Pontificia Universidad Católica del Ecuador

Facultad De Ingeniería

Escuela de Sistemas



TEMA:

Análisis del área comercial de una empresa proveedora de servicios de internet aplicando inteligencia de negocios.

AUTOR:

JOSE ANTONIO MONTENEGRO AGUIRRE

TRABAJO PREVIA A LA OBTENCIÓN DEL TÍTULO DE INGENIERO DE SISTEMAS DE LA INFORMACIÓN

QUITO, 26 de Mayo del 2023

DEDICATORIA

Es con gran orgullo y satisfacción que dedico esta tesis a las personas más importantes en mi vida, quienes han sido un pilar fundamental en mi camino académico y personal.

A mis padres, Emerson Montenegro y Raquel Aguirre, quienes siempre han sido mi guía y apoyo en todo momento. Gracias por inculcarme valores como la perseverancia, el esfuerzo y la dedicación, y por enseñarme que no hay límites cuando se tiene la determinación de lograr algo.

A mis hermanas, Barbara Montenegro y Lizeth Montenegro, por su incondicional apoyo y motivación en todo momento. Su presencia ha sido fundamental para mí, y estoy agradecido(a) por su amor y compañía.

A mi novia, Camila Cajas, por su presencia constante y por ser mi fuente de inspiración. Gracias por estar ahí en los momentos buenos y en los difíciles, y por alentarme a perseguir mis sueños.

A mis abuelos, Raúl Montenegro, Digna Castro, Guadalupe Villarreal y Rómulo Aguirre, que en paz descanse, por ser mi ejemplo de perseverancia, humildad y sabiduría. Siempre llevaré en mi corazón sus enseñanzas y su legado.

A mis tíos, por su constante apoyo y por ser una fuente de inspiración en mi vida.

Esta tesis es una muestra de mi gratitud y agradecimiento hacia todas las personas que han sido parte de mi camino. Espero que este logro les haga sentir tan orgullosos como yo me siento de pertenecer a una familia maravillosa y llena de amor.

Con todo mi amor y agradecimiento, José Antonio Montenegro Aguirre

AGRADECIMIENTO

Me dirijo a ustedes con profunda gratitud y emociones en el corazón para expresar mi sincero agradecimiento a todas las personas que han sido un apoyo fundamental en la realización de mi tesis y en general en la culminación de mi carrera.

En primer lugar, quisiera dar las gracias a mis padres, Emerson Montenegro y Raquel Aguirre, por su inquebrantable apoyo, paciencia y amor incondicional en todo momento. Sin su apoyo, no habría sido posible llegar hasta aquí y alcanzar mi meta académica.

También quisiera agradecer a mis hermanas, Barbara Montenegro y Lizeth Montenegro, quienes siempre estuvieron presentes para ofrecer su ayuda y motivación en los momentos más difíciles.

Asimismo, quiero reconocer el apoyo incondicional de mi novia, Camila Cajas. Su amor, paciencia y comprensión fueron fundamentales en este proceso. No habría podido lograrlo sin su presencia constante y su incondicional apoyo.

Agradezco también a mis abuelos, Raúl Montenegro, Digna Castro, Guadalupe Villarreal y Rómulo Aguirre, que en paz descanse, por ser una fuente de inspiración y sabiduría para mí. Su ejemplo y enseñanzas siempre estarán presentes en mi vida.

Finalmente, quisiera agradecer a mis tíos, quienes siempre estuvieron dispuestos a brindarme su ayuda y apoyo en todo momento.

Agradezco profundamente a todas las personas que han sido parte de este proyecto y han hecho posible la realización de mi tesis. Espero poder retribuirles todo lo que han hecho por mí en un futuro cercano.

Con afecto y gratitud, José Antonio Montenegro Aguirre

RESUMEN

Este es un proyecto que busca aplicar inteligencia de negocios en el área comercial de una empresa ficticia llamada "NetMax", proveedora de servicios de internet. En la cual se plantea la idea de que empresas deben adaptarse a los nuevos desafíos del mercado mediante la transformación digital, y que existen herramientas digitales que ofrecen tecnologías y metodologías para gestionar y analizar información.

El problema planteado es que las ventas de la empresa han disminuido en los últimos años, y se busca ayudar al área comercial mediante la implementación de herramientas de analítica de datos que les permitan interpretar información para tomar decisiones y plantear estrategias más precisas y acertadas, lo que podría mejorar las ventas y el posicionamiento de la empresa.

Se establecen objetivos generales y específicos para el proyecto, que incluyen identificar los principales indicadores comerciales relevantes para la empresa, crear un dashboard con Power BI para presentar la información de manera fácil y sencilla, y analizar la información representada en los gráficos del dashboard.

El alcance del proyecto se limita a analizar los datos del área comercial de la empresa y visualizar la información relevante para la toma de decisiones mediante la implementación de herramientas de BI en su versión sin costo. Se espera que la implementación de estas herramientas permita optimizar recursos y reducir costos, mejorando el rendimiento general de la empresa.

ÍNDICE

Tabla de contenido

ÍNDICE DE FIGURAS, GRÁFICOS Y TABLAS VII

ÍNDICE DE FIGURAS VII

CAPÍTULO I: INTRODUCCIÓN 1

1. MARCO DE REFERENCIA 1

1.1. Justificación 1

1.2. Planteamiento del problema 1

1.3. Objetivo General 2

1.4. Objetivos Específicos 2

1.5. Alcance 2

CAPÍTULO II: FUNDAMENTACIÓN TEÓRICA 3

2. Marco Teórico 3

2.1. Inteligencia de Negocios 3

2.2. Metodologías 3

2.2.1. Metodología Kimball 3

2.2.2. Scrum 4

2.2.3. Crisp-DM 4

2.3. Servicios de internet 5

2.4. Herramientas de BI 5

2.4.1. Microsoft Power BI 5
2.4.2. Tableu 6
2.5. KPI 6
2.6. Pyhton 6
2.7. Jupyter Notebook 7
2.8. Dashboard 7
2.9. DAX (Data Analysis Expressions) 8
CAPÍTULO III: METODOLOGÍA 9
3. Metodología del plan de tesis 9
3.1. Descripción de la empresa 9
3.2. Comprensión de los datos 10
3.3. Identificación de indicadores (KPIs) 12
3.4. Procesamiento de datos 13
3.5. Carga de datos 14
CAPÍTULO IV: DESARROLLO Y ANALISIS DEL DASHBOARD 15
4. Desarrollo y análisis del dashboard 15
4.1. Selección de la herramienta 15
4.2. Proceso ETL para Power BI 15
4.2.1. Extracción 15
4.2.2. Transformación 15
4.2.3. Carga 17

4.3. Construcción del Dashboard.....18
4.3.1. Diseño Visual.....18
4.3.2 Desarrollo técnico.....19
4.4. Análisis del dashboard.....21
CONCLUSIONES Y RECOMENDACIONES.....26
Conclusiones.....26
Recomendaciones.....27
BIBLIOGRAFÍA.....28
ANEXOS.....30
ANEXO 1: Códigos DAX.....30

ÍNDICE DE FIGURAS

Figura 1 *Verificación de datos en Power BI* 15

Figura 2 *Reemplazo de valores en Power Query*.....16

Figura 3 *Cambio de tipo de caracteres en Power Query*.....17

Figura 4 *Primera interfaz* 18

Figura 5 *Segunda Interfaz* 19

Figura 6 *Creación área de trabajo* 21

Figura 7 *Publicación del dashboard en Power BI*.....21

Figura 8 *Primer página del dashboard* 22

Figura 9 *Segunda página del dashboard* 23

Figura 10 *Tercer página del dashboard* 24

CAPÍTULO I: INTRODUCCIÓN

1. MARCO DE REFERENCIA

1.1. Justificación

En la actualidad las empresas deben tener una transformación digital para adaptarse a los nuevos desafíos y ser competitivos en el mercado, existen herramientas digitales que ofrecen un conjunto de tecnologías y metodologías que ayudan a las empresas a gestionar y analizar su información que permitirá identificar patrones, tendencias y oportunidades de negocio. Por este motivo la empresa quiere hacer un análisis de inteligencia de negocios en el área comercial analizando el comportamiento de sus clientes e identificando las necesidades y preferencias, que permitan mejorar sus ventas y posicionarse entre las mejores a nivel nacional como proveedores de servicios de internet.

