspectiva “Tiempo” para disponibilidad:

La nueva tabla tendrá el nombre “FECHA”.

Tendrá una clave principal con el nombre “idFecha”.

Se mantendrán los campos “Año”, “Trimestre”, “Mes”, “Semana” como se muestra en la Figura 19.

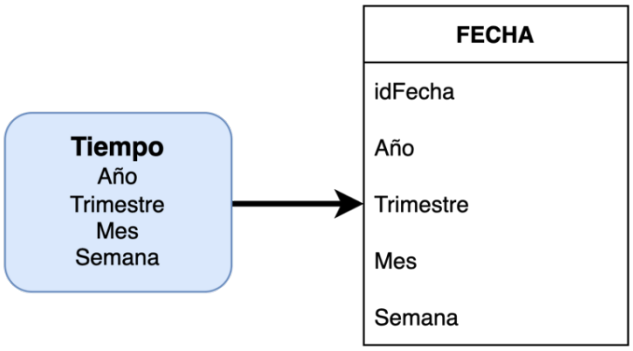


Figura 19 - Tabla de dimensiones Fecha.

- Perspectiva “Tiempo” para transacciones:

La nueva tabla tendrá el nombre “FECHA_HORA”.

Tendrá una clave principal con el nombre “idFecha”.

Se mantendrán los campos “Año”, “Mes”, “Dia”, “Hora”.

La tabla de dimensiones para FECHA_HORA se muestra en la Figura 20.

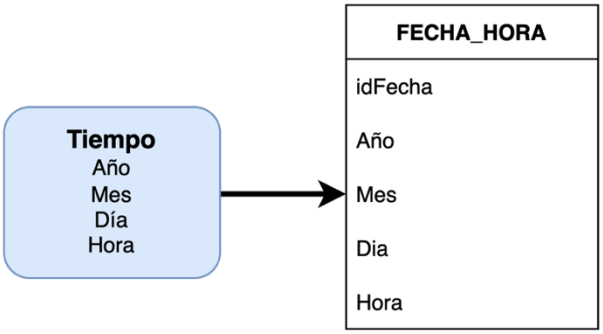


Figura 20 – Tabla de dimensiones FECHA_HORA

- Perspectiva “Sistema de reserva” para transacciones:

La nueva tabla tendrá el nombre “SISTEMA”.

Tendrá una clave principal con el nombre “idSistema”.

Se mantendrán los campos “Nombre”.

La tabla de dimensiones para SISTEMA se muestra en la Figura 21.

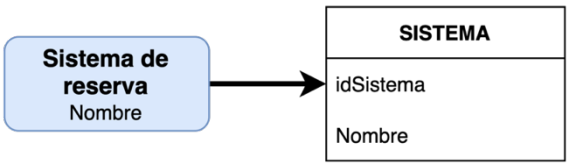


Figura 21 – Tabla de dimensiones SISTEMA.

2.4.3.3 Tabla de hechos

En este paso se define la tabla de hecho, que contendrán los indicadores de estudio.

- La tabla de hechos “DISPONIBILIDAD”.

La clave principal tiene relación con la dimensión “idFecha”.

Se crean los siguientes hechos “AppWeb”, “Envio” y “Recepcion”.

La tabla de hechos para DISPONIBILIDAD se muestra en la Figura 22.

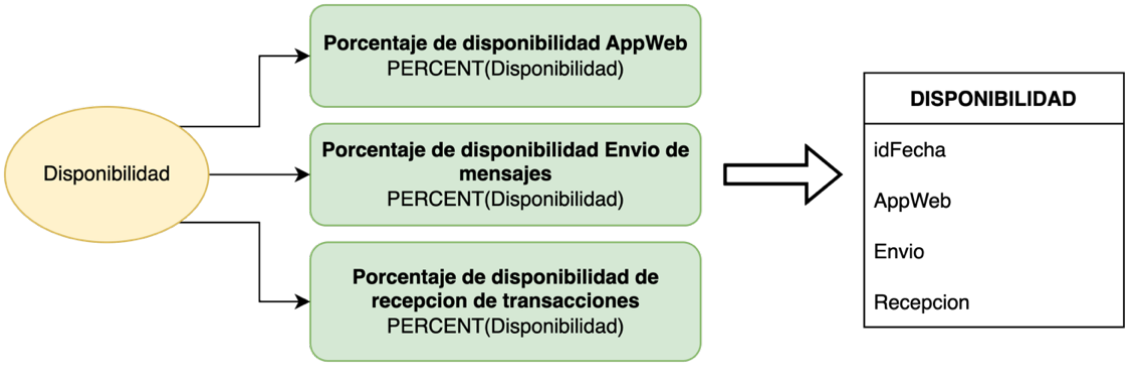


Figura 22 – Tabla de hechos para DISPONIBILIDAD

- La tabla de hechos “TRANSACCIONES”.

Las claves principales tendrán relación con “IdFechaHora” y “idSistema”

Se definen los siguientes hechos: “Transacciones”, “Almacenamiento”.

El diseño de la tabla de hechos “TRANSACCIONES” se muestra en la Figura 23.

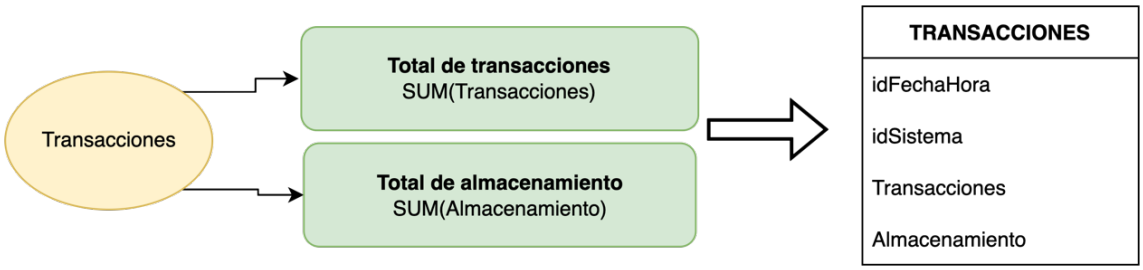


Figura 23 – Tabla de hechos TRANSACCIONES.

2.4.3.4 Uniones

En este paso, se realizan las uniones entre las tablas de dimensiones y las tablas de hechos como se muestra en la Figura 24 y Figura 25

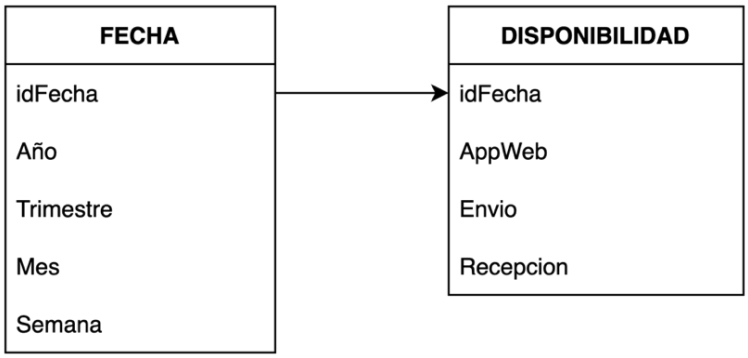


Figura 24 – Uniones para la tabla de hechos DISPONIBILIDAD.

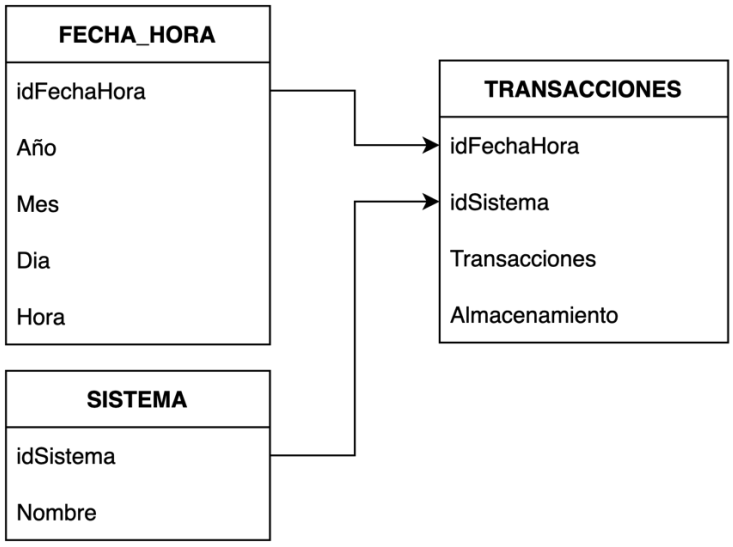


Figura 25 – Uniones para la tabla de hechos TRANSACCIONES.

2.4.4 Integración de datos

A continuación, se define la política de carga, actualización y los procesos que se van a llevar a cabo.

2.4.4.1 Carga inicial

El proceso de carga inicial se muestra en la Figura 26.

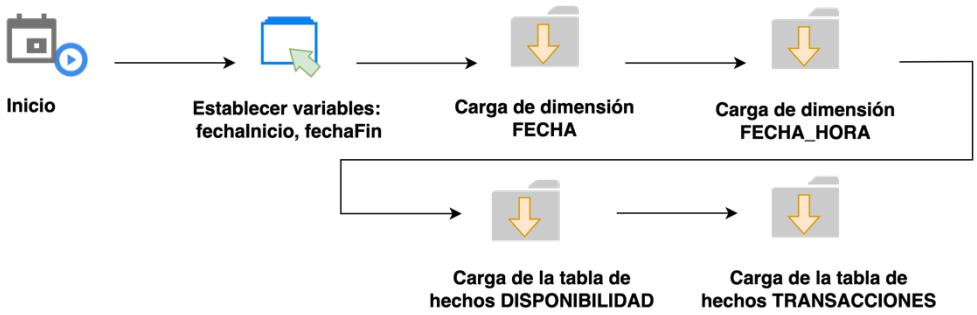


Figura 26 – Carga inicial para el Data Mart de desarrollo.

A continuación, se detallan los pasos a seguir en el proceso de carga inicial.

- Inicio: Inicia el proceso en el momento que se le indique.
- Establecer variables: Se establece la variable “fechaInicio” como la fecha de la primera suscripción y la variable “fechaFin” como la fecha actual.
- Carga dimensión FECHA: Para cargar esta dimensión se reutiliza el proceso detallado en la sección 2.3.4.1.
- Cargar dimensión FECHA_HORA: Similar al caso anterior que genera fechas en el rango desde la fecha de inicio de la fecha de finalización, este script tiene un rango de una hora. El formato para la columna “idFechaHora” se compone de “año” + “mes” + “dia” + “hora”. Cómo ejemplo del resultado de este script se muestra la Tabla 7.

Tabla 7 – Ejemplo de datos para la dimensión FECHA_HORA.

idFecha	Año	Mes	Dia	Hora
2022010100	2022	1	1	00
2022010101	2022	1	1	01
2022010102	2022	1	1	02

Resultado de este script se guarda en la tabla “FECHA_HORA”.

- Carga de la tabla de hechos “DISPONIBILIDAD”: La carga de esta tabla se compone de dos pasos como se muestra en la Figura 27

Figura 27 – Carga de la tabla de hechos disponibilidad.

- **Obtener datos OLTP**

En este paso se carga la información de las columnas “AppWeb”, “Envio”, “Recepcion” con la información de la tabla “microservices_status”, para ello se crea un script que genera todos los rangos de fechas semanales entre la “fechaInicio” y “fechaFinal”, luego para cada rango (fecha1, fecha2) se obtiene la siguiente consulta SQL:

SELECT COUNT (*) AS fallos FROM microservice_status WHERE (microservice_name = “appweb” or microservice_name = “webhook” OR microservice_name = “sms_out”) AND status = “error” AND date_time > fecha1 AND date_time <= fecha2 GROUP BY microservice_name.

El resultado de la consulta presenta la cantidad de veces que el servicio no estuvo disponible en el transcurso de la semana para: appweb (columna “AppWeb”), webhook (columna “Recepcion”) y sms_out (columna “Envio”), para conocer el porcentaje de disponibilidad se considera que se consulta a los diferentes microservicios cada 5 minutos, es decir que cada semana se consultan 2016 veces el estado de los microservicios, por lo tanto, se tiene la Ecuación 1.

$$Porcentaje = \left(1 - \frac{fallos}{2016}\right) * 100$$

Ecuación 1 – Cálculo del porcentaje de disponibilidad.

- **Cargar DISPONIBILIDAD**

Los resultados de los cálculos anteriores para disponibilidad de “AppWeb”, “Envio” y “Recepcion” se guardan en la tabla “DISPONIBILIDAD”.

- Carga de la tabla de hechos “TRANSACCIONES”: La carga de esta tabla se compone de dos pasos como se muestra en la Figura 28.

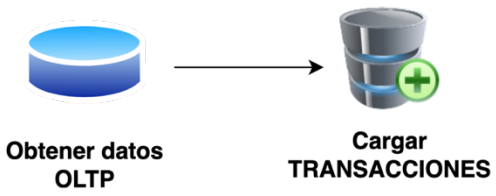


Figura 28 – Carga de la tabla de hechos DISPONIBILIDAD.

- **Obtener datos OLTP**

En esta etapa de carga la columna “Transacciones” con la siguiente consulta SQL:

SELECT system as idSystem, COUNT (*) as Transacciones FROM wh_webhook_data WHERE date_time > fechaInicio and date_time <= fechaFinal GROUP BY hour(date_time) , system

A continuación, se carga la columna “Almacenamiento” con las siguientes consultas SQL:

SELECT round ((data_length / 1024 / 1024), 2) "Size in MB" FROM information_schema.TABLES WHERE table_schema = "touroppgo" AND table_name = "wh_webhook_data";

- **Cargar TRANSACCIONES.**

Los resultados anteriores se almacenan en la tabla “TRANSACCIONES”.

2.4.4.2 Actualización

Las políticas actualización que se definieron son las siguientes:

- La información se actualizará todos los martes a las 12 de la noche.
- Para la tabla de hechos “DISPONIBILIDAD”, las columnas “AppWeb”, “Envio” y “Recepcion” se cargarán de manera incremental teniendo en cuenta la última fecha de actualización.
- Para la tabla de hechos “TRANSACCIONES”, las columnas “Transacciones” y “Almacenamiento” se cargarán de manera incremental de acuerdo con la última fecha de actualización.

El proceso ETL para la actualización es similar al de la carga con las siguientes consideraciones:

- Inicio: el proceso empieza todos los martes a las 12 pm
- La carga de las dimensiones “FECHA” y “FECHA_HORA” no es necesaria ya que al inicio se cargó datos hasta la finalización del año en curso.
- La carga de la dimensión “SISTEMA” sobre escribirá la información en la tabla.
- La carga de la tabla de hechos “TRANSACCIONES” llevará el mismo procedimiento que la carga inicial.

2.4.5 Indicadores dinámicos.

El data mart de transacciones está construido para responder a las preguntas sobre el comportamiento de las transacciones y el almacenamiento en el transcurso del tiempo mientras que para poder conocer el comportamiento futuro es necesario extender el data mart y tomar en cuenta indicadores dinámicos que permitan predecir las nuevas

transacciones, utilizando el modelo que se detalla en la sección 2.9.3, y predecir el almacenamiento de las bases de datos, utilizando el modelo que se detalla en la sección 2.9.4. Para esto se plantea el modelo que se muestra en la Figura 29.

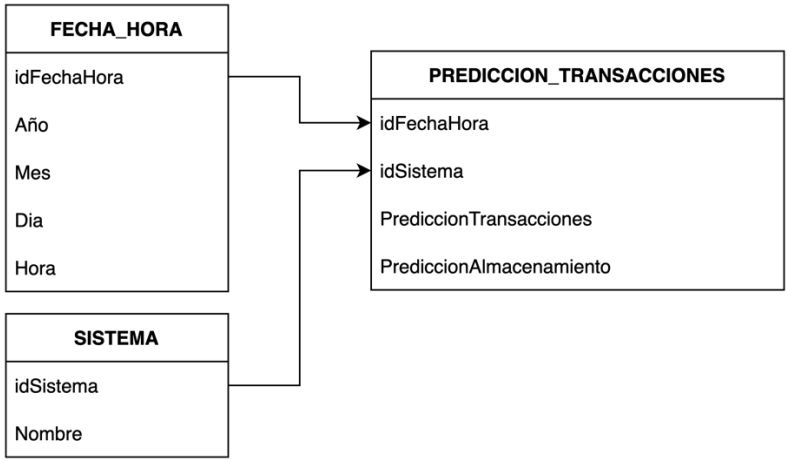


Figura 29 - Modelo para predicción de transacciones.

Este modelo "PREDICCION_TRANSACCIONES" cuenta con las siguientes características:

- La llave primaria "idFechaHora" tiene relación con la tabla de dimensión "FECHA_HORA".
- Se definen los hechos "PrediccionTransacciones" y "PrediccionAlmacenamiento".

El proceso de actualización de este modelo tiene los siguientes los pasos:

- Inicio: La actualización de la tabla "PREDICCION_TRANSACCIONES" se realiza inmediatamente después de la actualización de la tabla de hechos "TRANSACCIONES" ya que esta tabla es la fuente de datos para realizar las predicciones.
- Establecer variables: se establece la variable "fechaInicio" como la fecha actual y "fechaFin" como la fecha actual más 90 días.
- Predicción: La columna "Transacciones" y la columna "Almacenamiento" de la tabla de hechos "TRANSACCIONES" sirven como fuente de datos para alimentar los modelos detallados en la sección 2.9.3 y 2.9.4 respectivamente. Estos modelos devuelven los valores de predicciones diarias hasta la fecha final.
- Carga: Se limpian los registros existentes en la tabla "PREDICCION_TRANSACCIONES" para luego guardar los nuevos valores de predicciones.

2.5 Data Mart para marketing

2.5.1 Análisis de requerimientos

A continuación, se detallan los requerimientos que se recopilaron en conjunto con el equipo de marketing.

2.5.1.1 Identificar preguntas

1. ¿Cuántos leads está generando el equipo de marketing y cuál es la predicción hasta finalizar el año?

2.5.1.2 Análisis de perspectivas e indicadores

1. ¿Cuántos leads está generando el equipo de marketing y cuál es la predicción hasta finalizar el año?

- **Perspectivas:** Tiempo
- **Indicadores:** Total de leads.

2.5.1.3 Modelo conceptual

A continuación, se muestra el modelo conceptual para marketing en la Figura 30.

Figura 30 – Modelo conceptual para el área de marketing.

2.5.2 Análisis de OLTP

En este punto se analizan las fuentes de datos para establecer la forma de cálculo de los indicadores y la relación con el modelo conceptual planteado en el paso anterior.

2.5.2.1 Conformar indicadores

Las operaciones para calcular los indicadores en función de los hechos y funciones es la siguiente:

- **“Total de Leads”**
  - Hechos: Leads.
  - Función de sumarización: SUM

El indicador “Total de leads” representa el total de leads que ha generado el equipo de marketing hasta la fecha.

2.5.2.2 Establecer correspondencias

En la Figura 31 se muestran las fuentes de datos y las relaciones que tiene con el modelo conceptual.

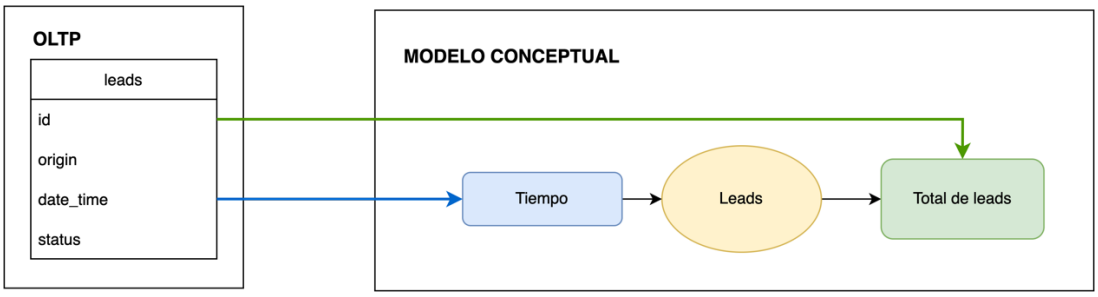


Figura 31 – Correspondencias entre los OLTP y el modelo conceptual.

Las relaciones identificadas son las siguientes:

- La columna “date_time” de la tabla leads se relaciona con la perspectiva “Tiempo”.
- La columna “id” de la tabla leads se relaciona con el indicador “Total de leads”.

2.5.2.3 Nivel de granularidad

A continuación, se van a seleccionar los campos que concentran cada perspectiva.

- Perspectiva “Tiempo” para disponibilidad: los datos que se van a emplear para el análisis son los siguientes:
  - Año
  - Trimestre
  - Mes
  - Semana

2.5.2.4 Modelo conceptual ampliado

Luego de analizar el nivel de granularidad se presenta el modelo conceptual ampliado en la Figura 32.

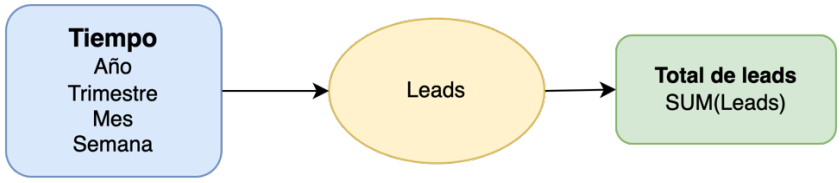


Figura 32 – Modelo conceptual ampliado para marketing.

2.5.3 Modelo lógico

A continuación, se conforma el modelo lógico teniendo como base el modelo conceptual que ha sido creado.

2.5.3.1 Tipo de modelo lógico

Para este Data Mart de marketing se utilizará un esquema de tipo estrella dado que el modelo cuenta con una tabla de dimensión y una de hechos.

2.5.3.2 Tabla de dimensiones

En este paso se diseñan las tablas de las dimensiones que conforman la Data Mart.

- Perspectiva “Tiempo”:

La nueva tabla tendrá el nombre “FECHA”.

Tendrá una clave principal con el nombre “idFecha”.

Se mantendrán los campos “Año”, “Trimestre”, “Mes”, “Semana”.

En la tabla Figura 33 se muestra la tabla de dimensiones FECHA.

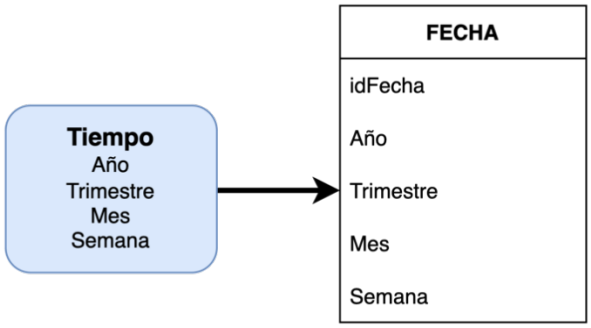


Figura 33 – Tabla de dimensiones FECHA.

2.5.3.3 Tabla de hechos

En este paso se define la tabla de hecho que contendrá el indicador de estudio.

- La tabla de hechos “LEADS”.

La clave principal tiene relación con la dimensión “idFecha”.

Se crean los siguientes hechos. “Leads”.

El diseño de la tabla de hechos “LEADS” se muestra en la Figura 34.

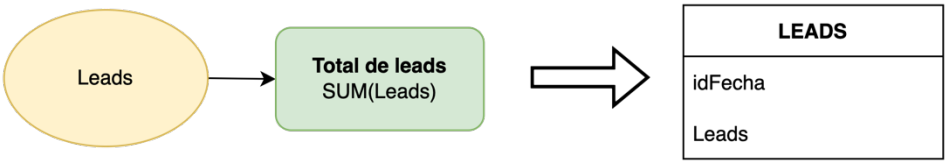


Figura 34 – Tabla de hechos LEADS

2.5.3.4 Uniones

En este paso, se realizan las uniones entre las tablas de dimensiones y las tablas de hechos como se muestra en la Figura 35.

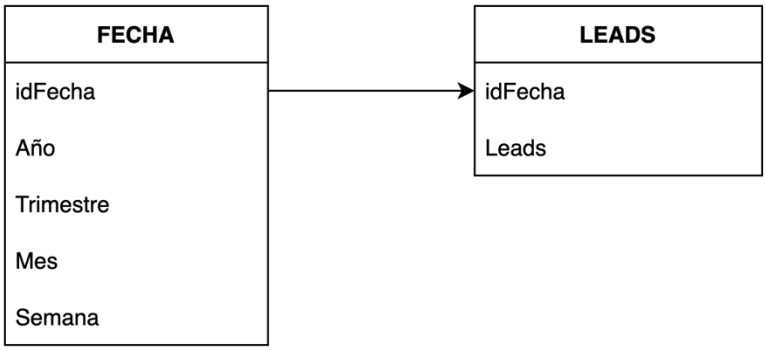


Figura 35 – Uniones en el Data Mart de marketing

2.5.4 Integración de datos

A continuación, se define la política de carga, actualización y los procesos que se van a llevar a cabo.

2.5.4.1 Carga inicial

El proceso de carga inicial se muestra en la Figura 36.

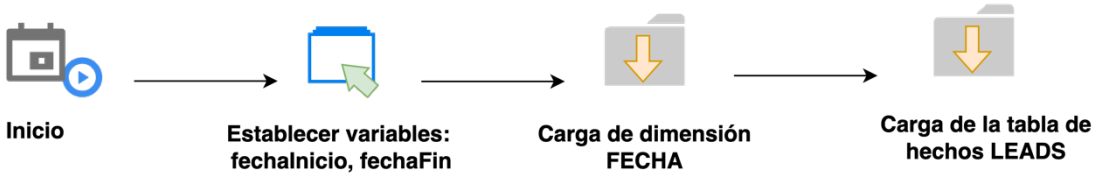


Figura 36 – Carga inicial para el Data Mart de ventas.

Los pasos para seguir en el proceso de carga inicial son los siguientes:

- Inicio: Inicia el proceso en el momento que se le indique.
- Establecer variables: Se establece la variable “fechaInicio” como la fecha de la primera suscripción y la variable “fechaFin” como la fecha actual.
- Carga dimensión FECHA: Para cargar esta dimensión se reutiliza el proceso definido en la sección 2.3.4.1.
- Carga de la tabla de hechos “LEADS”: La carga de esta tabla se compone de dos pasos como se muestra en la Figura 37.

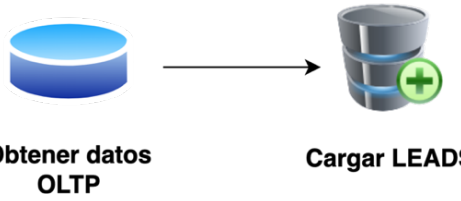


Figura 37 – Carga de la tabla de hechos LEADS

- **Obtener datos OLTP**

En este paso se carga la información de la columna “Leads” con la información de la tabla “leads”, para ello se crea un script que genera todos los rangos de fechas semanales entre la “fechaInicio” y “fechaFinal”, luego para cada rango (fecha1, fecha2) se obtiene la siguiente consulta SQL:

SELECT COUNT (*) AS Leads FROM leads WHERE date_time > fecha1 AND date_time <= fecha2
El resultado de esta consulta presenta la cantidad de leads que se recibieron durante cada semana hasta la fecha.

- **Cargar LEADS**

Los resultados anteriores se guardan en la tabla “LEADS”

2.5.4.2 Actualización

Las políticas de actualización que se definieron son las siguientes:

- La información se actualizará todos los martes a las 12 de la noche.
- Para la tabla de hechos “LEADS”, la columna “Leads” se cargarán de manera incremental teniendo en cuenta la fecha de la última actualización.

EL proceso ETL para la actualización tendrá las siguientes consideraciones:

- Inicio: el proceso empieza todos los martes a las 12 pm
- La carga de las dimensiones “FECHA” y “FECHA_HORA” no es necesaria ya que al inicio se cargó datos hasta la finalización del año en curso.
- La carga de la tabla de hechos “LEADS” tendrá el mismo proceso que en la carga inicial.

2.5.5 Indicadores dinámicos.

El data mart de leads está construido para responder a las preguntas sobre el comportamiento de los leads en el transcurso del tiempo mientras que para poder conocer el comportamiento futuro es necesario extender el data mart y tomar en cuenta indicadores dinámicos que permitan predecir los nuevos leads, utilizando el modelo que se detalla en la sección 2.9.5. Para esto se plantea el modelo que se muestra en la Figura 38

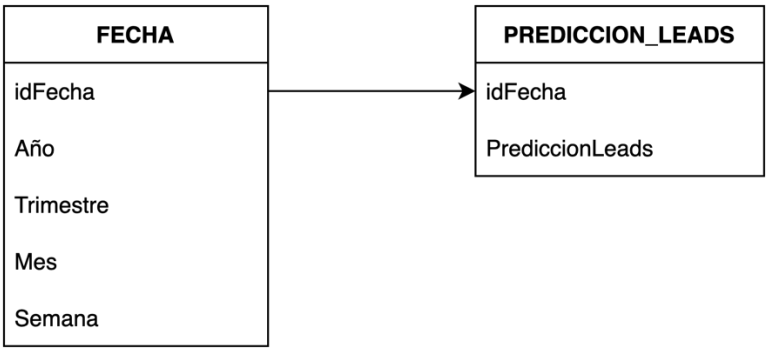


Figura 38 - Modelo para predicción de leads.

