

Bootcamp de Análisis y Visualización de datos nivel intermedio

Informe del proyecto

**Análisis de desempeño operativo de servicios de Cajeros
Automáticos**

Equipo 12

Integrantes:

Fredy Abril

Bladimir Escobar

2024

Contenido

Comprensión del negocio	3
Objetivo general.....	3
Metodología.....	3
Objetivos Específicos.....	3
Calendario de actividades	4
Comprensión de los datos	5
Descripción de datos	5
Análisis de valores faltantes	7
Análisis de Valores Atípicos	7
Análisis de la Columna TIEMPO_SOLUCIÓN_LVL_6	8
Preparación de datos	10
Cambios en los valores negativos en la columna TIEMPO_RESOLV	10
Eliminación de la Columna TIEMPO_SOLUCIÓN_LVL_6.....	10
Manejo de datos faltantes	11
Creación de la columna ANHO para año de solicitudes	11
Manejo de Valores Nulos en la Columna NOMBRE_PROVEEDOR	13
Corrección de la Marca de Cajero Automático "BELLTECH"	14
Modelado	15
Resumen operativo de cajeros automáticos	15
Seguimiento de tickets.....	16
Análisis de solicitudes de primera y segunda línea.....	17
Evaluación y despliegue	19
conclusión.....	19
Recomendaciones.....	19

Comprensión del negocio

Objetivo general

El objetivo del proyecto es analizar los datos relacionados con las solicitudes y tiempos de respuesta de proveedores para identificar patrones, detectar anomalías y mejorar los procesos operativos. En particular, se busca comprender y mejorar la eficiencia de los tiempos de solución a diferentes niveles del servicio.

Metodología

Para llevar a cabo este proyecto, se ha elegido la metodología CRISP-DM que consta de las siguientes fases:

FASES

COMPRENSIÓN DEL NEGOCIO	Entender los objetivos y requisitos del proyecto
COMPRENSIÓN DE LOS DATOS	Recopilar, explorar y familiarizarse con los datos disponibles.
PREPARACIÓN DE LOS DATOS	Limpieza de datos y hacer las correcciones para llevar los datos más limpios para el modelado
MODELADO	Creación de visualizaciones en power BI
EVALUACIÓN	Evaluar el modelo para asegurar que cumple con los objetivos de negocio
DESPLIEGUE	Documentación de las conclusiones y recomendaciones para futuras iteraciones del proyecto, y realizar la presentación del proyecto

Objetivos Específicos

Evaluar la Calidad de los datos:

- Identificar y cuantificar valores faltantes y atípicos en las columnas clave.
- Determinar la precisión y consistencia de los datos registrados.

Mejorar la Integridad de los datos:

- Implementar estrategias de imputación para valores faltantes.
- Corregir valores atípicos e inconsistentes.

Analizar el Desempeño de los proveedores:

- Examinar los tiempos de respuesta y solución para identificar áreas de mejora.
- Evaluar el impacto de la calidad del proveedor en los tiempos de resolución.

Proporcionar Recomendaciones de mejora:

- Desarrollar estrategias basadas en el análisis de datos para optimizar los procesos operativos.
- Sugerir cambios en la captura y manejo de datos para aumentar la calidad de la información.

Calendario de actividades

A continuación, se detalla el calendario de actividades diseñado para este proyecto, siguiendo la metodología CRISP-DM. Este calendario asegura que todas las tareas se completen de manera organizada y dentro de los plazos establecidos.

FASES		ACTIVIDAD	SESION	FECHA
Fase 1	Comprensión del negocio	Reunión con el compañero de equipo donde se establece las expectativas, roles y responsabilidades.	1	22/05/2024
		Identificar los requisitos y las restricciones del proyecto y la metodología del proyecto.	2	22/05/2024
		Crear un calendario de actividades eb donde se detallara las tareas y tiempos estimados para cada fase de la metodología.	2	22/05/2024
		Definir los objetivos, el alcance y necesidades del proyecto, clarificando los objetivos del proyecto y delimitando el alcance del trabajo	3	23/05/2024
		Identificar una fuente de datos confiable que se ajusten a los requerimientos esto podría incluir bases de datos gubernamentales, bases de datos académicas, conjuntos de datos disponibles en línea, datos de empresas privadas	4	24/05/2024
Fase 2	Comprensión de los datos	Recopilación de datos iniciales y comprensión de su estructura, formato y calidad	5	27/05/2024
		Descripción de los datos donde se describa las características y detalles de los datos utilizados.	5	27/05/2024
		Análisis exploratorio de los datos considerando la visualización de los datos para identificar patrones, tendencias o anomalías de manera más efectiva	6	28/05/2024
		Verificación de la calidad de los datos identificación valores atípicos, datos faltantes y errores, para garantizar la integridad y confiabilidad de los datos.	7	29/05/2024
		Evaluar la relevancia de las variables disponibles para los objetivos del proyecto	8	30/05/2024
		Documentar las observaciones preliminares sobre los datos y las posibles limitaciones	8	30/05/2024
Fase 3	Preparación de los datos	Limpieza de los datos: Durante esta etapa, es importante abordar problemas como datos faltantes, valores atípicos, errores de entrada y duplicados, utilizando técnicas como la imputación de datos para rellenar valores faltantes, y eliminar o corregir valores atípicos y duplicados.	9	31/05/2024
		Identificando los KPIs para el análisis y el modelado	9	31/05/2024
		Transforma los datos para que sean adecuados para el modelado en Power BI. Esto puede incluir la estandarización de variables, la normalización de datos, la codificación de variables categóricas y la creación de nuevas variables derivadas.	9	31/05/2024
Fase 4	Modelado	Preparación de los datos en Power BI, cargando los datos del Notebook	10	1/06/2024
		Selección de las visualizaciones y técnicas de análisis adecuadas para los objetivos del proyecto	10	4/06/2024
		Creación de modelos de datos en Power BI	11	4/06/2024
		Diseño de paneles y visualizaciones en Power BI	11	5/06/2024
		Validación y refinamiento de modelos en Power BI	11	5/06/2024
Fase 5	Evaluación	Evaluación del modelo	12	6/06/2024
		Evaluación de la efectividad de los modelos y análisis realizados.		6/06/2024
		Análisis de la calidad y la relevancia de los resultados obtenidos.		6/06/2024
		Revisión de los resultados con los interesados y recopilación de retroalimentación		6/06/2024
Fase 6	Despliegue	Documentación de las conclusiones y recomendaciones para futuras iteraciones del proyecto.	13	7/06/2024
		publicación del proyecto en Power BI Service		7/06/2024
		Creación de informe final		7/06/2024
		Presentación del proyecto		7/06/2024

