Caso de clasificación (EDA)

Sebastián Macías Gallego
Codign Dojo - Ciencia de Datos
2022

Problema

Proyectar

Desde el área de mantenimiento es neceario realizar la proyección de costos anual, para ello se requiere conocer la inversión en los diferentes sistemas de los activos.

Automatizar

El sistema de facturación actual no permite la clasificación de los sistemas o sistemas intervenido, es una clasificación manual y solo se ha realizado para una parte del historial de facturación.

Clasificar

El reto es asegurar que los sistemas de la facturación nueva e histórica sea clasificada de manera adecuada.

Análisis de los desafíos

Desafío 1

Cantidad de categorías

- Más de 200 activos.
- 21 categorías para clasificar.
- Dataframe con un total de 17 columnas y 134887 filas.

Desafío 2

Datos faltantes

- Hasta un 40% de la información.

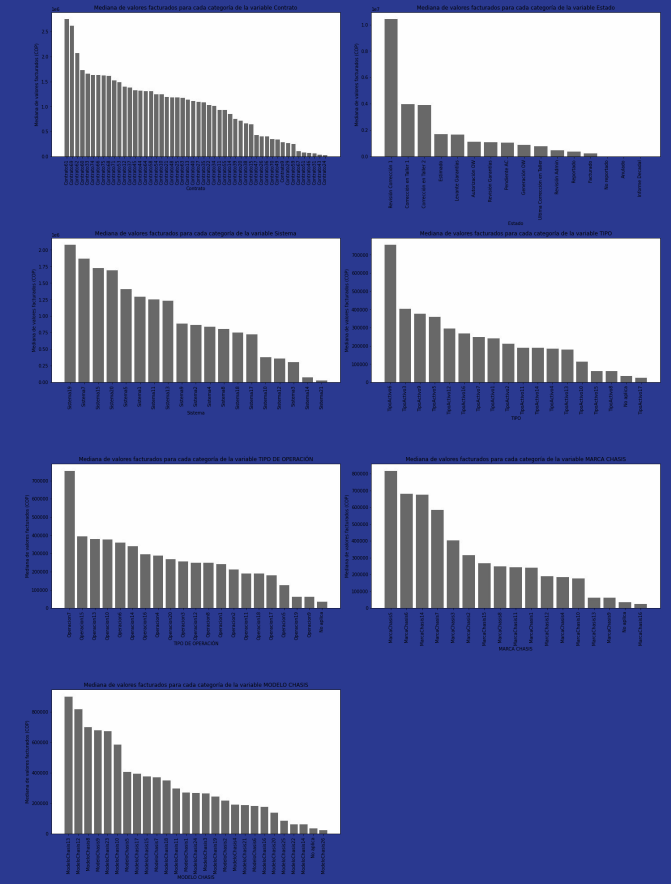
Desafío 3

Valores atípicos

- Rangos de valores que van 0.0 COP hasta los 1.0 e7 COP

Solución

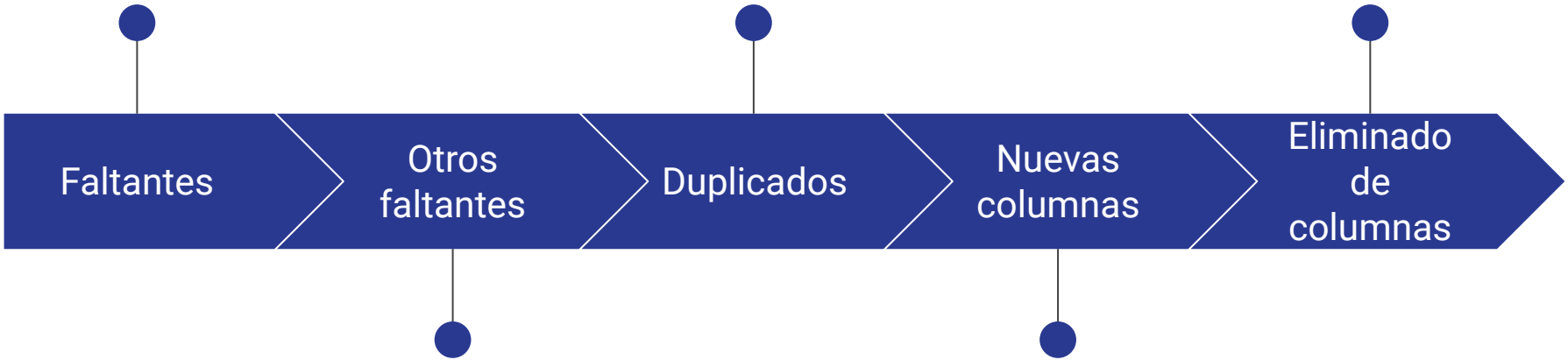
Procesado de información, contextualización del problema.