Este modelo “PREDICCION_LEADS” cuenta con las siguientes características:

- La llave primaria “idFecha” tiene relación con la tabla de dimensión “FECHA”.
- Se define el hecho “PrediccionLeads”.

El proceso de actualización de este modelo tiene los siguientes los pasos:

- Inicio: La actualización de la tabla “PREDICCION_LEADS” se realiza inmediatamente después de la actualización de la tabla de hechos “LEADS” ya que esta tabla es la fuente de datos para realizar las predicciones.
- Establecer variables: se establece la variable “fechaInicio” como la fecha actual y “fechaFin” como 31 de diciembre del año en curso.
- Predicción: La columna “Leads” de la tabla de hechos “LEADS” sirve como fuente de datos para alimentar el modelo de la sección 2.9.5. Este modelo devuelve las predicciones semanales hasta la fecha final.
- Carga: Se limpian los registros existentes en la tabla “PREDICCION_LEADS” para luego guardar los nuevos valores de predicciones.

2.6 Data Mart para clientes finales

2.6.1 Análisis de requerimientos

En este apartado se busca analizar a los clientes finales de la empresa para poder conocer sus características principales y la forma en la que interactúan con el servicio para entender cómo mejorarlo.

2.6.1.1 Identificar preguntas

1. ¿De qué países provienen nuestros clientes finales?
2. ¿Cuáles son las características principales de los clientes finales?

Entre las características principales que serían de utilidad para poder identificar oportunidades de mejora se analizan: El género del cliente (Masculino o Femenino), si aceptó o no recibir el servicio, si calificó o no su actividad.

3. ¿Qué tipos de clientes finales se pueden distinguir?

Los tipos de clientes se identificarán de acuerdo a la forma en la que los mismos interactúan con el servicio, estos tipos se van a determinar en base a un modelo de inteligencia artificial que agrupe los clientes finales en función de la información proporcionada sobre la interacción que tuvieron los mismos con el sistema.

4. ¿Cuáles son los usuarios que tienen en mayor medida los mejores tipos de clientes finales?

En base al análisis anterior lo que se busca es conocer cuáles son los usuarios que poseen los mejores clientes finales.

2.6.1.2 Análisis de perspectivas e indicadores

1. ¿De qué países provienen nuestros clientes finales?

- **Perspectivas:** Ubicación
- **Indicadores:** Total de clientes finales.

2. ¿Cuáles son las características principales de los clientes finales?

- **Perspectivas:** Características
- **Indicadores:** Total de clientes finales.

3. ¿Qué tipos de clientes finales se pueden distinguir?

- **Perspectivas:** Tipo
- **Indicadores:** Total de clientes finales.

4. ¿Cuáles son los usuarios que tienen en mayor medida los mejores tipos de clientes finales?

- **Perspectivas:** Usuario, Tipo
- **Indicadores:** Total de clientes finales

2.6.1.3 Modelo conceptual

A continuación, en la Figura 39 se muestra el modelo conceptual para el Data Mart para el análisis de clientes finales.

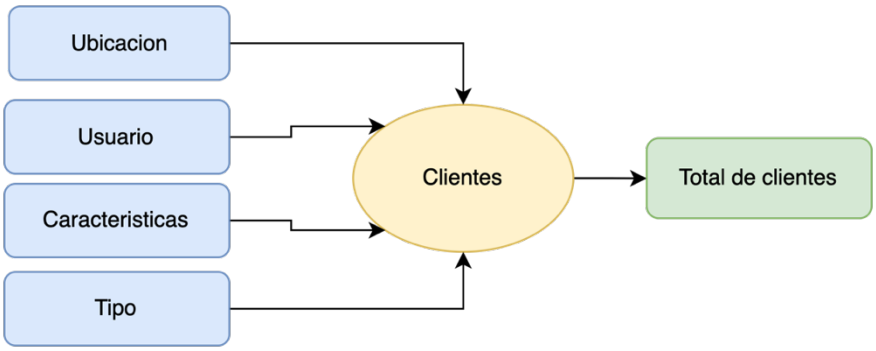


Figura 39 – Modelo conceptual para clientes finales.

2.6.2 Análisis de OLTP

A continuación, se analizan las fuentes de datos para establecer el cálculo de los indicadores y la relación con el modelo conceptual anteriormente definido.

2.6.2.1 Conformar indicadores

La forma en la que se van a calcular los respectivos indicadores en función de los hechos y las funciones es la siguiente:

- **"Total de clientes finales"**
  - Hechos: Clientes finales
  - Función de sumarización: SUM

2.6.2.2 Establecer correspondencias

En la Figura 40 se muestran las fuentes de datos OLTP y las relaciones que tiene con el modelo conceptual.

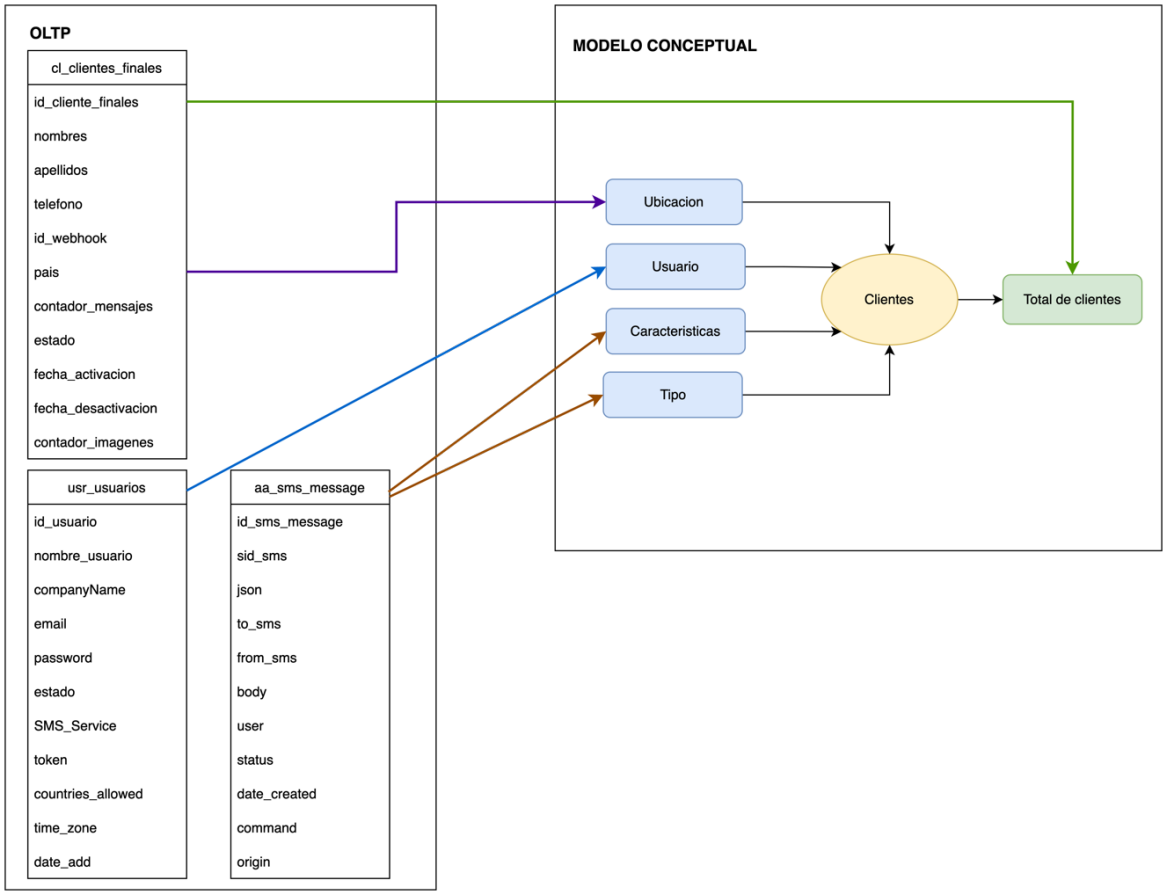


Figura 40 – Correspondencias entre los OLTP y el modelo conceptual de clientes finales.

Las relaciones que se identificaron son las siguientes:

- La columna “país” de la tabla “cl_clientes_finales” se relaciona con la perspectiva “Ubicación”.
- La tabla “usr_usuarios” se relaciona con la perspectiva “Usuario”.
- La tabla “aa_sms_message” se relaciona con las perspectivas “Caracteristicas” y “Tipo”. Cabe mencionar que el indicador “Tipo” se determinará con ayuda del modelo de clasificación de clientes finales descrito en la sección 2.9.7.
- La columna “id_cliente_finales” de la tabla “cl_clientes_finales” se relaciona con el indicador “Total de clientes”.

2.6.2.3 Nivel de granularidad

A continuación, se van a seleccionar los campos que concentran cada perspectiva.

Perspectiva “Ubicacion”: los datos que se consideran son los siguientes:

- Pais.

Perspectiva “Usuario”: los datos a considerar son los siguientes:

- Nombre.
- Sistema

Perspectiva “Caracteristicas”: se consideran los siguientes datos.

- Genero
- Aceptacion
- Califica

Perspectiva “Tipo”: con ayuda del modelo de la sección 2.9.7. se identificaron los siguientes tipos de clientes finales.

- Comun
- Normal
- Ideal

2.6.2.4 Modelo conceptual ampliado

Luego de analizar el nivel de granularidad, en la Figura 41 se presenta el siguiente modelo conceptual ampliado:

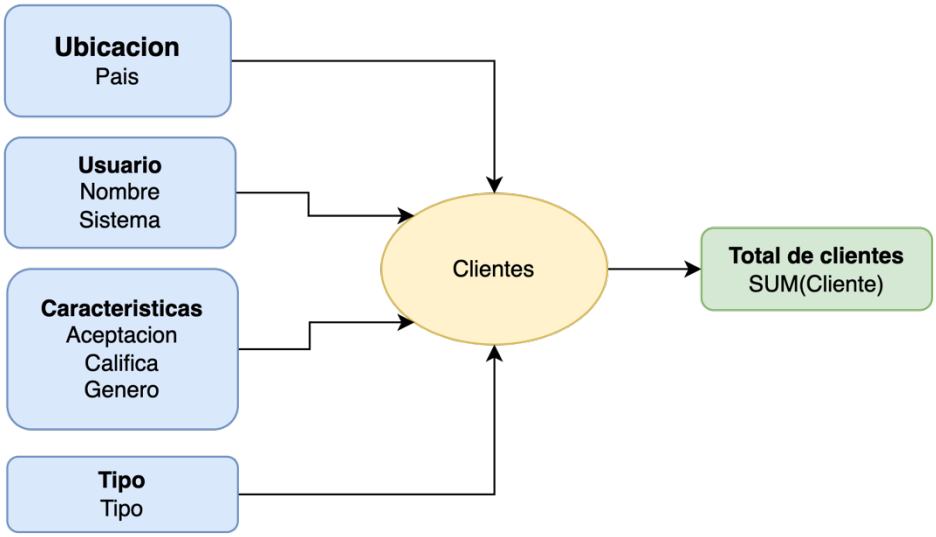


Figura 41 – Modelo conceptual ampliado para clientes finales.

2.6.3 Modelo lógico

A continuación, se conforma el modelo lógico teniendo como base el modelo conceptual que ha sido creado.

2.6.3.1 Tipo de modelo lógico

Para este Data Mart de desarrollo se utilizará un esquema de tipo estrella dado que es un modelo muy simple.

2.6.3.2 Tabla de dimensiones

En este paso se diseñan las tablas de las dimensiones que conforman la Data Mart.

- Perspectiva “Ubicacion”:

La nueva tabla tendrá el nombre “UBICACION”.

Tendrá una clave principal con el nombre “idUbicacion”

Tendrá también la columna: Pais, como se muestra en la Figura 42.

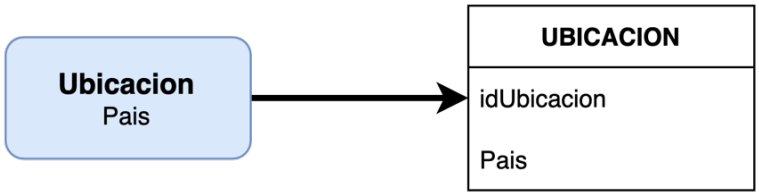


Figura 42 – Tabla de dimensión UBICACION

- Perspectiva “Usuario”:

La tabla tendrá el nombre: “USUARIO”

La tabla tendrá la clave principal con el nombre “idUsuario”

La tabla tendrá las columnas: “Nombre” y “Sistema” como se muestra en la Figura 43.

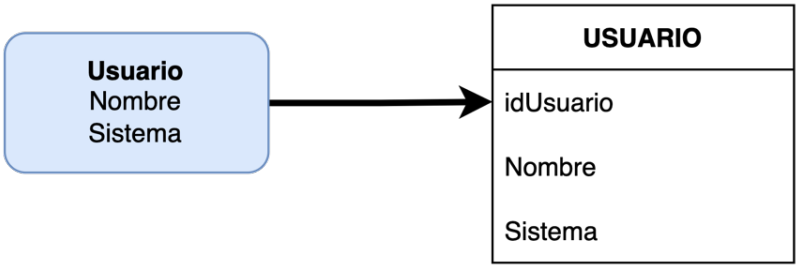


Figura 43 – Tabla de dimensión USUARIO

- Perspectiva “Caracteristicas”:

La tabla tendrá el nombre: “CARACTERISTICA”

La tabla tendrá la clave principal con el nombre “idCaracteristica”

La tabla tendrá las siguientes columnas: “Aceptacion”, “Califica” y “Genero” como se muestra en la Figura 44.

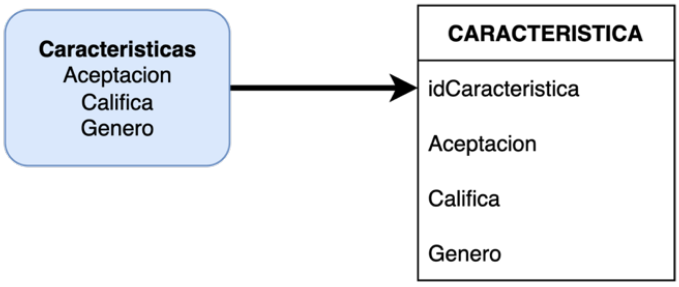


Figura 44 – Tabla de dimensión CARACTERISTICA

- Perspectiva “Tipo”:

La tabla tendrá el nombre: “TIPO”

La tabla tendrá la clave principal con el nombre “idTipo”

La tabla tendrá las siguientes columnas: “Tipo” como se muestra en la Figura 45.

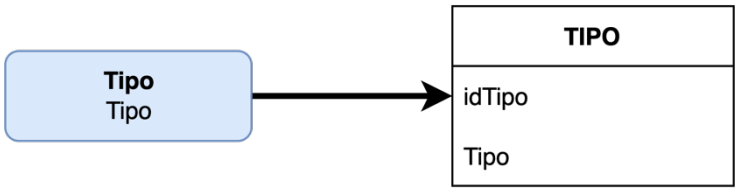


Figura 45 – Tabla de dimensión TIPO.

2.6.3.3 Tabla de hechos

En este paso se define la tabla de hecho, que contendrán los indicadores de estudio.

- La tabla de hechos “Total de clientes”.

La clave principal tiene relación con la dimensión “idUbicacion”, “idUsuario”, “idCaracteristica”.

Se crean los siguientes hechos: “CLIENTES” como se muestra en la Figura 46.

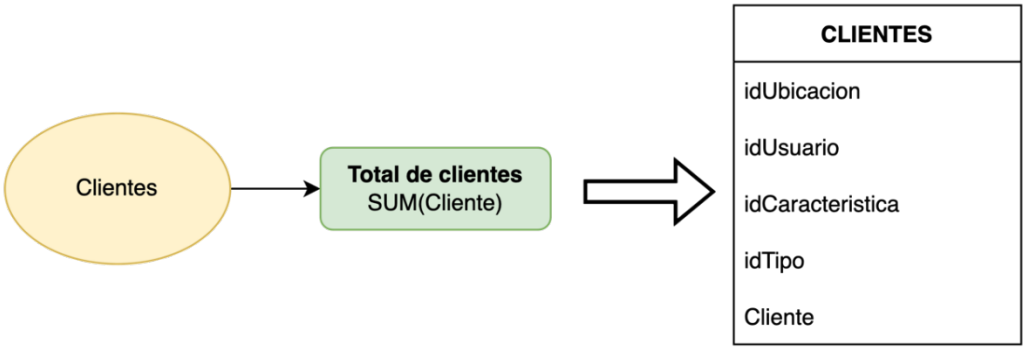


Figura 46 – Tabla de hechos CLIENTES.

2.6.3.4 Uniones

En este paso, se realizan las uniones entre las tablas de dimensiones y las tablas de hechos como en la Figura 47.

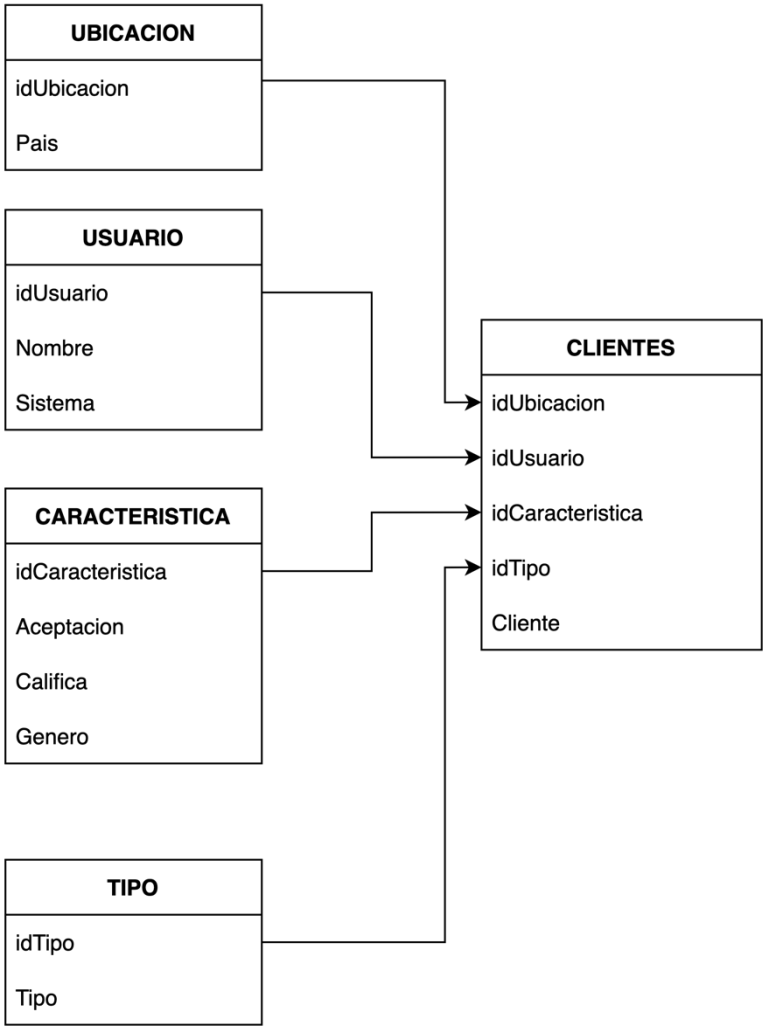


Figura 47 – Uniones para el Data Mart de clientes finales.

2.6.4 Integración de datos

A continuación, se define la política de carga, actualización y los procesos que se van a llevar a cabo.

2.6.4.1 Carga inicial

El proceso de carga inicial se muestra en la Figura 48.

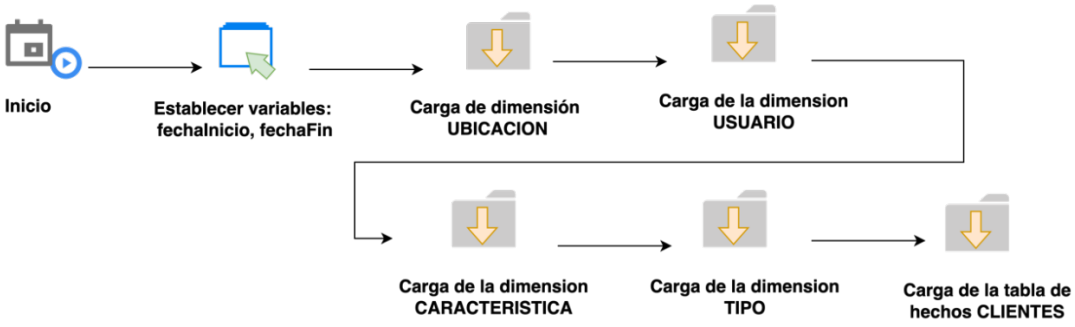


Figura 48 – Carga inicial del Data Mart de clientes finales.

A continuación, se detallan los pasos a seguir en el proceso de carga inicial.

- Inicio: Inicia el proceso en el momento que se le indique.
- Establecer variables: Se establece la variable “fechaInicio” como la fecha de la primera suscripción y la variable “fechaFin” como la fecha actual.
- Carga dimensión UBICACION: Esta tabla se compone de una lista de todos los códigos de país que existen.
- Carga dimensión USUARIO: Para esta tabla se realiza la siguiente consulta SQL:

SELECT id_usuario as idUsuario, usuario as Usuario FROM usr_usuarios.

El resultado se guarda en la tabla USUARIO.

- Carga dimensión CARACTERISTICA: Para esta tabla se realiza un script que genera las posibles combinaciones para las columnas “Aceptacion” (1, 0), “Califica” (1,0), “Genero” (1,0)

Como ejemplo de resultado se muestra la Tabla 8.

Tabla 8- Datos para la dimensión CARACTERISTICA.

idCaracterisitca	Aceptacion	Califica	Genero
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

Este resultado se guarda en la tabla “CARACTERISTICA”

- Carga dimensión TIPO: Para esta tabla se cargan los siguientes valores: “Comun”, “Normal”, “Ideal” como se muestra la Tabla 9.

Tabla 9 – Datos para la dimensión TIPO.

idTipo	Nombre
1	Comun
2	Normal
3	Ideal

Este resultado se guarda en la tabla “TIPO”

- Carga de la tabla de hechos “CLIENTES”: La carga de esta tabla se compone de dos pasos como se muestra en la Figura 49.

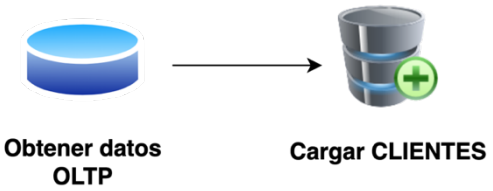


Figura 49 – Carga de la tabla de hechos CLIENTES.

2.6.4.2 Obtener datos OLTP

En este paso se carga la información de la columna “Clientes” con la información de la tabla “cl_clientes_finales”, para ello se crea un script que genera todos los rangos de fechas semanales entre la “fechaInicio” y “fechaFinal”, luego para cada rango (fecha1, fecha2) se obtiene la siguiente consulta SQL:

SELECT id_cliente_final AS Cliente, país AS idUbicacion, usuario as idUsuario FROM cl_clientes_finales WHERE fecha_activacion > fecha1 AND fecha_activacion <= fecha2;

El resultado de esta consulta presenta un identificador de cada cliente final, A continuación, se corre un script que tiene la tarea de determinar el clúster para el “Tipo” al que pertenece el cliente según sus características haciendo uso del modelo de inteligencia artificial de la sección 2.9.7.

2.6.4.3 Cargar CLIENTES

Los resultados anteriores se guardan en la tabla “CLIENTES”

2.6.4.4 Actualización

Las políticas actualización que se definieron son las siguientes:

- La información se actualizará todos los martes a las 12 de la noche.
- La dimensión “UBICACIÓN” no se debe volver a cargar dado que en la carga inicial ya contendrá todos los posibles valores.
- La dimensión “USUARIO” se cargará incrementalmente, es decir se aumentarán los nuevos usuarios que se suscriban al servicio de la empresa.
- La dimensión “CARACTERISTICA” no se debe volver a cargar dado que en la carga inicial se le pasa todos los posibles casos.
- La dimensión “TIPO” no se debe volver a cargar dado que en la carga inicial ya contendrá todos los posibles valores.
- La tabla de hechos “CLIENTES” se cargará incrementalmente considerando aquellos clientes finales nuevos que aparecen después de la última actualización.

2.7 Data Mart para usuarios

2.7.1 Análisis de requerimientos

En esta sección se busca analizar a los usuarios que se han suscrito al servicio de la empresa para poder conocer sus características principales.

2.7.1.1 Identificar preguntas

1. ¿De qué países provienen nuestros usuarios?
2. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el volumen de mensajes utilizados y transacciones recibidas?
3. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el rating con el que le han calificado sus usuarios?

2.7.1.2 Análisis de perspectivas e indicadores

1. ¿De qué países provienen nuestros usuarios?
- **Perspectivas:** Ubicación
- **Indicadores:** Total de usuarios

2. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el volumen de mensajes utilizados y transacciones recibidas?
- **Perspectivas:** Volumen
- **Indicadores:** Total de usuarios

3. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el rating con el que le han calificado sus usuarios?
- **Perspectivas:** Rating
- **Indicadores:** Total de usuarios

2.7.1.3 Modelo conceptual

A continuación, en la Figura 50 se muestra el modelo conceptual para el Data Mart para el análisis de usuarios.

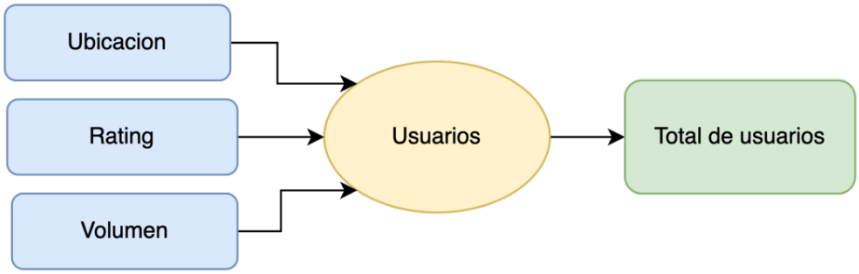


Figura 50 – Modelo conceptual para usuarios.

2.7.2 Análisis de OLTP

A continuación, se analizan las fuentes de datos para establecer el cálculo de los indicadores y la relación con el modelo conceptual anteriormente definido.

2.7.2.1 Conformar indicadores

La forma en la que se van a calcular los respectivos indicadores en función de los hechos y las funciones es la siguiente:

- **"Total de usuarios"**
  - Hechos: Usuarios
  - Función de sumarización: SUM

2.7.2.2 Establecer correspondencias

En la Figura 51 se muestran las fuentes de datos OLTP y las relaciones que tiene con el modelo conceptual.

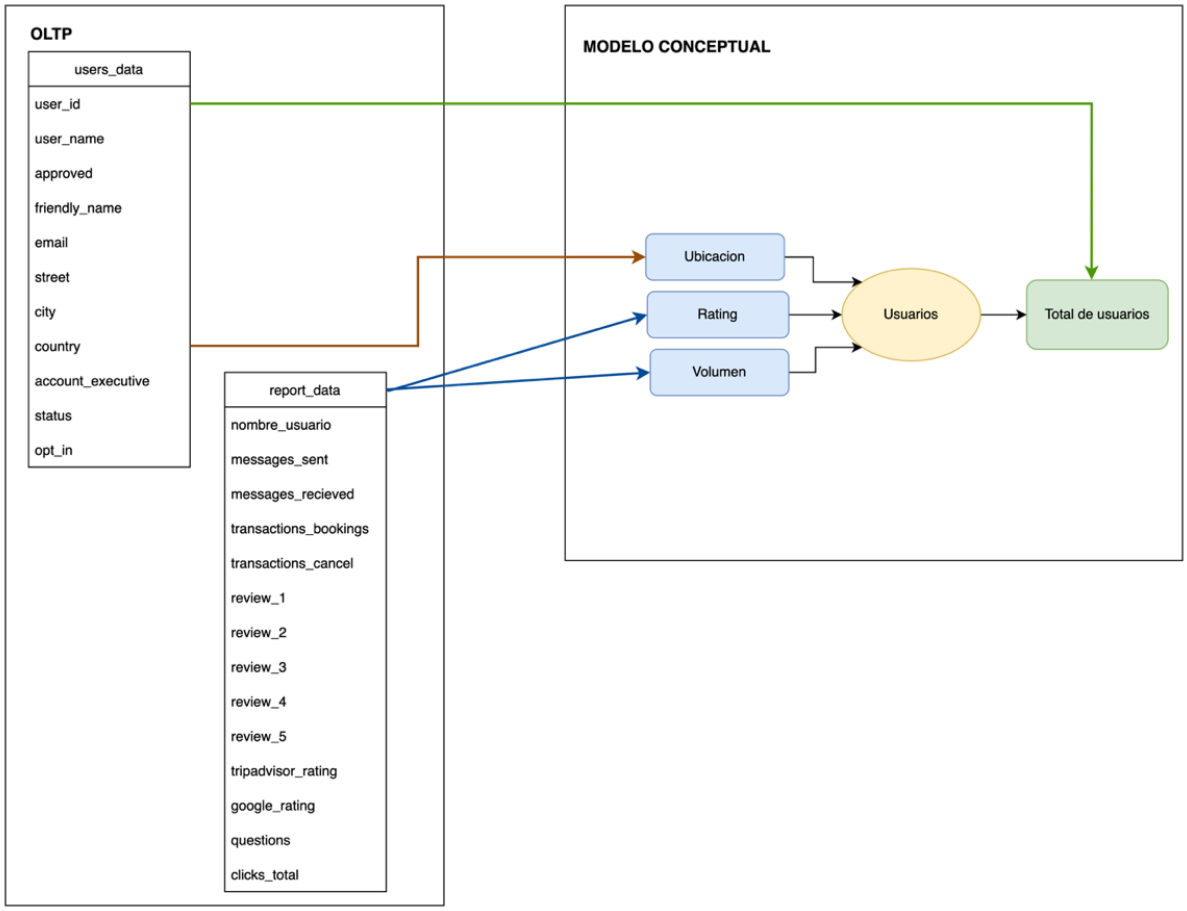


Figura 51 – Correspondencias entre los OLTP y el modelo conceptual de usuarios.

