

INFORME DEL PROCESO ETL y EDA DEL PROYECTO CONCESIONARIA

1. Introducción

Este informe resume el proceso de Exploración y Limpieza de Datos (EDA y ETL) aplicado al proyecto de análisis de ventas de vehículos en Power BI. El objetivo fue transformar datos crudos en un modelo analítico limpio y normalizado, listo para crear un dashboard profesional basado en un esquema copo de nieve

2. Fuentes de Datos ETL

Se trabajó con una tabla principal de hechos que en un principio contenía más de 26 columnas (ventas de vehículos) la cual mediante google cloud python y pandas se logró normalizar y extraer varias tablas ,una de Hechos (Ventas Vehiculos) y múltiples tablas de dimensiones normalizadas tales como: marca, modelo, ciudad,, color, combustible, condición, método de pago,segmento, sucursal, transmisión y vendedor.

Se aplicaron transformaciones en Power Query:

1. - Corrección de tipos de dato
2. - Limpieza de valores negativos y mayores a 100 reemplazados por la mediana.
3. - Normalización del dataset en tablas de dimensiones.
4. - Eliminación de duplicados y claves inconsistentes.
5. - Integración mediante relaciones desde las dimensiones hacia la tabla de hechos. (MODELO DE DATOS)
6. El modelo final sigue un esquema copo de nieve compuesto por:
 - 1 - Una tabla de hechos: ventas_vehiculos.
 - 2- Múltiples tablas de dimensiones como marca, modelo, ciudad, et
 - 4 - Conclusión.
7. Tras la aplicación del proceso ETL, los datos quedaron completamente limpios, normalizados Listas para el EDA

3. Proceso EDA

Análisis exploratorio de patrones de Datos ; se identificaron:

1. - Valores nulos en columnas clave (Precio, Año_Vehiculo, Edad_Comprador).
2. - Errores de formato como textos 'Error' en campos numéricos.
3. - Valores negativos en precio.
4. - Inconsistencias entre dimensiones (marcas asociadas a modelos incorrectos).
5. - Valores atípicos en precio, año del vehículo y Edad de los clientes
6. Se encontró inconsistencias en modelo marca , vendedor ventas ,

4. Modelado y Dashboard

- a. Modelado.
- b. Tabla Calendario
- c. Medidas Dax.
- d. Gráficos y Visuales .