~98% : Actividad de mantenimiento general (Lavado), asociado a un solo sistema, el 21

779, lo que corresponde a **menos del 0.5%** del total de datos, por lo que fueron borrados

Columna de número de OT, fecha de entrega, fecha de inicio y terminación, número de activo ajustado.



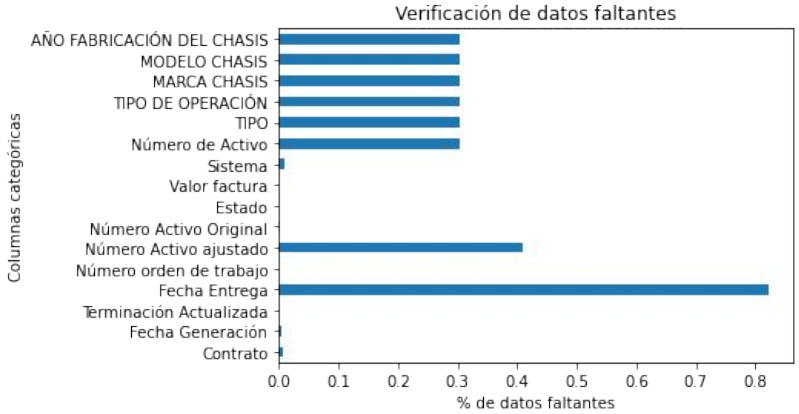
Filas eliminadas representaban **menos del 2%** de la data restante

- Tiempo de actividad aproximado en días, con cero para las actividades de mantenimiento general..
- Identificación de actividad general

Puesta en práctica

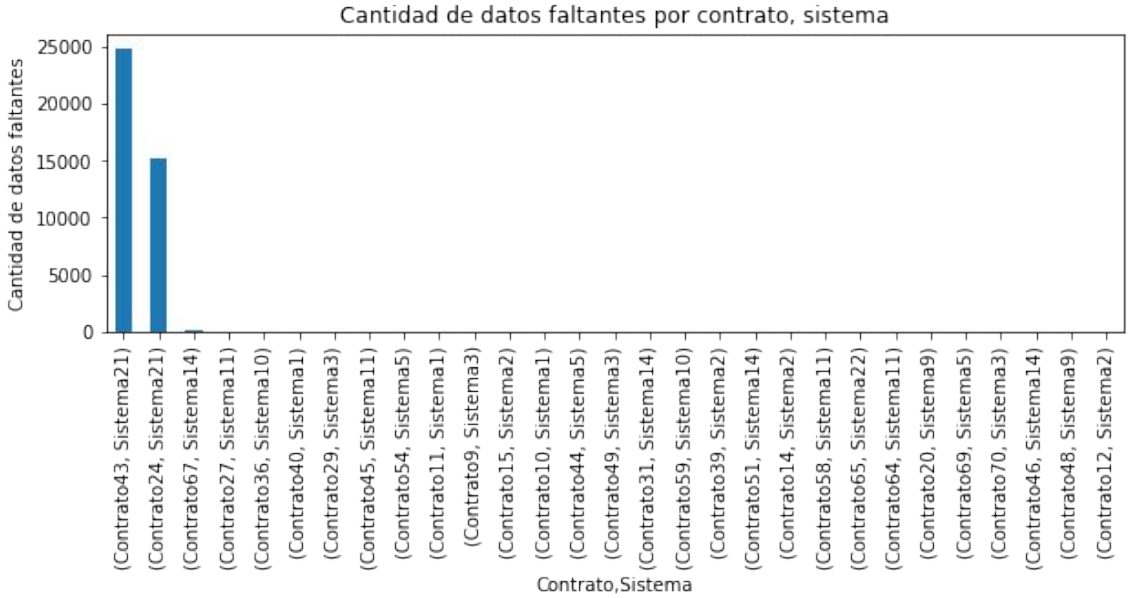
~98% : Actividad de mantenimiento general (Lavado), asociado a un solo sistema, el 21