1.2. Planteamiento del problema

La empresa ha detectado que sus ventas han empezado a descender en los últimos años es por esto por lo que, con las herramientas de analítica de datos, se busca ayudar a la empresa específicamente al área comercial a tener a su alcance información que pueda ser interpretada de manera fácil y sencilla con la cual puedan tomar decisiones y plantear estrategias que sean

mucho más precisas y acertadas que ayuden al crecimiento y al cumplimiento de metas establecidos dentro de la empresa.

Actualmente dentro de la empresa se tienen el pensamiento de que la implementación de nuevas herramientas puede llegar a ser muy costosas e innecesarias, pero con la implementación de esta propuesta se podrá demostrar que las herramientas de BI en su versión sin costo pueden ayudar a optimizar recursos, lo que puede reducir costos y mejorar el rendimiento general de la empresa.

1.3. Objetivo General

Analizar el área comercial de una empresa proveedora de servicios de internet aplicando inteligencia de negocios.

1.4. Objetivos Específicos

- Identificar los principales indicadores comerciales relevantes para la empresa "NetMax" proveedora de servicios de internet.
- Realizar un dashboard con Power BI que permita a la empresa tomar decisiones con la información presentada.
- Analizar la información representada en las gráficas del dashboard.

1.5. Alcance

Analizar los datos del área comercial en la empresa (ficticia) "Netmax" proveedora de servicios de internet, aplicando inteligencia de negocios se podrán identificar los principales indicadores para el área comercial y con el desarrollo del dashboard se podrá evaluar la efectividad de las estrategias de precios, también al implementar el uso de la herramienta de Power BI nos permitirá realizar un proceso ETL con los datos actuales que tiene el área comercial de la empresa y visualizar la información más relevante que el área comercial requiera para tomar decisiones.

CAPÍTULO II: FUNDAMENTACIÓN TEÓRICA

2. Marco Teórico

2.1. Inteligencia de Negocios

La inteligencia de negocios se refiere al proceso de recopilar, analizar y presentar información para ayudar a las organizaciones a tomar decisiones basadas en información confiable y accesible. En la actualidad, la inteligencia de negocios representa una ventaja competitiva ya que esta te permite tener una visión integral de la empresa y utilizarla para implementar estrategias, eliminar las ineficiencias y adaptarte rápidamente a los cambios del mercado.

Es importante tener en cuenta que esta es una definición actual de inteligencia de negocios ya que originalmente surgió en la década de 1960 como un sistema de intercambio de información entre organizaciones, pero se desarrolló aún más en la década de 1980 junto con modelos informáticos para la toma de decisiones y convertir los datos en información antes de transformarse en una oferta específica de parte de equipos de BI con soluciones de servicio dependientes de TI. Las soluciones de inteligencia de negocio modernas priorizan el análisis flexible, datos confiables y la velocidad para obtener información (Tableu, s.f)

2.2. Metodologías

2.2.1. Metodología Kimball

La Metodología Kimball se enfoca en las misiones empresariales y en construir modelos analíticos para alcanzar los objetivos estratégicos planteados. La correcta implementación de esta metodología es crítica para tomar decisiones bien fundamentadas basadas en información de calidad que contribuyan al éxito estratégico de la organización. Además, la Metodología de

Inteligencia de Negocios de Ralph Kimball brinda instrucciones claras y sencillas para la implementación de esta metodología, lo que permite verificar su facilidad de aplicación.

En la actualidad dentro de las empresas se establecen planes estratégicos que permitan cumplir con los objetivos de negocio y de esta forma posicionarse dentro del mercado, debido a esto surge la importancia de aplicar un modelo de Inteligencia de Negocios de Ralph Kimball ya que esta metodología propone directrices que permiten construir estos modelos y permiten a futuro ser incorporados en nuevas organizaciones. (Forero-Castañeda Deivy Alexander, Sánchez-Garcia Jorge Armando, 2021, p.6)

2.2.2. Scrum

La metodología Scrum puede aplicarse en la inteligencia de negocios para mejorar la eficiencia y la calidad del proceso de análisis y toma de decisiones, ya que Scrum es una metodología ágil que se centra en la entrega temprana y continua de valor a los usuarios. En el contexto de la inteligencia de negocios, esto significa que se pueden identificar rápidamente los requerimientos del usuario y generar prototipos de informes, cuadros de mando y visualizaciones para mejorar la toma de decisiones" (Nascimento et al., 2019, p. 114)

2.2.3. Crisp-DM

La metodología CRISP-DM es un proceso iterativo y estructurado que consta de seis fases, además se ha convertido en un estándar para proyectos de minería de datos, desde hace veinte años, ha sido una fuente de inspiración para otros estándares, como SEMMA de SAS o ASUM-DM de IBM, y ha dado lugar a variantes que se adaptan a diferentes industrias o tipos de proyectos.

Sin embargo, en los últimos años, los proyectos de ciencia de datos han experimentado cambios que plantean ciertas limitaciones a la metodología CRISP-DM. Como resultado, se están

considerando nuevas formas de aplicar la metodología, como añadir nuevas fases o aplicarla solo a ciertas partes de los proyectos dependiendo de lo que se requiera. (Haya, 2021)

2.3. Servicios de internet

Los servicios de internet son brindados por diferentes empresas, conocidas como proveedores, que ofrecen conexión a internet a los usuarios a través de redes. Son responsables de proporcionar el acceso a internet y permitir que los usuarios realicen diversas actividades en línea, como enviar correos electrónicos, hacer compras y realizar investigaciones, entre otras.

Los servicios de internet pueden ser de diferentes tipos, incluyendo comerciales, sin fines de lucro, privados o de propiedad comunitaria. En cualquier caso, los proveedores de servicios de internet funcionan como una puerta de entrada a internet y a todas las posibilidades que ofrece. (WISP, 2021)

2.4. Herramientas de BI

2.4.1. Microsoft Power BI

Es una herramienta de inteligencia empresarial y visualización de datos para particulares y empresas que ofrece a los usuarios una variedad de herramientas para agregar, analizar, visualizar y compartir datos de manera segura y confiable. Su interfaz de usuario es muy intuitiva y similar a la de Excel y su integración con otros productos de Microsoft la convierte en una herramienta de autoservicio muy útil que requiere poca capacitación inicial.

Power BI tiene una versión gratuita y una de pago, que ofrece características adicionales para los usuarios que lo necesiten. Esta herramienta cuenta con una versión de escritorio, versión en la nube e incluso con una aplicación móvil.

En general, Power BI es una herramienta eficaz para reunir y procesar datos de una variedad de fuentes, utilizando gráficos y visualizaciones atractivas y fáciles de procesar. Esto permite a los usuarios generar y compartir informes útiles de lo que está sucediendo en la

empresa. Además, se puede conectar a una gran variedad de fuentes de datos, desde hojas de cálculo de Excel básicas hasta bases de datos y aplicaciones en la nube. (Bello, E, 2022)

2.4.2. Tableu

Es una plataforma de visualización de datos que se creó en 2003 y que utiliza representaciones visuales para analizar y presentar información. Esta herramienta forma parte de la inteligencia de negocios, ya que ayuda a visualizar los datos estudiados y facilita la toma de decisiones, ya que unifica la información en la empresa.