Comprensión de los datos

Descripción de datos

El conjunto de datos consta de 162,424 filas y 48 columnas, el cual registra las solicitudes y los tickets de servicio de cajeros automáticos. A continuación, se presenta una descripción detallada de las columnas:

Nombre Columna	Descripción	
MINUTA_ID	Identificador único de la minuta	numero
Fecha Solicitud	Fecha en la que se realizó la solicitud	fecha
Hora Solicitud	Hora en la que se realizó la solicitud	texto
Dia Solicitud	Día en el que se realizó la solicitud	texto
Jornada Solicitud	Jornada en la que se realizó la solicitud (por ejemplo, mañana, tarde, noche y madrugada)	texto
Categoria Minuta	Categoría a la que pertenece la minuta	texto
Punto ID	El identificador único del punto donde se encuentra el cajero automático asociado a la minuta.	numero
ATM	El código de identificación del cajero automático.	numero
Marca ATM	Marca del cajero automático	texto
Modelo ATM	Modelo del cajero automático	texto
ATM VIP	Indicador de si el cajero automático es VIP	texto
Nombre ATM	Nombre asignado al cajero automatico	texto
PUNTO_TIPO_UBICACION	El tipo de ubicación del punto donde se encuentra el cajero automático por ejemplo si es urbano o rural.	texto
PUNTO_CIUADADOP	El código de la ciudad operativa del punto	texto
Ciudad Operativa	Ciudad en donde se realizó la solicitud y la operación	texto
PUNTO_LAT	Latitud de la ciudad	numero
PUNTO_LON	Longitud de la ciudad	numero
Fecha Asignación	Fecha en la que se asignó la solicitud	fecha
Hora Asignación	Hora en la que se asignó la solicitud	texto
Dia Asignación	Día en el que se asignó la solicitud	texto

Nombre Categoría	Hace referencia al tipo de operación que se realizó, distinguiendo entre operaciones de primera línea y operaciones de segunda línea. Esto significa que clasifica las solicitudes en función del nivel de complejidad o del tipo de tarea que se llevó a cabo.	texto
Subcategoría	Proporciona una descripción más detallada o específica de la naturaleza de la solicitud o minuta	texto
Novedad	Descripción de la subcategoría	texto
MINUTA_DESCRIP	Registros detallados de solicitudes de servicio, problemas técnicos, tareas operativas u otros eventos que requieren seguimiento	texto
Nombre Proveedor	Nombre del proveedor asignado	texto
TRAZABILIDAD	Muestra la trazabilidad de la solicitud	texto
FECHA_ACCES	Fecha de acceso	fecha
ESTADO_REQ	Estado de la solicitud en numero	numero
NOMBRE_ESTADO_REQ	Describe el estado de la solicitud	fecha
HORA_RESOLV	La hora de resolución de la solicitud.	fecha
Tiempo de Respuesta	Tiempo de respuesta de la solicitud	numero
Tiempo de Respuesta Tarde	Tiempo de respuesta de la solicitud por la tarde por parte de la empresa	numero
Tiempo Respuesta Proveedor Tarde	Tiempo de respuesta de la solicitud por la tarde por parte del proveedor	numero
TIEMPO_RESP_TOTAL	Tiempo de respuesta total de la solicitud en minutos	numero
Hora Llegada a Sitio	Hora Llegada a Llegada al sitio para realizar la solicitud requerida	texto
Fecha Llegada a Sitio	Fecha Llegada a Llegada al sitio para realizar la solicitud requerida	fecha
Fecha Cierre Ticket	Fecha Cierre ticket Cierre la solicitud	fecha
Hora Cierre Ticket	Hora Cierre ticket Cierre la solicitud	texto
TIEMPO_RESOLV	El tiempo de resolución de la solicitud en minutos	numero
HORAS_FRACCION	Horas de resolución de la solicitud	numero
Tiempo Solución lvi 6	El tiempo de solución del nivel 6. en minutos	numero
Tiempo Espera al Proveedor	El tiempo de espera al proveedor.	numero
TIEMPO_REQUERIDO_CLI	El tiempo requerido por el cliente.	numero
MINUTA_TICKETID	Identificador único de la minuta del ticket	texto