Las relaciones que se identificaron son las siguientes:

- La columna “user_id” de la tabla “users_data” se relaciona con el indicador “Total de usuarios”.
- La tabla “report_data” se relaciona con las perspectivas “Rating” y “Volumen” dado que la información de esta tabla será la fuente de datos para alimentar los modelos de clustering para determinar el grupo al que corresponde el usuario de acuerdo a su rating (sección 2.9.9) y su volumen (sección 2.9.10).

2.7.2.3 Nivel de granularidad

A continuación, se van a seleccionar los campos que concentran cada perspectiva.

Perspectiva “Ubicacion”: los datos que se consideran son los siguientes:

- País.

Perspectiva “Rating”: con ayuda del modelo de la sección 2.9.9 se identificaron los siguientes tipos de usuarios en base al rating “Bueno” y “Excelente”. Los datos que se consideran son:

- Rating

Perspectiva “Volumen”: con ayuda del modelo de la sección 2.9.10 se identificaron los siguientes tipos de usuarios en base al volumen: “Pequeño”, “Mediano” y “Grande”. Los datos que se consideran son:

- Volumen

2.7.2.4 Modelo conceptual ampliado

Luego de analizar el nivel de granularidad, se presenta el siguiente modelo conceptual ampliado:

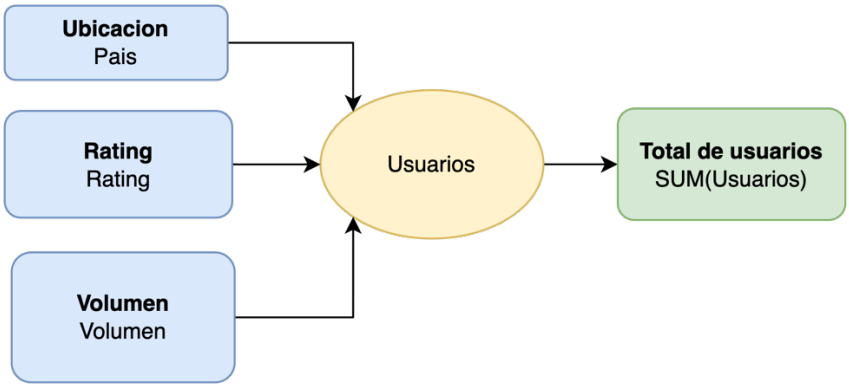


Figura 52 – Modelo conceptual ampliado para usuarios.

2.7.3 Modelo lógico

A continuación, se conforma el modelo lógico teniendo como base el modelo conceptual que ha sido creado.

2.7.3.1 Tipo de modelo lógico

Para este Data Mart de desarrollo se utilizará un esquema de tipo estrella dado que es un modelo simple y la cantidad de perspectivas e indicadores es reducida.

2.7.3.2 Tabla de dimensiones

En este paso se diseñan las tablas de las dimensiones que conforman la Data Mart.

- Perspectiva “Ubicacion”:

La nueva tabla tendrá el nombre “UBICACION”.

Tendrá una clave principal con el nombre “idUbicacion”

Tendrá también la columna: País como se muestra en la Figura 53

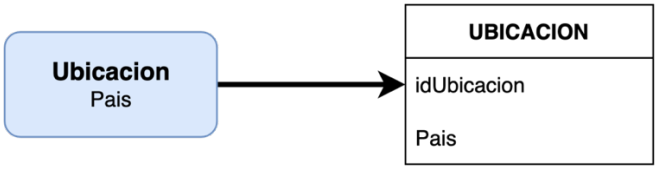


Figura 53 – Tabla de dimensión UBICACIÓN.

- Perspectiva “Rating”:

La tabla tendrá el nombre: “RATING”

La tabla tendrá la clave principal con el nombre “idRating”

La tabla tendrá la siguiente columna: “Rating” como se muestra en la Figura 54.

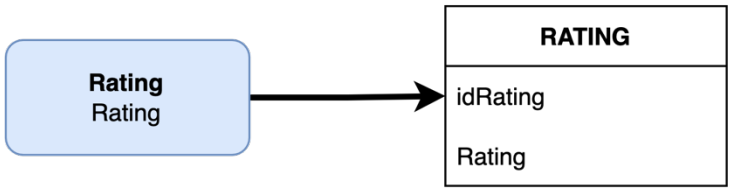


Figura 54 - Tabla de dimensión RATING.

- Perspectiva “Volumen”:

La tabla tendrá el nombre: “VOLUMEN”

La tabla tendrá la clave principal con el nombre “idVolumen”

La tabla tendrá la siguiente columna: “Volumen” como se muestra en la Figura 55.

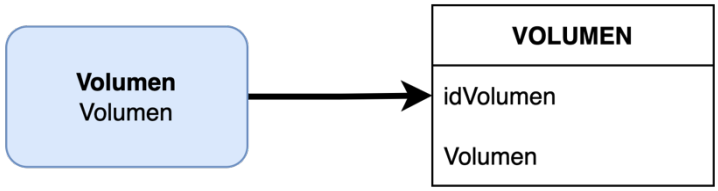


Figura 55 – Tabla de dimensión VOLUMEN

2.7.3.3 Tabla de hechos

En este paso se define la tabla de hecho, que contendrán los indicadores de estudio.

- La tabla de hechos “Total de usuarios”.

La clave principal tiene relación con la dimensión “idUbicacion”, “idRating”, “idVolumen”.

Se crean los siguientes hechos: “USUARIOS”. Como se muestra en la Figura 56.

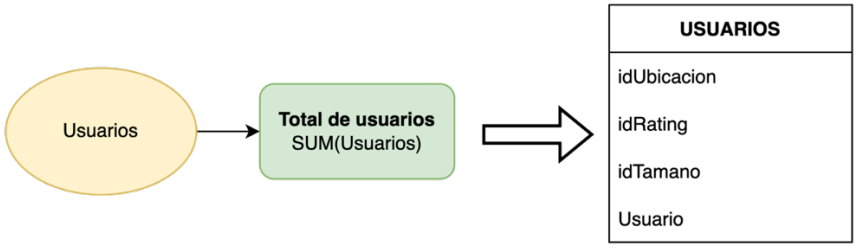


Figura 56 – Tabla de hechos USUARIOS.

2.7.3.4 Uniones

En este paso, se realizan las uniones entre las tablas de dimensiones y las tablas de hechos como se muestra en la Figura 57.

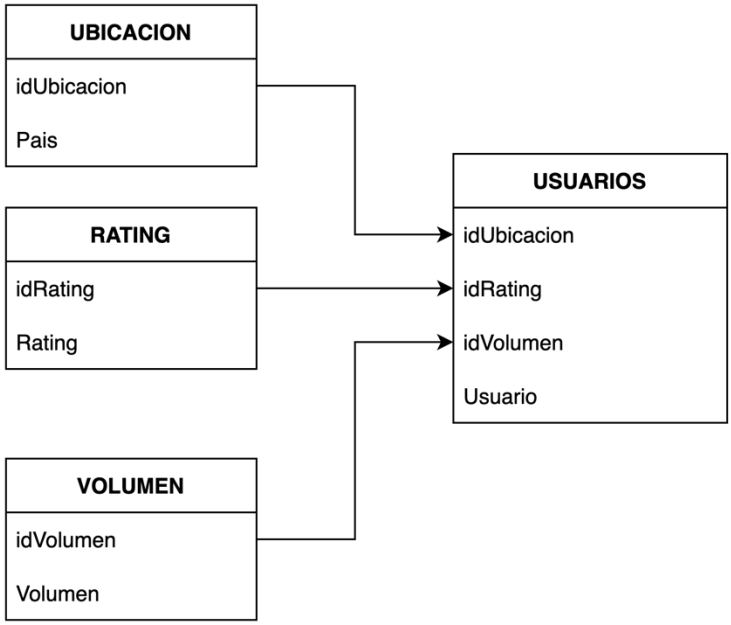


Figura 57 – Uniones para el Data Mart de usuarios

2.7.4 Integración de datos

A continuación, se define la política de carga, actualización y los procesos que se van a llevar a cabo.

2.7.4.1 Carga inicial

El proceso de carga inicial se muestra en la Figura 58.

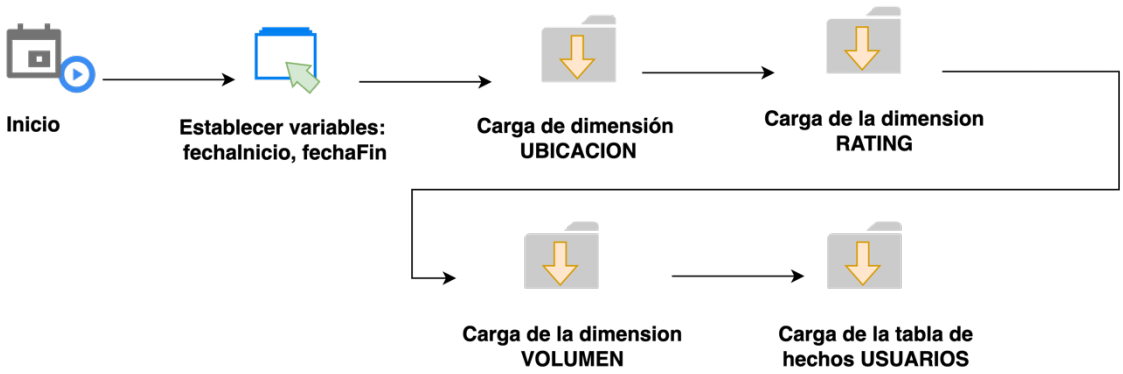


Figura 58 – Carga inicial para el Data Mart de usuarios.

A continuación, se detallan los pasos a seguir en el proceso de carga inicial.

- Inicio: Inicia el proceso en el momento que se le indique.
- Establecer variables: Se establece la variable “fechaInicio” como la fecha de la primera suscripción y la variable “fechaFin” como la fecha actual.
- Carga dimensión UBICACION: Esta tabla se compone de una lista de todos los códigos de país que existen.

- Carga dimensión RATING: Para esta tabla se cargan los valores “Bueno” y “Excelente” como se muestra en la Tabla 10.

Tabla 10 – Datos para la dimensión RATING.

idRating	Rating
1	Bueno
2	Excelente

- Carga dimensión VOLUMEN: Para esta tabla se guardan los valores “Pequeno”, “Mediano” y “Grande” como se muestra en la Tabla 11.

Tabla 11 – Datos para la dimensión VOLUMEN.

idVolumen	Volumen
1	Pequeño
2	Mediano
3	Grande

- Carga de la tabla de hechos “USUARIOS”: La carga de esta tabla se compone de dos pasos como se muestra en la Figura 59.

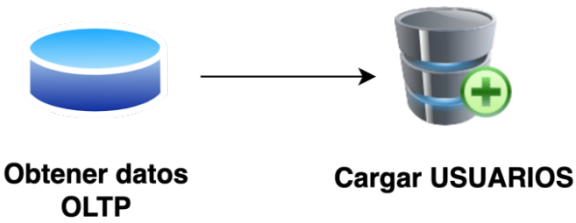


Figura 59 – Carga de la tabla de hechos USUARIOS.

2.7.4.2 Obtener datos OLTP

En este paso se carga la información de la columna “Usuario” con la información de la tabla “users_data”, para ello se crea un script que consulta entre la “fechaInicio” y “fechaFinal” la siguiente consulta SQL:

```
SELECT user_id AS Usuario, country AS idUbicacion FROM users_data
WHERE approved > fechaInicio AND approved <= fechaFinal;
```

El resultado de esta consulta presenta un identificador de cada usuario, A continuación, se corre un script que tiene la tarea de determinar el clúster para el “Rating” que se detalla más a profundidad en la sección 0 y el “Volumen” al que pertenece el cliente según sus

características haciendo uso del modelo de inteligencia artificial como que se detalla en la sección 2.9.10.

2.7.4.3 Cargar USUARIOS

Los resultados anteriores se guardan en la tabla “USUARIOS”

2.7.4.4 Actualización

Las políticas actualización que se definieron son las siguientes:

- La información se actualizará todos los martes a las 12 de la noche.
- La dimensión “UBICACIÓN” no se debe volver a cargar dado que en la carga inicial ya contendrá todos los posibles valores.
- La dimensión “RATING” no se debe volver a cargar dado que en la carga inicial ya contendrá todos los posibles valores.
- La dimensión “VOLUMEN” no se debe volver a cargar dado que en la carga inicial ya contendrá todos los posibles valores.
- La tabla de hechos “USUARIOS” se cargará incrementalmente considerando aquellos nuevos usuarios que se suscriban después de la última actualización.

2.8 Data Mart para chatbot

2.8.1 Análisis de requerimientos

2.8.1.1 Identificar preguntas

1. ¿Cuáles son los temas principales de las preguntas que el chatbot no entendió?

Esta pregunta es de gran importancia para poder conocer cuáles son los temas sobre los que los clientes finales preguntan con más frecuencia y que el chatbot no les puede responder. De esta manera se buscar identificar los temas de las preguntas para poder entrenar mejor al chatbot.

2. ¿Cuáles son los comandos más utilizados por el chatbot en sus respuestas?

Esta pregunta permite identificar el uso de los diferentes comandos de chatbot.

2.8.1.2 Análisis de perspectivas e indicadores

1. ¿Cuáles son los temas principales de las preguntas que el chatbot no entendió?

- **Perspectivas:** Tema, Tiempo
- **Indicadores:** Total de preguntas

3. ¿Cuáles son los comandos más utilizados por el chatbot en sus respuestas?

- **Perspectivas:** Comando, Tiempo
- **Indicadores:** Total de respuestas

2.8.1.3 Modelo conceptual

A continuación, en la Figura 60 se muestra el modelo conceptual para el Data Mart para el análisis del chatbot.

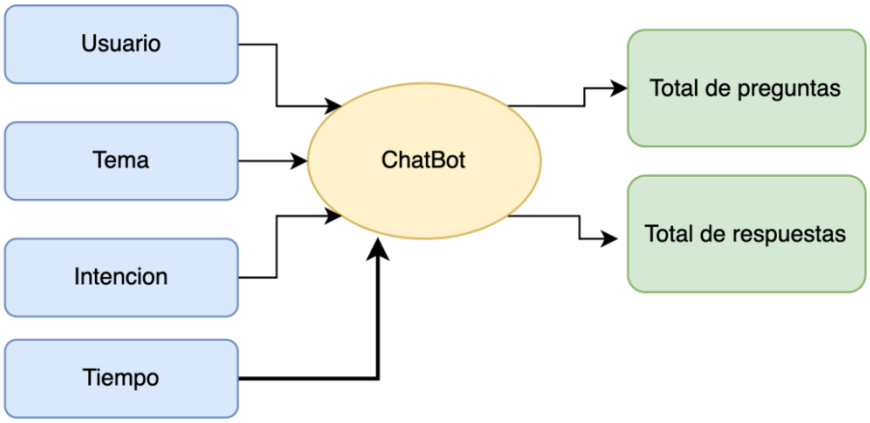


Figura 60 – Modelo conceptual para chatbot.

2.8.2 Análisis de OLTP

A continuación, se analizan las fuentes de datos para establecer el cálculo de los indicadores y la relación con el modelo conceptual anteriormente definido.

2.8.2.1 Conformar indicadores

La forma en la que se van a calcular los respectivos indicadores en función de los hechos y las funciones es la siguiente:

- **"Total de preguntas"**
  - Hechos: Preguntas
  - Función de sumarización: SUM
- **"Total de respuestas"**
  - Hechos: Respuestas
  - Función de sumarización: SUM

2.8.2.2 Establecer correspondencias

En la Figura 61 se muestran las fuentes de datos OLTP y las relaciones que tiene con el modelo conceptual.

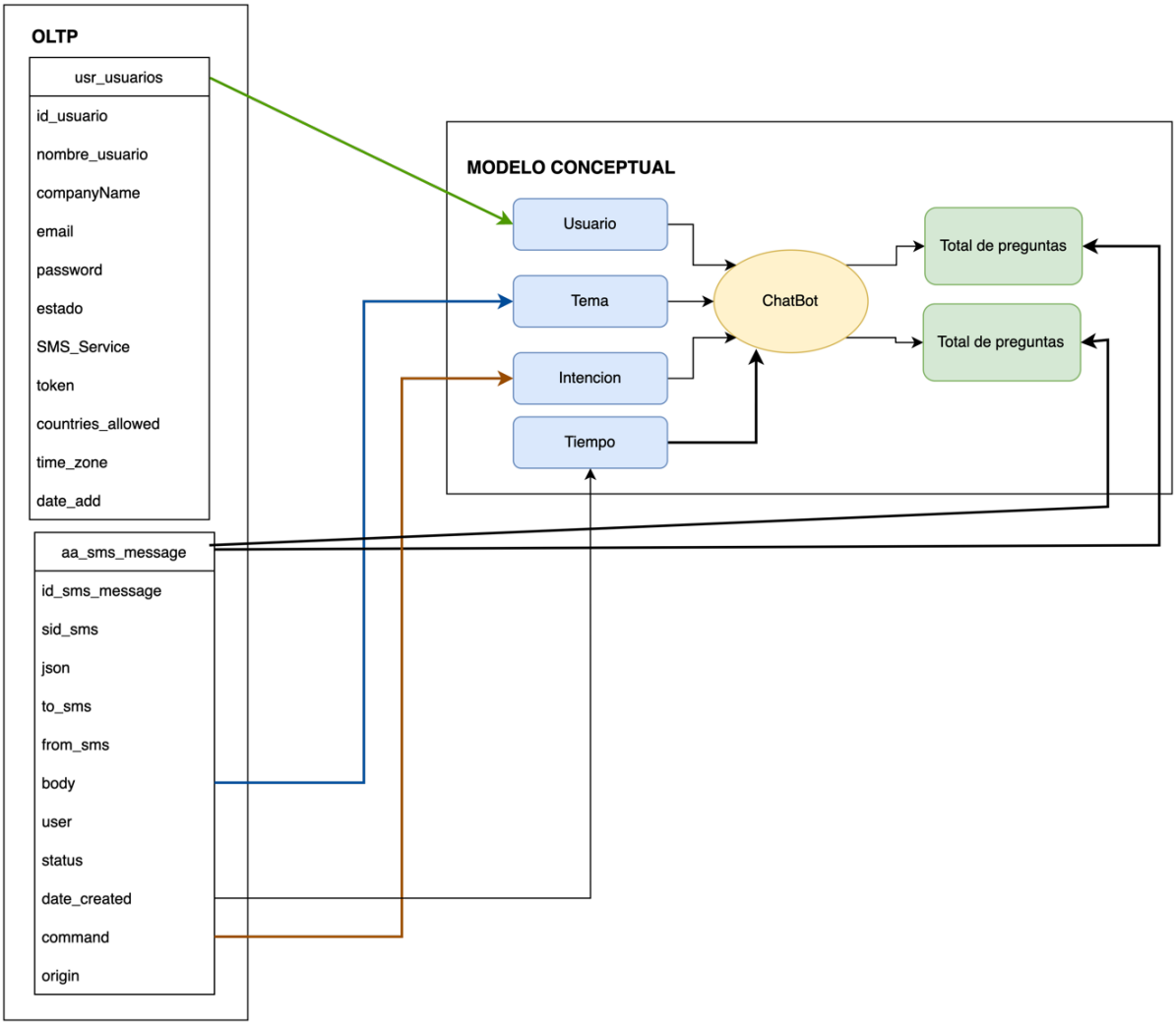


Figura 61 – Correspondencias entre los OLTP y el modelo conceptual.

Las relaciones que se identificaron son las siguientes:

- La tabla “usr_usuarios” se relaciona con la perspectiva “Usuarios”
- La columna “body” de la tabla “aa_sms_message” se relaciona con la perspectiva “Tema”.
- La columna “command” de la tabla “aa_sms_message” se relaciona con la perspectiva “Intencion”.
- La columna “date_created” de la tabla “aa_sms_message” se relaciona con la perspectiva “Tiempo”.

2.8.2.3 Nivel de granularidad

A continuación, se determinan los campos que se requieren en cada perspectiva.

Perspectiva “Usuario”: los datos a considerar son los siguientes:

- Nombre.

- Sistema

Perspectiva “Tema”: los campos a considerar son los siguientes:

- Tópico

Perspectiva: “Intencion”: los campos a considerar son los siguientes:

- Nombre

Perspectiva “Tiempo” para disponibilidad: los datos que se van a emplear para el análisis son los siguientes:

- Año
- Trimestre
- Mes
- Semana

2.8.2.4 Modelo conceptual ampliado

Luego de analizar el nivel de granularidad, se presenta en la Figura 62, el siguiente modelo conceptual ampliado:

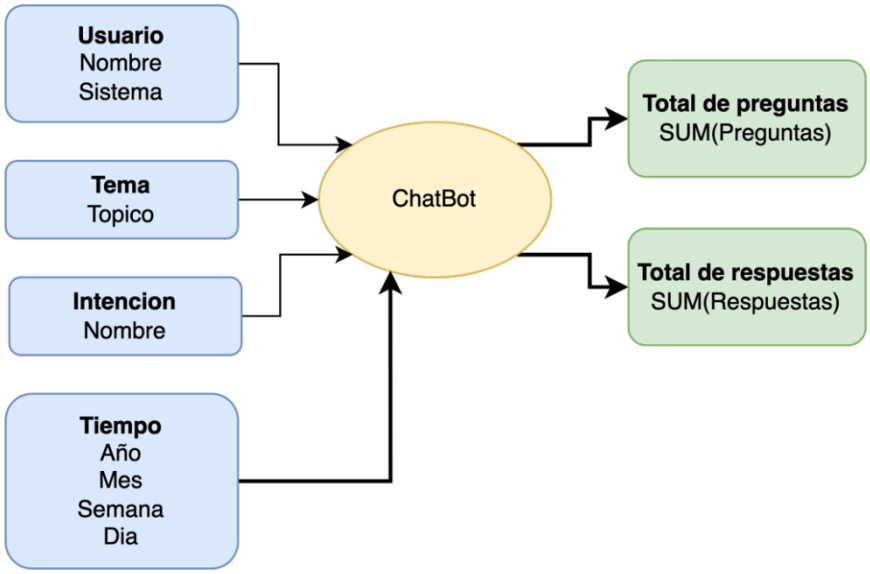


Figura 62 – Modelo conceptual ampliado para chatbot.

2.8.3 Modelo lógico

A continuación, se conforma el modelo lógico teniendo como base el modelo conceptual que ha sido creado.

2.8.3.1 Tipo de modelo lógico

Para este Data Mart de desarrollo se utilizará un esquema de tipo estrella dado que es un modelo simple y la cantidad de perspectivas e indicadores es reducida.

2.8.3.2 Tabla de dimensiones

En este paso se diseñan las tablas de las dimensiones que conforman la Data Mart.

- Perspectiva “Usuario”:

La tabla tendrá el nombre: “USUARIO”

La tabla tendrá la clave principal con el nombre “idUsuario”

La tabla tendrá las columnas: “Nombre” y “Sistema” como se muestra en la Figura 63

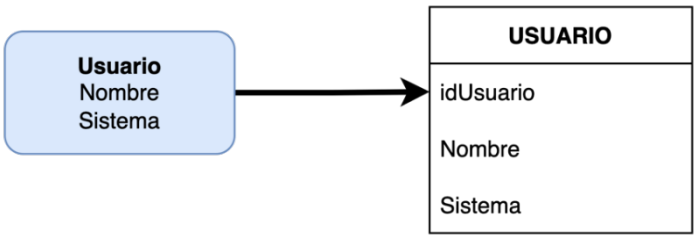


Figura 63 – Tabla de dimensión USUARIO.

- Perspectiva “Tema”:

La tabla tendrá el nombre: “TEMA”

La clave principal de la tabla tendrá el nombre “idTema”

Además, la tabla tendrá la columna: “Topico” como se muestra en la Figura 64

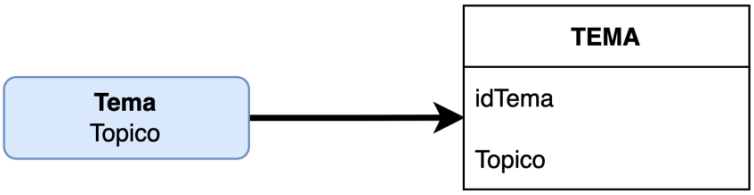


Figura 64 – Tabla de dimensión TEMA.

- Perspectiva “Intencion”:

La tabla tendrá el nombre “INTENCION”

La clave principal de la tabla tendrá el nombre: “idIntencion”

La tabla tendrá la siguiente columna “Nombre” como se muestra en la Figura 65.

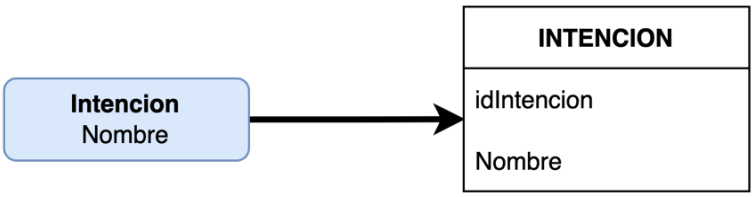


Figura 65 – Tabla de dimensión INTENCION.

- Perspectiva “Tiempo”:

La nueva tabla tendrá el nombre “FECHA”.

Tendrá una clave principal con el nombre “idFecha”.

Se mantienen los campos “Año”, “Trimestre”, “Mes” y “Semana” como se muestra en la Figura 66.

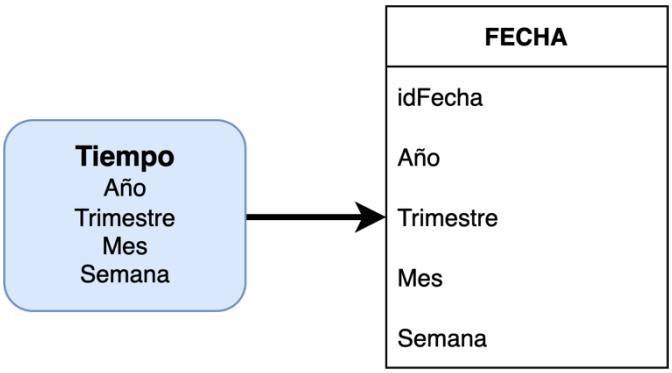


Figura 66 – Tabla de dimensión FECHA.

2.8.3.3 Tabla de hechos

En este paso se define la tabla de hechos, que contendrá los indicadores de estudio.

Se divide en dos tablas de hechos “CHATBOT_P” Y “CHATBOT_R”.

- La tabla de hechos “CHATBOT_P”.

La clave principal tiene relación con la dimensión “idUsuario”, “IdTema” y se crea el hecho “Preguntas” como se muestra en la Figura 67.

- La tabla de hechos “CHATBOT_R”.

La clave principal tiene relación con la dimensión “idUsuario”, “idIntencion” y se crea el hecho “Respuestas” como se muestra en la Figura 67.