Tableau está diseñado para las personas, independientemente de su trabajo o función en la empresa. Esta herramienta es accesible para todos, y su objetivo es poner el poder de los datos en manos de las personas para que puedan tomar decisiones informadas y mejorar la eficiencia en la empresa. (KEEPCODING, 2022)

2.5. KPI

También conocido en español como indicador clave de desempeño, se trata de un conjunto de medidas utilizadas en cualquier sector o área empresarial para resumir la información sobre la eficacia y productividad de las acciones realizadas en un negocio, con el objetivo de tomar decisiones y determinar qué estrategias han sido más efectivas en la consecución de objetivos específicos en un proceso o proyecto determinado. (Logicalis, 2019)

2.6. Pyhton

Python es un lenguaje de programación de nivel elevado que puede ser empleado para construir una amplia variedad de aplicaciones. Este lenguaje es empleado en muchas áreas gracias a su sencillez y a sus amplias posibilidades, ya que facilita trabajar con inteligencia artificial, big data, machine learning y data science, entre muchos otros campos. Este es un lenguaje interpretado lo que significa que no es preciso compilarlo para ejecutar aplicaciones escritas en este lenguaje. En su lugar, las aplicaciones se ejecutan directamente en la

computadora utilizando un programa llamado intérprete, lo que evita la necesidad de "traducirlo" al lenguaje de máquina. (Universidades Santander, 2021)

2.7. Jupyter Notebook

Es una herramienta de software que fue presentada en el año 2015 por el Proyecto Jupyter, una organización sin fines de lucro. Esta aplicación cliente-servidor tiene la capacidad de producir documentos web en formato JSON que siguen un esquema versionado y que contienen una secuencia ordenada de celdas de entrada y salida. En estas celdas, se puede incorporar diverso contenido como código, texto en formato Markdown, fórmulas matemáticas, ecuaciones, así como contenido multimedia, también conocido como Rich Media. La aplicación se ejecuta desde un cliente web en cualquier navegador estándar, siendo necesario previamente instalar y ejecutar el servidor Jupyter Notebook en el sistema. Los documentos creados en Jupyter pueden ser exportados a diversos formatos, como HTML, PDF, Markdown o Python, y también pueden ser compartidos con otros usuarios por correo electrónico, a través de Dropbox o GitHub, o mediante el visor integrado de Jupyter Notebook. (IONOS, 2019)

2.8. Dashboard

Es una herramienta de gestión de información que permite monitorear, analizar y presentar visualmente los indicadores clave de rendimiento, métricas y datos fundamentales que son esenciales para seguir el estado de una empresa.

Podemos entender el dashboard como un resumen que recopila datos de diversas fuentes en un solo lugar y los presenta de manera clara y fácil de entender, lo que permite que la información más importante sea destacada de forma evidente. La implementación de un dashboard ayuda a simplificar el análisis de la información que se tiene dentro de una empresa ya que, al sumergirnos en el mundo de los datos, podemos enfrentarnos a la posibilidad de perder la visión general debido a la sobrecarga de información, lo que puede impedir que apliquemos las métricas adecuadas. (D. Ortiz, 2022)

2.9. DAX (Data Analysis Expressions)

Es un lenguaje desarrollado por Microsoft que sirve para recopilar funciones, operadores y constantes que se pueden usar en una fórmula o expresión para calcular y devolver uno o varios valores, está diseñado para ser sencillo y fácil de aprender ya que es muy similar al nombre de las funciones de Excel, este lenguaje puede ser utilizado principalmente en Power Pivot, Power Query y Power BI. (Audox, 2018)

CAPÍTULO III: METODOLOGÍA

3. Metodología del plan de tesis

3.1. Descripción de la empresa

Empresa: NetMax

Misión: Proporcionar servicios de internet de alta calidad y acceso confiable a la red mundial de información, ofreciendo soluciones personalizadas y tecnología de vanguardia para satisfacer las necesidades de nuestros clientes.

Visión: Ser líderes en el mercado de servicios de internet, brindando a nuestros clientes una experiencia en línea excepcional con tecnología de última generación y un servicio al cliente excepcional, para ser reconocidos como el proveedor de internet preferido en nuestra área de operación.

Objetivos estratégicos

Mejorar la calidad del servicio: Establecer medidas y procesos para garantizar la alta calidad del servicio y una conexión estable y confiable para los clientes, lo que aumentará la satisfacción del cliente y la fidelidad.

Fomentar la innovación: Fomentar la innovación y el desarrollo de nuevos productos y servicios de internet que satisfagan las necesidades y expectativas de los clientes y que ayuden a la empresa a mantener su posición de liderazgo en el mercado.

Fortalecer la marca: Fortalecer la presencia de la marca y aumentar su reconocimiento en el mercado, lo que permitirá a la empresa mejorar su posición competitiva y atraer nuevos clientes.

3.2. Comprensión de los datos

Los datos fueron obtenidos de Kaggle de una empresa proveedora de internet norteamericana (“AT&T”) ya que estos cuentan con una licencia abierta, los cuales pueden son de libres de uso, distribución y modificación.

Este dataset consta de las siguientes columnas:

- address_full: La dirección postal completa de un hogar que buscamos.
- incorporated_place: La ciudad incorporada a la que pertenece la dirección.
- major_city: La ciudad en la que se encuentra la dirección.
- State: El estado en el que se encuentra la dirección.
- Lat: La latitud de la dirección. Desde OpenAddresses o NYC Open Data.
- Lon:La longitud de la dirección. Desde OpenAddresses o NYC Open Data.
- block_group: El grupo de bloques Census de la dirección, a partir de 2019. De la API Census Geocoder basada en laty lon.
- collection_datetime: La marca de tiempo de Unix en la que se usó la dirección para consultar el sitio web del proveedor.
- Provider: El proveedor de servicios de internet.
- speed_down: La velocidad de descarga anunciada más barata para la dirección.
- speed_up: La velocidad de carga anunciada más barata para la dirección.
- speed_unit: La unidad de velocidad. Esto siempre es en megabits por segundo (Mbps).
- Price: El costo en USD del plan de internet anunciado más barato para la dirección.
- Technology: El tipo de tecnología (fibra o no fibra) que se utiliza para ofrecer el plan de internet más económico.

- Package: El nombre del plan de internet más barato.
- fastest_speed_down: La velocidad de descarga anunciada del paquete más rápido. Este suele ser el mismo que el plan más barato si speed_downes inferior a 200 Mbps.
- fastest_speed_price: La velocidad de carga anunciada del paquete de internet más rápido para la dirección.
- Fn: El nombre del archivo de respuestas de la API desde donde se analizó este registro. Para ser utilizado para la resolución de problemas. Las respuestas de la API se alojan externamente en AWS s3.
- redlining_grade: El grado de línea roja, fusionado de Mapping Inequality basado en el laty londe la dirección.
- race_perc_non_white: El porcentaje de personas de color (que no son blancos no hispanos) en el grupo de bloque del censo del destinatario expresado como una proporción. Obtenido de la Encuesta sobre la comunidad estadounidense de 5 años de 2019.
- median_household_income: El ingreso familiar promedio en el grupo de bloques del censo de las direcciones. Obtenido de la Encuesta sobre la comunidad estadounidense de 5 años de 2019.
- income_lmi median_household_incomedividido: por el ingreso familiar promedio de la ciudad (obtenido de la Oficina del Censo de EE. UU.).
- income_dollars_below_median: El ingreso familiar promedio de la ciudad menos el median_household_income.
- ppl_per_sq_mile: Las personas por milla cuadrada se utilizan para determinar la densidad de población. Obtenido de archivos de formas TIGER de 2019 de la Oficina del Censo de EE. UU.

- n_providers: El número de otros competidores cableados en el grupo de bloques del Censo de las direcciones. Obtenido del Formulario 477 de la FCC.
- internet_perc_broadband: El porcentaje de la población que ya está suscrita a banda ancha en un grupo de bloques censales de direcciones expresado como una proporción.

3.3. Identificación de indicadores (KPIs)

Con los campos que ofrece este conjunto de información sobre la ubicación, la velocidad y el precio de la conexión a Internet, así como datos demográficos y socioeconómicos de las áreas de servicio. Algunos de los principales indicadores comerciales que se pueden identificar con estos campos son:

Geográficos: Las variables "major_city", "incorporated_place" y "state", pueden utilizarse para identificar las ciudades, estados que podrían estar más interesados en los servicios de internet en una determinada área. Esto puede ayudar a los proveedores a orientar sus campañas publicitarias y a desarrollar paquetes de servicios que se adapten mejor a las necesidades de los clientes.

Velocidad de descarga y carga: Las variables "speed_down" y "speed_up" pueden utilizarse para mostrar la velocidad promedio de descarga y carga ofrecida en una determinada área. Se pueden comparar estas velocidades con las otras ciudades para evaluar la competitividad.

Precio: La variable "price" proporciona información sobre el precio promedio de los paquetes de internet ofrecidos en una determinada área. Esto puede ayudar a evaluar su posición en el mercado y a identificar oportunidades de crecimiento.

Tecnología: La variable "technology" indica el tipo de tecnología utilizada para proporcionar el servicio de internet en una determinada área. Esto puede ayudar a los proveedores a identificar las tecnologías más populares y a planificar su infraestructura futura.

Competencia: La variable "n_providers" indica el número de proveedores de servicios de internet que operan en una determinada área. Esto puede ayudar a los proveedores a identificar la competencia y a desarrollar estrategias para diferenciarse de otros proveedores.