KTC	KTC	texto
PUNTO_ESTADO	Estado de la solicitud en numero	numero
ESTREQ_DESCRIPCION	Proporciona una descripción de las operaciones que realizaron los técnicos de mantenimiento al finalizar el servicio en los cajeros automáticos.	texto
dia	Dia de la operación	numero

Análisis de valores faltantes

Porcentaje de valores faltantes

Columna	Porcentaje de valores faltantes
Tiempo Solución lvi 6	87%
Tiempo Respuesta Proveedor Tarde	82%
Nombre proveedor	60%
Tiempo de respuesta	56%
Tiempo Espera al Proveedor	51%
Tiempo de Respuesta Tarde	44%

Estos hallazgos revelan una falta significativa de datos en varias columnas, lo cual puede afectar la capacidad de analizar y comprender ciertos aspectos del proyecto.

Análisis de Valores Atípicos

Se identificó en la columna TIEMPO_RESOLV un valor mínimo negativo de -1426 como valor mínimo.

Este valor negativo sugiere un error en la recolección o registro de datos lo cual es inusual ya que los tiempos de resolución no pueden ser negativos.

A continuación, se presenta un resumen de los hallazgos:

- Número total de filas: 162424
- Número de valores negativos: 527
- Porcentaje de valores negativos: 0.32%

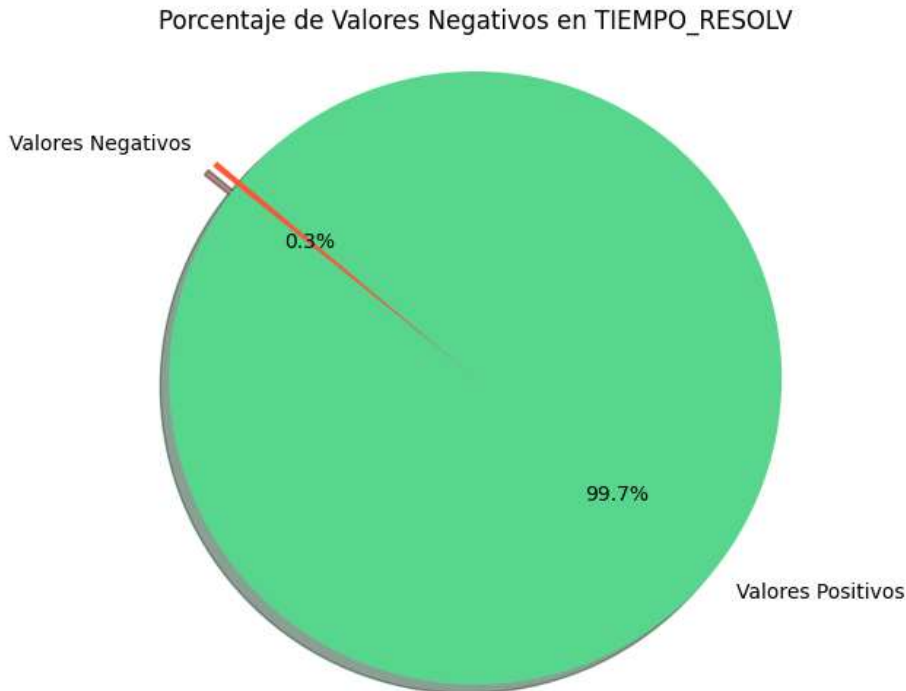


imagen 1

Se deben considerar las siguientes acciones para tratar los valores negativos:

Investigación Adicional:

Es necesario investigar la razón detrás de estos valores negativos para comprender si son errores de entrada, problemas en la recolección de datos o tienen algún significado especial.

Posibles Causas:

Los valores negativos podrían ser el resultado de errores en la captura de datos o problemas técnicos durante la recolección.

También podrían tener un significado particular según el contexto del sistema o proceso que generó los datos.

Análisis de la Columna TIEMPO_SOLUCIÓN_LVL_6

La columna TIEMPO_SOLUCIÓN_LVL_6 fue analizada para comprender el tiempo de solución en cierto nivel del proceso. A continuación, se presentan los hallazgos y consideraciones:

Valores Vacíos o Nulos: Se encontró que el 87% de los registros en esta columna están vacíos o tienen valores nulos. Este alto porcentaje de datos faltantes puede

deberse a diversas razones, como problemas en la recopilación de datos o falta de registro.