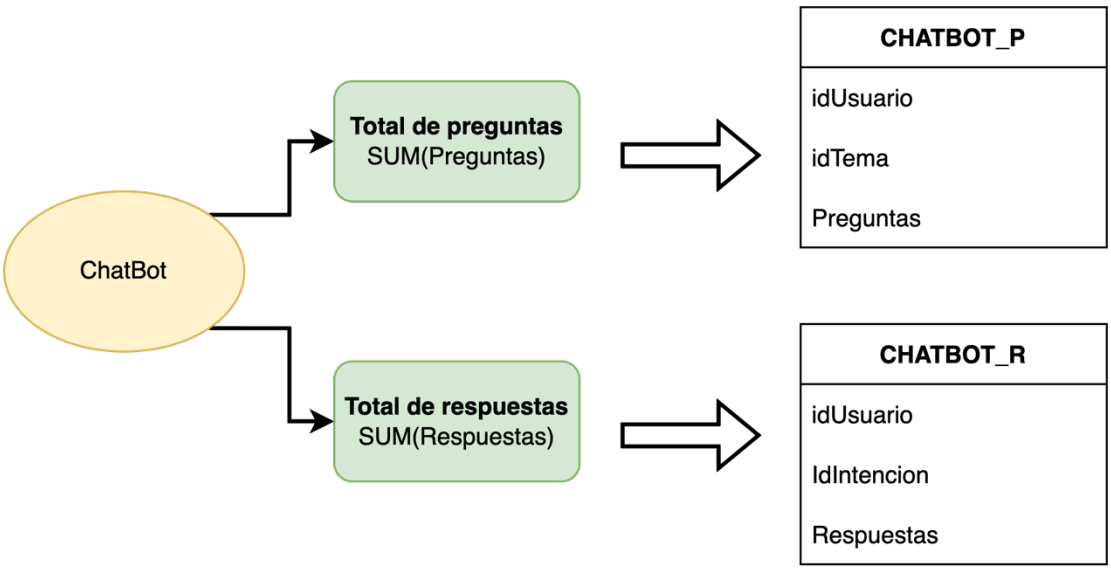


Figura 67 – Tablas de hechos CHATBOT_P y CHATBOT_R

2.8.3.4 Uniones

En este paso se realizan las uniones entre las tablas de dimensiones y las tablas de hechos, el resultado se muestra en la Figura 68 y Figura 69.

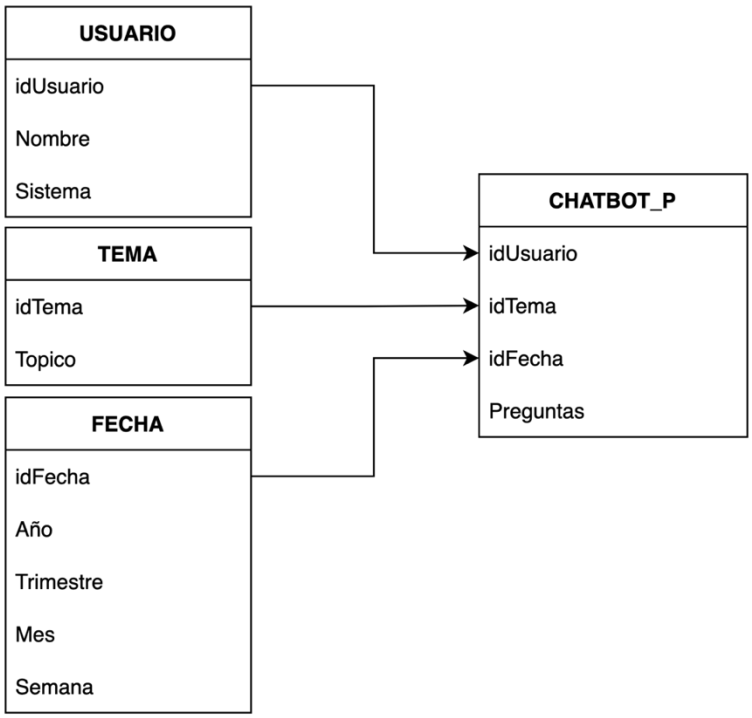


Figura 68 – Uniones para el Data Mart de chatbot preguntas.

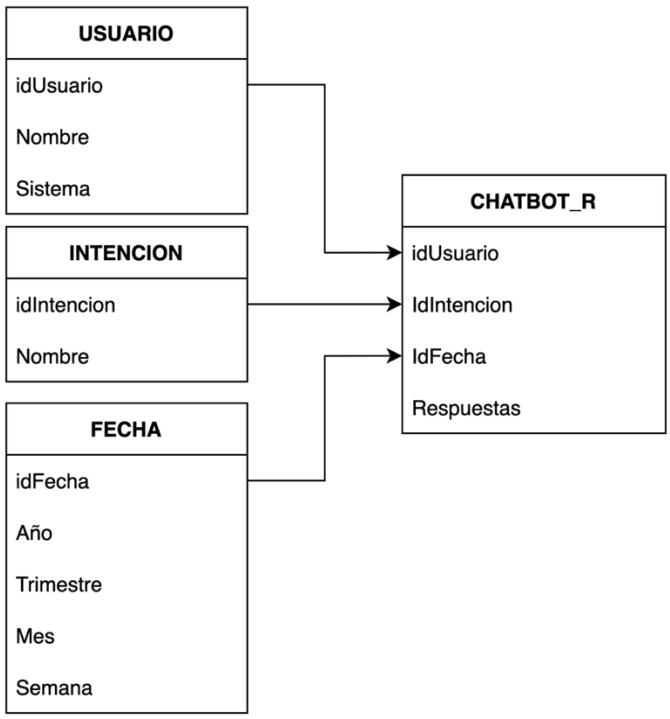


Figura 69 - Uniones para el chatbot respuestas.

2.8.4 Integración de datos

A continuación, se define la política de carga, actualización y los procesos que se van a llevar a cabo.

2.8.4.1 Carga inicial

El proceso de carga inicial se muestra en la Figura 70.

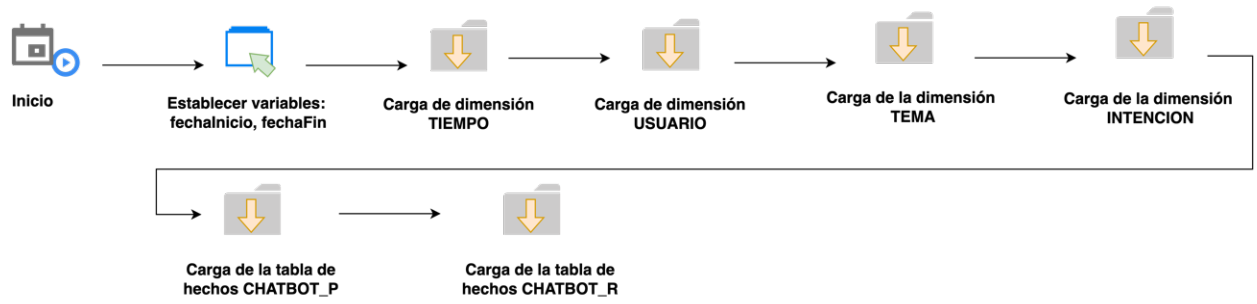


Figura 70 – Carga inicial para el Data Mart de chatbot

A continuación, se detallan los pasos a seguir en el proceso de carga inicial.

- Inicio: Inicia el proceso en el momento que se le indique.
- Establecer variables: Se establece la variable “fechaInicio” como la fecha de la primera suscripción y la variable “fechaFin” como la fecha actual.
- Carga dimensión FECHA: Para la carga de la dimensión fecha se creó un script que genera fechas en un rango desde la fecha de inicio hasta la fecha de finalización (fin del año en curso) con un rango semanal. El formato para la columna “idFecha” se compone de “año + semana”. Un ejemplo del resultado de este script, se muestra la Tabla 12.

Tabla 12 – Ejemplo de datos para la carga de la dimensión FECHA.

idFecha	Año	Trimestre	Mes	Semana
202201	2022	1	1	1
202202	2022	1	1	2
202203	2022	1	1	3

El resultado de este script se guarda en la tabla “FECHA”.

- Carga dimensión USUARIO: Para esta tabla se realiza la siguiente consulta SQL:

SELECT id_usuario as idUsuario, usuario as Usuario FROM usr_usuarios.

El resultado se guarda en la tabla USUARIO.

- Carga de la tabla de hechos CHATBOT_P: Para cargar esta información se realizan un script que ejecuta los siguientes pasos:

Realiza la siguiente consulta SQL:

SELECT body AS preguntas, user AS idUsuario, date_created as fecha FROM aa_sms_message WHERE command = “@fallback” and type = “in” and date_created > fechaInicio AND date_created <= fechaFin

La consulta anterior (Consulta1: “idUsuario”, “preguntas”, “fecha”) devuelve todos los mensajes que enviaron los clientes finales entre el periodo de tiempo seleccionado y que el chatbot no entendió.

El siguiente paso es enviar cada resultado de la consulta anterior (la columna ”body”) al modelo de inteligencia artificial de la sección 2.9.11 que devolverá el tema sobre el que trata la pregunta. El resultado anterior tendrá las siguientes columnas: (Consulta2: “idUsuario”, “preguntas” y “tema”).

A continuación, se agrupan los resultados de la (Consulta2) tomando en cuenta los temas que se encuentran en el resultado anterior y se guardan en la tabla de dimensión “TEMA”.

Al consultar la dimensión TEMA se obtiene un nuevo resultado (Consulta3: “idTema”, “Tema”)

Al unir las Consulta2 y Consulta3 teniendo en cuenta la columna “tema” se obtiene un resultado como el siguiente (Consulta4: “idUsuario”, “idTema”, “preguntas”, “tema”). Como último paso se convierte la columna “fecha” a su correspondiente “idFecha” y se elimina la columna “tema” y el resultado final se guarda en la tabla “CHATBOT_P”

- Cargar dimensión INTENCION: Para cargar esta tabla se realiza el siguiente proceso.

Se realiza la siguiente consulta SQL.

SELECT command AS Nombre FROM aa_sms_message GROUP BY command;

El resultado de la consulta se guarda en la tabla INTENCION.

- Carga de la tabla de hechos CHATBOT_R: Para cargar información en esta tabla se ejecuta un script que realiza los siguientes pasos:

Realiza la siguiente consulta SQL:

SELECT user AS idUsuario, command AS intencion, body AS respuesta, date_created as fecha FROM aa_sms_message WHERE type = “out” and date_created > fechaInicio AND date_created <= fechaFin.

La consulta anterior (Consulta1: “idUsuario”, “intencion”, “respuesta”, “fecha”) devuelve todas las respuestas que realizo el chatbot con las respectivas intenciones detectadas en el periodo de tiempo seleccionado.

A continuación, se hace un join entre la Consulta1 con la tabla de dimensión “INTENCION” para añadir la columna “idIntencion”. Como paso final se transforma la columna “fecha” a su correspondiente “idFecha” y el resultado se guarda en la tabla de hechos CHATBOT_R.

2.8.4.2 Actualización

Las políticas de actualización que se definieron son las siguientes:

- La información se actualizará todos los martes a las 12 de la noche.
- La dimensión “USUARIO” se cargará de manera incremental siguiendo el mismo proceso que la carga inicial.
- La dimensión “TEMA” también se cargará de manera incremental, es decir que los nuevos temas de conversación que el chatbot no entendió se irán anexando a la tabla “TEMA”.
- La carga de las dimensiones “FECHA” y “FECHA_HORA” no es necesaria ya que al inicio se cargó datos hasta la finalización del año en curso.
- La tabla de hechos “CHATBOT_P” se carga de manera considerando la última fecha de actualización hasta la fecha actual
- La dimensión “INTENCION” se cargará de manera incremental, es decir que las nuevas intenciones de conversación se insertan en la tabla “INTENCION”.
- La tabla de hechos “CHATBOT_R” se carga de manera incremental, es decir que se carga considerando la última fecha de actualización hasta la fecha actual.

2.9 Herramientas y modelos de inteligencia artificial.

En esta sección, se detalla el proceso de construcción de las diferentes herramientas y modelos de inteligencia artificial diseñados para generar conocimiento nuevo para la empresa y que sirven de apoyo en los procesos de carga y actualización de datos en los distintos Data Mart.

2.9.1 Predicción de suscripciones.

El objetivo de esta herramienta es que permita predecir el número de nuevas suscripciones a lo largo del tiempo. Los datos disponibles para esta tarea se detallan en la Tabla 13.

Tabla 13 – Datos para la predicción de suscripciones.

Campo	Tipo	Descripción
semana	numérico	Representa la semana del año
suscripciones	numérico	Representa el valor total de suscripciones en la semana

Al graficar los valores reales se obtiene una curva como la que se muestra en la Figura 71.

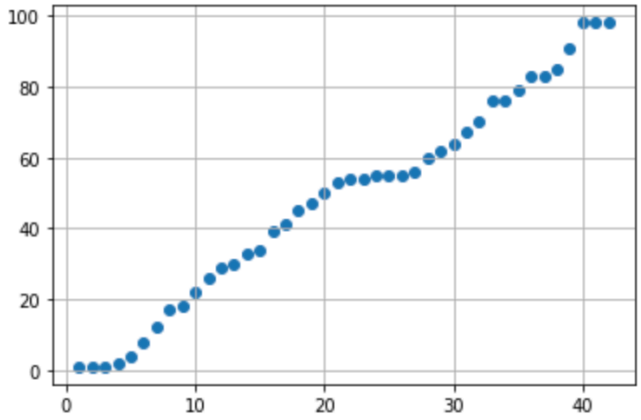


Figura 71 – Grafica de datos reales de suscripciones.

En la Figura 71, en el eje ‘x’ se presenta el número de la semana y en el eje ‘y’ el total de suscripciones acumuladas. Considerando que no se dispone de una cantidad alta de datos, es conveniente considerar los modelos de regresión lineal o polinomial para determinar la curva más cercana a los valores reales. Para ello, se utiliza el método poly1d de la librería Numpy de Python entrenando modelos de grado 1,2 y 3, el resultado se muestra en la Figura 72.

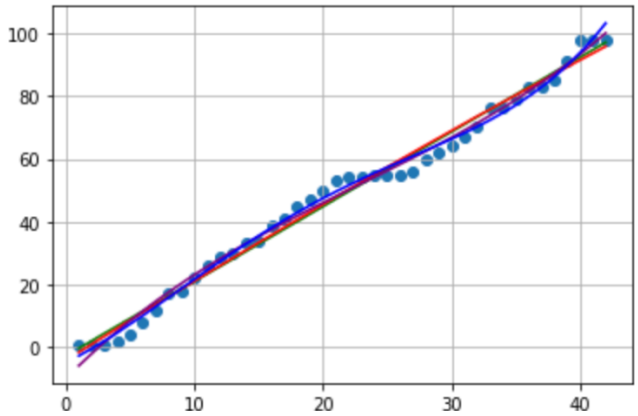


Figura 72 – Grafica de curvas aproximadas a las suscripciones.

Para elegir la curva con la mejor aproximación se utiliza el método “R-cuadrado ajustado” que indica el porcentaje de semejanza de la misma con los valores reales, al calcular este dato se obtienen los resultados de la Tabla 14.

Tabla 14 – Comparación de resultados para predicción de suscripciones.

Grado	R-cuadrado ajustado
1	0.986
2	0.986
3	0.990

En la Tabla 14, se observa que la curva con mejor aproximación a los valores reales es de grado 3, pero con escaza diferencia con los modelos de grado 1 y 2. Dado que el modelo de grado 1 alcanza un 98% de similitud se lo considera adecuado para la predicción de suscripciones, el resultado de este modelo se muestra en la Figura 73.

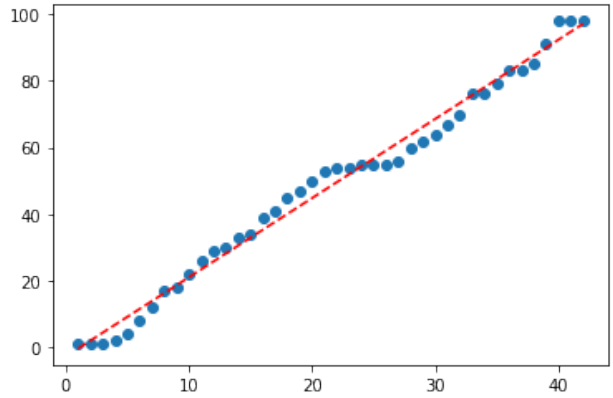


Figura 73 – Modelo de predicción de suscripciones.

Para poder aplicar este modelo se realizó un script que tiene las siguientes características:

- Está escrito en lenguaje Python.
- Recibe todos los datos de las suscripciones hasta la fecha actual y el número de valores que se desea predecir.
- Se reentrena el modelo con la nueva información antes de realizar una nueva predicción.
- La respuesta es un arreglo con la predicción de los valores futuros.

2.9.2 Predicción de cancelaciones.

Esta herramienta se diseña con la finalidad de predecir el número de cancelaciones de suscripciones a lo largo del tiempo. Los datos disponibles para esta tarea se detallan en la Tabla 15.

Tabla 15 – Datos para la predicción de cancelaciones.

Campo	Tipo	Descripción
semana	Numérico	Representa la semana del año
cancelaciones	Numérico	Representa el valor total de cancelaciones en la semana

Al graficar los valores reales se obtiene una gráfica como la Figura 74.

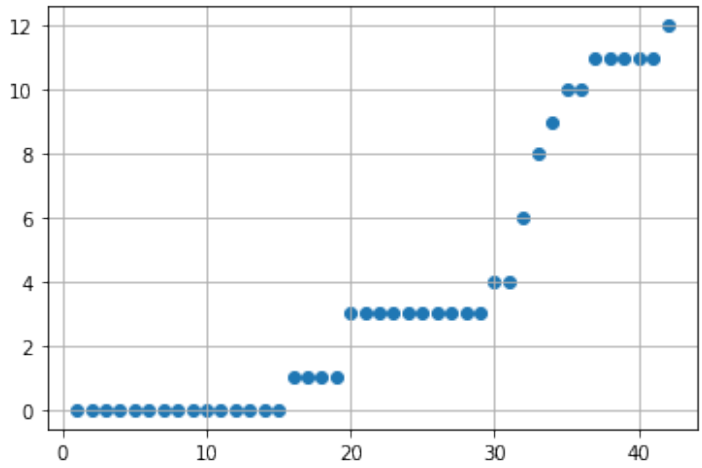


Figura 74 – Grafica de datos reales de cancelaciones.

En la Figura 74, el eje ‘x’ representa el número de la semana y el eje ‘y’ representa el total de cancelaciones por cada semana. El número de datos disponibles es bajo por lo que se considera utilizar modelos de regresión lineal o polinomial para descubrir una curva aproximada que permita predecir los valores de cancelaciones futuros. Para ello se utiliza el método poly1d de la librería Numpy de Python y se entrenan modelos de grado 1, 2 y 3 y se observa el resultado en la Figura 75.

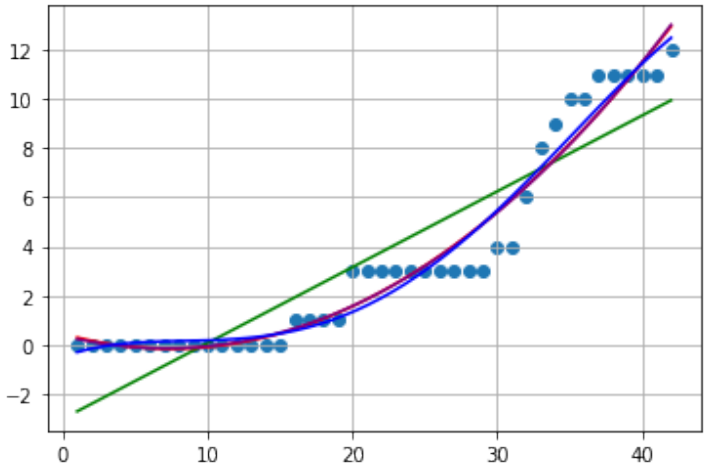


Figura 75 – Grafico de curvas aproximadas.

A continuación, se escoge la curva con la mejor aproximación utilizando el método “R-cuadrado ajustado” de donde se obtienen los resultados de la Tabla 16.

Tabla 16 – Comparación de resultados para predicción de cancelaciones.

Grado	R-cuadrado
1	0.826
2	0.9519
3	0.9507

En la Tabla 16, se observa que la curva con mejor aproximación a los valores reales es aquella que provienen de un modelo de grado 2 como se muestra en la Figura 76.

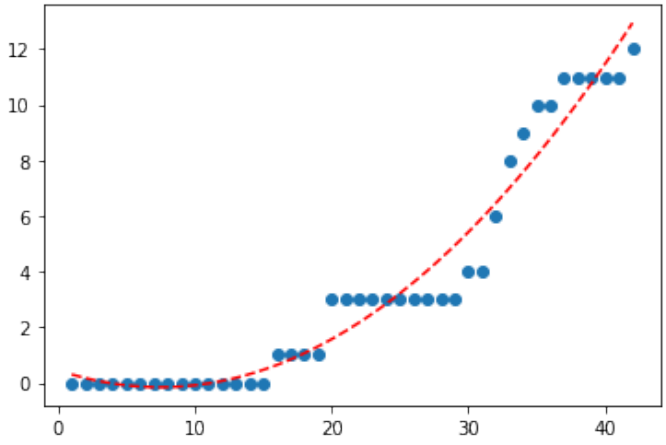


Figura 76 – Modelo de predicción de cancelaciones.

Para aplicar este modelo para la predicción cancelaciones se desarrolló un script con las siguientes caracteristicas:

- Está escrito en lenguaje Python.
- Recibe los datos de las cancelaciones hasta la fecha actual y el número de nuevos valores que se desea predecir.
- El modelo se reentrena con la nueva información antes de realizar la predicción.
- La respuesta es un arreglo con las predicciones realizadas.

2.9.3 Predicción del total de transacciones

El objetivo de este modelo es predecir el total de nuevas de transacciones diarias a lo largo del tiempo. La información con la que se cuenta para esta tarea son los valores de transacciones acumulados por cada día.

Al graficar esta información tenemos un resultado como el de la Figura 77.

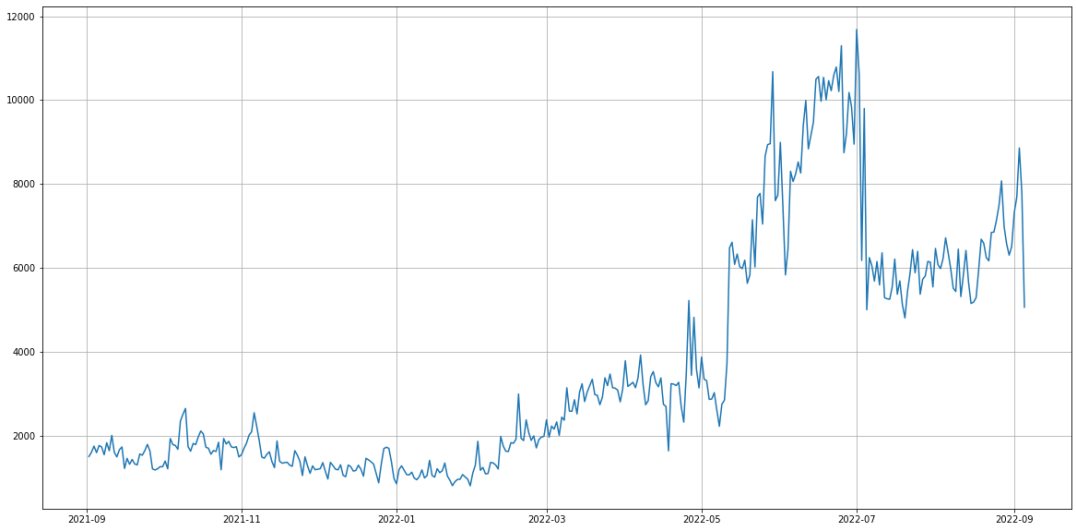


Figura 77 – Gráfico de transacciones en el tiempo.

Está claro a partir de la gráfica que hay un aumento general en la tendencia.
Según [18], para analizar series temporales existen varias metodologías entre las cuales destaca ARIMA (Autoregressive Integrated Moving Average) el cuál es un modelo estadístico utilizado para la predicción de series de tiempo. Combina tres modelos diferentes:

1. Modelo autorregresivo (AR): Un modelo que utiliza valores pasados de la serie de tiempo para predecir valores futuros. Supone que el valor futuro de la serie de tiempo es una función lineal de sus valores pasados, con algo de ruido agregado.
2. Modelo integrado (I): Un modelo que se utiliza para hacer que una serie de tiempo no estacionaria sea estacionaria. Una serie de tiempo es estacionaria si la media, la varianza y la auto covarianza son constantes a lo largo del tiempo. Las series de tiempo no estacionarias se pueden hacer estacionarias diferenciando la serie de tiempo. Esto se hace restando el valor en un momento anterior del valor actual.
3. Modelo de promedio móvil (MA): Un modelo que utiliza el término de error (la diferencia entre el valor predicho y el valor observado) de los puntos de tiempo anteriores para predecir el valor futuro.

El modelo ARIMA se especifica con tres parámetros: p, d y q. El parámetro p es el número de observaciones de retraso incluidas en el modelo, d es el grado de diferenciación (el número de veces que se han restado los valores pasados de los datos) y q es el tamaño de la ventana de promedio móvil.

Antes de empezar a trabajar con un modelo es necesario identificar si la serie temporal es estacionaria o no. Para ello vamos a verificar haciendo uso de la prueba de Dickey Fuller

que indica que la serie se considera estacionaria si el valor de p es bajo y los valores críticos en intervalos de confianza del 1, 5 y 10 % están muy cercanos a las estadísticas ADF.

Los resultados de la prueba Dickey Fuller se muestran en la Tabla 17.

Tabla 17 – Resultado de la prueba Dickey Fuller para la muestra original.

Ítem	Valor
ADF Statistic	-1.262324
p-value	0.646228
#Lags Used	15.000000
Number of Observations Used	353.000000
Critical Value (1%)	-3.449011
Critical Value (5%)	-2.869763
Critical Value (10%)	-2.571151

Las estadísticas ADF están lejos de los valores críticos y el valor de p es mayor que 0.05. Por lo tanto, se concluye que la serie no es estacionaria.

En este caso es posible hacer que la serie se vuelva estacionaria restando la media móvil. Después de realizar esta operación, los resultados de la prueba Dickey Fuller son los de la Tabla 18.

Tabla 18 – Resultados de la prueba Dickey Fuller sin su media móvil.

Ítem	Valor
ADF Statistic	-3.262324
p-value	0.04646228
#Lags Used	15.000000
Number of Observations Used	353.000000
Critical Value (1%)	-3.449011
Critical Value (5%)	-2.869763
Critical Value (10%)	-2.571151

El valor de p es menor que 0.05 y las estadísticas ADF se encuentran cerca de los valores críticos por lo tanto es posible continuar con el proceso de entrenamiento del modelo.

A continuación, se procede a entrenar un modelo usando la herramienta Google Colab que provee un espacio de trabajo eficiente para esta tarea. Para determinar los parámetros p, d y q para esta serie se debe considerar las funciones ACF (función de correlación

automática) y PACF (función de correlación automática parcial). ACF indica la correlación entre el momento actual y las observaciones en los momentos anteriores, esta correlación permite obtener el número óptimo de términos MA que también es el orden del modelo. PACF indica la correlación entre observaciones realizadas en dos puntos de tiempo y tiene en cuenta la influencia de otros puntos de datos, usar PACF sirve para determinar la cantidad óptima de términos del modelo AR que es también el orden del modelo. Al graficar estas funciones se obtienen las Figura 78 y Figura 79.

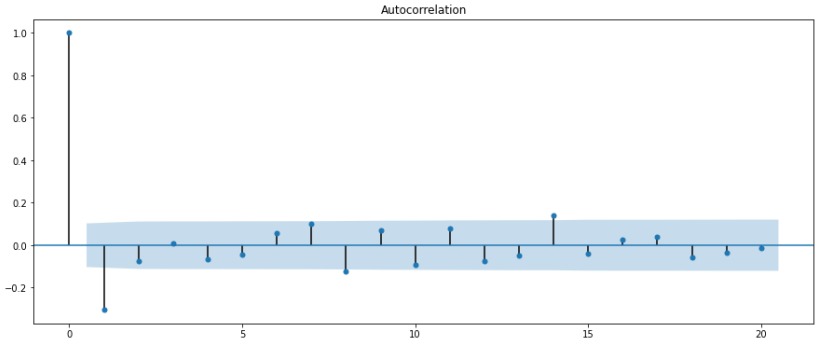


Figura 78 – Gráfica de autocorrelación.

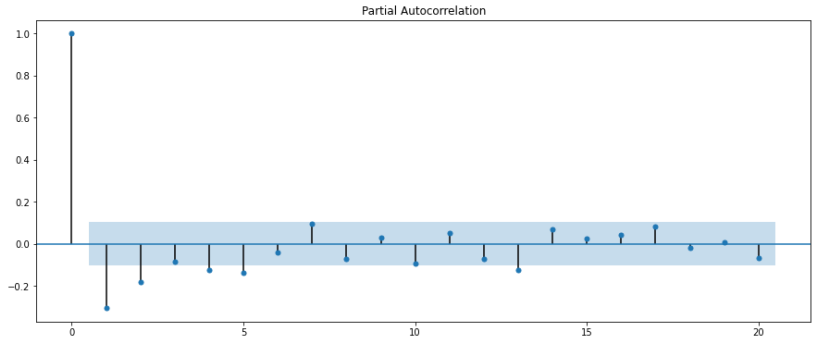


Figura 79 – Gráfica de autocorrelación parcial.