Los datos faltantes corresponden a las caracterísiticas de los activos



Faltantes

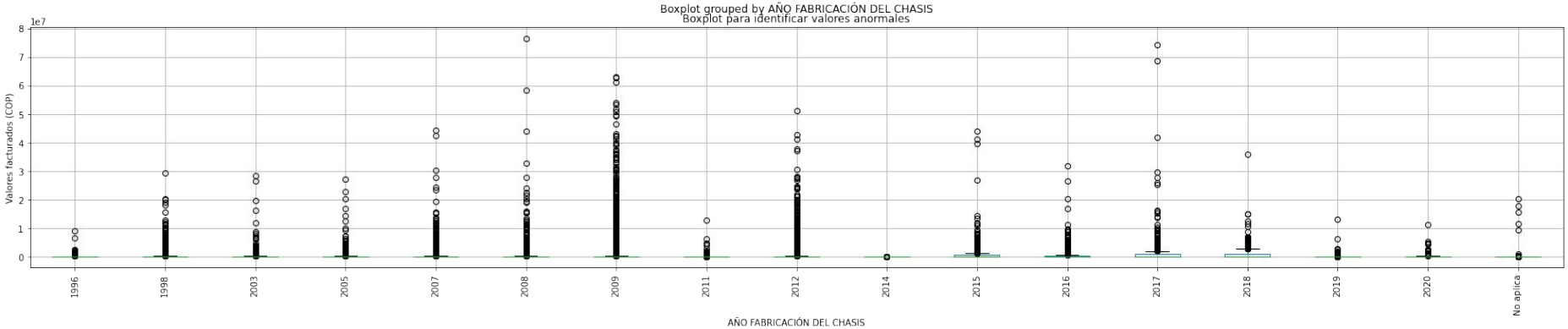
Otros faltantes

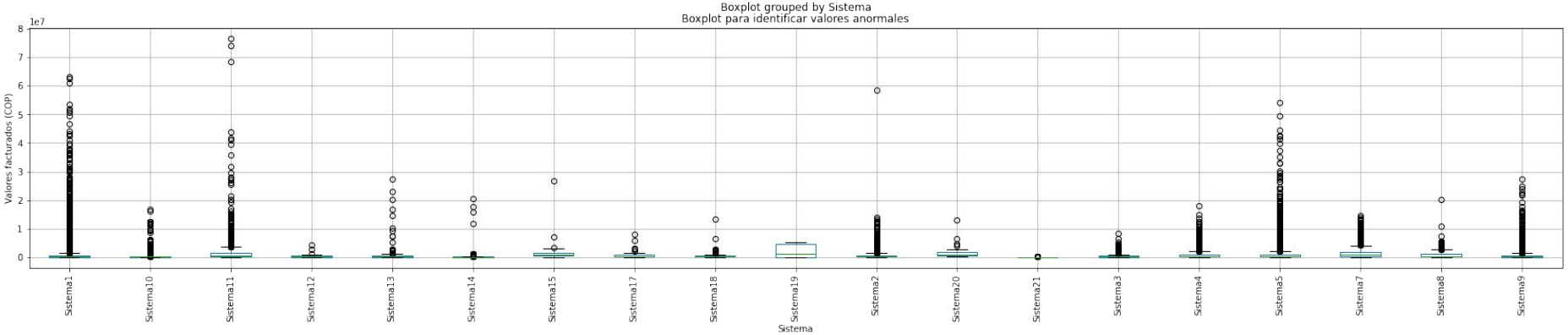