Porcentaje de Valores vacios en la columna TIEMPO_SOLUCIÓN_LVL_6

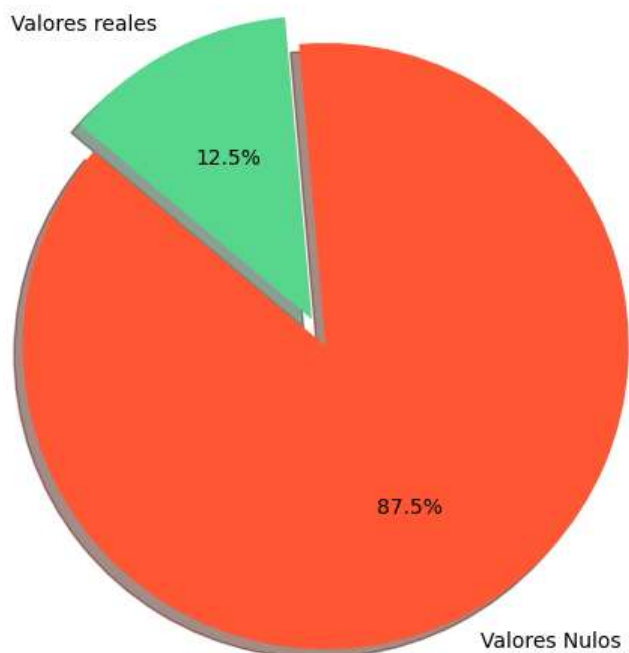


imagen 2

Inconsistencia en los Datos Existentes: Además de los valores vacíos, se observó que los datos existentes en esta columna no corresponden a los valores esperados o reales. Esto sugiere posibles errores en la captura o registro de los datos. Parece que los datos presentes en esta columna vienen de la columna “TIEMPO_RESOLV” lo cual no debería ser así debido a que este tiempo se toma de otra manera.

Impacto en el Análisis:

La presencia de valores vacíos y datos inconsistentes en la columna “TIEMPO_SOLUCIÓN_LVL_6” puede tener un impacto significativo en nuestro análisis, ya que limita nuestra capacidad para obtener conclusiones precisas sobre el tiempo de solución en este nivel del proceso.

Consideraciones Adicionales:

Origen de los Valores Incorrectos: Es importante investigar la razón detrás de los datos incorrectos existentes en esta columna. Esto podría implicar errores en la captura de datos, problemas técnicos, o definiciones incorrectas de los procesos.

Preparación de datos

Cambios en los valores negativos en la columna TIEMPO_RESOLV

Para abordar los valores negativos identificados en la columna 'TIEMPO_RESOLV', se han tomado las siguientes medidas:

- Se identificaron un total de 527 valores negativos en un conjunto de datos que contiene 162424 registros.
- Estos valores negativos representan aproximadamente el 0.32% del total de registros en la columna.
- Se decidió cambiar estos valores negativos por cero para corregir esta anomalía.

Se utilizó el siguiente código para reemplazar los valores negativos por cero en la columna "TIEMPO_RESOLV":

```
df_eventos_cajeros.loc[df_eventos_cajeros['TIEMPO_RESOLV'] < 0,  
'TIEMPO_RESOLV'] = 0
```

Eliminación de la Columna TIEMPO_SOLUCIÓN_LVL_6

Dado que más del 80% de los valores en esta columna son nulos y los datos existentes no son representativos de los tiempos reales de solución, se tomó la decisión de eliminar esta columna del conjunto de datos. A continuación, se explican las razones:

Alto Porcentaje de Valores Nulos:

El 87% de los valores en esta columna están ausentes, lo que indica una falta significativa de datos. Esto dificulta cualquier análisis significativo sobre esta variable.

Datos No Correspondientes:

Los datos existentes en esta columna no representan los tiempos de solución reales, los datos presentes en esta columna vienen de la columna "TIEMPO_RESOLV" lo cual no debería ser así debido a que este tiempo se toma de otra manera; Esto puede llevar a interpretaciones incorrectas o sesgadas de los datos.

Dada la falta de información útil y la naturaleza no representativa de los datos existentes, la eliminación de esta columna no afectará negativamente nuestro análisis y nos ayudará a mantener la calidad y la integridad de nuestros datos.

Manejo de datos faltantes

Durante el proceso de análisis de los datos, se encontraron valores faltantes en varias columnas del conjunto de datos. Se aplicó un proceso de imputación para manejar estos valores ausentes. A continuación, se detalla cómo se abordó esta situación:

Identificación de Valores Faltantes

Se identificaron un total de 30 columnas con valores faltantes en el conjunto de datos.

- **Columnas Cuantitativas:** Para las columnas que contenían datos cuantitativos, los valores nulos se reemplazaron por 0.
- **Columnas Categóricas:** Para las columnas categóricas, los valores nulos se mantuvieron como 'NULL' para indicar la ausencia de información.

Razones para el Manejo de Valores Faltantes:

Consistencia en los Datos:

La presencia de valores faltantes puede afectar la integridad y la consistencia de nuestro análisis. Por lo tanto, se decidió abordar estos valores ausentes de manera uniforme en todas las columnas.

Preservación de la Información:

Reemplazar los valores nulos por 0 en columnas cuantitativas nos permite preservar la información sobre la ausencia de datos, especialmente en casos donde los valores nulos pueden tener un significado específico.

Mantener 'NULL' en columnas categóricas nos permite distinguir entre valores ausentes y categorías válidas.

El manejo adecuado de los valores faltantes asegura la integridad de nuestros datos y nos permite realizar análisis más precisos y significativos.

Creación de la columna ANHO para año de solicitudes

Con el objetivo de mejorar la legibilidad y facilitar la manipulación de fechas en el conjunto de datos, se ha creado una nueva columna llamada "ANHO", la cual contendrá únicamente el año de las solicitudes. Esta decisión se tomó por las siguientes razones:

Legibilidad Mejorada:

Al extraer solo el año de las fechas de las solicitudes, se simplifica la visualización de la información y se mejora la comprensión de los datos.