3.4. Procesamiento de datos

Esta es una parte muy importante ya que es donde se mejorará la calidad de los datos al limpiarlos, normalizarlos y estandarizarlos, lo cual ayudará a reducir errores y aumentar la precisión de los análisis y decisiones basadas en los datos.

Para hacer el procesamiento de los datos se utilizará Jupyter Notebook y Python 3.

Primero se importan las librerías necesarias que permitirá cargar y analizar los datos con lo que se va a trabajar.

Después se carga el conjunto de datos en un objeto de dataframe llamado "df", lo que permite trabajar con los datos del archivo.

Además, se utiliza un método de Pandas que se utiliza para mostrar los tipos de datos de las columnas del dataframe, lo que permite verificar la consistencia y de ser necesario la adecuación de los tipos de datos.

También se procede a contar el número de valores faltantes en cada columna del dataframe, lo que permite a los usuarios identificar qué columnas tienen valores faltantes y tratarlos adecuadamente antes de realizar un análisis o modelado de los datos.

Luego se imputan los valores faltantes en la variable “speed _unit”, para ello se utiliza el método “mode” para encontrar el valor más común en la columna e insertarlo en los valores faltantes.

Asimismo, se elimina la columna “redlinig_grade” ya que la mayoría de los datos que contiene son valores faltantes y no será utilizada para el análisis.

Por último, se procede a descargar el dataframe “datos_limpios.csv” con los cambios realizados en los anteriores gráficos. Ver Anexo N°1.

3.5. Carga de datos

Con los datos extraídos y procesados proceden a ser cargados para la elaboración del dashboard en la herramienta Power Bi, para continuar con el análisis del área comercial. Ver Anexo N°2.

CAPÍTULO IV: DESARROLLO Y ANALISIS DEL DASHBOARD

4. Desarrollo y análisis del dashboard

4.1. Selección de la herramienta

Para el desarrollo del dashboard se va a utilizar la herramienta de Power BI en su versión gratuita porque tiene una amplia gama de herramientas de visualización de datos, lo que permite a los usuarios crear visualizaciones atractivas y fáciles de entender. Esto es importante en un dashboard, ya que los usuarios necesitan ver rápidamente las tendencias y patrones en los datos. También permite a los usuarios interactuar con los datos en el dashboard, lo que les permite profundizar en los detalles y descubrir nuevas tendencias y patrones.

En resumen, Power BI es una excelente opción para crear un dashboard debido a su capacidad de integrar, visualizar y analizar datos de diversas fuentes, así como su capacidad para permitir la interacción, personalización y colaboración

4.2. Proceso ETL para Power BI

Es importante realizar un proceso ETL, ya que a pesar de haber realizado el procesamiento de datos con Python se debe verificar que los datos se carguen de manera correcta en la herramienta de Power BI ya que al momento que se hace la transformación esta puede llegar a no reconocer algunos caracteres o no identificar el tipo de carácter de cada campo, lo cual puede producir errores al realizar el dashboard.

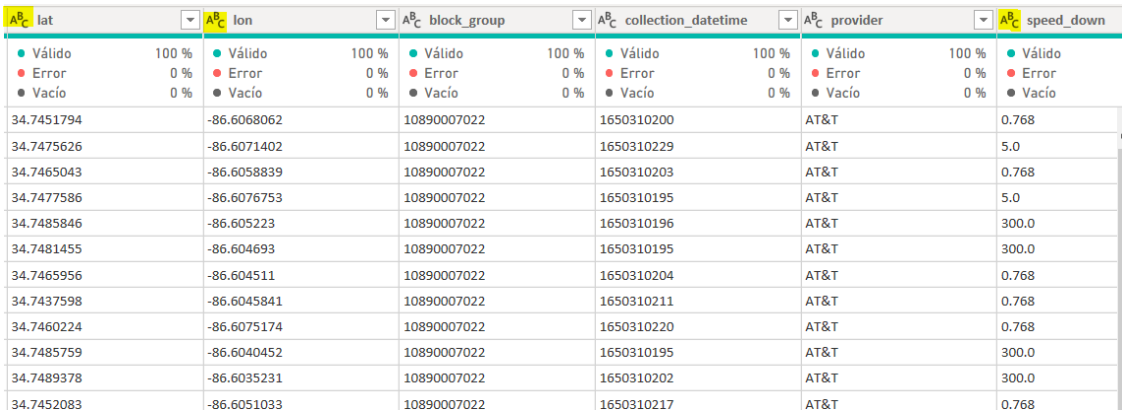
4.2.1. Extracción

En esta fase, se utilizan los datos de la ilustración 8 de nuestro archivo csv procesado en Python. El objetivo es recopilar los datos necesarios para su posterior revisión en Power Query.

4.2.2. Transformación

Figura 1

Verificación de datos en Power BI

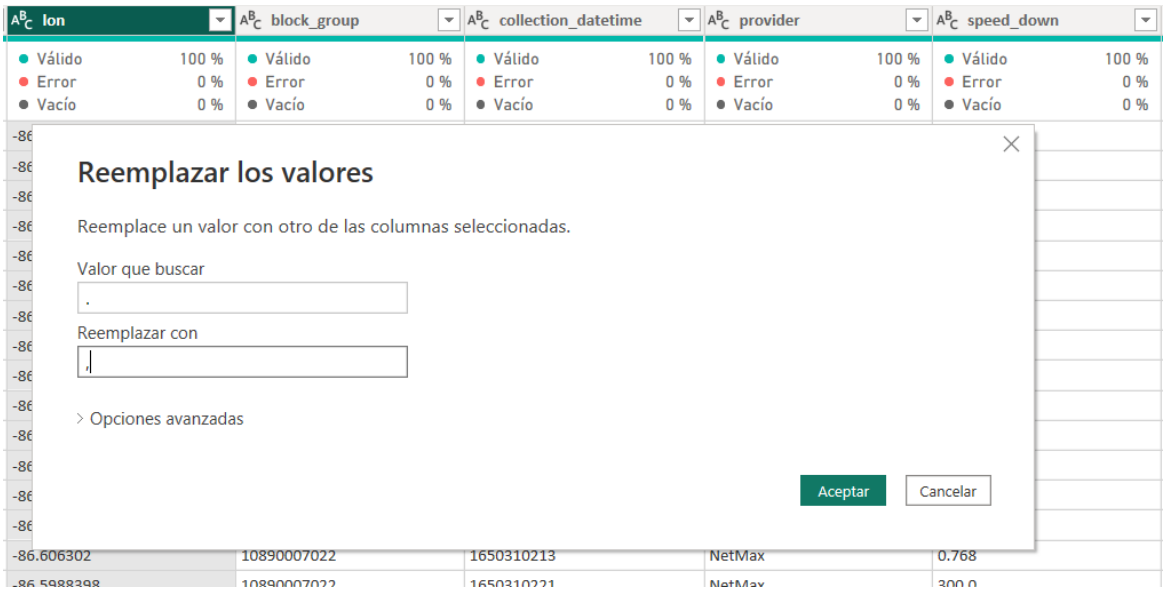


lat	lon	block_group	collection_datetime	provider	speed_down
Válido 100 %	Válido 100 %	Válido 100 %	Válido 100 %	Válido 100 %	Válido
Error 0 %	Error 0 %	Error 0 %	Error 0 %	Error 0 %	Error
Vacío 0 %	Vacío 0 %	Vacío 0 %	Vacío 0 %	Vacío 0 %	Vacío
34.7451794	-86.6068062	10890007022	1650310200	AT&T	0.768
34.7475626	-86.6071402	10890007022	1650310229	AT&T	5.0
34.7465043	-86.6058839	10890007022	1650310203	AT&T	0.768
34.7477586	-86.6076753	10890007022	1650310195	AT&T	5.0
34.7485846	-86.605223	10890007022	1650310196	AT&T	300.0
34.7481455	-86.604693	10890007022	1650310195	AT&T	300.0
34.7465956	-86.604511	10890007022	1650310204	AT&T	0.768
34.7437598	-86.6045841	10890007022	1650310211	AT&T	0.768
34.7460224	-86.6075174	10890007022	1650310220	AT&T	0.768
34.7485759	-86.6040452	10890007022	1650310195	AT&T	300.0
34.7489378	-86.6035231	10890007022	1650310202	AT&T	300.0
34.7452083	-86.6051033	10890007022	1650310217	AT&T	0.768

Nota. En Power Query se observa que varios campos del dataset no son reconocidos con el tipo de carácter correcto, ya que los campos lat, lon, speed_down, entre otros son campos de carácter numérico o decimal, que deben ser corregidos para el análisis.