En las Figura 78 y Figura 79 los cuadros azules representan los umbrales de importancia, las líneas verticales representan los valores ACF y PACF en un momento dado y solo se consideran significativas las líneas verticales que superan los cuadros azules.
Considerando lo anterior, de la gráfica ACF se obtiene que el orden optimo es 1, de la gráfica de PACF el orden optimo es 2 y el orden de diferenciación se puede ajustar en 1.

Luego de entrenar el modelo con los datos de entrada, se procede a probar el modelo al realizar una predicción. En la Figura 80 se muestran los datos reales en color amarillo, los datos predichos en azul y el intervalo de confianza del 95% en gris.

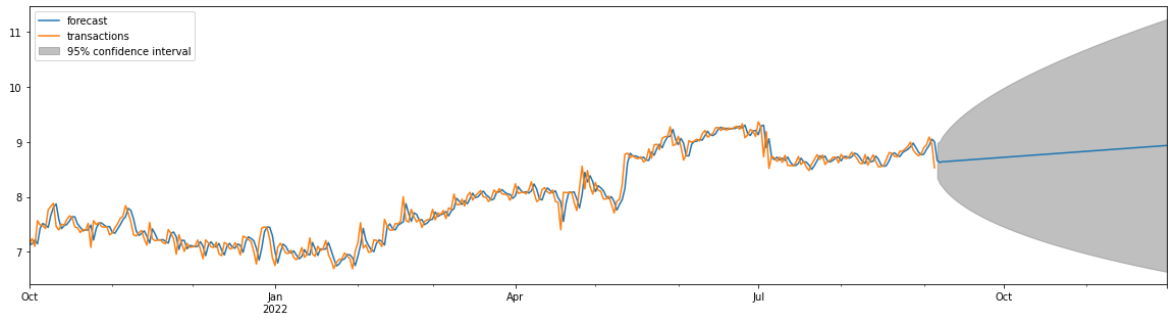


Figura 80 – Resultado del modelo de predicción de series temporales.

Por último, para poder realizar las predicciones de una manera más cómoda se descarga el modelo entrenado utilizando la librería Pickle. La librería Pickle permite guardar el modelo en formato .sav el cual se puede cargar, reentrenar y utilizar posteriormente para realizar predicciones.

Una vez que se tiene el modelo entrenado es necesario hacer que el modelo esté disponible para poder ser reutilizado. Para ello se realizó un script en Python que recibe un archivo .csv con la información más actualizada, reentrena el modelo con la nueva información y realiza una predicción de los valores futuros.

2.9.4 Predicción del total de almacenamiento

El objetivo principal de este modelo es poder predecir el porcentaje de almacenamiento que se alcanzará en el futuro para los volúmenes que contienen las bases de datos. Los datos disponibles para esta tarea son los que se detallan en la Tabla 19.

Tabla 19 – Datos disponibles para predicción de almacenamiento.

Campo	Tipo	Descripción
día	numérico	Representa el día del año
almacenamiento	Numérico	Representa el porcentaje total de almacenamiento utilizado

Al graficar el porcentaje de consumo de almacenamiento se obtiene una curva como la Figura 81.

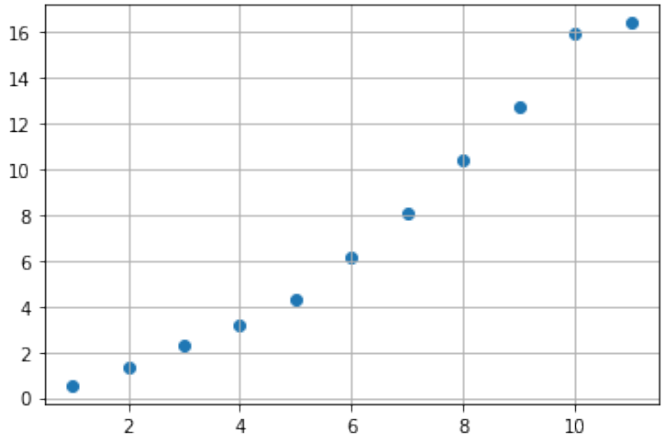


Figura 81 – Grafico de consumo de almacenamiento.

En la Figura 81, el eje ‘x’representa el tiempo en meses y el eje ‘y’ representa el porcentaje de consumo del almacenamiento. Teniendo en cuenta que los datos disponibles son minimos se consideran los modelos lineal y polinomial para encontrar la curva que más se aproxime a los valores reales. Utilizando el método poly1d de la librería Numpy de Python se entrenan modelos de grado 1, 2 y 3 obteniendo los resultados de la Figura 82.

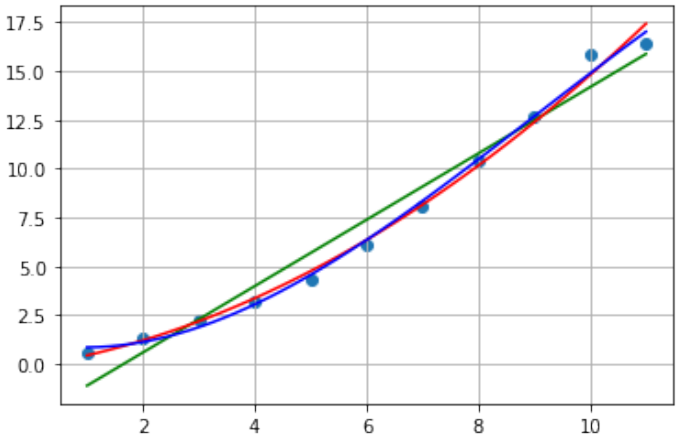


Figura 82 – Curvas aproximadas al porcentaje de almacenamiento.

Para verificar cual es la mejor aproximación se utiliza el método “R-cuadrado ajustado” de donde se obtienen los resultados de la Tabla 20.

Tabla 20 – Comparación de resultados para predicción de almacenamiento.

Grado	**R-cuadrado**
1	0.959
2	0.989
3	0.991

Con ayuda de la Tabla 20 se determina que el modelo más adecuado para esta tarea es de grado 2 y el resultado se muestra en la Figura 83.

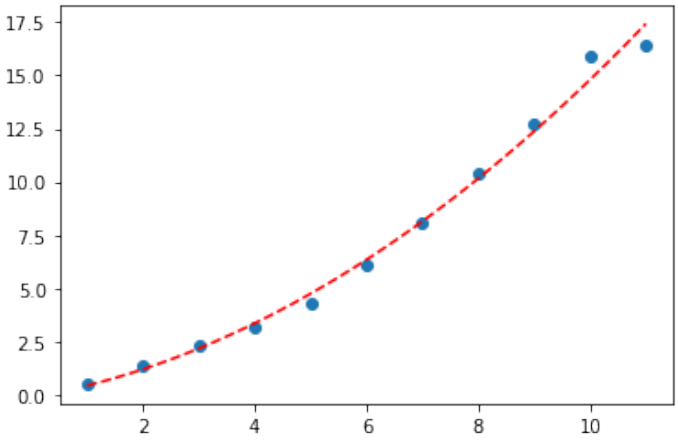


Figura 83 – Modelo para predicción de almacenamiento.

Para la aplicación de este modelo de los valores futuros de porcentaje de almacenamiento, se escribió un script con las siguientes características:

- Está escrito en lenguaje Python.
- El script recibe la información de porcentaje de almacenamiento hasta la fecha actual y el número de predicciones deseadas.
- El modelo se reentrena con la nueva información antes de realizar la predicción.
- La respuesta es un arreglo con las predicciones realizadas.

2.9.5 Predicción del total de leads

El objetivo principal de este modelo es poder predecir la cantidad de leads que va a lograr conseguir el equipo de marketing a lo largo del tiempo. Los datos disponibles para esta tarea se muestran en la Tabla 21.

Tabla 21 – Datos para la predicción de Leads.

Campo	Tipo	Descripción
semana	Numérico	Representa la semana del año
leads	Numérico	Representa el valor total de leads obtenidos

Al graficar los valores reales se obtiene una curva como la que se muestra en la Figura 84.

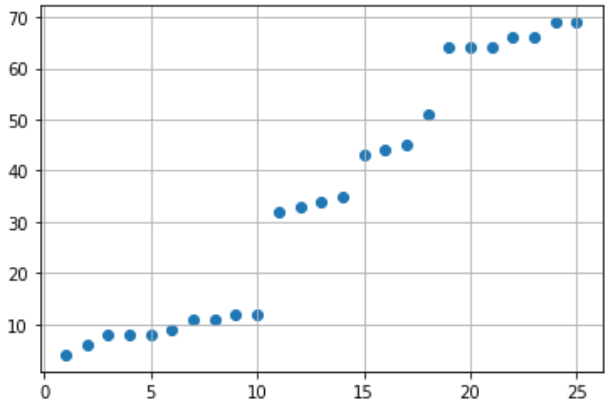


Figura 84 – Grafica de Leads por semana.

En la Figura 84, el eje 'x' representa el número de la semana y el eje 'y' representa el total de leads obtenidos por cada semana. Tomando en cuenta que la cantidad de datos es reducida, se consideran los modelos de regresion lineal y polinomial para determinar la curva mas aproximada utilizando el método poly1d de la librería Numpy de Python, se entrenaron modelos de grado 1, 2 y 3 y se observa el resultado en la Figura 85.

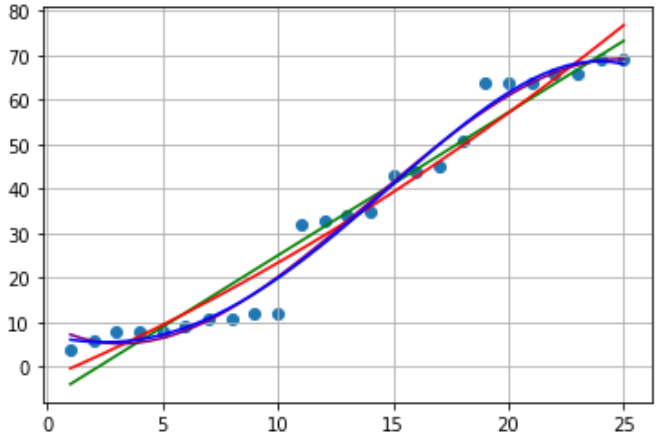


Figura 85 – Grafica de curvas aproximadas para leads.

Para determinar la curva con mejor aproximación se utiliza el método "R-cuadrado ajustado" de donde se obtienen los resultados de la Tabla 22.

Tabla 22 – Comparación de resultados para predicción de Leads.

Grado	R-cuadrado
1	0.947
2	0.950
3	0.931

De la Tabla 22 se deduce que la curva más aproximada a los valores reales proviene de un modelo de grado 2 y el resultado de este modelo se muestra en la Figura 86.

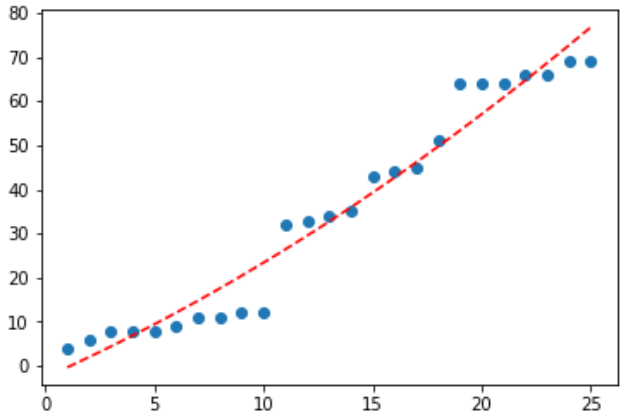


Figura 86 – Modelo para predicción de Leads.

Para la aplicación de este modelo para la predicción de los valores futuros de leads se realizó un script con las siguientes características.

- Está escrito en lenguaje Python.
- Recibe los datos de Leads hasta la fecha actual y el número de nuevos valores que se desea predecir.
- El modelo se reentrena con la nueva información antes de realizar la predicción.
- La respuesta es un arreglo con las predicciones realizadas.

2.9.6 Modelo de clasificación de texto.

La información que contienen las transacciones sobre los clientes finales es reducida y se desea poder extraer información adicional. En este caso puntual, no se dispone de la información del género de los clientes finales y se desea poder inferir a partir del nombre del mismo.

HuggingFace es un repositorio público online que contiene una gran variedad de modelos de inteligencia artificial pre entrenados como es el caso del modelo “Cameron/BERT-rtgender-opgender-annotations” que está especializado en clasificar texto con la intención de poder identificar el género de una persona en base al nombre de la misma [19]. A continuación, se muestran algunos resultados de este modelo en la Tabla 23.

Tabla 23 – Respuestas del modelo de clasificación de texto para género.

Nombre	Género
Daniela	1
Marco	0
Jasmine	1
German	0

Para poder utilizar este modelo de una forma recurrente se creó un script de Python que importa el modelo, lo inicializa y permite utilizarlo a través de un endpoint a manera de API. Por último, este script está montado dentro de la imagen de Docker para poder ser ejecutado correctamente. La imagen de Docker necesaria para ejecutar este modelo de clasificación de texto es provista por HuggingFace y contiene todas las herramientas listas para poder ejecutar los modelos que provee de una manera fácil y cómoda. Esta imagen de Docker esta aprovisionada con CPU y GPU Pytorch backend. [20]

2.9.7 Modelo de clasificación de clientes finales.

En este caso se desea conocer los grupos en los que se pueden clasificar los clientes finales en base a la forma en la que interactúan con el sistema de la empresa.
Para lograr esto se dispone de los datos de la Tabla 24.

Tabla 24 – Datos para clasificación de clientes finales.

Campo	Tipo	Descripción
gender	bool	Es el género del cliente final (0 hombre, 1 mujer)
sentMessages	number	El número de mensajes que el cliente final escribió
recievedMessages	number	El número de mensajes que el cliente final recibió
questionsAsked	Number	El número de preguntas que realizo
fallbackCount	number	El número de preguntas que el chatbot no le entendió
requestCancel	bool	1 si solicitó detener el servicio, 0 en caso contrario
rateTour	bool	1 si llego a calificar su tour, 0 en caso contrario

Como primer paso se debe determinar en cuantos grupos se pueden clasificar los clientes finales, para esto se utiliza el método Elbow que consiste en trazar la variación explicada en función del número de grupos y escoger el codo de la curva como el número de grupos a utilizar. El resultado de este método se muestra en la Figura 87.

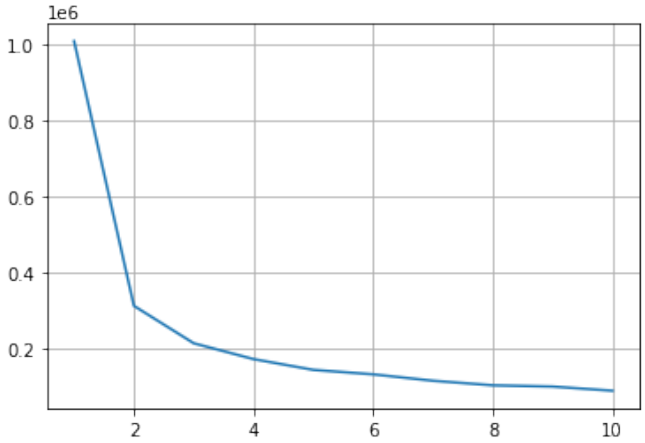


Figura 87 – Curva de Elbow para clasificación de clientes finales.

En base a la Figura 87 el número de grupos adecuado para en análisis es 3.
Para entrenar el modelo se hace de la herramienta Google Colab y el algoritmo Kmeans de la librería Sklearn de Python.

Una vez que el modelo ha sido entrenado, se desea tener una representación visual de los grupos identificados, para esto se utilizó la técnica PCA (Principal component analysis) que también forma parte de la librería Sklearn. Luego de aplicar PCA, se obtiene una representación de los grupos como se muestra en la Figura 88.

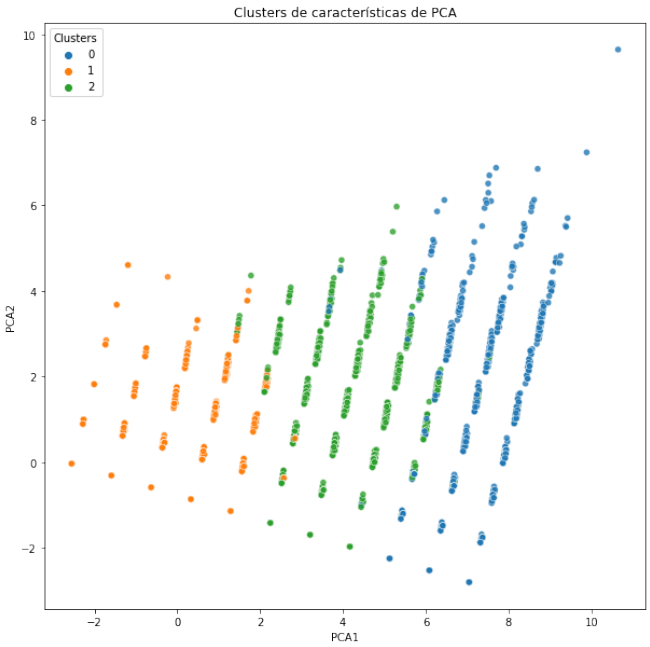


Figura 88 – Gráfica PCA para clientes finales.

Los 3 grupos que se identifican con este modelo de clasificación tienen las siguientes características.

Cluster0 (Ideal):

- Es un grupo pequeño que contiene el 5,04% de los clientes finales.
- En promedio recibió 10 mensajes y escribió 4.
- El 30% de las personas califica su actividad

Es decir que este grupo contiene el prototipo de clientes ideales que presentan una interacción alta con el sistema.

Cluster1 (Común):

- El grupo contiene el 78.04% de todos los clientes finales.
- En promedio se le enviaron 3 mensajes y respondió solo 1.
- El 6% de las personas califica su actividad

En este grupo se encuentran clientes finales cuya interacción con el sistema es muy baja.

Cluster2 (Normal):

- El grupo contiene el 16.92% de todos los clientes finales.
- Los clientes de este grupo recibieron 7 mensajes y respondieron a 2.
- El 20% de las personas califica su actividad.

En este grupo se encuentran clientes finales cuya interacción con el sistema es buena, es decir realizan preguntas e interactúan con el sistema.
Una vez que el modelo está listo, se utiliza la librería Pickle de Python para poder guardar el modelo entrenado.

Para poder utilizar de manera recurrente este modelo y poder realizar nuevas predicciones se realizó un script que recibe como parámetros las características del cliente final y devuelve el grupo al que pertenece.
Para que el modelo esté disponible se montó el script dentro de un Docker que contiene las librerías y herramientas para ejecutar el script de manera adecuada.

2.9.8 Modelo de análisis de sentimientos.

Como parte de la interacción de los clientes finales con el servicio, se solicita al mismo que por favor califique su experiencia del 1 al 5 donde 5 es más positivo. Los clientes finales tienen la libertad de escribir su calificación o también pueden escribir una reseña sobre el tour. A continuación, se muestran algunos ejemplos de estas reseñas.

- “5555555❤️❤️❤️”
- “Experience with our driver. Phil, a 5. The trip itself, not so great.”
- “4.5 just because waiting for the helicopter to fly us out of the Canyon took too long”
- “5! Great, friendly guides, beautiful city, fun bikes!”
- “Not happy at all. “

Como se puede apreciar en los ejemplos anteriores, no se puede analizar directamente estos comentarios y por ello nace la necesidad de contar con un modelo de inteligencia artificial que permita convertir estos reviews en una escala numérica de 1 a 5.
La comunidad de HuggingFace ofrece el modelo “nlptown/bert-base-multilingual-uncased-sentiment” que es un modelo multilingue entrenado para realizar analisis de sentimientos en reviews en lenguajes como: inglés, alemán, frances y español y predice el sentimiento del review en una escala de 1 a 5. [21]

Para poder utilizar este modelo fue necesario crear una script en python que importe el modelo, lo cargue y permita acceder al mismo a través de un endpoint, de esta manera se puede enviar el texto del review previemante estandarizado y el script devolverá el resultado del review en una escala del 1 al 5.

A continuación, se muestran algunos resultados obtenidos en la Tabla 25.

Tabla 25- Resultados del modelo de análisis de sentimientos.

Review	Valor
5555555❤️❤️❤️	5
Experience with our driver. Phil, a 5. The trip itself, not so great.	3
5! Great, friendly guides, beautiful city, fun bikes!	5
Not happy at all.	1

2.9.9 Modelo de clasificación de usuarios en base al rating

El objetivo de este modelo es poder clasificar a los usuarios en base al rating que le realizaron los clientes finales y también el rating general que maneja en Google Reviews y TripAdvisor.

La información con la que se cuenta para entrenar este modelo se detalla en la Tabla 26.

Tabla 26 – Datos para clasificación de usuarios respecto al rating.

Nombre	Tipo	Descripción
Rating	Numerico	La calificación promedio de sus clientes finales
GoogleRating	Numerico	La calificación en Google reviews
tripRating	Numerico	La calificación en TripAdvisor
Fallbacks	Numerico	La cantidad de fallbacks
Questions	Numerico	La cantidad de preguntas recibidas
Clicks	Numerico	El número de clics que se dieron a los enlaces que envió a sus clientes

Para identificar el número de grupos óptimo se utiliza la técnica de Elbow. El resultado de aplicar esta técnica se muestra en la Figura 89.

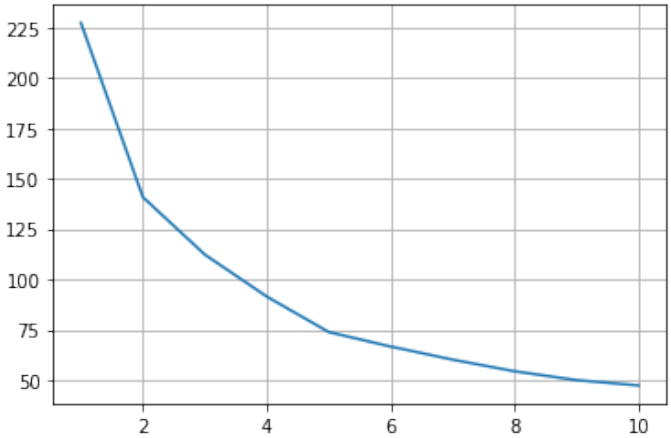


Figura 89 – Curva de Elbow para clasificación de usuarios por rating.

A partir de la Figura 89 se aprecia que n=2 tiene un punto alto de inflexión, n=3, n=4 y n=5 no tienen cambios apreciables por lo cual se determina que los grupos óptimos es n=2.

Para alimentar el modelo de inteligencia artificial se utiliza en la información recopilada de los clientes finales, se utiliza el modelo Kmeans de la librería Sklearn.
Para visualizar los grupos encontrados utilizan la técnica de análisis de componentes principales con la cual se puede visualizar de una mejor manera. El resultado se muestra en la Figura 90.

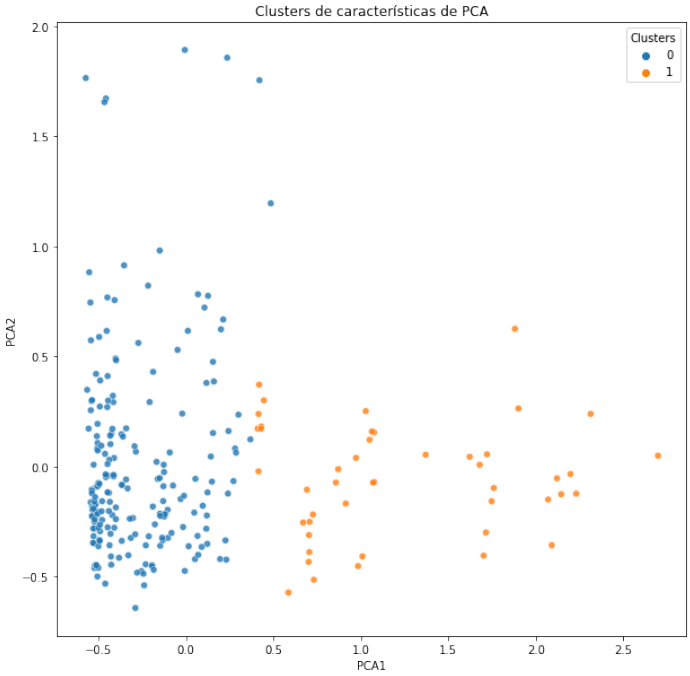


Figura 90 – Gráfica de características PCA para usuarios.

Los grupos encontrados tienen las siguientes características:

Cluster0 (Excelente):

- La calificación promedio de sus clientes es 4.83
- La calificación promedio que mantienen en Google es 4.82
- La calificación promedio qué mantienen TripAdvisor es 4.82

Cluster1 (Bueno):

- La calificación promedio de sus clientes es 4.5
- La calificación promedio en Google está arriba de 3.7
- La calificación promedio en TripAdvisor y está arriba de 4.

De igual manera este modelo se guarda utilizando la librería Pickle.

Para ser disponible este modelo se escribió un script en Python que permite cargar el modelo y lo hace accesible a través de un endpoint con el cual es posible enviarle las características del usuario y el servicio de volverá el grupo al que pertenece.

Por último, este servicio se despliega contenedor de Docker el cual contiene las herramientas y librerías necesarias para que se ejecute correctamente.

2.9.10 Modelo de clasificación de usuarios en base al volumen

El objetivo principal de este modelo será conocer los grupos que se pueden obtener a partir de la información relacionada al volumen del usuario.

Para analizar esta información se detallan los datos disponibles en la Tabla 27.

Tabla 27- Datos para la clasificación de usuarios respecto al volumen.

Campo	tipo	Descripción
Messages	Numérico	De mensajes total utilizados.
Bookings	Numérico	El número de reservas recibidas
Cancel	Numérico	El número de cancelaciones recibidas
Reviews	Numérico	El número de Reviews recibidos
Fallback	Numérico	El número de preguntas que no entendió el bot
Questions	Numérico	El número de preguntas que le realizó sus clientes
clicks	Numérico	Número de clips que realizaron sus clientes a sus enlaces

Para conocer el número ideal de grupos se utiliza en la técnica de Elbow, El resultado se muestra en la Figura 91.

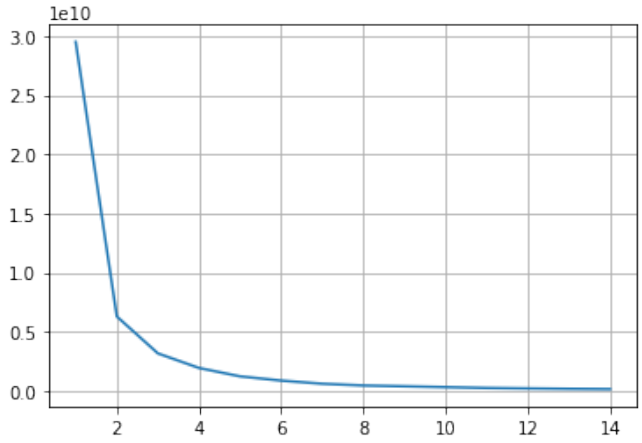


Figura 91 – Curva de Elbow para clasificación de usuarios respecto al volumen.

Para representar los grupos identificados se utiliza la técnica de componentes principales como se muestra en la Figura 92.

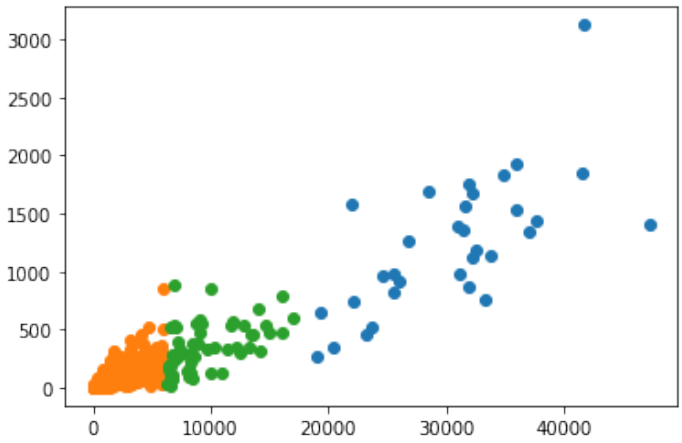


Figura 92 – Grafica PCA para usuarios respecto al volumen.

Las características de los grupos encontrados son las siguientes:

Cluster0 (Grande):

- El número de mensajes utilizados al mes superan los 30,000

Cluster1 (Pequeño):

- El número de mensajes utilizados al mes superan los 1800.

Cluster2 (Mediano):

- El número de mensajes utilizados del mes supera los 9500.

Una vez que no se utiliza la librería Pickle de Python para poder guardar y descargar el modelo entrenado.