Manipulación con datetime:

Al utilizar la librería datetime de Python para manipular fechas, es más conveniente trabajar con datos que contienen solo el año.

Revisión de Datos del Proveedor

Se identificó que la marca de cajero automático "BELLTECH" es en realidad un proveedor, por lo que se procedió a reemplazar los valores correspondientes para reflejar la información correcta.

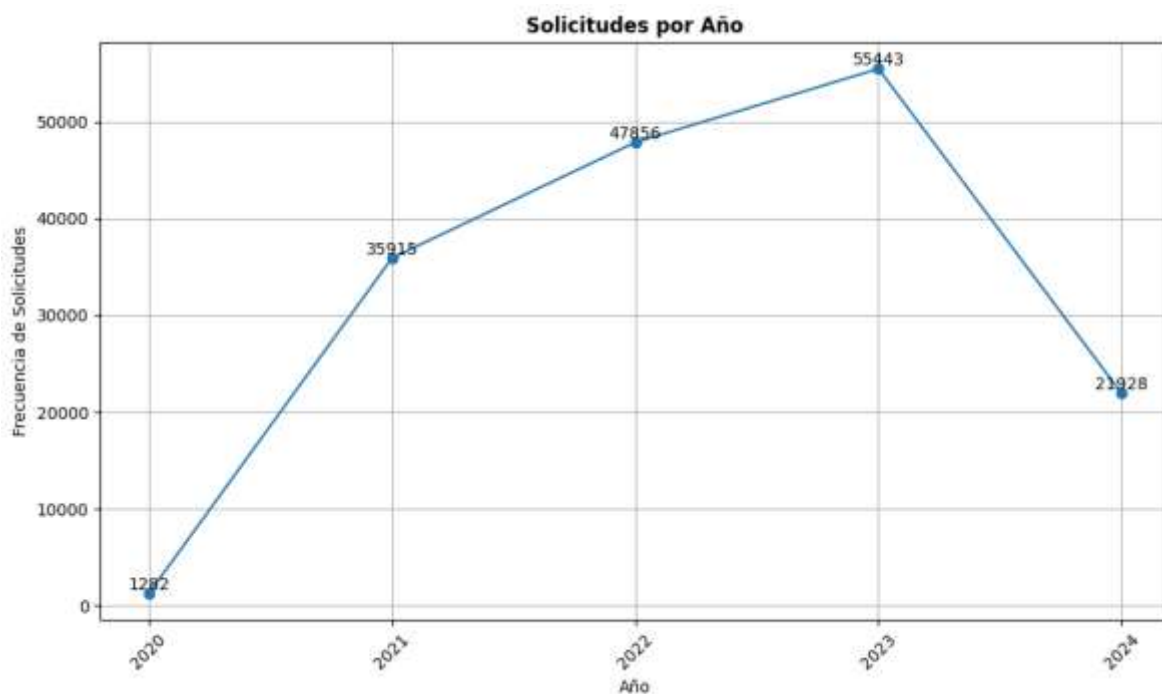


imagen 3

Manejo de Valores Nulos en la Columna NOMBRE_PROVEEDOR

Durante el análisis de los datos, se identificaron valores nulos en la columna "NOMBRE_PROVEEDOR". Estos valores nulos pueden tener un impacto negativo en nuestro análisis y en la interpretación de los datos. A continuación, se describe cómo se abordó esta situación:

Estrategia de Imputación

Dado que eliminar las filas con valores nulos podría reducir significativamente el tamaño de nuestro conjunto de datos y perder información relevante, se decidió aplicar una estrategia de imputación para llenar estos valores faltantes. La estrategia utilizada fue la siguiente:

Relleno con el Término 'Desconocido':

En lugar de eliminar las filas, se optó por rellenar los valores nulos en la columna NOMBRE_PROVEEDOR con el término 'Desconocido'.

Al utilizar el término 'Desconocido', estamos reconociendo explícitamente la falta de información específica sobre el proveedor asociado a estos eventos.

Razones para la Estrategia Utilizada

Preservación de Datos:

Rellenar los valores nulos con 'Desconocido' nos permite conservar todas las filas de datos disponibles y evitar la pérdida de información relevante.

Reconocimiento de la Falta de Información:

Al utilizar 'Desconocido', se reconoce claramente que no disponemos de información específica sobre el proveedor en esos casos.

