IT ACADEMY

Exploración Anual de **Ventas**, **Productos** y **Consumidores** y **Predicción de Envíos** con Machine Learning

Por **Sara Gutierrez**

Tabla de Contenidos

IT ACADEMY

Proyecto Final **Sara Gutierrez**

Objetivos del Proyecto

- Hacer una exploración analítica de las ventas por años
- Hacer un modelo predictivo con Machine Learning

IT ACADEMY

Proyecto Final **Sara Gutierrez**

¿Cómo lo hicimos?

Análisis Exploratorio

DATA COLLECTION

DATA MANAGEMENT

EXPLORATORY DATA ANALYST

IT ACADEMY

Proyecto Final **Sara Gutierrez**

DATA COLLECTION

A través de la API de Kaggle
Al cual podéis acceder aquí

```python
from kaggle.api.kaggle_api_extended import KaggleApi

# Autenticación
api = KaggleApi()
api.authenticate()

# Descargar el dataset
api.dataset_download_files('rohitsahoo/sales-forecasting', path='.', unzip=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9800 entries, 0 to 9799
Data columns (total 18 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9800 non-null   int64
 1   Order ID       9800 non-null   object
 2   Order Date     9800 non-null   object
 3   Ship Date      9800 non-null   object
 4   Ship Mode      9800 non-null   object
 5   Customer ID    9800 non-null   object
 6   Customer Name  9800 non-null   object
 7   Segment        9800 non-null   object
 8   Country        9800 non-null   object
 9   City           9800 non-null   object
 10  State          9800 non-null   object
 11  Postal Code    9800 non-null   object
 12  Region         9800 non-null   object
 13  Product ID     9800 non-null   object
 14  Category       9800 non-null   object
 15  Sub-Category   9800 non-null   object
 16  Product Name   9800 non-null   object
 17  Sales          9800 non-null   float64
dtypes: float64(1), int64(1), object(16)
memory usage: 1.3+ MB
None
```

DATA MANAGEMENT

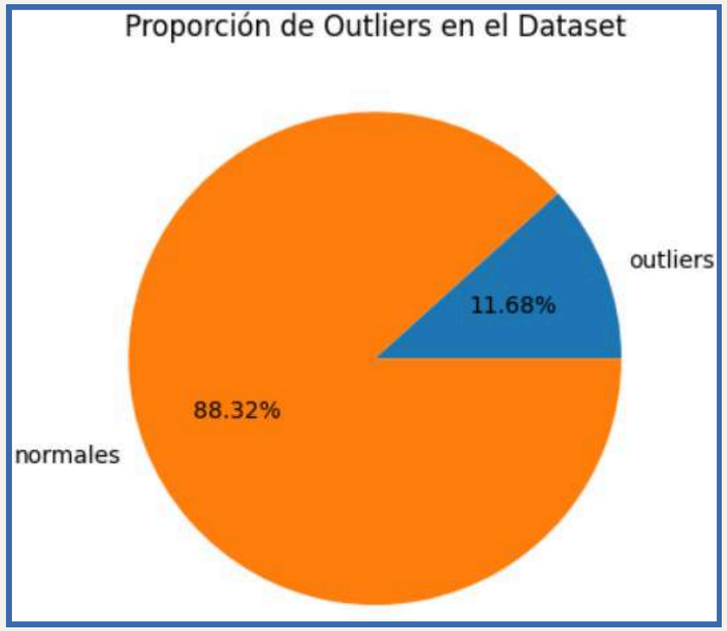
- Nuestro dataset estaba muy limpio
- Cambiamos algunos tipos y creamos dos nuevas columnas
- Substituimos los nulos