Figura 2

Reemplazo de valores en Power Query

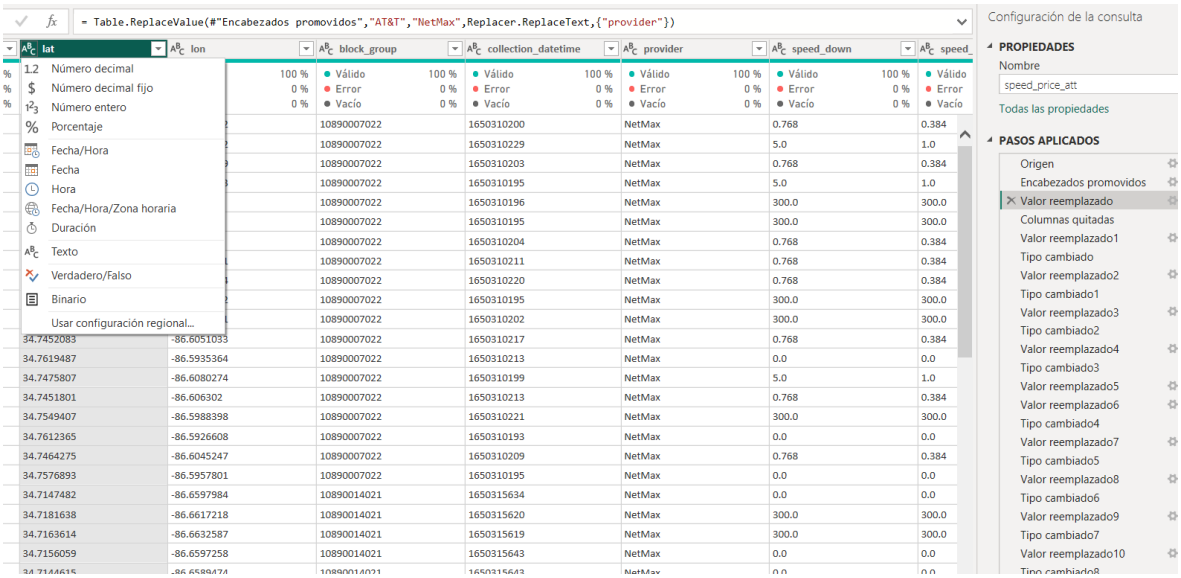


Nota. Se identifica que los campos numéricos no son reconocidos por Power Query debido a que toma en cuenta la configuración regional de la computadora, en la cual el separador de decimales es la coma y en los datos del dataset el separador de decimales es el punto. Es

por esto por lo que se reemplaza puntos por comas en los campos numéricos con decimales, también se puede corregir este error cambiando la configuración regional en Power BI

Figura 3

Cambio de tipo de caracteres en Power Query



Nota. Como se observa en la ilustración Power Query tiene la función de cambiar al carácter que se requiera, en este caso a tipo numérico con decimal, también en la ilustración se observa a la derecha todos los pasos aplicados en cada uno de los campos para tener una información óptima para el desarrollo del dashboard.

4.2.3. Carga

Una vez que se verifica que los datos transformados están correctos, se procede a crear visualizaciones y desarrollar el Dashboard. Power BI permite almacenar los datos en su modelo de datos interno, que incluye tablas, relaciones y medidas.

4.3. Construcción del Dashboard

Para el desarrollo del dashboard se toma en cuenta los indicadores (KPIs) que se identificó en el capítulo tres ya que los gráficos que se generen en Power BI deben satisfacer las necesidades del área comercial de la empresa

4.3.1. Diseño Visual

Para la presentación visual de la información de la empresa se decidió realizar cuatro páginas, en las cuales se plantea tener dos interfaces diferentes, la primera en la que se presente una pantalla de inicio donde se describa brevemente el área comercial y la segunda interfaz que se utilizara para las tres páginas donde se realizaran todos los gráficos para analizar el área.

Figura 4

Primera interfaz

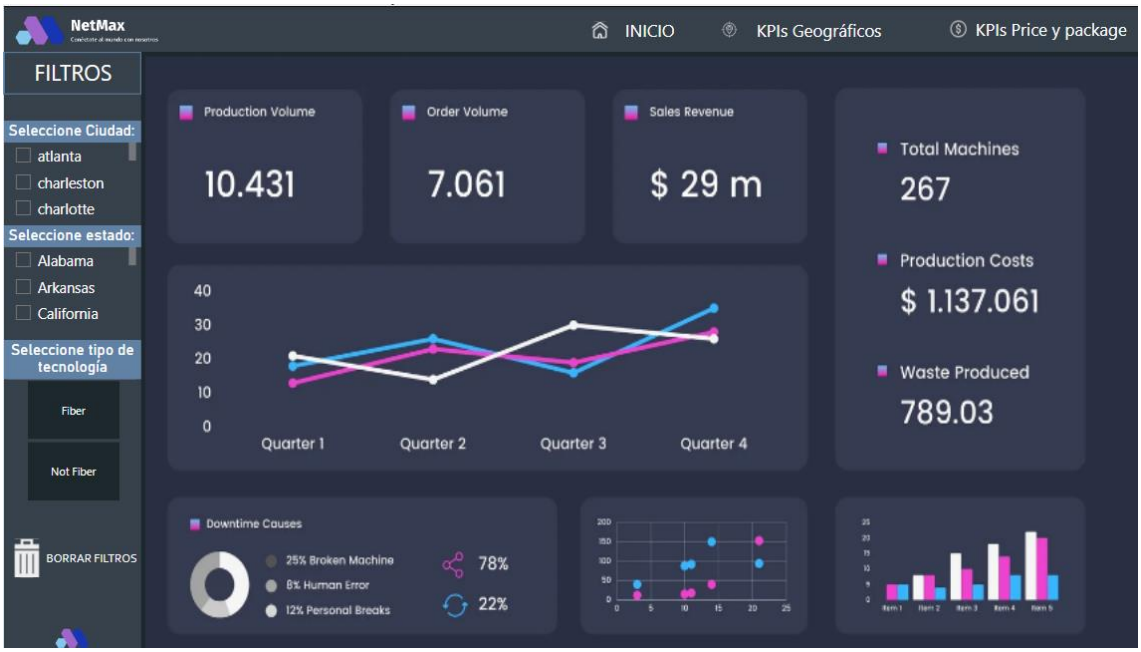


Nota. La pantalla de inicio del dashboard es la primera interfaz que los usuarios ven cuando acceden al reporte. Su propósito es proporcionar los objetivos clave del área comercial de la empresa. La interfaz se divide en diferentes secciones, a la izquierda están los botones los

cuales dirigen a las páginas del dashboard cada uno de los cuales muestra datos e indicadores específicos.

Figura 5

Segunda Interfaz



Nota. Esta es el segundo prototipo de interfaz que será utilizada para las tres páginas del dashboard, la cual cuenta con los filtros en la parte izquierda, también cuenta con una barra superior donde se encuentran los botones que permitirán navegar entre las páginas y por último en la parte central están los gráficos que se implementar según se requiera en función de los KPIs ya establecidos

4.3.2 Desarrollo técnico

Una vez que se ha diseñado la interfaz visual, se procede a la implementación técnica del dashboard. En las cuales se implica el uso de DAX (Data Analysis Expressions) para la creación de dashboard.

