



# Proyecciones de Ventas

Análisis realizado por Mateo Rodríguez

# Reporte de avances

Qué cubre este reporte

---

**01** Base de datos proporcionada

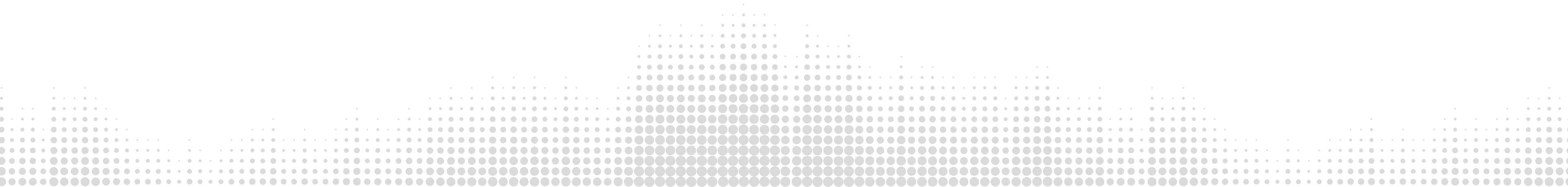
**02** Variables a considerar

**03** Gráficos relevantes

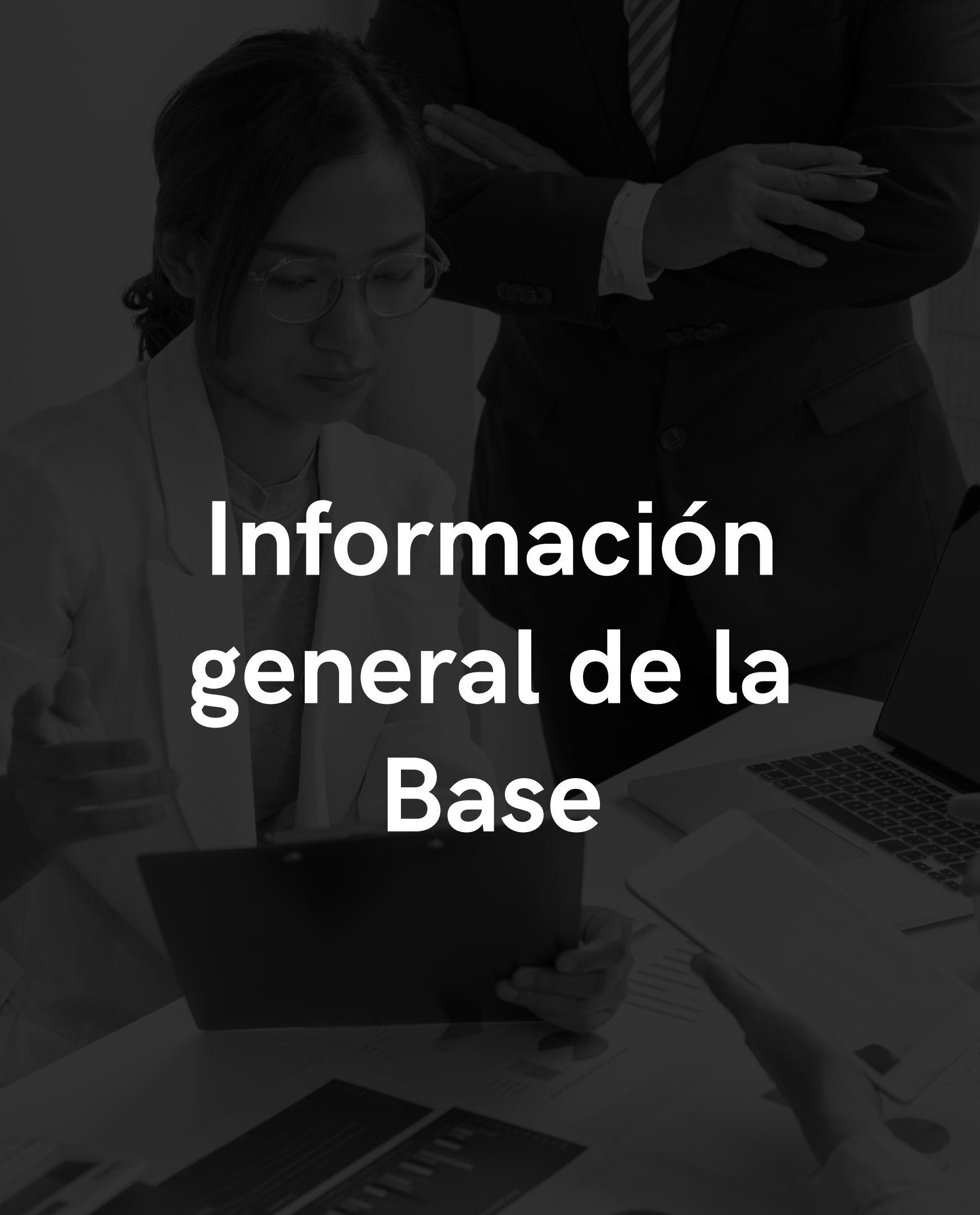
**04** Algoritmos empleados

**05** Conclusiones

**06** Recomendaciones







# Información general de la Base