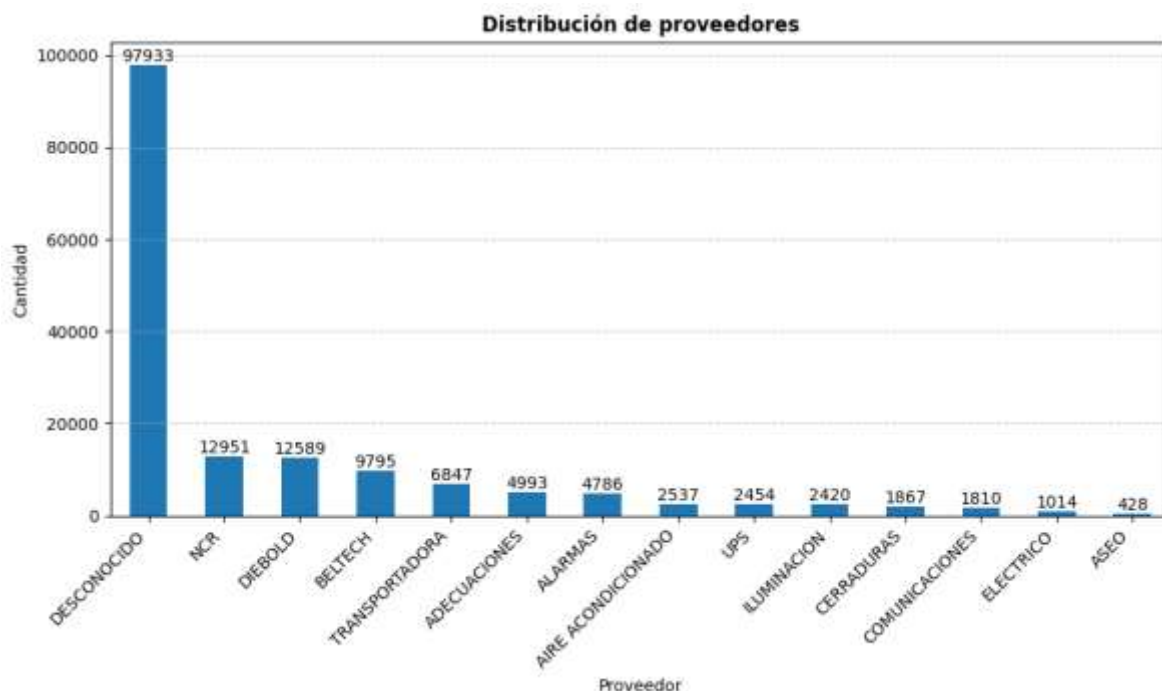


imagen 4

En la imagen 4 se puede observar un alto nivel del proveedor desconocido debido a la falta de información específica sobre el proveedor asociado a estos registros.

La falta de información específica o registros adecuados sobre los proveedores llevó a que la mayoría de las filas fueran etiquetadas como 'Desconocido'.

Esta representación nos ayuda a reconocer claramente la falta de información específica sobre el proveedor en una gran parte del conjunto de datos.

Corrección de la Marca de Cajero Automático "BELLTECH"

Durante el análisis de los datos, se identificó que la marca de cajero automático "BELLTECH" en realidad corresponde a un proveedor en lugar de una marca de cajero. Por lo tanto, se realizó una corrección para reflejar con precisión esta información en el conjunto de datos. A continuación, se detalla este proceso:

Identificación del Error

Durante la revisión de los datos, se encontró que la marca de cajero automático "BELLTECH" estaba incorrectamente registrada como una marca de cajero.

Acción Tomada

Se decidió corregir este error reemplazando los valores correspondientes en el conjunto de datos para reflejar con precisión que "BELLTECH" es un proveedor en lugar de una marca de cajero.

Razones para la Corrección

Exactitud de los Datos:

Es importante mantener la precisión en la información registrada en el conjunto de datos para evitar interpretaciones incorrectas.