Para poder utilizar este modelo de manera recurrente y realizar nuevas predicciones, se realizó un script en Python que expone un endpoint tipo API que recibe como parámetros las características mencionadas anteriormente del usuario y devuelve el grupo al que pertenece.

Cómo paso final se monta dentro de un Docker que contiene las librerías y herramientas necesarias para ejecutar en escribir de manera adecuada.

2.9.11 Modelo de extracción de palabras clave

Éste modelo está enfocado a utilizarse para entender los temas principales que se tratan en las preguntas que el chatbot no entendió.

Las preguntas que realizan los usuarios tienen formas muy variadas por lo que no se puede analizar directamente las preguntas que realizan los usuarios, se propone utilizar moderna de extracción de palabras clave para poder obtener los temas principales de estas preguntas. Para esta tarea se utiliza el modelo de inteligencia artificial “yanekyuk/bert-

uncased-keyword-extractor" éste modelo está entrenado para poder identificar las palabras más importantes en un texto dado [22]

Para poder utilizar este modelo fue necesario crear una script en Python que importe el modelo, lo cargue y permita acceder al mismo a través de una endpoint, de esta manera se puede enviar el texto de las preguntas realizadas por las personas y devuelve las palabras clave en la pregunta.

Además, se montó el script en una imagen de Docker suministrada por la comunidad de HuggingFace para poder ejecutarlo sin problemas.

A continuación, en la Tabla 28 se muestran algunos resultados obtenidos.

Tabla 28 – Resultados del modelo de extracción de palabras clave.

Pregunta	Palabras Clave
Good morning, we're scheduled for a hoover dam tour pickup at Aria 9:20am this morning. Can i confirm if it"s on time?	Hoover, aria, time
Hey, you are meeting us at the AIRPORT. We arrive Monday, 12:54 pm September 27	Meeting, airport
How do I reserve transport back to the airport?	Transport, airport
Where do I need to go when We arrive today at Cabo airport to find you? Do I need to call or do anything at arrival?	Cabo, airport, arrival

2.10 Implementación de los servicios.

En esta sección se detalla la manera en la que se despliegan todos estos microservicios que componen el proyecto.

A continuación, en la Figura 93 se muestra un esquema general de la implementación de los servicios principales.

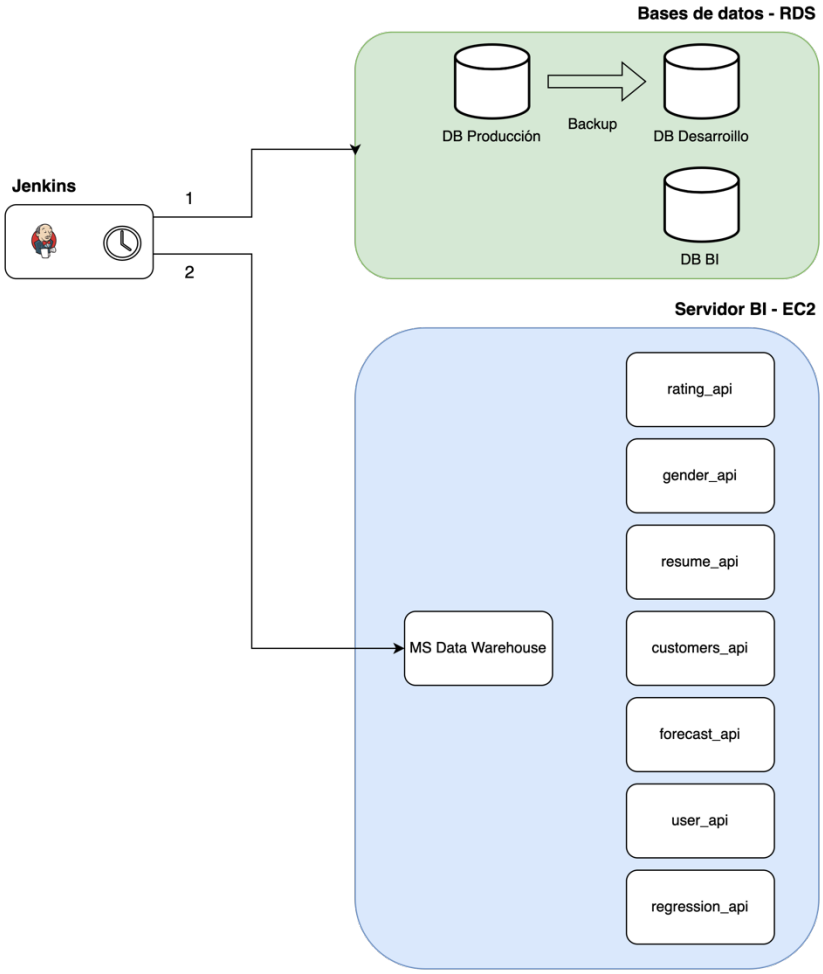


Figura 93 – Diagrama general de la implementación del proyecto.

A continuación, se detallan los componentes principales del proyecto:

2.10.1 Jenkins

Es un software de automatización que nos permite crear tareas y automatizar la ejecución de las mismas en este caso se ha utilizado Jenkins para las siguientes tareas:

- Cómo dictan las políticas de actualización de los diferentes Data Mart, los procesos tienen que iniciar el proceso de actualización de datos cada martes a las 12 de la noche por lo tanto crea una tarea Jenkins que se dispara en este horario.
- La primera parte de la tarea consiste en realizar un backup desde la base de datos de producción hacia la base de datos de desarrollo.
- Una vez finalizado el backup de la base de datos, el servidor envía una señal de arranque al microservicio “MS Data Warehouse” ubicado en el servidor BI

2.10.2 RDS

La base de datos donde se alojan las tablas creadas para el Data Warehouse es de tipo MySQL. Estas bases de datos son administradas por el servicio RDS de AWS.

2.10.3 Servidor BI

Esta es una instancia EC2 administrada por AWS en la cual se despliegan todos los servicios creados haciendo uso de la herramienta docker compose. Los microservicios creados se detallan a continuación:

- **data_warehouse:** Este es el microservicio principal que es el responsable de recibir la orden de inicio desde la tarea automática en Jenkins y realiza los procesos de carga y actualización de los diferentes Data Mart haciendo uso de las API auxiliares que se diseñaron y crearon en el proyecto.
- **rating_api:** Este microservicio permite utilizar el modelo de análisis de sentimientos descrito en la sección 2.9.8, es decir, recibe como parámetro el texto de la reseña que escribió el cliente final sobre la actividad que realizo y responde la calificacion entre 1 y 5.
- **gender_api:** Este microservicio tiene la función de exponer el modelo de clasificación de texto detallado en la sección □, para esto el microservicio expone un endpoint al que se le envía como parámetro el nombre del cliente final y devuelve el género del mismo.
- **resume_api:** El microservicio expone el modelo de extracción de palabras clave de la sección 2.9.11 a través de un endpoint que recibe como parámetro la pregunta que realizó el cliente final y que el chatbot no entendió, en la respuesta devuelve las palabras clave de la pregunta.
- **customers_api:** Este microservicio se emplea para poder utilizar el modelo de clasificación de clientes finales que se desarrolló en la sección 2.9.7 el cual recibe las características del cliente y retorna el grupo al que pertenece.
- **forecast_api:** Este modelo se usa para utilizar el modelo de predicción de transacciones desarrollado en la sección 2.9.3 que recibe la información de transacciones recibidas y devuelve la predicción de transacciones futuras.
- **user_api:** Este microservicio está diseñado para poder utilizar los modelos de clasificación en base al rating y al volumen descritos en la sección 2.9.9 y 2.9.10 para lo cual el microservicio expone un endpoint que recibe como parámetros las características del usuario y responde el clúster de rating y de volumen al que pertenece dicho usuario.

- **regresion_api:** Este microservicio se diseñó para realizar las predicciones de los modelos de predicción de suscripciones, cancelaciones y leads desarrollados en las secciones 2.9.1, 2.9.2 y de manera que para cada caso expone un endpoint que recibe la información actualizada y devuelve la predicción correspondiente.

2.11 Despliegue

La etapa de desarrollo y pruebas se implementó en un computador con las siguientes características: 16GB de RAM, procesador Core i7 con 1TB de memoria SD.

Una vez que todos los microservicios se levantan dentro del computador de pruebas usando Docker Compose se registró el consumo de recursos utilizando el comando “docker system df” y se obtuvieron los resultados de la Tabla 29.

Tabla 29 – Recursos necesarios para desplegar el sistema completo.

Item	Valor
Imágenes de docker	16
Contenedores	7
Tamaño de los volúmenes	24,94 GB
Máximo consumo de memoria RAM	4,5 GB

Los tipos de instancias T2 que provee AWS se detallan en la Tabla 30 [23].

Tabla 30 – Tipos de instancias EC2 disponibles en AWS.

Tipo de instancia	CPU virtual	Créditos	Memoria GiB
t2.nano	1	3	0,5
t2.micro	1	6	1
t2.small	1	12	2
t3.medium	2	24	4
t2.large	2	36	8

Teniendo en cuenta la Tabla 29 y Tabla 30, el servidor EC2 más adecuado para realizar esta implementación es la instancia t2.large + una memoria SSD de 30GB.

2.12 Visualización de datos

La visualización de datos es una parte primordial en este proyecto ya que permitirá apreciar de una manera más compacta, cómoda y sencilla todos los resultados obtenidos durante

el proceso de aplicación de herramientas de inteligencia de negocios. Cómo se mencionó en la sección 1.4.3, la herramienta seleccionada para crear el tablero de visualización de los datos es Google Data Studio.

2.12.1 Elección de los gráficos.

A continuación, se eligen los gráficos más apropiados para presentar la información en función de cada pregunta de negocio.

2.12.1.1 Tablero de ventas

La fuente de datos para este tablero es el Data Mart construido en la sección 2.3.

1. ¿Cómo están creciendo las suscripciones en el transcurso del año y cuál es la predicción de crecimiento hasta finalizar este año?

La pregunta implica conocer la tendencia de las suscripciones en el tiempo, por lo tanto, de elige un gráfico de líneas que dibuje el número de suscripciones acumuladas a lo largo de cada semana con la siguiente Dimensión y Métrica.

- Dimensión: Se considera la columna “Semana” de la tabla “FECHA”.
- Métrica: La información se obtiene de las columnas “Suscripciones” y “PrediccionSuscripciones” de las tablas de hechos “SUSCRIPCIONES” y “PREDICCION_SUSCRIPCIONES”.

Se identifican los siguientes KPI’s:

- El número de nuevos usuarios en la última semana.
- El total de usuarios conseguidos hasta la fecha
- El promedio de usuarios nuevos por semana.

2. ¿Cuántos clientes están cancelando sus suscripciones y cuál es la predicción de cancelaciones hasta finalizar este año?

Similar a la pregunta anterior, se trata de comprender la tendencia de cancelaciones en el tiempo por lo cual se elige un gráfico de líneas con la siguiente Dimensión y Métrica.

- Dimensión: Se considera la columna “Semana” de la tabla “FECHA”.
- Métrica: La información se obtiene de las columnas “Cancelaciones” y “PrediccionCancelaciones” de las tablas de hechos “SUSCRIPCIONES” y “PREDICCION_SUSCRIPCIONES”.

Adicionalmente, se identificaron los siguientes KPI’s:

- El número de cancelaciones en la última semana.
- El total de cancelaciones a lo largo del año.
- El promedio de cancelaciones por semana.

3. ¿Cómo se comporta el promedio de ingresos a lo largo del tiempo?

Esta pregunta también implica conocer la tendencia del promedio de ingreso mensual por lo que se elige un gráfico de líneas.

- Dimensión: Se considera la columna “Semana” de la tabla “FECHA”.
- Métrica: La información se obtiene de las columnas “IngresoPromedio” de la tabla de hechos “SUSCRIPCIONES”.

Así mismo, se identificó el siguiente KPI:

- El promedio de ingreso total

Por otra parte, se busca analizar las preguntas anteriores filtrando las fechas en las que ocurrieron por lo cual se añade un control de filtro por periodo.

El Dashboard para Ventas diseñado se muestra en la Figura 94.

2.12.1.2 Tablero de desarrollo

La fuente de datos para este tablero es el Data Mart construido en la sección 2.4

1. ¿Cómo se ha comportado la disponibilidad del servicio en el transcurso del tiempo?

Esta pregunta implica conocer la tendencia de la disponibilidad de los servicios por lo que se elige un gráfico de líneas.

- Dimensión: Se considera la columna “Semana” de la tabla “FECHA”.
- Métrica: La información se obtiene de las columnas “AppWeb”, “Envio”, “Recepcion” de la tabla de hechos “DISPONIBILIDAD”.

Los KPI’s que se identificaron son los siguientes:

- Porcentaje de disponibilidad de la aplicación web hasta la fecha.
- Porcentaje de disponibilidad del servicio de envío de mensajes hasta la fecha.
- Porcentaje de disponibilidad del servicio de recepción de transacciones hasta la fecha.

2. ¿Cuándo será necesario realizar un escalamiento horizontal de la infraestructura?

En esta pregunta se desea conocer la tendencia de las transacciones a lo largo del tiempo, por lo cual se elige una gráfica de serie temporal que presente la siguiente Dimensión y Métrica:

- Dimensión: Se considera la columna “Dia” de la tabla “FECHA_HORA”.
- Métrica: La información se obtiene de las columnas “Transacciones” y “PrediccionTransacciones” de las tablas de hechos “TRANSACCIONES” y “PREDICCION_TRANSACCIONES”.

3. ¿Cuándo será necesario escalar la base de datos?

En esta pregunta se desea conocer la tendencia del porcentaje de almacenamiento ocupado a lo largo del tiempo, por lo cual se elige una gráfica lineal que presente la siguiente Dimensión y Métrica:

- Dimensión: Se considera la columna “Mes” de la tabla “FECHA_HORA”.
- Métrica: La información se obtiene de las columnas “Almacenamiento” y “PrediccionAlmacenamiento” de las tablas de hechos “TRANSACCIONES” y “PREDICCION_TRANSACCIONES”.

4. ¿Cuáles son las horas del día en las que el uso del sistema es más intensivo?

En esta pregunta se desea conocer la tendencia de transacciones a lo largo del tiempo, por lo cual se elige una gráfica lineal que presente la siguiente Dimensión y Métrica:

- Dimensión: Se considera la columna “Hora” de la tabla “FECHA_HORA”.
- Métrica: La información se obtiene de la columna “Transacciones” de la tabla de hechos “TRANSACCIONES”.

Por otra parte, se busca analizar las preguntas anteriores para cada sistema de reserva por lo cual se añade un control de lista desplegable para los sistemas de reserva. El dashboard completo que se diseñó se presenta en la Figura 98 y Figura 99.

2.12.1.3 Tablero de marketing

La fuente de datos para este tablero es el Data Mart construido en la sección 2.5.

1. ¿Cuántos leads está generando el equipo de marketing y cuál es la predicción hasta finalizar el año?

Esta pregunta implica conocer la tendencia de leads generados por lo cual, se elige un gráfico de líneas.

- Dimensión: Se considera la columna “Semana” de la tabla “FECHA”.

- Métrica: La información se obtiene de la columna “Leads” de la tabla de hechos “LEADS”.

Los KPI’s que se identificaron son:

- El número de leads generados en la última semana.
- El número de leads totales generados hasta la fecha.
- El promedio de leads conseguidos por cada semana.

También se busca analizar las preguntas anteriores filtrando las fechas en las que ocurrieron y se añade un control de filtro por periodo. El dashboard completo para el área de marketing se muestra en la Figura 103.

2.12.1.4 Tablero para análisis de clientes finales.

La fuente de datos para este tablero es el Data Mart construido en la sección 2.6.

1. ¿De qué países provienen nuestros clientes finales?

Esta pregunta trata de realizar una comparación entre los diferentes países de los cuales provienen los clientes finales. Para esto es conveniente considerar una tabla y un gráfico de mapa geográfico.

- Dimensión: Se considera la columna “País” de la tabla “UBICACION”.
- Métrica: La información se obtiene de la columna “Cliente” de la tabla de hechos “CLIENTES”.

2. ¿Cuáles son las características principales de los clientes finales?

Esta pregunta pretende conocer la comparación entre las diferentes características de los clientes finales, por esto se consideran los gráficos de pastel con las siguientes Dimensiones y métricas.

- Dimensión: Se considera la columna “Aceptacion”, “Califica” y “Genero” de la tabla “CARACTERISITICA”.
- Métrica: La información se obtiene de la columna “Cliente” de la tabla de hechos “CLIENTES”.

3. ¿Qué tipos de clientes finales se pueden distinguir?

Esta pregunta pretende conocer la comparación entre los diferentes tipos de clientes finales, por esto se consideran un gráfico de barras con las siguientes dimensiones y métricas.

- Dimensión: Se considera la columna “Aceptacion”, “Califica” y “Genero” de la tabla “CARACTERISITICA”.
- Métrica: La información se obtiene de la columna “Cliente” de la tabla de hechos “CLIENTES”.

4. ¿Cuáles son los usuarios que tienen en mayor medida los mejores tipos de clientes finales?

En este caso se busca analizar la composición de los clientes finales de cada usuario, para realizar una visualización adecuada se considera el grafico de barras apiladas al 100% con las siguientes dimensiones y métricas:

- Dimensión: Se considera la columna “Nombre” de la tabla “USUARIO”.
- Métrica: La información se obtiene de la columna “Comun”, “Normal” e “Ideal” de la tabla “TIPO”.

El dashboard completo para el análisis de clientes finales se muestra en la Figura 104.

2.12.1.5 Tablero para análisis de usuarios

La fuente de datos para este tablero es el Data Mart construido en la sección 2.7.

1. ¿De qué países provienen nuestros usuarios?

Esta pregunta trata de realizar una comparación entre los diferentes países de los cuales provienen los clientes finales. Para esto es conveniente considerar una tabla y un gráfico de mapa geográfico.

- Dimensión: Se considera la columna “País” de la tabla “UBICACION”.
- Métrica: La información se obtiene de la columna “Cliente” de la tabla de hechos “USUARIOS”.

2. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el volumen de mensajes utilizados y transacciones recibidas?

Esta pregunta pretende conocer la comparación entre los diferentes grupos de volúmenes de los usuarios, por esto se consideran los gráficos de pastel con las siguientes Dimensiones y métricas.

- Dimensión: Se considera la columna “Pequeno”, “Medio” y “Grande” de la tabla “VOLUMEN”.
- Métrica: La información se obtiene de la columna “Usuario” de la tabla de hechos “USUARIOS”.

3. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el rating con el que le han calificado sus usuarios?

En esta pregunta se desea conocer la comparación entre los grupos de rating de los usuarios, por esto se consideran los gráficos de pastel con las siguientes Dimensiones y métricas.

- Dimensión: Se considera la columna “Bueno” y “Excelente” de la tabla “RATING”.
- Métrica: La información se obtiene de la columna “Usuario” de la tabla de hechos “USUARIOS”.

4. ¿Cuáles son los mejores usuarios?

En esta pregunta se busca conocer la composición de los usuarios en función los grupos encontrados para el volumen y el rating, para poder visualizar de una manera simple se elige un gráfico de rectángulos con las siguientes dimensiones y métricas.

- Dimensión: Se consideran las tablas “RATING” y “VOLUMEN”.
- Métrica: La columna “Usuario” de la tabla de hechos “USUARIOS”.

El dashboard completo para el análisis de usuarios se muestra en la Figura 109.

2.12.1.6 Tablero para análisis de chatbot.

La fuente de datos para este tablero es el Data Mart construido en la sección 2.8.

1. ¿Cuáles son los temas principales de las preguntas que el chatbot no entendió?

Para responder la pregunta se busca analizar la comparación entre los diferentes temas que el chatbot no entendió y para tener una visualización más general se utiliza el grafico de nubes de palabras con las siguientes dimensiones y métricas.

- Dimensión: Se considera la columna “Topico” la tabla “TEMA”.
- Métrica: La columna “Preguntas” de la tabla de hechos “CHATBOT_P”.

Además, también se usa una tabla con mapa de calor para poder analizar visualmente los temas más comunes.

2. ¿Cuáles son los comandos más utilizados por el chatbot en sus respuestas?

Para responder la pregunta se busca analizar la comparación entre los diferentes comandos (intenciones) que el chatbot utilizó y una gráfica de barras horizontales

- Dimensión: Se considera la columna “Nombre” la tabla “INTENCION”.
- Métrica: La columna “Preguntas” de la tabla de hechos “CHATBOT_R”.

También se busca analizar las preguntas anteriores filtrando las fechas en las que ocurrieron y se añade un control de filtro por periodo. El dashboard completo para el análisis del chatbot se muestra en la Figura 113.

3 RESULTADOS Y DISCUSIÓN

En esta sección se van a detallar y analizar los resultados que se obtuvieron para las diferentes preguntas planteadas en las diferentes áreas de análisis.

A continuación, se presentan los dashboards que se construyeron en la herramienta Google Data Studio, usando como fuente los datos obtenidos en los distintos Data Mart.

La interfaz está dividida en 8 páginas que se describen a continuación.

- Pagina 1. Presentación.
- Página 2. Tablero para el área de ventas.
- Páginas 3 y 4. Tablero para el área de desarrollo.
- Pagina 5. Tablero para el área de marketing.
- Página 6. Tablero para análisis de clientes.
- Página 7. Tablero para análisis de usuarios
- Página 8. Tablero para análisis de chatbot.

3.1 Resultados para el área de ventas

A continuación, se presenta el tablero generado para el área de ventas en la Figura 94.

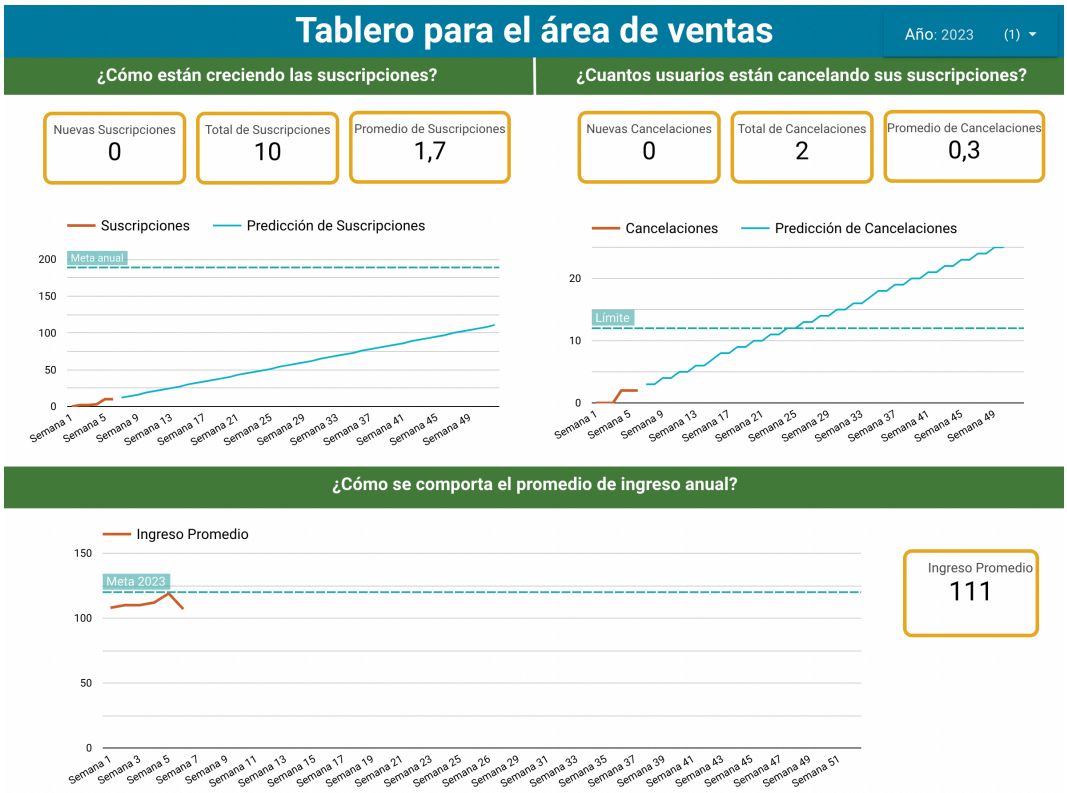


Figura 94 – Dashboard para el área de ventas

A continuación, se responde cada una de las preguntas de negocio.

1. ¿Cómo están creciendo las suscripciones en el transcurso del año y cuál es la predicción de crecimiento hasta finalizar este año?

Para responder a estas preguntas de negocio visualmente se muestra en la Figura 95.

Figura 95 – Análisis de suscripciones.

La primera parte muestra algunos KPI importantes en el área de ventas.

- El KPI “Nuevas Suscripciones” indica el número de nuevas suscripciones en la semana. En este caso indica que en esta semana no se han conseguido nuevos usuarios

- El KPI “Total de Suscripciones” indica el total de suscripciones acumuladas en el año. En este caso en total en lo que va de año se han conseguido 10 usuarios.

- El KPI “Promedio de suscripciones” indica el promedio de nuevas suscripciones por semana. En este caso se indica que en promedio se han conseguido 1.7 nuevos usuarios cada nueva semana

También se muestra una gráfica de líneas que presenta la siguiente información:

- En el eje x muestra las semanas del año.

- En el eje y se tienen dos líneas, en color rojo se muestran los valores reales de las suscripciones acumuladas a lo largo de cada semana. En color azul se muestra la predicción de nuevas suscripciones.
- La línea punteada representa la meta anual que se había planteado para el año (189 nuevas suscripciones).

Luego de analizar la gráfica se pueden determinar que, al ritmo de crecimiento actual, no va a ser posible alcanzar el objetivo anual. Lo cual implica que es necesario tomar medidas para que esto sea posible. Como recomendaciones para lograr este objetivo se mencionan las siguientes:

- Añadir más personas al equipo de ventas.
- Participar en ferias de turismo para que la empresa tenga más visibilidad y además que se puedan interactuar con nuevos usuarios potenciales.
- Incrementar la exposición de la empresa en redes sociales.
- Crear campañas y descuentos para fechas como black friday y navidad.

2. ¿Cuántos clientes están cancelando sus suscripciones y cuál es la predicción de cancelaciones hasta finalizar este año?

Para responder visualmente a esta pregunta se muestra en la Figura 96.

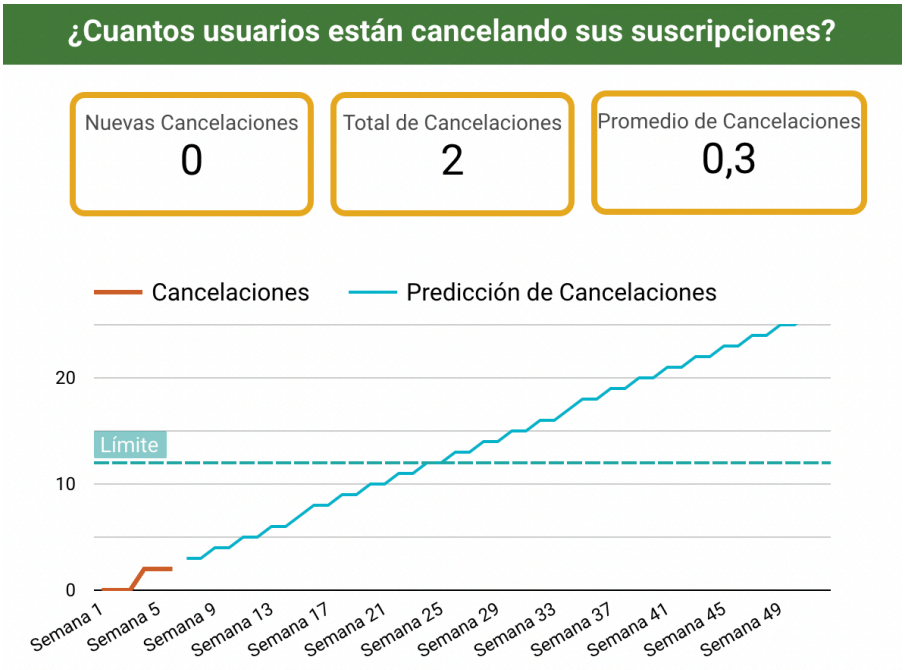


Figura 96 – Análisis de cancelaciones.

En la primera parte se muestran algunos indicadores clave KPI, que son de utilidad para poder conocer el estado de los usuarios que deciden dejar el servicio.

- El KPI “Nuevas Cancelaciones” indica el número de usuarios que decidieron cancelar su suscripción, en este caso se muestra que ningún usuario canceló su suscripción en esta semana.

- El KPI “Total de Cancelaciones” indica el número de cancelaciones en lo que va del año, en este caso son 2.

- El KPI “Promedio de Cancelaciones” indica la tasa de cancelaciones por semana, en este caso son 0.3 por semana.

También, se incluye una gráfica de líneas que contiene la siguiente información:

- En el eje x la semana del año.

- En el eje y número total de cancelaciones.

- La línea roja representa las cancelaciones reales.

- La línea azul representa una predicción de cancelaciones.