Filas eliminadas representaban **menos del 2%** de la data restante

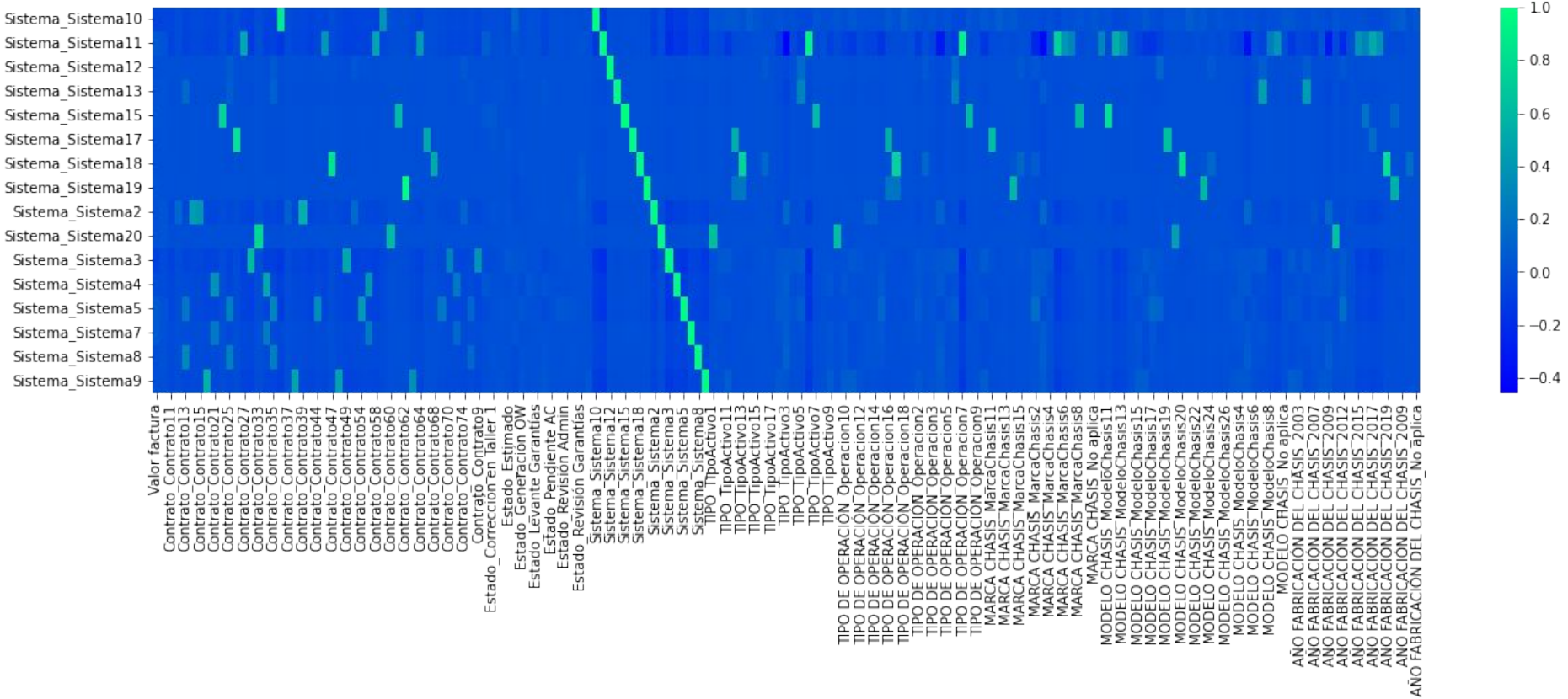
Datos atípicos

Rangos de valores que van 0.0 COP hasta los 1.0 e7 COP





Correlación



Impacto

DataSet listo para procesamiento.