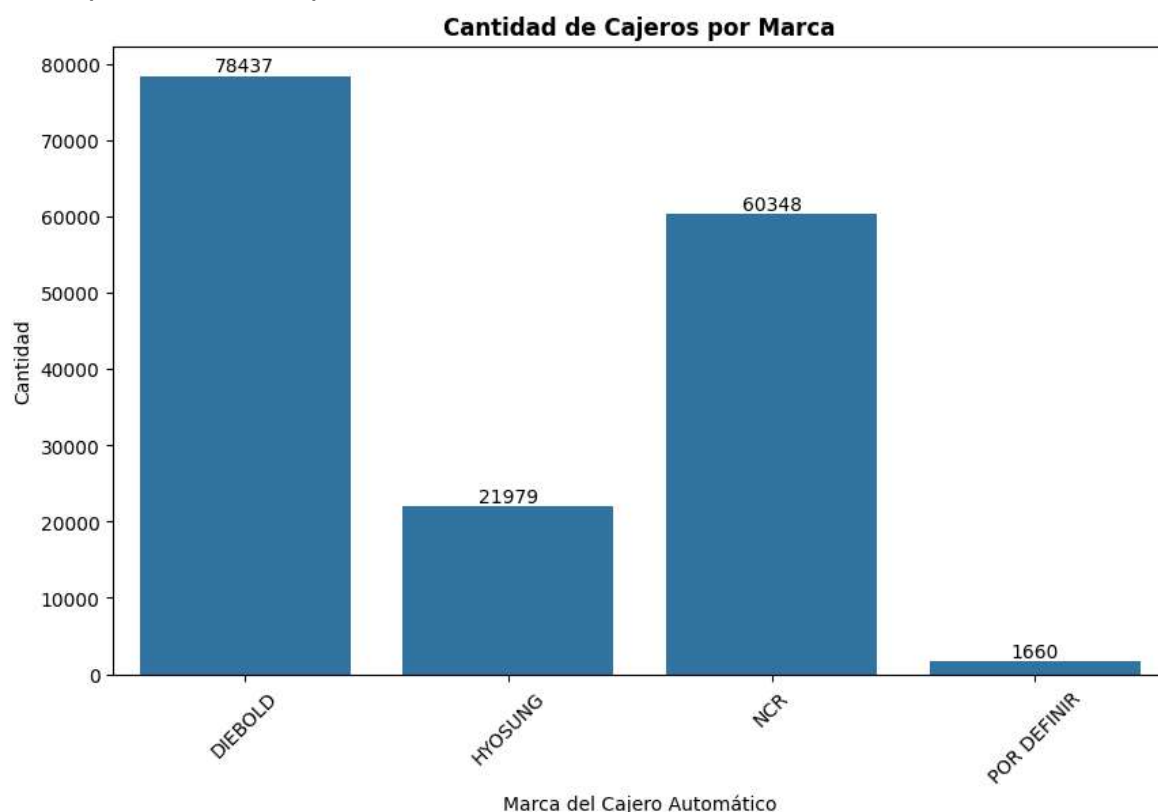


imagen 5

Modelado

Resumen operativo de cajeros automáticos

Se realizó un resumen operativo de los cajeros automáticos, permitiendo una visualización rápida y detallada del rendimiento y la operación de los mismos. A continuación, se presenta un análisis de diferentes aspectos clave de los cajeros automáticos, incluyendo el número de cajeros, total de tickets, categorías, ubicaciones y mucho más.



Imagen 6

Seguimiento de tickets

Se realizó un reporte detallado de los tickets de soporte o incidencias, permitiendo analizar el flujo de trabajo, el estado actual de los tickets y la eficiencia en la resolución de problemas. A continuación, se presenta un análisis exhaustivo de los tickets, incluyendo su estado, tiempos de resolución, categorías más comunes y

más.

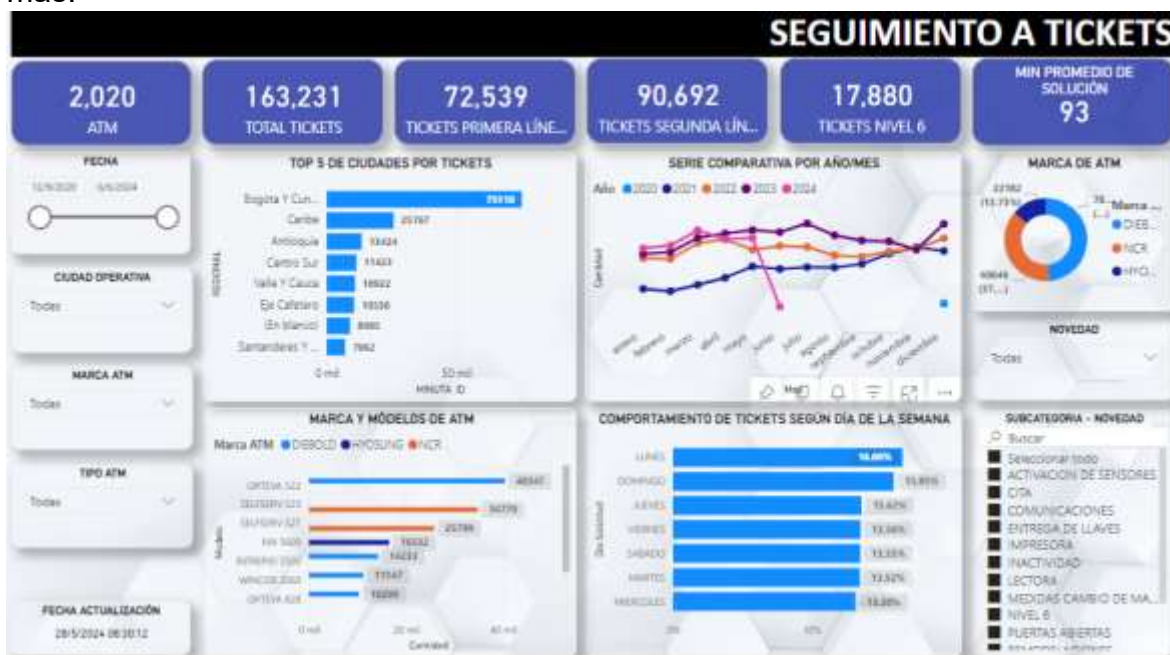


imagen 7

Analisis de solicitudes de primera y segunda linea

El análisis de primera línea y segunda línea permite entender cómo se distribuye la carga de trabajo entre los equipos de soporte de nivel 1 (primera línea) y nivel 2 (segunda línea), así como la eficiencia en la resolución de problemas en cada

nivel.