- La línea punteada representa el límite de cancelaciones.

Cómo se puede observar en la Figura 96, la predicción indica que en la semana número 24 se va a llegar al límite de posibles cancelaciones y no se debería producir ninguna más para cumplir con el objetivo. Según la predicción es posible que al finalizar el año se lleguen a 25 cancelaciones lo cual es un indicador negativo para la empresa.

Cómo medidas para mitigar que sigan creciendo el número de cancelaciones se proponen las siguientes:

- Realizar un acompañamiento a los usuarios que estén teniendo un resultado bajo con el servicio prestado por la empresa ya sea por una configuración defectuosa o que no sea acorde.

- Crear promociones para incentivar a continuar la suscripción al servicio enfocado a aquellas empresas que notifican su intención de cancelar.

3. ¿Cómo se comporta el promedio de ingreso a lo largo del tiempo?

Para responder a esta pregunta se cuenta con la Figura 97.

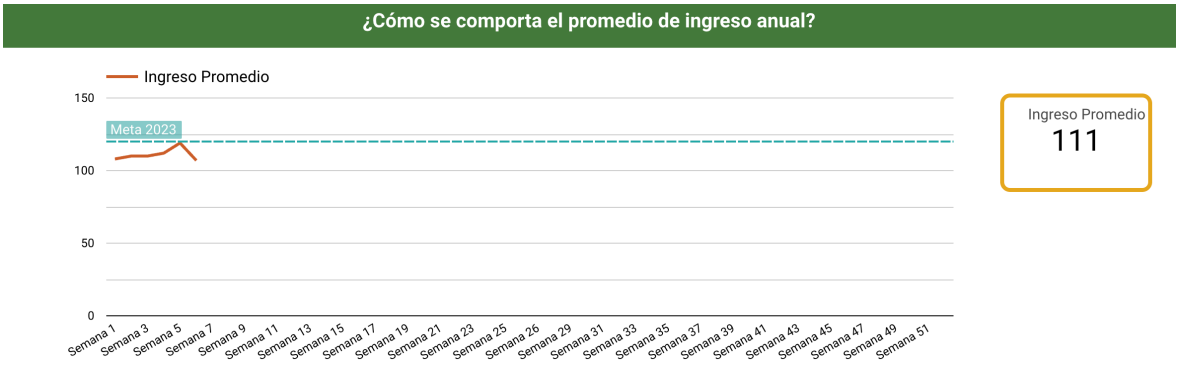


Figura 97 – Análisis del ingreso promedio.

Por una parte, se muestra un indicador cuyo valor indica el promedio de las suscripciones de todos los clientes, en este caso son $111 que está por debajo de la meta

También se incluyó una gráfica de líneas que cuenta con los siguientes componentes:

- En el eje X las semanas del año.
- En el eje Y el ingreso promedio.
- La línea roja representa el ingreso promedio a lo largo del tiempo.
- La línea punteada representa el objetivo del año 2022.

Cómo se puede apreciar en la gráfica el ingreso promedio se mantiene por debajo de la meta anual lo cual indica que los clientes nuevos han contratado planes bajos.

3.2 Resultados para el área de desarrollo

El Tablero realizado para el área de desarrollo se muestra en la Figura 98 y Figura 99.

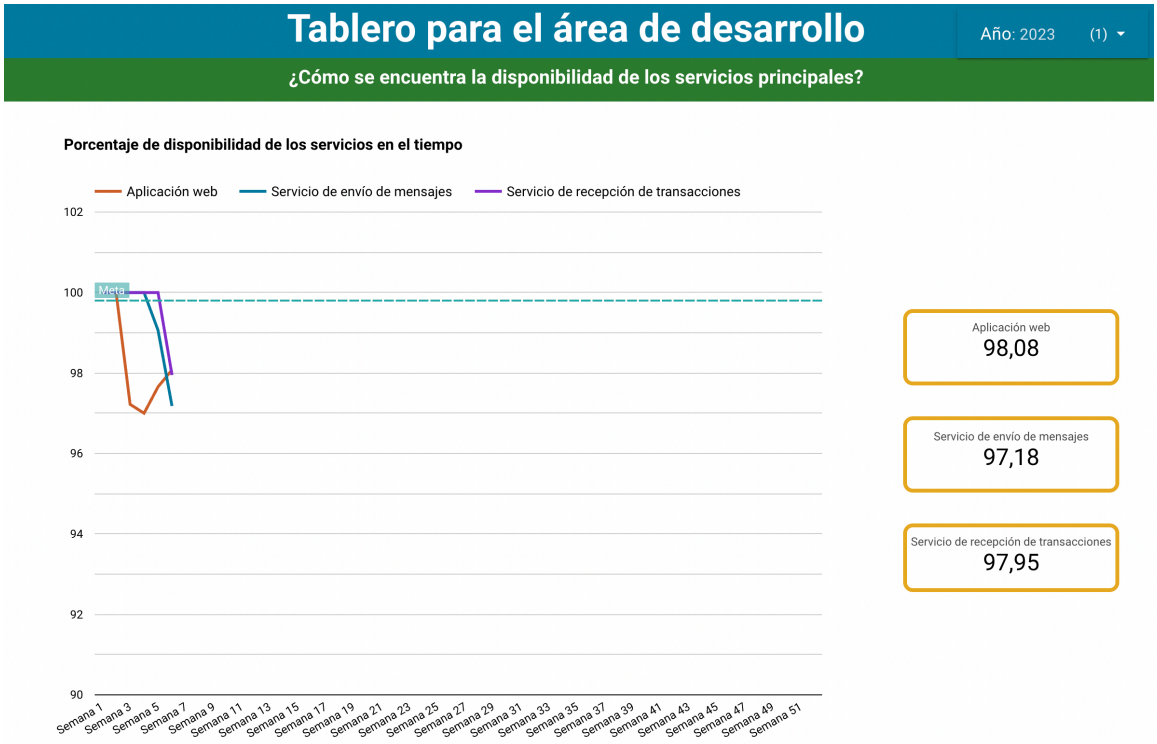


Figura 98 – Dashboard para análisis de la disponibilidad de los servicios.

Figura 99 – Dashboard para desarrollo

1. ¿Cómo se ha comportado la disponibilidad del servicio en el transcurso del tiempo?

Esta pregunta se puede responder en base a la Figura 98.

Se muestran los indicadores de la disponibilidad anual de la aplicación web, el servicio de recepción de transacciones y el servicio de envió de mensajes.

La grafica de lineas tiene los siguientes componentes:

- En el eje x se tienen las semanas del año
- En el eje y se tiene el porcentaje de disponibilidad de los servicios
- La línea de color rojo representa la disponibilidad de la aplicación web.
- La línea de color celeste representa la disponibilidad del servicio de recepción de nuevas transacciones.
- La línea de color purpura representa la disponibilidad del servicio de envió de mensajes.

Se puede notar que los servicios analizados no están cumpliendo la meta de disponibilidad semanal, ya que todos están por debajo del 99% de disponibilidad. Esto es un indicador negativo que indica que se han estado produciendo varias caídas o latencias en los

servicios principales de la empresa. Como recomendaciones para poder mitigar estos indicadores se presentan las siguientes:

- Analizar las causas de los errores en los microservicios correspondientes y corregir los mismos.
- Acelerar los procesos de reacción cuando un microservicio se encuentra caído, para esto sería conveniente considerar la migración de estos servicios a la plataforma de Kubernetes que dispone la empresa para que mejore la orquestación de los microservicios.

2. ¿Cuándo será necesario realizar un escalamiento horizontal de la infraestructura?

Esta pregunta se responde con la Figura 100, que se muestra a continuación.

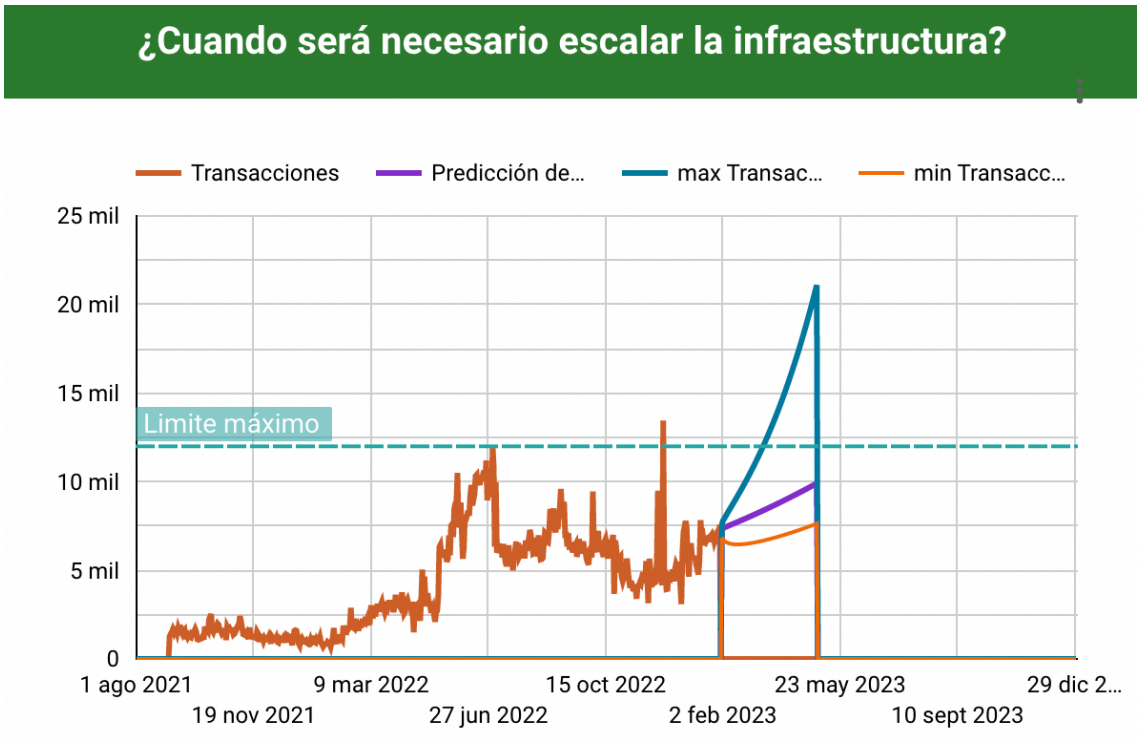


Figura 100 – Análisis de transacciones.

La grafica de serie temporal cuenta con los siguientes componentes.

- En el eje x se presentan la fecha.
- En el eje y se tienen los valores de transacciones diarias.
- La línea de color roja presenta el total de transacciones recibidas cada día del año.
- La predicción de transacciones futuras devuelve un intervalo de confianza representado por el valor predicho dibujado en la línea purpura, el valor mínimo dibujado por la línea purpura y el valor máximo representado por la línea azul.

- Las líneas horizontales representan el límite que no se deberían superar, caso contrario el sistema presentaría muchas latencias en el servicio.

De las líneas de predicciones se puede observar que es posible que antes de finalizar el mes de febrero el número de transacciones que se reciben superen los límites soportados por la infraestructura actual y que sea necesario trabajar en el escalamiento de los servicios.

Cabe señalar que, en el pasado, la cantidad de transacciones llegó a superar los límites y esto provocó varias caídas e intermitencias en el sistema, esto se produjo debido a un error de configuración que aceptaba las transacciones de usuarios que ya no estaban activos, esto se solucionó filtrando estas transacciones.

3. ¿Cuándo será necesario escalar la base de datos?

Esta pregunta se responde con la Figura 101.

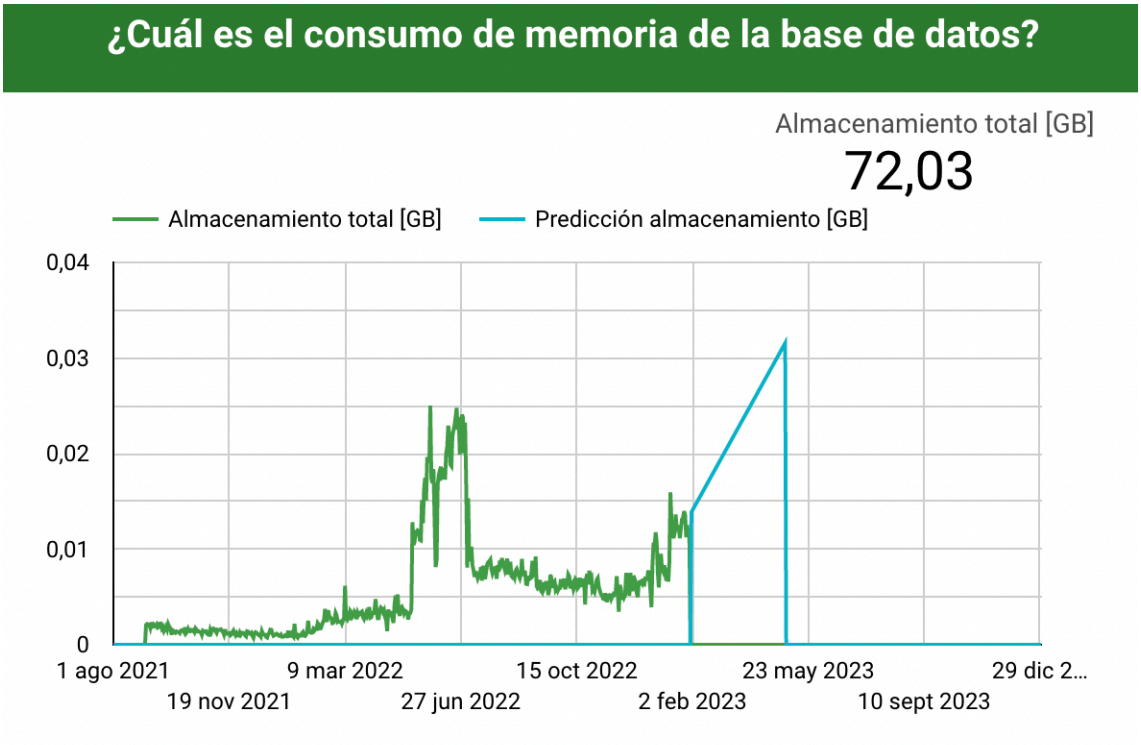


Figura 101 – Análisis de porcentaje de almacenamiento ocupado.

La grafica tiene los siguientes componentes:

- En el eje x se muestran la línea de tiempo del año.
- En el eje y el porcentaje de volumen ocupado.

- La línea verde representa el volumen nuevo utilizado diariamente y la línea azul la tendencia de la predicción del almacenamiento.

Dependiendo el sistema de reserva, la gráfica cambia, pero en la mayoría se puede notar que el volumen alcanzado no superará los límites en el corto plazo.

4. ¿Cuáles son las horas del día en las que el uso del sistema es más intensivo?

Figura 102 – Análisis de transacciones por hora.

De la Figura 102, se puede concluir que en porcentaje la hora del día con menor cantidad de transacciones recibidas son las 5 A.M. en horario central.

3.3 Resultados en el área de marketing

1. ¿Cuántos leads está generando el equipo de marketing y cuál es la predicción hasta finalizar el año?

Figura 103 – Dashboard para el área de marketing.

Éste Dashboard se compone de los siguientes componentes:

- El indicador “Nuevas intenciones de compra” muestra el total en la semana.
- El indicador “Total acumulado” muestra el total de intenciones de compra conseguidos durante todo el año.
- El indicador “Promedio semanal” muestra la cantidad promedio de intenciones de compra por cada semana.

La gráfica de líneas tiene las siguientes características:

- En el eje x se muestra las semanas del año.
- En el eje y se muestran los leads
- La línea roja presenta los valores reales de intenciones de compra.
- La línea azul presenta los valores de predicción de intenciones de compra.

Cómo se puede apreciar en la Figura 103 el número de leads que se predice conseguir hasta la finalización del año (134) va a superar el objetivo anual que son 76. Este es un buen indicador para la empresa ya que seguramente se va a superar el objetivo.

3.4 Resultados en el análisis de clientes finales

El dashboard para el análisis de los clientes finales se muestra en la Figura 104.

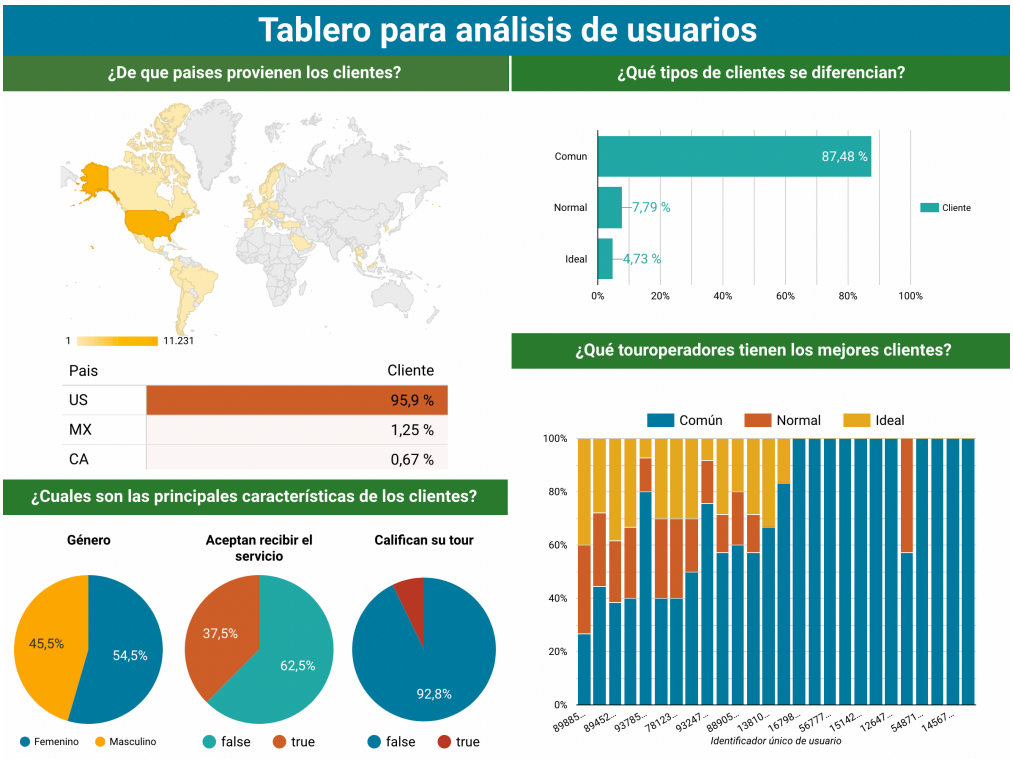
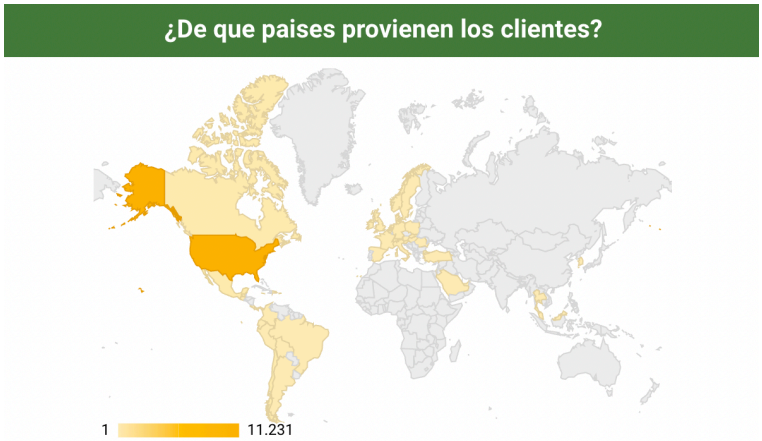


Figura 104 – Dashboard para análisis de clientes finales.

1. ¿De qué países provienen nuestros clientes finales?

Para responder esta pregunta se tiene la gráfica y tabla de la Figura 105.



Pais	Cliente
US	95,9 %
MX	1,25 %
CA	0,67 %

Figura 105 – Análisis de clientes finales por ubicación.

En la Figura 105, se puede apreciar qué los clientes finales provienen principalmente de Estados Unidos con un amplio 92.96%, Canadá con solo el 5.84% y el resto de los países con un porcentaje mínimo. Esto significa que existe un mercado muy amplio por explorar para la empresa en el resto de países europeos.

2. ¿Cuáles son las características principales de los clientes finales?

Para responder a esta pregunta se presenta la Figura 106.

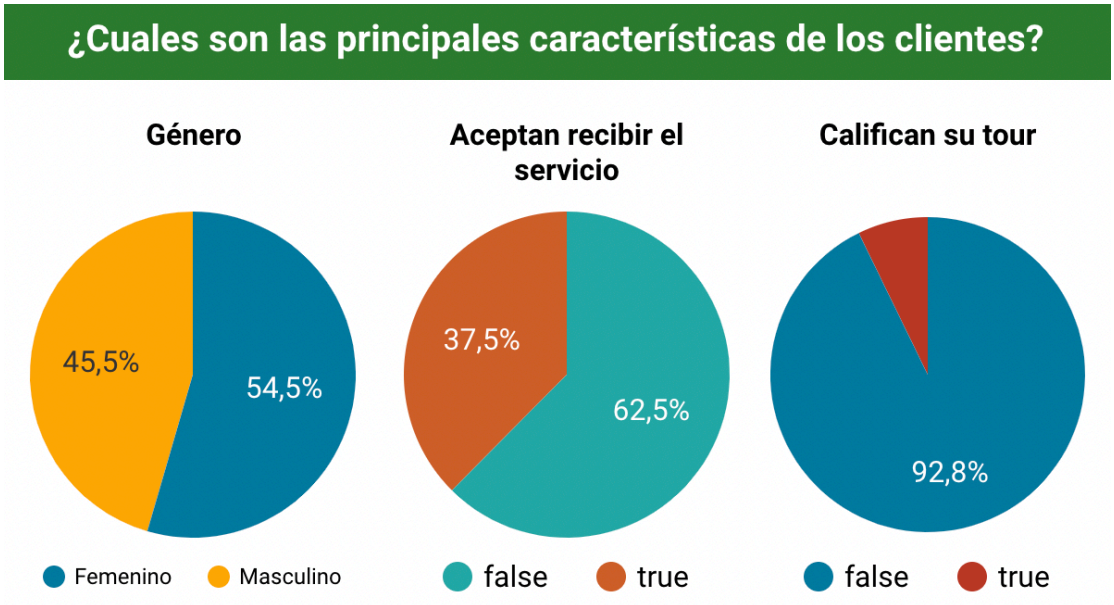


Figura 106 – Análisis de clientes finales por sus características.

Se puede identificar que las principales características de los clientes finales son las siguientes:

- Género: Los clientes finales son principalmente mujeres con un 54,5%
- Clientes que aceptan recibir el servicio: El 37.5% de los clientes finales no acepta recibir el servicio.

Lo anterior indica que existe una importante oportunidad de mejora que puede estar enfocada en temas como: La redacción de los mensajes, la longitud del mensaje, el tono de los mensajes ("formal", "casual") o el uso de emojis.

- Clientes que llegan a calificar la actividad que realizaron: Aquí también existe un indicador negativo ya que el 92.8% de los clientes finales no la califica la actividad que realizó.

Se le debe dar una importancia alta ya que esta acción de calificar el tour ayuda a los clientes finales a mejorar su posición en Google o en TripAdvisor lo cual hace que los usuarios estén más satisfechos con el servicio.

3. ¿Qué tipos de clientes finales se pueden distinguir?

Para responder a esta pregunta se presenta la Figura 107.

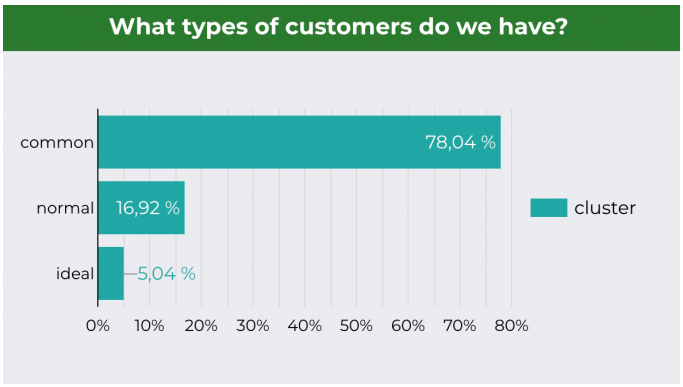


Figura 107 – Análisis de clientes finales de acuerdo con el tipo.

Los tipos de clientes finales que se pueden identificar son los siguientes:

- Ideal: corresponde al 5.04% del total
- Normal: corresponde al 16.92%
- Común: corresponde al 78.04%

Los resultados anteriores indican que

Considerando las características de los grupos que se identificaron gracias al modelo de clasificación de clientes finales en la sección 2.9.7, se puede notar que los clientes finales que presentan una interacción alta con el sistema son muy bajos.

En este caso las oportunidades de mejora que se pueden identificar están orientadas a entrenar al chatbot con nuevos temas de conversación y que el chatbot pueda emitir información más relevante sobre las actividades.

4. ¿Cuáles son los usuarios que tienen en mayor medida los mejores tipos de clientes finales?

Para responder esta pregunta se presenta la Figura 108

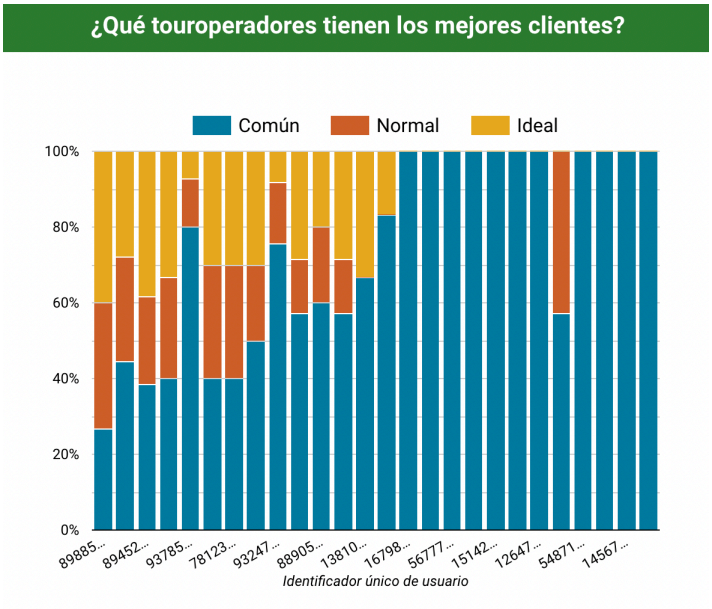


Figura 108 – Análisis de usuarios de acuerdo con sus clientes.

La Figura 108, presenta porcentajes acumulados de los tipos de clientes finales para los distintos usuarios.

En color azul se presentan los clientes finales “Comunes”, en color rojo están los clientes finales “Normales” y en amarillo están los clientes finales “Ideales”.

De esta manera se puede determinar los usuarios que tienen los mejores clientes finales, esto es de gran importancia ya que permite analizar las configuraciones de las cuentas de estos usuarios para poder conocer cómo ayudar a mejorar a los demás usuarios.

3.5 Resultados en el análisis de usuarios

El dashboard que se generó para el análisis de los usuarios se muestra en la Figura 109.

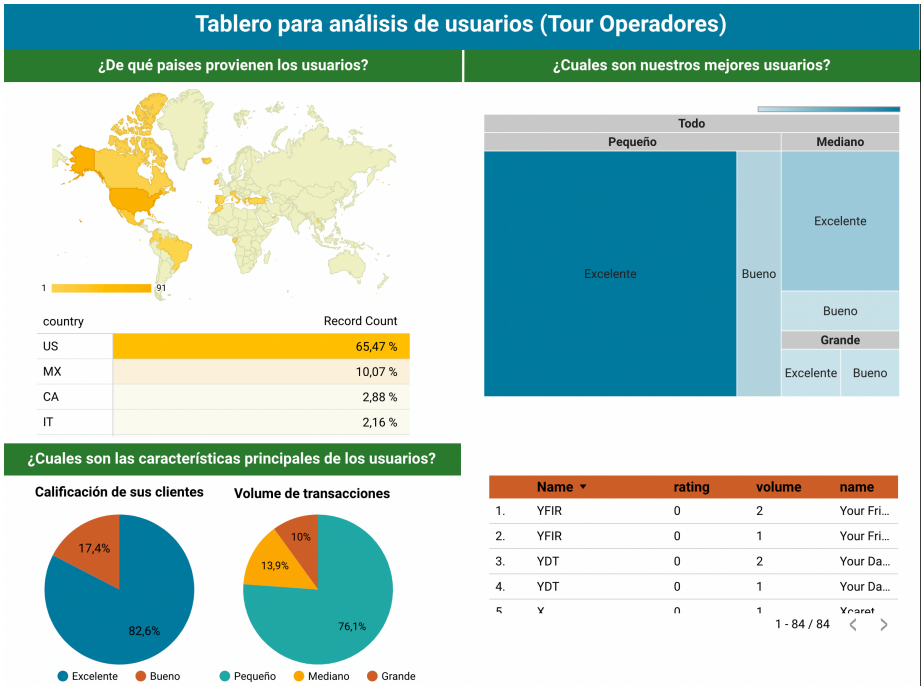


Figura 109 – Dashboard para el análisis de usuarios.