A continuación, se detalla las funciones DAX utilizadas para el desarrollo técnico:

A. Rankx: Esta medida utiliza la función RANKX para calcular el rango de una variable llamada "Recuento clientes" en la columna "incorporated_place" de la tabla "datos_limpios". El parámetro "DESC" indica que se ordene en orden descendente y "Skip" indica que se salten empates en el rango. Ver Anexo N°3
B. Recuento clientes: Esta medida utiliza la función COUNTA para contar el número de valores no vacíos en la columna "provider". Calcula el recuento de clientes basado en la cantidad de proveedores presentes en los datos. Ver Anexo N°3
C. Recuento tech: Esta medida utiliza la función COUNTA para contar el número de valores no vacíos en la columna "technology". Calcula el recuento de tecnologías basado en la cantidad de tecnologías presentes en los datos. Ver Anexo N°3
D. Suma price: Esta medida utiliza la función SUM para sumar los valores de la columna "price". Calcula la suma total de los precios. Ver Anexo N°3
E. Tech no fiber: Esta medida utiliza la función COUNTROWS junto con la función FILTER para contar el número de filas en la tabla "datos_limpios" donde la columna "technology" es igual a "Not Fiber". Calcula el recuento de las filas que tienen tecnología "No Fiber". Ver Anexo N°3
F. Tech fiber: Esta medida utiliza la función COUNTROWS junto con la función FILTER para contar el número de filas en la tabla "datos_limpios" donde la columna "technology" es igual a "Fiber". Calcula el recuento de las filas que tienen tecnología "Fiber". Ver Anexo N°3
G. Parámetro selección: Esta expresión define una lista de opciones para un parámetro. Cada opción tiene una etiqueta y un valor. En este caso, las opciones son "$" y "# Clientes", y cada una está vinculada a una medida específica ("suma price" y "Recuento clientes", respectivamente) de la tabla "datos_limpios". Esto se puede utilizar para seleccionar qué medida se mostrará en un informe o visualización, dependiendo de la opción seleccionada

por el usuario. El número después del nombre de la medida indica el índice de la opción en la lista. Ver Anexo N°3.

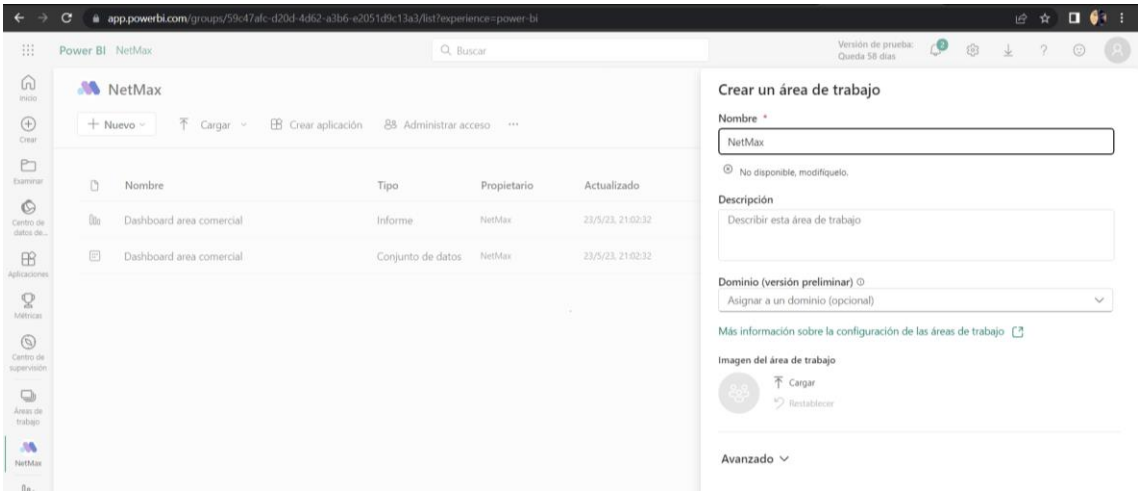
4.4. Publicación del dashboard

La herramienta de Power BI cuenta con su versión en la nube en la cual se puede publicar subir los reportes o dashboard para compartir la información y visualizaciones interactivas con otros usuarios de la empresa.

Para publicar se necesita abrir la aplicación de Power BI en cualquier navegador con el correo y contraseña.

Figura 6

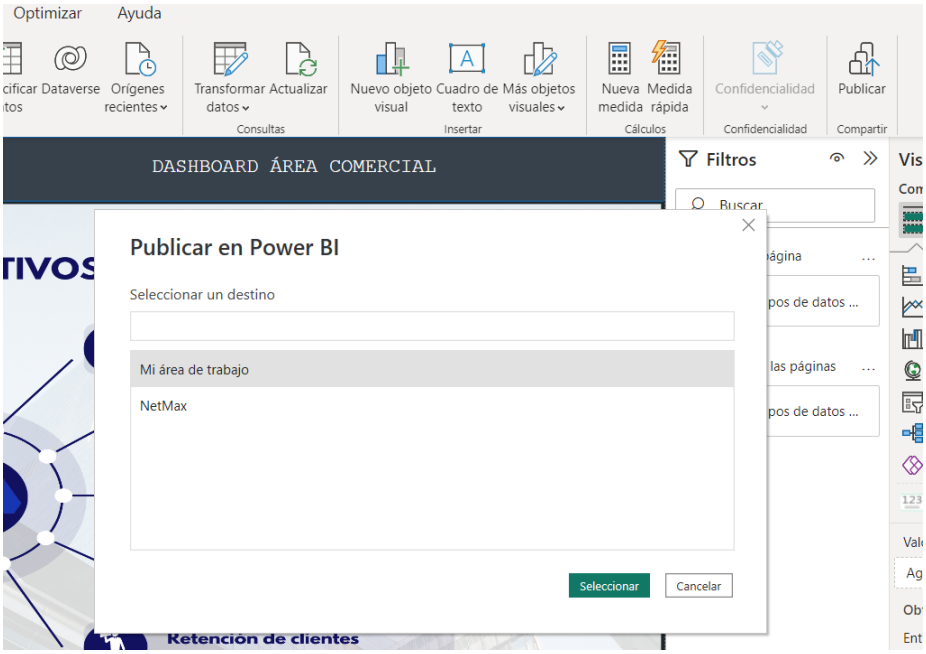
Creación área de trabajo



Nota. Se crea el área de trabajo donde se publicará el dashboard creado anteriormente para la facilidad de su acceso.

Figura 7

Publicación del dashboard en Power BI

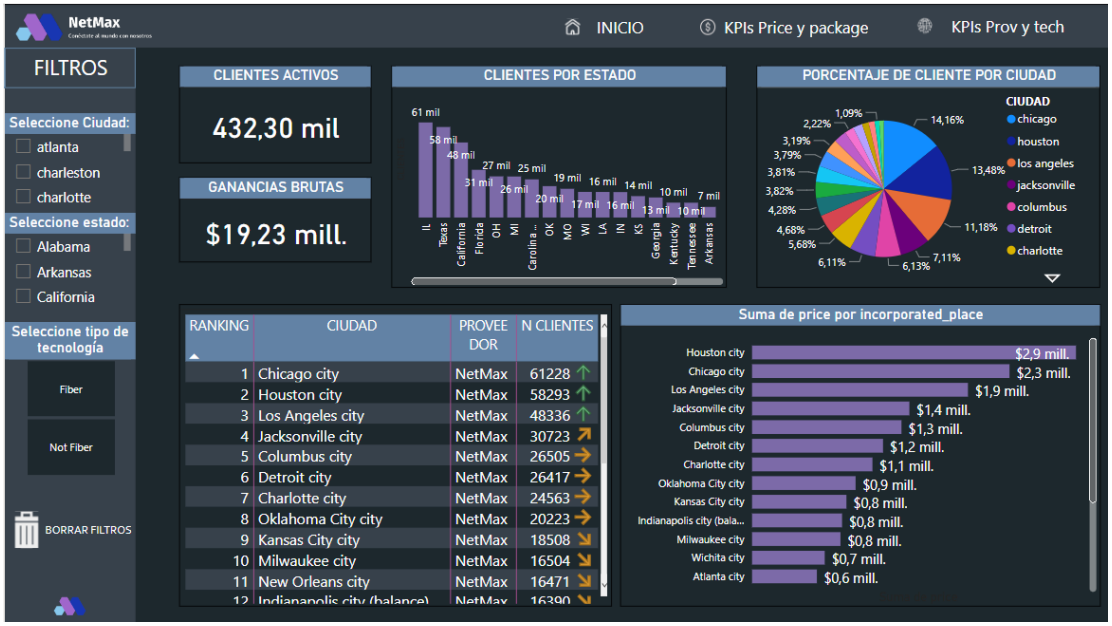


Nota. Se abre el dashboard desarrollado en la aplicación desktop de Power BI y se selecciona el área de trabajo creada para publicar el dashboard. Ver Anexo N 4.