## Depurar base de datos (NaN)

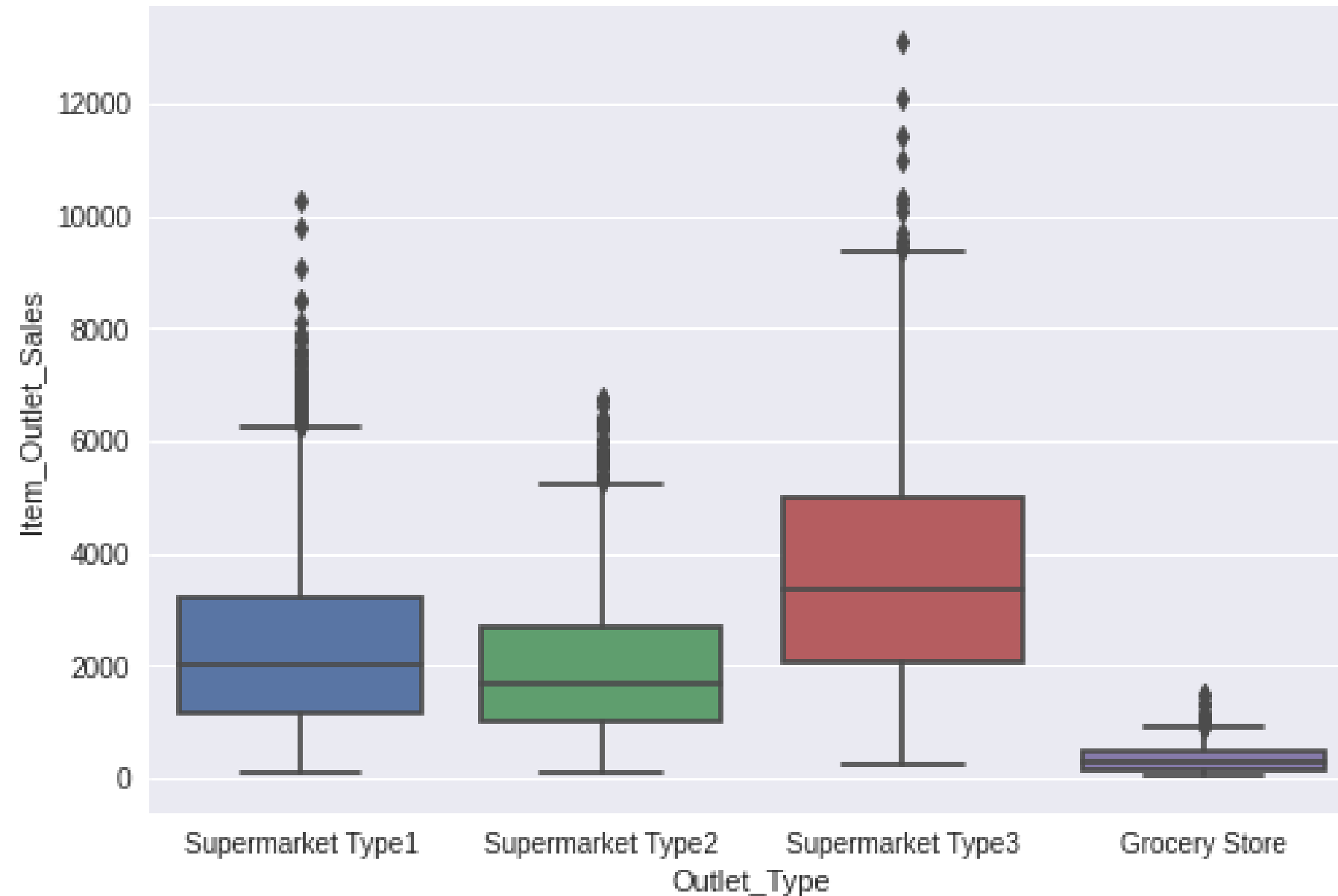
Se realizó la depuración de la base de datos eliminando las columnas con datos vacíos. Esto eliminó los datos de varios outlets (OUT 10, OUT 45, OUT 17)

## Codificación de variables categóricas

Las variables categóricas (nominales y ordinales) fueron codificadas para poder ejecutar los algoritmos y tener una mayor precisión en los resultados.