1. ¿De qué países provienen nuestros usuarios?

Para responder a esta pregunta se presenta la Figura 110.

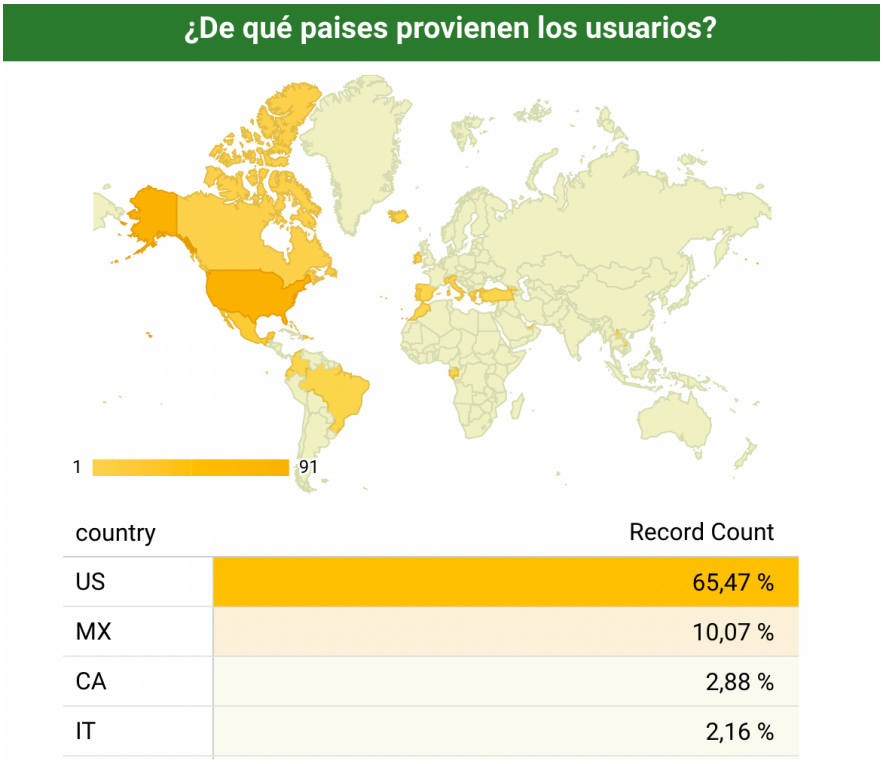


Figura 110 – Análisis de usuarios de acuerdo con la ubicación.

Se puede notar que los suscriptores provienen principalmente de Estados Unidos, México y Canadá.

2. ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el volumen de mensajes utilizados y transacciones recibidas? Y ¿Qué tipos de usuarios se pueden distinguir de acuerdo con el rating con el que le han calificado sus usuarios?

La pregunta se puede responder con la Figura 111.

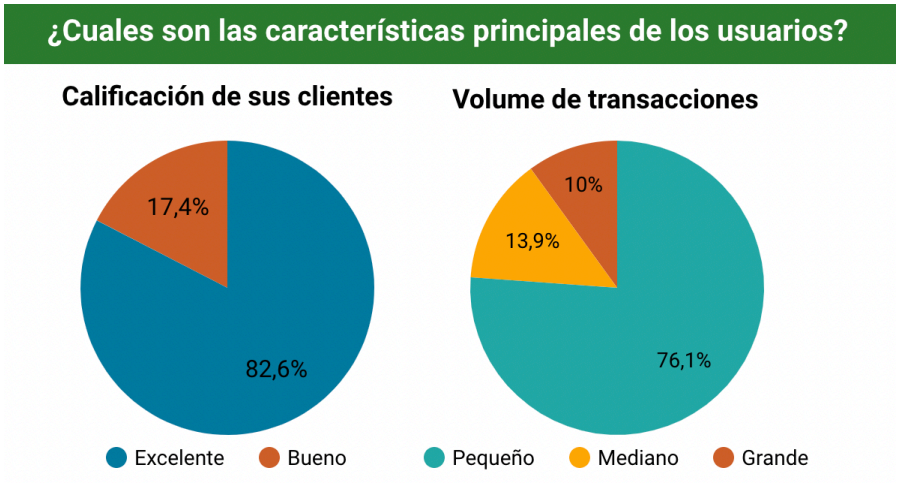


Figura 111 – Análisis de usuarios de acuerdo con sus características.

Rating

Se identificaron los usuarios “buenos” con el 17,4% y “excelentes” con el 82,6% de participación.

El resultado anterior es un indicador muy bueno porque permite apreciar que los usuarios están recibiendo buenas calificaciones de parte de los clientes finales y en parte también es gracias al servicio de la empresa ya que se incentiva a los usuarios a calificar las actividades y cuando la reseña es positiva se le pide que deje un comentario en Google o TripAdvisor. De esta manera se comprueba que el servicio está ayudando a los usuarios.

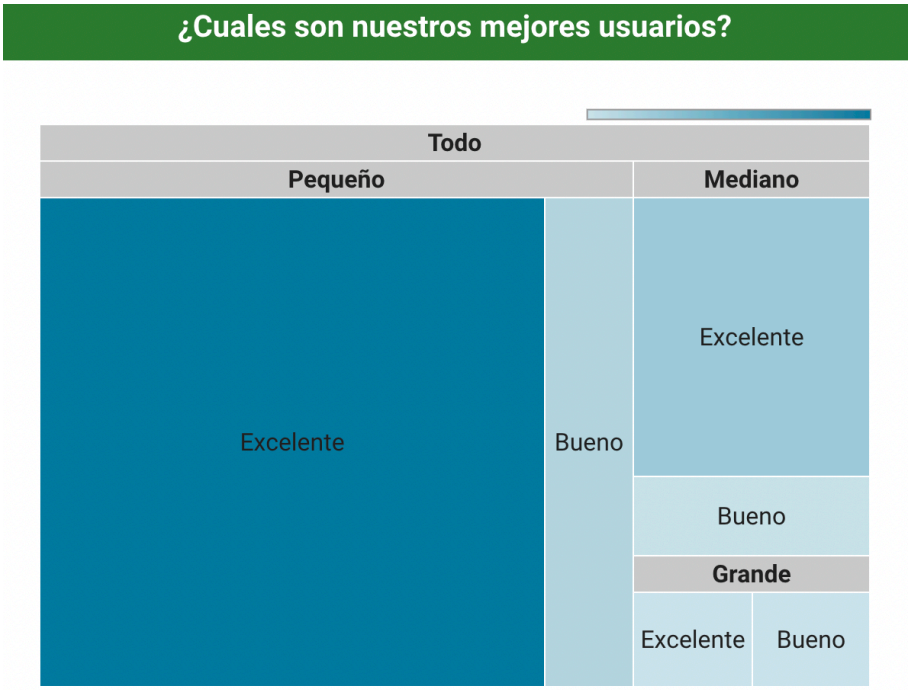
Volumen

Para esta categoría se identificaron los usuarios “Pequeños” con 76,1% de participación, “Medianos” con 13.9% de participación y “Grandes” con 76.1% de participación.

El resultado anterior indica que principalmente existen usuarios pequeños en relación con el número de transacciones recibidas y el número de mensajes enviados.

3. ¿Cuáles son los mejores usuarios?

Esta pregunta se puede responder con la herramienta de la Figura 112.



	Name ▼	rating	volume	name
1.	YFIR	0	2	Your Fri...
2.	YFIR	0	1	Your Fri...
3.	YDT	0	2	Your Da...
4.	YDT	0	1	Your Da...
5	X	0	1	Xcaret

1 - 84 / 84

Figura 112 – Identificar usuarios de acuerdo con sus características.

Con esta herramienta es posible conocer los diferentes tipos de usuarios, para ello se selecciona el cuadro correspondiente a las características que se desean analizar, en el caso anterior se buscan los usuarios con las calificaciones “Excelentes” y de tamaño “Grande” con lo cual en la tabla de arriba se muestran los usuarios que cumplen con estas características.

3.6 Resultados en el análisis del Chatbot

El Dashboard para el análisis del chatbot se muestra en la Figura 113.

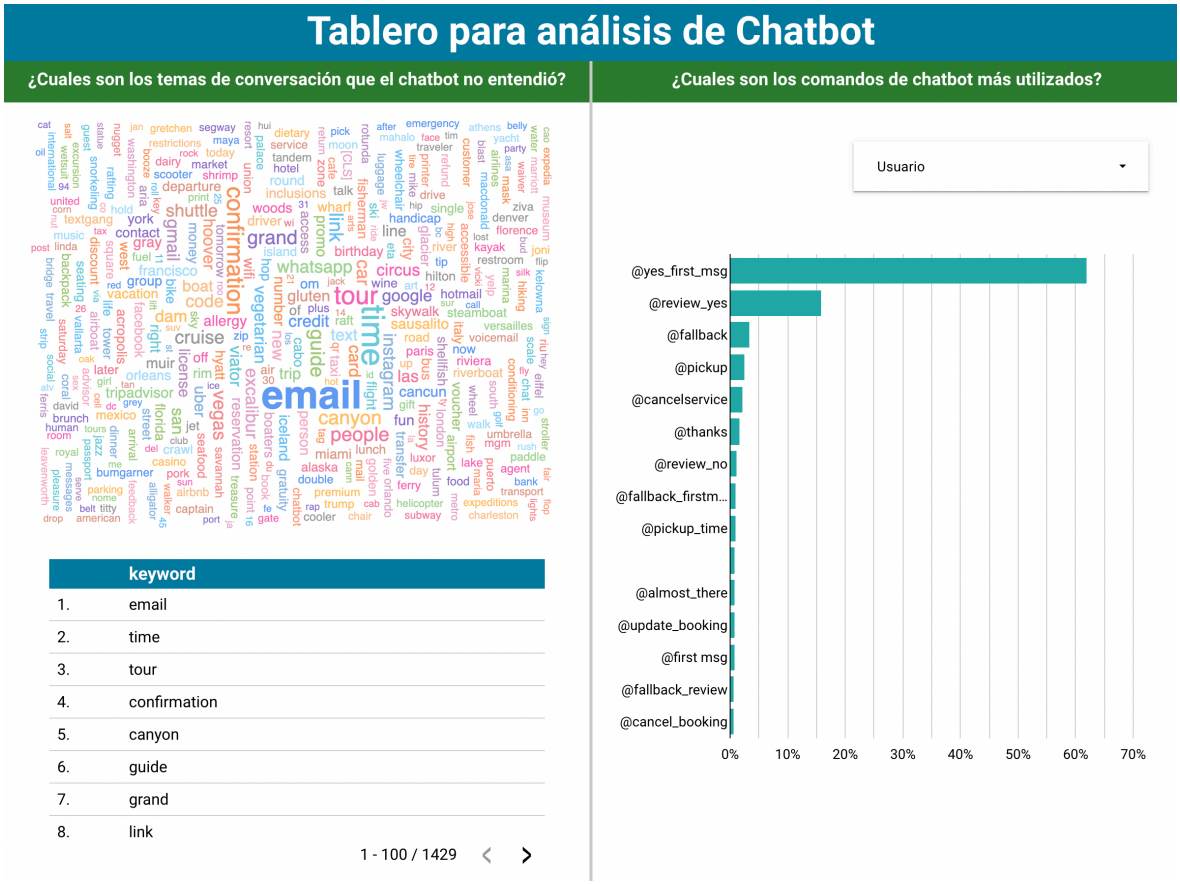


Figura 113 – Dashboard para análisis de chatbot.

1. ¿Cuáles son los temas principales de las preguntas que el chatbot no entendió?

Para poder responder esta pregunta se presenta la Figura 114.

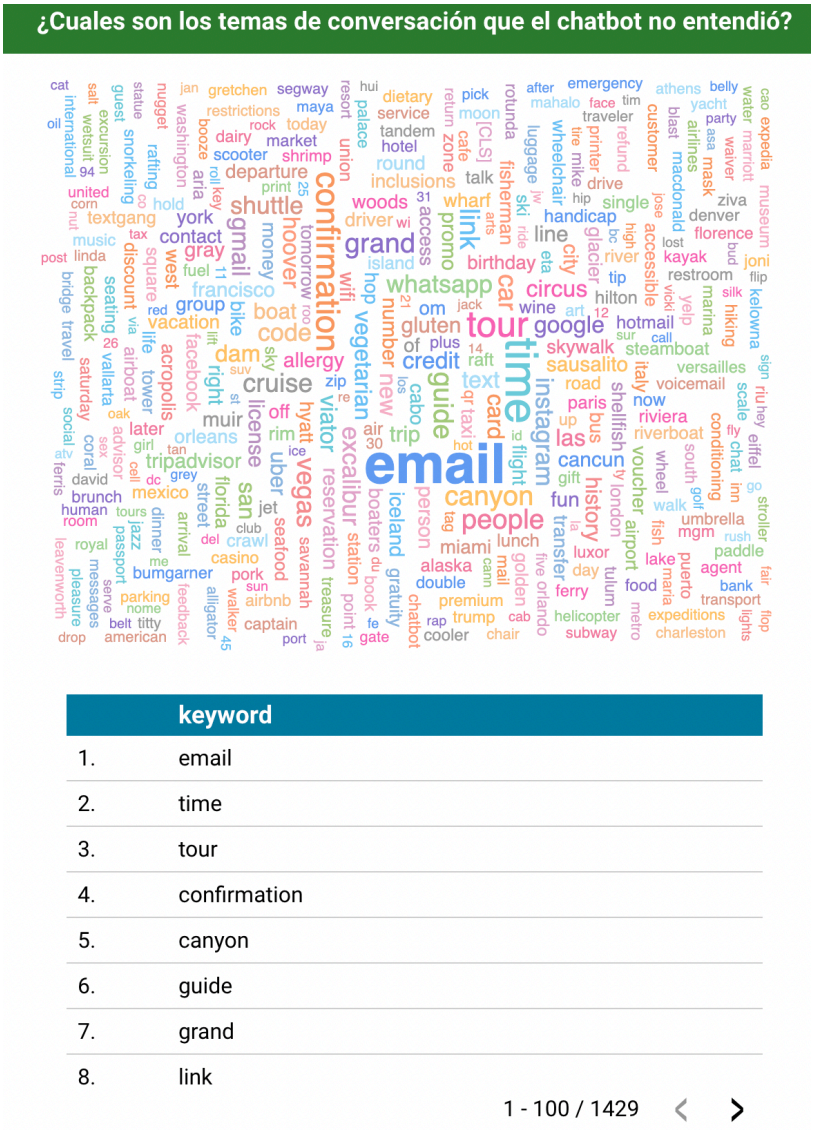


Figura 114 – Análisis de tópicos que el chatbot no entendió.

Los componentes de esta visualización son los siguientes:

- Una nube de palabras que resalta las palabras más comunes en la serie de datos con la que se le alimenta.
- Una tabla de calor que muestra las palabras más comunes.

En la Figura 114 se puede notar que los temas de conversación que los clientes finales le preguntaron al chatbot y no entendió son los siguientes: “email”, “time”, “tour”, “confirmation”, etc. Esta información es muy valiosa para poder entrenar al chatbot con estos nuevos temas de conversación con lo cual se consigue que el chat emita menos fallbacks y que la calidad del servicio hacia los clientes finales mejore.

2. ¿Cuáles son los comandos más utilizados por el chatbot en sus respuestas?

La respuesta a esta pregunta se muestra en la Figura 115.

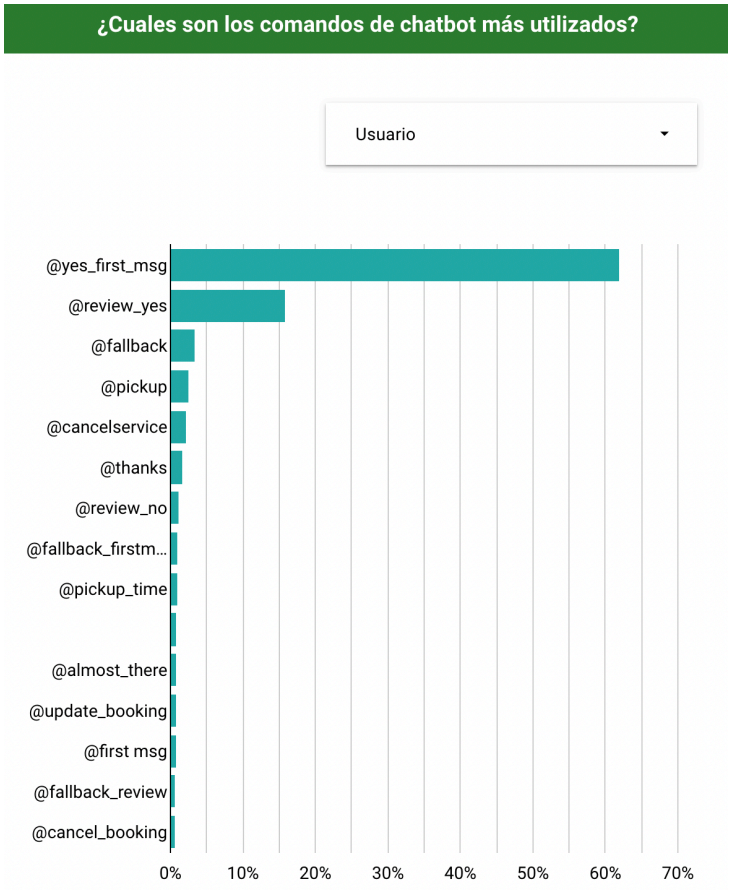


Figura 115 – Comandos de chatbot más comunes.

En la Figura 115 se muestran los comandos más utilizados por el chatbot y se puede notar que uno de ellos es @fallback lo cual indica que hay muchos temas de conversaciones que el chatbot no está entendiendo y esto debe ser tratado para mejorar el servicio según lo descubierto en la pregunta anterior. Por otra parte, también se tiene el comando @thanks entre los más utilizados, lo cual indica que el servicio brindado por la empresa crea el sentimiento de gratitud en los clientes finales.

4 CONCLUSIONES

- Se implementaron herramientas de inteligencia de negocios para la empresa TourOpp que permitió responder las preguntas de negocio más importantes que presentó cada área lo cual le aportó mucho valor y la posibilidad de identificar sus puntos fuertes y también las oportunidades de mejora para los puntos débiles.

- El uso de la metodología Hefesto 2.0 ayudó mucho en el proceso de planificación, creación e implementación de los diferentes Data Mart para las áreas más importantes de la empresa como son desarrollo de software, ventas, marketing, análisis de clientes finales, usuarios y su chatbot. De esta forma el proceso se llevó a cabo de una forma ordenada y eficiente.

- La aplicación de inteligencia de negocios se desarrolló utilizando herramientas de uso gratuito tal y como se plantearon en las restricciones proporcionadas por la empresa, las herramientas que se seleccionaron brindaron las funcionalidades necesarias para llevar a cabo todo el proyecto desde la construcción de los Data Mart, los procesos de carga y actualización y también el proceso de visualización de datos.

- Entre las ventajas más importantes de este proyecto se destacan que los costos asociados a este proyecto son mínimos ya que solo cuenta el costo del servidor donde se despliegan los microservicios creados y que se puede personalizar cada paso del proceso de acuerdo a la realidad y necesidades de la empresa.

- El uso de modelos de inteligencia artificial presentó un aporte valioso al descubrir conocimiento nuevo a partir de la información disponible en la empresa. Para lograr esto se utilizaron modelos de inteligencia artificial desarrollados por la comunidad de HuggingFace y también se entrenaron modelos propios según las necesidades de la empresa.

- El uso de las herramientas Docker y Docker Compose permitió que el despliegue de todos los servicios que se desarrollaron en este proyecto se realice de una

manera sencilla y eficiente, de esta manera se pudieron ejecutar servicios creados en diferentes lenguajes (JavaScript y Python), con diferentes versiones de librerías y dependencias y no fue necesario realizar las configuraciones específicas que requieren todos los microservicios en un mismo servidor.

- La herramienta de visualización de datos se realizó utilizando la herramienta gratuita Google Data Studio que permitió generar un reporte de calidad, que presenta graficas para las ideas que se intentan transmitir de acuerdo a los datos disponibles, que es dinámico, que genera valor y le ayuda a tomar mejores decisiones a la empresa.

RECOMENDACIONES

- La empresa TourOpp está en constante cambio y crecimiento, lo cual implica que este proyecto debe continuar actualizándose y trabajando para poder seguir mejorando en función de las futuras necesidades de la empresa.

- Es importante considerar la realimentación de las personas que utilicen el dashboard para conocer si cumple con los objetivos planteados y en el caso de que no estén alineados realizar las correcciones necesarias.

- Para obtener información verídica es necesario conocer a detalle que significa cada dato en cada fuente de información ya que es muy fácil caer en falsas interpretaciones de los mismos y producir errores.

- Es necesario tener presente que las personas que utilizan el dashboard deben tener permisos limitados de lectura ya que si los permisos no están limitados podrían modificar o eliminar los componentes del dashboard.

- Es importante mantener la calidad de los datos desde los orígenes de los mismos para mejorar la calidad de este proyecto dado que datos incompletos, nulos o erróneos son inconvenientes para el análisis.

TRABAJO FUTURO

- Actualmente el proyecto contempla las preguntas de negocio más relevantes, pero existe la posibilidad seguir aumentando o modificando el mismo de acuerdo a las necesidades empresariales.

- La creación de reportes es otro aspecto de gran relevancia para la empresa, este proyecto logró integrar varias fuentes de datos y descubrir información nueva para la empresa, pero puede ser importante crear reportes que permitan recopilar dicha información y compartirla con sus usuarios.

REFERENCIAS BIBLIOGRÁFICAS

[1] S. M. Kumar and M. Belwal, “Performance dashboard: Cutting-edge business intelligence and data visualization,” *Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017*, pp. 1201–1207, 2018, doi: 10.1109/SmartTechCon.2017.8358558.

[2] P. R. M. de Andrade and S. Sadaoui, “Improving business decision making based on KPI management system,” *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, vol. 2017-Janua, pp. 1280–1285, 2017, doi: 10.1109/SMC.2017.8122789.

[3] T. Lähteenmäki, “Optimal sales team structure in software business,” 2017, [Online]. Available: https://aaltodoc.aalto.fi/handle/123456789/26680

[4] R. D. Bernabeu, “Hefesto Data Warehouseing,” p. 146, 2010, [Online]. Available: http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/hefesto-metodologia-propia-para-la-construccion-un-data-warehhttp://www.dataprix.com/data-warehousing-y-metodologia-hefesto/ii-hefesto-metodologia-propia-para-la-construccion-un-data-wa

[5] G. E. Silva Peñafiel, V. M. Zapata Yánez, K. P. Morales Guamán, and L. M. Toaquiza Padilla, “Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones,” *Ciencia Digital*, vol. 3, no. 3.4., pp. 397–418, Sep. 2019, doi: 10.33262/cienciadigital.v3i3.4..922.

[6] G. R. Rivadera, “La metodología de Kimball para el diseño de almacenes de datos (Data warehouses),” *Cuadernos de Ingeniería*, no. 5, pp. 56–71, 2010, [Online]. Available: http://revistas.ucasal.edu.ar/index.php/CI/article/view/169

[7] T. Brown, “Data Warehouse Implementation with the SAS ® System.”

[8] O. Hutsulyak, “Node.js vs Python — What to Choose in 2022,” *TechMagic*, 2021. https://www.techmagic.co/blog/node-js-vs-python-what-to-choose/ (accessed Mar. 09, 2022).

[9] “Home - Docker.” https://www.docker.com/ (accessed Sep. 14, 2022).

[10] “Overview | Docker Documentation.” https://docs.docker.com/compose/ (accessed Sep. 14, 2022).

[11] “Jenkins.” https://www.jenkins.io/ (accessed Sep. 14, 2022).

[12] Amazon, “AWS | Servicio de Bases de Datos Relacionales,” *AWS Amazon*, 2019. https://aws.amazon.com/es/rds/ (accessed Mar. 10, 2022).

[13] “Google Data Studio Overview.” https://datastudio.google.com/overview (accessed Sep. 14, 2022).

[14] “14 Best Types of Charts and Graphs for Data Visualization [+ Guide].” https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization (accessed Dec. 25, 2022).

[15] “Te damos la bienvenida a Colaboratory - Colaboratory.” https://colab.research.google.com/ (accessed Sep. 16, 2022).

[16] U. Khalid, L. E. Okafor, and K. Burzynska, “Does the size of the tourism sector influence the economic policy response to the COVID-19 pandemic?,” *Current Issues in Tourism*, vol. 24, no. 19, pp. 2801–2820, 2021, doi: 10.1080/13683500.2021.1874311.

[17] E. Y. Nikolskaya, V. A. Lepeshkin, E. A. Blinova, I. P. Kulgachev, and S. V. Ilkevich, “Improvement of digital technology in the tourism sector,” *Journal of Environmental Management and Tourism*, vol. 10, no. 6, pp. 1197–1201, 2019, doi: 10.14505/jemt.v10.6(38).01.

[18] G. Edward and G. Meirion Jenkins, “SERIES TEMPORALES, MODELO ARIMA METODOLOGÍA DE BOX-JENKINS”.

[19] “BERT.” https://huggingface.co/docs/transformers/main/en/model_doc/bert#transformers.BertForSequenceClassification (accessed Dec. 19, 2022).

[20] “huggingface/transformers-pytorch-gpu - Docker Image | Docker Hub.” https://hub.docker.com/r/huggingface/transformers-pytorch-gpu (accessed Dec. 19, 2022).

[21] “nlptown/bert-base-multilingual-uncased-sentiment · Hugging Face.” https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment (accessed Dec. 19, 2022).

[22] “yanekyuk/bert-uncased-keyword-extractor · Hugging Face.” https://huggingface.co/yanekyuk/bert-uncased-keyword-extractor (accessed Dec. 19, 2022).

[23] “Tipos de instancias de Amazon EC2 - Amazon Web Services.” https://aws.amazon.com/es/ec2/instance-types/ (accessed Dec. 27, 2022).

ANEXOS

ANEXO A

Uno de los pasos iniciales para el proyecto es identificar las preguntas de negocio que la empresa necesita responder, para esto se organizaron reuniónes con el equipo de desarrollo, ventas y gerencia en las cual se recopilaron una serie de ideas que sirvieron de base para poder construir las preguntas de negocio que se abordan a lo largo del proyecto. A continuación, se detallan las ideas iniciales que se recopilaron con los diferentes equipos.

Desarrollo:

- Analizar los fallbacks (preguntas que el chatbot no entiende) y su influencia en el sistema.
- Analizar el uso de los comandos del chatbot (las intenciones de las preguntas que realizan los usuarios y que el chatbot está entrenado para detectar y responder), verificar el porcentaje de uso de los comandos del chatbot verificar si están resolviendo o no los problemas para los que se plantearon.
- Estimar los tiempos que tarda el procesamiento de los diferentes microservicios.
- Generar proyecciones sobre la demanda del servicio, estar listos para poder brindar el servicio a más clientes.
- Recopilar los meses o temporadas con mayor actividad.
- Analizar los picos horarios de los consumos de la aplicación.
- Agrupar los diferentes clientes finales por sexo, país, estado, número de mensajes en promedio para extraer características de los grupos.

Gerencia:

- Estudiar si nuestros clientes finales que menos mensajes reciben si ponen más reviews que aquellos que más mensajes reciben
- Mapear las preguntas que el chatbot no entiende para crear nuevos comandos, aumentar la satisfacción del cliente y poder enviar más mensajes.

- Descubrir si mientras más mensajes se envían a un cliente final, más preguntas realizan.
- Descubrir si mientras más mensajes se envían, mejor es el rating o review
- Mientras más corto o conciso es el primer mensaje hay más personas que dicen ok al servicio.
- Si el operador está en el extranjero es más probable que acepte recibir el servicio.
- Cuando los clientes finales reciben muchos fallbacks hay mayor probabilidad de que decidan dejar de recibir el servicio.
- Cuando el tour operador utiliza nuestro servicio, los reviews aumentan y también la calidad de estos y la también la satisfacción de los clientes finales
- Revisar cuantas personas vuelven a comprar otro producto del mismo turoperador en una fecha distinta
- Mejorar la infraestructura actual con los datos recopilados. Es decir, determinar si mañana se van a necesitar más servidores, etc
- Revisar si se cumple con la oferta de valor. Que tan cerca estamos de cumplir con la oferta de valor.
- Otra oferta de valor de la empresa es poder reducir el tiempo que los tours operadores responden a las preguntas frecuentes.

Ventas:

- Obtener información de las cuentas de los usuarios y los mensajes enviados.
- Cuál es el impacto de activar funcionalidades no utilizadas para el envío de mensajes.
- Incremento de reviews mediante la estrategia de discount codes
- Usar rebrandly para poder conocer cuantas veces se da click en los enlaces
- Conteo de reviews