4.5. Análisis del dashboard

Figura 8

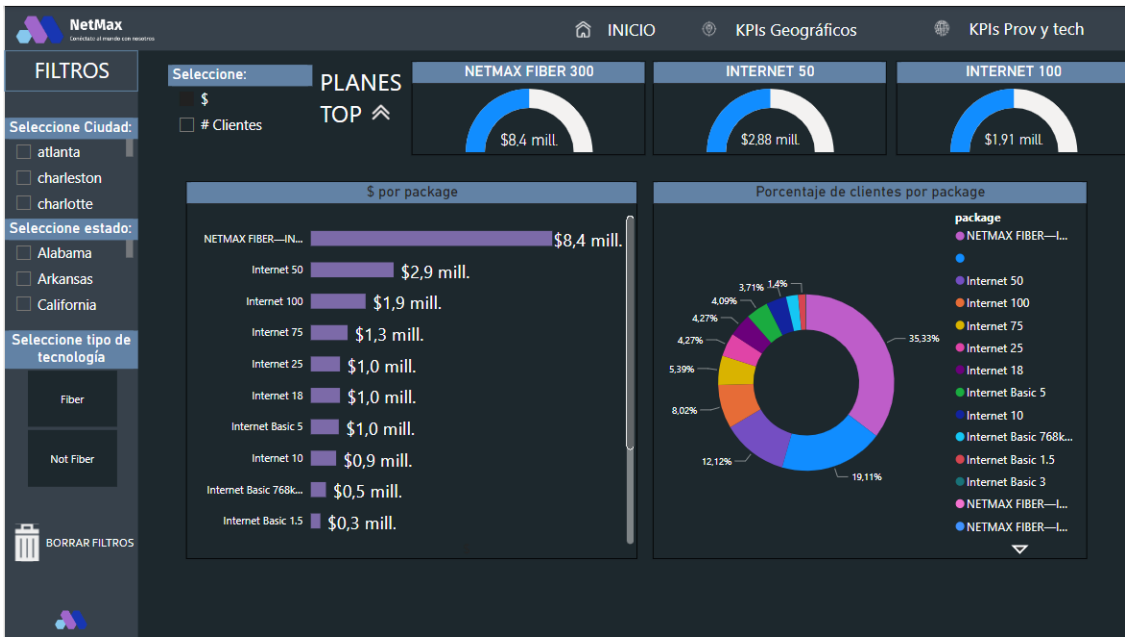
Primer página del dashboard



Nota. Está primera página observamos dos gráficos de barras, una tabla y un gráfico de pastel que demuestra en que zonas geográficas donde la empresa tiene mayor acogida y mayores ganancias, las ciudades que tienen mayor cantidad de clientes y ganancia son Houston, Chicago y Los Ángeles, también se identifica las ciudades donde se tiene menor cantidad de clientes son Jackson, Charleston y Huntsville, en estas ciudades se deberían aplicar estrategias para captar nuevos clientes, además en el dashboard se observa dos KPIs en la cual podemos observar los clientes activos y las ganancias brutas de la empresa, en las cuales se deberían establecer metas para incentivar al área comercial a mejorar la venta de planes y ganancias. Además, los filtros nos permiten hacer comparación entre varias ciudades.

Figura 9

Segunda página del dashboard



Nota. En la segunda página del dashboard se observa los paquetes que ofrece la empresa, en la parte superior se encuentran tres medidores con la información de los mejores planes con respecto a cuánto ingresos genera cada plan y cuantos clientes poseen cada plan, los cuales son “NetMax Fiber 300”, “Internet 50” y “Internet 100”, también se situaron un gráfico

de barras y un gráfico de anillo en los cuales se observan todos los planes que dispone la empresa y se identifica que existen varios que no generan buenos ingresos a la empresa como son los planes “NetMax Fiber Internet 500”, “NetMax Fiber Internet 100”, “NetMax Fiber Internet 2000”, para los que se deberían realizar una investigación exhaustiva sobre los posibles clientes para comprender sus necesidades, preferencias, comportamientos de compra y ajustar los planes para que mejore su nivel de aceptación.

Figura 10

Tercer página del dashboard



Nota. En la última página del dashboard se observa dos KPIs con información del promedio de velocidad de descarga y de subida de los planes ofertados por la empresa los cuales son muy buenos porque se observa que en promedio están por los 110 Mbps de subida y 126 Mbps de descarga, también se presentas dos gráficos de barras en la primera de ellas se hace una comparación del promedio de velocidad de descarga y de subida de los planes en cada estado lo que ayuda a descubrir en que ciudades hay un bajo promedio de velocidad de descarga y de subida, para implementar mejor infraestructura y brindar un mejor servicio a los clientes, en

el segundo gráfico de barras hace un comparación de la tecnología instalada en cada estado, la cual ayuda a saber las preferencias de los clientes.

Además, hay un mapa en el dashboard donde se observa varias burbujas en las cuales se pueden identificar donde hay mayor cantidad de competidores y también en que estados se puede aprovechar para abrir nuevas sedes y expandir la oferta de los servicios prestados por la empresa. Y por último hay un gráfico de anillo en el cual se identifica que la mayoría de los clientes no tiene fibra, se debería realizar mejoras de infraestructura e instalar fibra ya que es mucho más rápida y puede ayudar a captar más clientes.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- La compresión de los datos obtenidos proporciono información valiosa para identificar los KPIs de la empresa “NetMax” que tiene como objetivo mejorar la calidad del servicio, fomentar la innovación y fortalecer su marca. Los KPIs identificados de precio, geografía, velocidad de descarga y carga, tecnología y competencia son fundamentales para desarrollar dashboard coherentes ya que con ellos se identifico que la velocidad de descarga y carga anunciada más barata para una dirección no siempre coincide con la velocidad más rápida disponible. Esta información es importante para ofrecer a los clientes opciones adecuadas y transparentes en términos de velocidad y precio. También que la tecnología utilizada para ofrecer los planes de internet más económicos puede variar entre fibra y no fibra. Esto indica que la empresa está aprovechando diferentes tecnologías para adaptarse a las necesidades de sus clientes y brindarles opciones de conexión confiables y eficientes. Además, el análisis de precios reveló que los planes de internet más económicos varían en costo, lo que puede influir en la decisión de los clientes al elegir un proveedor de servicios. Es importante para NetMax mantener una oferta competitiva en términos de precio para atraer y retener a los clientes. Y por último la disponibilidad de datos precisos sobre las direcciones, las velocidades de internet y los precios permite a la empresa evaluar la calidad del servicio en diferentes áreas geográficas y tomar medidas para mejorar donde sea necesario.
- La implementación de herramientas de inteligencia de negocios en este caso el desarrollo del dashboard en Power BI ayudo a recopilar y analizar grandes volúmenes de datos de manera rápida y eficiente, lo que permitirá a la organización tomar decisiones basadas

en información actualizada y precisa. Esto puede conducir a una toma de decisiones más informada y fundamentada, lo que mejora las operaciones y los resultados comerciales.

- El análisis de los gráficos del dashboard ofreció una forma de comunicar datos complejos de manera clara y concisa. Permiten visualizar patrones, tendencias y relaciones entre variables, lo que facilita la comprensión de la información por parte del área comercial.

Recomendaciones

- Para desarrollar un dashboard se debe tomar en cuenta varios factores ya que este debe ser claro y fácil de entender. Evita el exceso de información que puede confundir a los usuarios y utiliza un diseño limpio y ordenado. Utiliza gráficos y visualizaciones adecuadas para comunicar la información de manera concisa.
- Mantenerse al tanto de las nuevas tecnologías y tendencias del mercado: La industria de servicios de internet está en constante evolución, por lo que es crucial que NetMax esté al tanto de las nuevas tecnologías y tendencias emergentes. Esto puede implicar explorar opciones como la conectividad 5G, la expansión de la banda ancha y la implementación de soluciones innovadoras para satisfacer las necesidades futuras de los clientes.
- Mejorar la centralización de la información, ya que se puede plantear hacer un data warehouse que proporcione una visión más completa y estructurada de los datos, mejorando la calidad y consistencia de estos, ofrece un rendimiento optimizado para consultas y permitiendo realizar un análisis multidimensional respaldando la toma de decisiones estratégicas. Estos beneficios hacen que un data warehouse sea una solución poderosa para gestionar y aprovechar los datos de manera efectiva de la empresa.