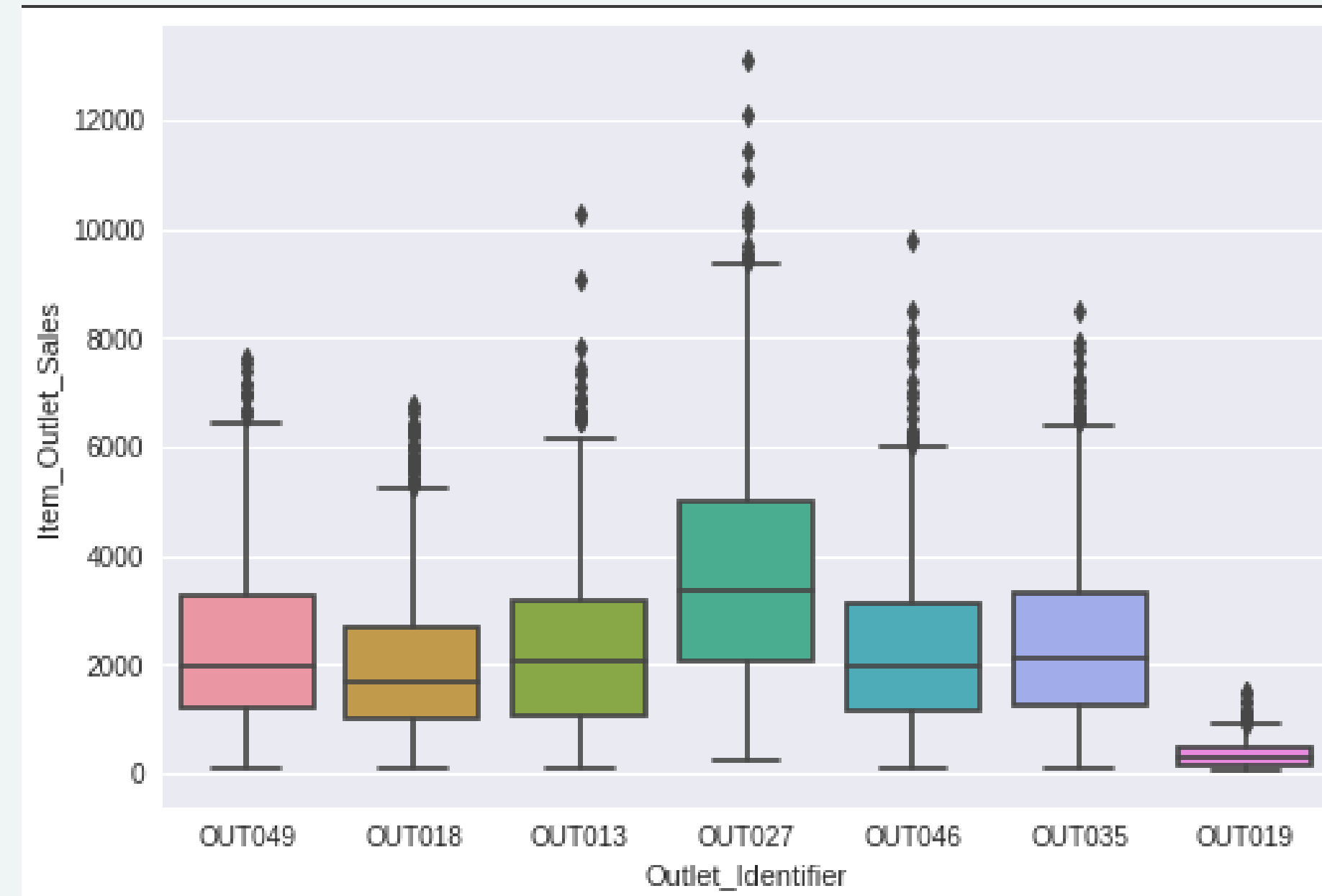
## Corrección de datos

La variable Fat Content fue clasificada de una mejor manera (Low Fat y Regular)



## Ventas por tipo de Outlet

Analizar la distribución de los datos tomando en consideración las variables Item\_Outlet\_Sales (Variable a predecir) y el tipo de Outlet



## Ventas por Outlet

Analizar la distribución de los datos tomando en consideración las variables Item\_Outlet\_Sales (Variable a predecir) y el número del Outlet.

# Algoritmos empleados



KNN Regressor - RandomForestRegressor

---

## KNN Regressor

Score: 0.67

Para construir el modelo se consideró las siguientes variables: 'Outlet\_type\_labels', 'Item\_MRP', 'Outlet\_type\_labels', 'Location\_labels', 'ID\_Label', 'Item\_type\_label'. No se empleo división para este modelo.

## RF Regressor

Score: 0.92

Se consideraron las mismas variables que para el modelo KNN pero con un mejor resultado: 'Outlet\_type\_labels', 'Item\_MRP', 'Outlet\_type\_labels', 'Location\_labels', 'ID\_Label', 'Item\_type\_label'. División de la base (Train y Test)