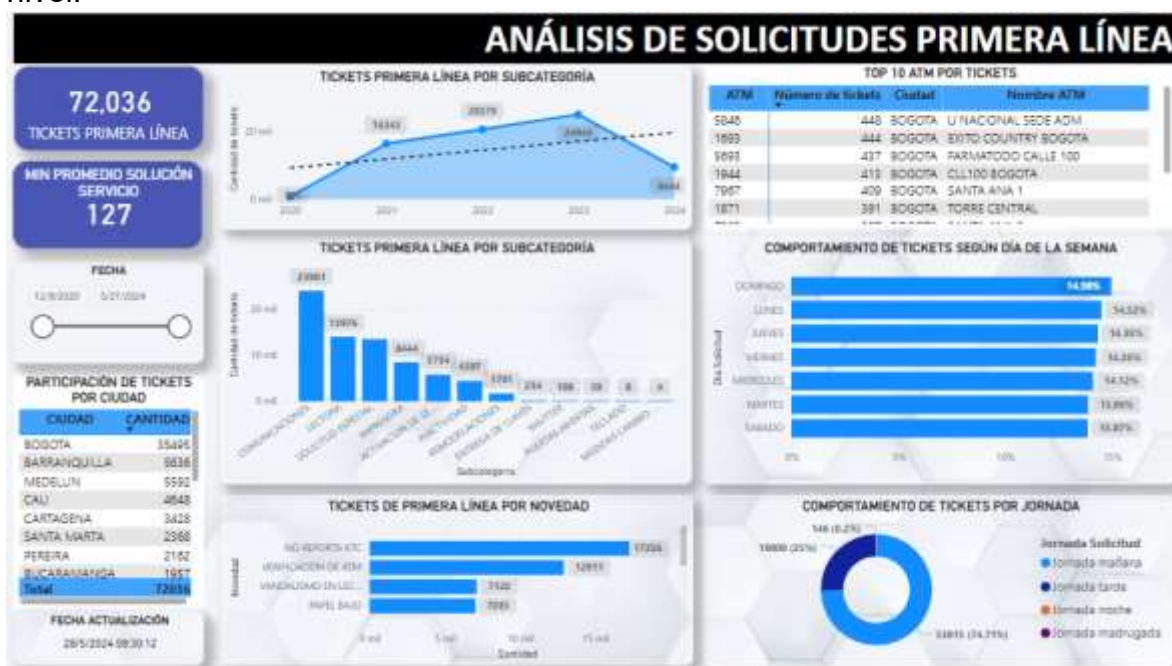


imagen 8

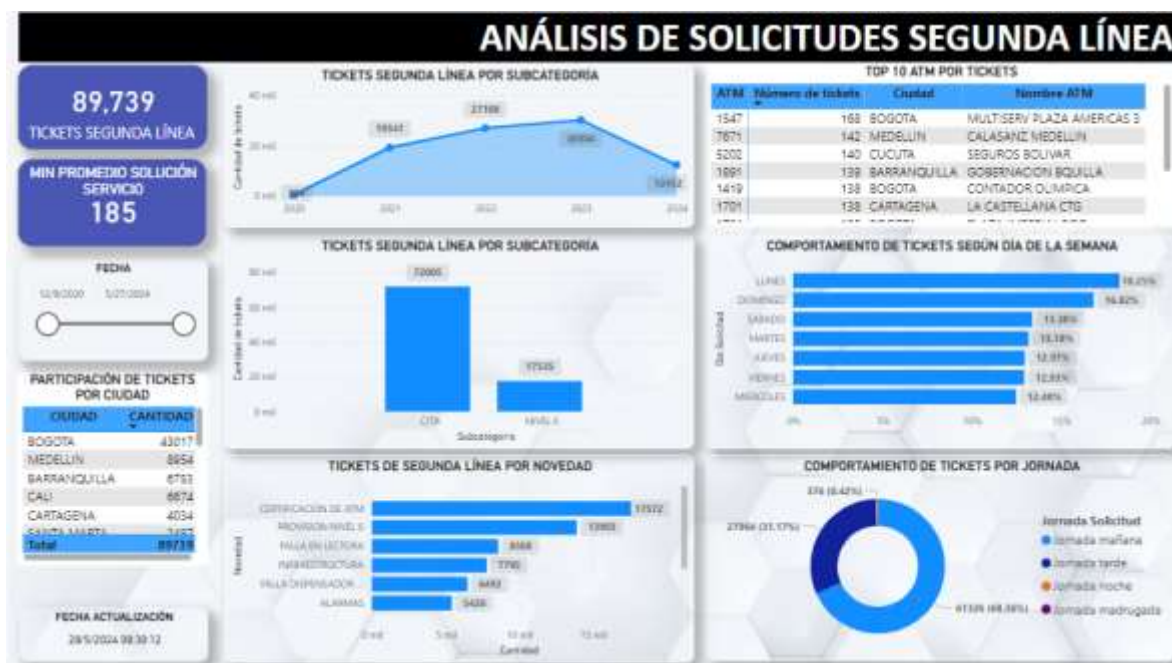


imagen 9

Evaluación y despliegue

Conclusión

En conclusión, el análisis de la satisfacción del cliente realizado en este proyecto ha revelado importantes problemas de calidad en los datos, incluyendo la presencia de valores faltantes y atípicos. Sin embargo, mediante la aplicación de técnicas de preparación de datos y modelado, se han podido identificar oportunidades de mejora en la eficiencia y eficacia de los procesos operativos.

A pesar de estos desafíos, el análisis ha permitido identificar oportunidades de mejora en la eficiencia y eficacia de los procesos operativos, como la optimización de los tiempos de respuesta y solución, la mejora de la calidad de los datos y la implementación de estrategias de mejora continua.

En resumen, este proyecto ha demostrado la importancia de la calidad de los datos y la necesidad de aplicar técnicas de análisis y preparación de datos para mejorar la eficiencia y eficacia de los procesos operativos. Se recomienda continuar monitoreando y mejorando la calidad de los datos para asegurar que se tomen decisiones informadas y se mejore la satisfacción del cliente.

Recomendaciones

Se recomienda continuar monitoreando y mejorando la calidad de los datos para tomar decisiones informadas y mejorar la satisfacción del cliente.