BIBLIOGRAFÍA

¿Qué es el Business Intelligence? Tu guía para la inteligencia de negocios y por qué es importante. (s. f.). Tableau. https://www.tableau.com/es-mx/learn/articles/business-intelligence

Forero-Castañeda Deivy Alexander, Sánchez-Garcia Jorge Armando (2021). Introducción a la inteligencia de negocios basada en la metodología Kimball. Revista Tecnol.Investig. Academia TIA, ISSN: 23448288, 9 (1), pp. 5-17. Bogotá-Colombia

Eisenmann, L., Fernandes, L. G., & Silva, A. B. (2018). A Method for the Adoption of Agile Methodologies in Business Intelligence Projects. In Advances in Intelligent Systems and Computing (pp. 95-105). Springer.

¿Qué son los servicios de internet? (2021, noviembre 19). WISP - Wireless Internet Service Provider. https://wisp.com.mx/servicio-de-internet/que-son-los-servicios-de-internet/

Bello, E. (2022). ¿Qué es Microsoft Power BI? Todo lo que tienes que saber. *Thinking for Innovation*. https://www.iebschool.com/blog/microsoft-power-bi-analitica-usabilidad/

¿Qué es Tableau y cómo funciona? (2022, marzo 16). *KeepCoding Tech School*. https://keepcoding.io/blog/que-es-tableau/

Ortiz, D., & Cyberclick. (2022, septiembre 27). *¿Qué es un dashboard y para qué se usa? (2023)*. Cyberclick.es. https://www.cyberclick.es/numerical-blog/que-es-un-dashboard

Logicalis. (2019, 6 de noviembre). KPIs: qué son, para qué sirven y por qué y cómo utilizarlos. Blog de Logicalis. Recuperado el 4 de abril de 2023, de https://blog.es.logicalis.com/analytics/kpis-qu%C3%A9-son-para-qu%C3%A9-sirven-y-por-qu%C3%A9-y-c%C3%B3mo-utilizarlos

Guía fundamental del Análisis DAFO. (2015, noviembre 25). Infoautonomos. https://www.infoautonomos.com/plan-de-negocio/analisis-dafo/

Haya, P. (2021, noviembre 29). La metodología CRISP-DM en ciencia de datos - IIC. Instituto de Ingeniería del Conocimiento. https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/

Python: qué es y por qué deberías aprender a utilizarlo. (s/f). Becas-santander.com. Recuperado el 4 de mayo de 2023, de https://www.becas-santander.com/es/blog/python-que-es.html

Jupyter Notebook: documentos web para análisis de datos, código en vivo y mucho más. (s/f). IONOS Digital Guide. Recuperado el 4 de mayo de 2023, de https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/jupyter-notebook/

¿Qué es DAX? (2018, enero 30). DAX - Capacitación, Soporte y Asesoría. https://www.audox.cl/dax/microsoft-dax/

ANEXOS

ANEXO 1: Procesamiento python

```
import pandas as pd

import numpy as np

df = pd.read_csv("speed_price_att.csv")

df.head(5)
```

	address full	incorporated place	major city	state	lat	lon	block group	collection c
0	2406 Country Club Ave NW, Huntsville AL 35816	Huntsville city	huntsville	AL	34.745179	-86.606806	10890007022	165
1	1902 Oglesby Dr NW, Huntsville AL 35816	Huntsville city	huntsville	AL	34.747563	-86.607140	10890007022	165
2	2312 Cardinal Ave NW, Huntsville AL 35816	Huntsville city	huntsville	AL	34.746504	-86.605884	10890007022	165
3	1903 Oglesby Dr NW, Huntsville AL 35816	Huntsville city	huntsville	AL	34.747759	-86.607675	10890007022	165
4	1905 Canterbury CirNW, Huntsville AL 35816	Huntsville city	huntsville	AL	34.748585	-86.605223	10890007022	165

5 rows x 26 columns

df.dtypes

```
address_full                  object
incorporated_place            object
major_city                    object
state                         object
lat                          float64
lon                          float64
block_group                    int64
collection_datetime            int64
provider                      object
speed_down                   float64
speed_up                     float64
speed_unit                    object
price                        float64
technology                    object
package                       object
fastest_speed_down           float64
fastest_speed_price          float64
fn                            object
redlining_grade               object
race_perc_non_white          float64
income_lmi                   float64
ppl_per_sq_mile              float64
n_providers                  float64
income_dollars_below_median  float64
internet_perc_broadband      float64
median_household_income        int64
dtype: object
```

df.isnull().sum()

```
address_full                      0
incorporated_place                0
major_city                        0
state                             0
lat                               0
lon                               0
block_group                       0
collection_datetime               0
provider                          0
speed_down                        0
speed_up                          0
speed_unit                    82600
price                         82600
technology                    82600
package                       82600
fastest_speed_down                0
fastest_speed_price               0
fn                                0
redlining_grade              246027
race_perc_non_white               0
income_lmi                    17661
ppl_per_sq_mile                7965
n_providers                    7969
income_dollars_below_median   17661
internet_perc_broadband          849
median_household_income            0
dtype: int64
```

most_common_speed_unit = df['speed_unit'].mode()[0]

df['speed_unit'] = df['speed_unit'].fillna(most_common_speed_unit)

num_nulls = df["speed_unit"].isnull().sum()

print("El número de valores nulos en la variable 'speed_unit' es:", num_nulls)

El número de valores nulos en la variable 'speed_unit' es: 0

df = df.drop(['redlining_grade'], axis=1)

df.columns

```
Index(['address_full', 'incorporated_place', 'major_city', 'state', 'lat',
       'lon', 'block_group', 'collection datetime', 'provider', 'speed down',
       'speed_up', 'speed_unit', 'price', 'technology', 'package',
       'fastest_speed_down', 'fastest_speed_price', 'fn',
       'race_perc_non_white', 'income_lmi', 'ppl_per_sq_mile', 'n_providers',
       'income_dollars_below_median', 'internet_perc_broadband',
       'median_household_income'],
      dtype='object')
```

ANEXO 2: Descarga de datos procesados

df.to_csv('datos_limpios.csv', index=False)

df

	address full	incorporated place	major city	state	lat	lon	block group	colk
626J	2600 S Taylor St, Little Rock AR 72204	Little Rock city	little rock	AR	34.728118	−92.336431	51 190019003	
6262	2413 S Fillmore St, Little Rock AR 72204	Little Rock city	little rock	AR	34.729582	−92.336928	51 190019003	
6263	2104 S Taylor St, Little Rock AR 72204	Little Rock city	little rock	AR	34.732527	−92.336248	51 190019003	
6264	2217 S Fillmore St, Little Rock AR 72204	Little Rock city	little rock	AR	34.731330	−92.336849	51 190019003	
6266	2212 Fair Park Blvd, Little Rock AR 72204	Little Rock city	little rock	AR	34.731401	−92.335189	51 190019003	
432292	155 S 76th St, Milwaukee WI 53214	Milwaukee city	milwaukee	WI	43.030372	−88.007733	550790J 27001	
432293	231 S 79th St, Milwaukee WI 53214	Milwaukee city	milwaukee	WI	43.028795	−88.011388	550790J 27001	
432296	206 N 79th St, Milwaukee WI 53213	Milwaukee city	milwaukee	WI	43.033209	−88.010972	550790J 27001	

ANEXO 3: Códigos DAX

A.
```
Rankx = RANKX(ALLSELECTED(datos_limpios[incorporated_place]),[Recuento clientes],,DESC,Skip)
```
B.
```
Recuento clientes = COUNTA([provider])
```
C.
```
Recuento tech = COUNTA([technology])
```
D.
```
suma price = SUM([price])
```
E.
```
tech  no fiber = COUNTROWS(FILTER(datos_limpios,(datos_limpios[technology] = "Not Fiber")))
```
F.
```
tech fiber = COUNTROWS(FILTER(datos_limpios,(datos_limpios[technology] = "Fiber")))
```
G.
```
Parámetro seleccion = {
                ("$", NAMEOF('datos_limpios'[suma price]), 0),
                ("# Clientes", NAMEOF('datos_limpios'[Recuento clientes]), 1)
```

ANEXO 4: Dashboard

https://app.powerbi.com/links/HiA3M34wmp?ctid=54474fb3-6359-40c1-b726-5d56a7dd2976&pbi_source=linkShare&bookmarkGuid=1932ec2c-ff4b-465d-b7de-3338588861ec