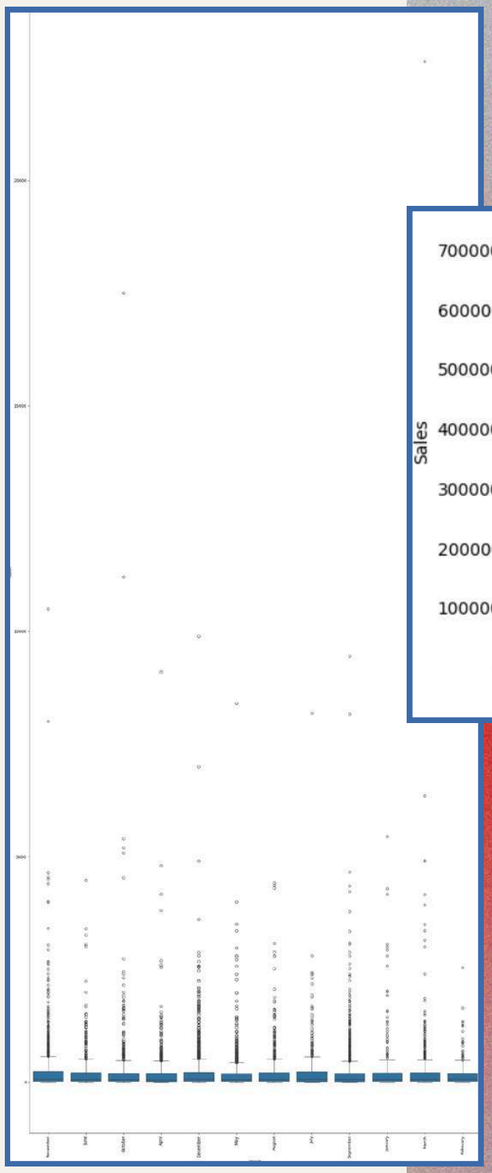
IT ACADEMY

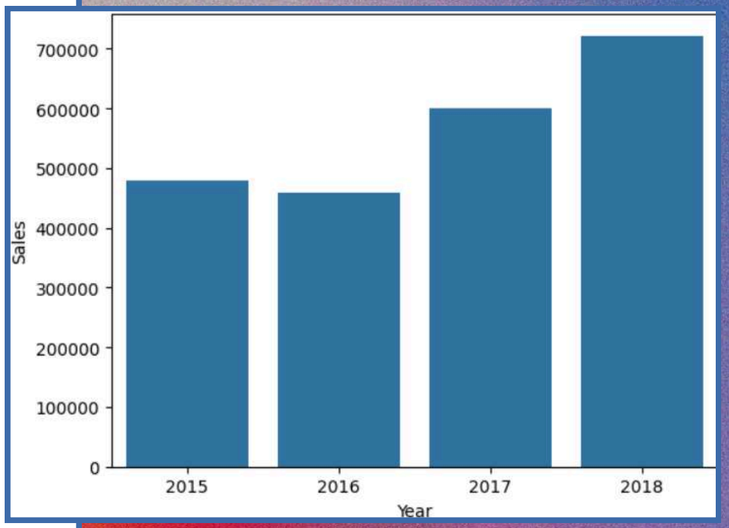
Proyecto Final **Sara Gutierrez**

EXPLORATORY DATA ANALYST

- Comprobamos si había patrones en los outliers.
- Decidimos trabajar con ellos porque son ventas reales.







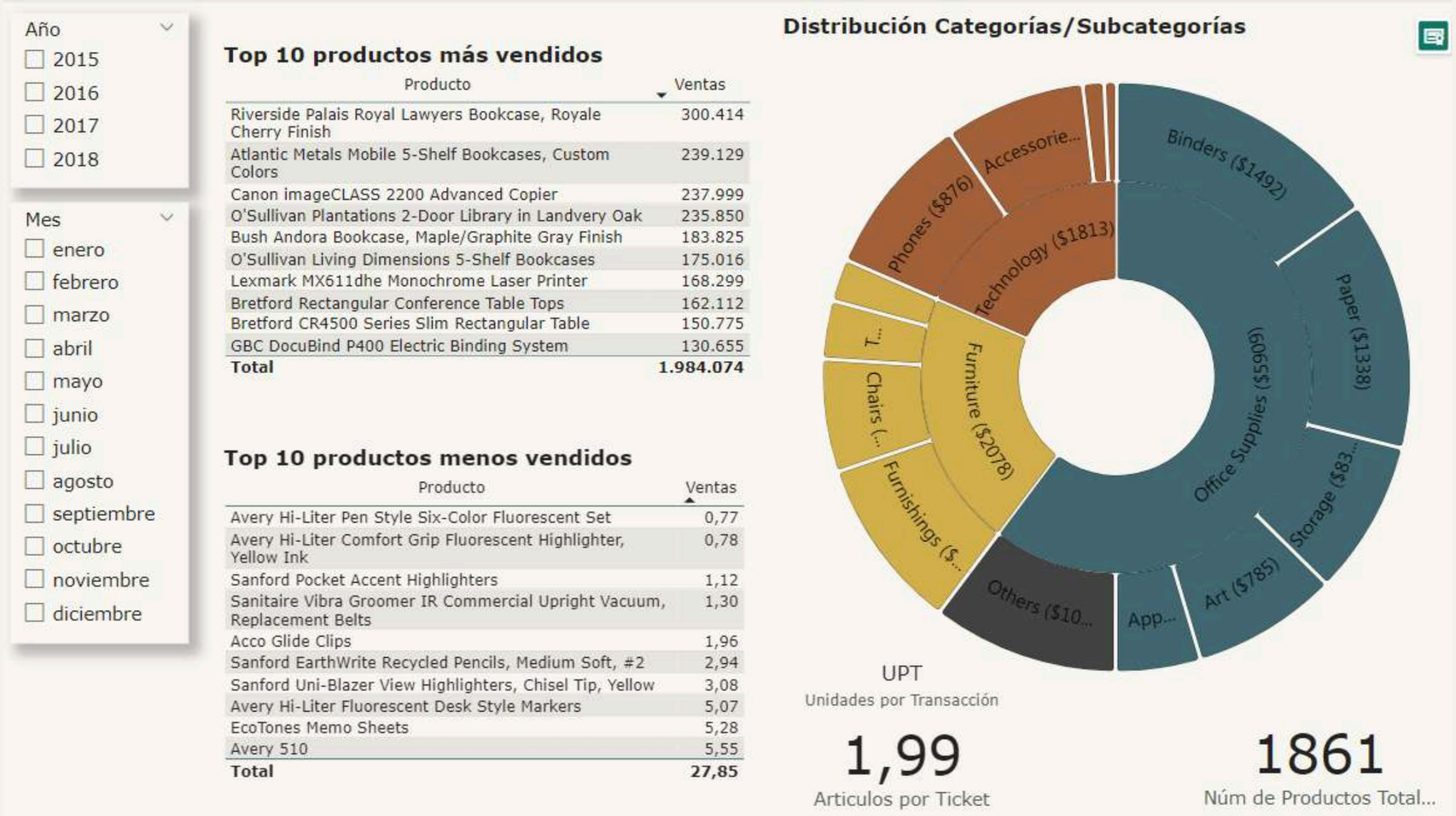
IT ACADEMY

Proyecto Final **Sara Gutierrez**

Resultados

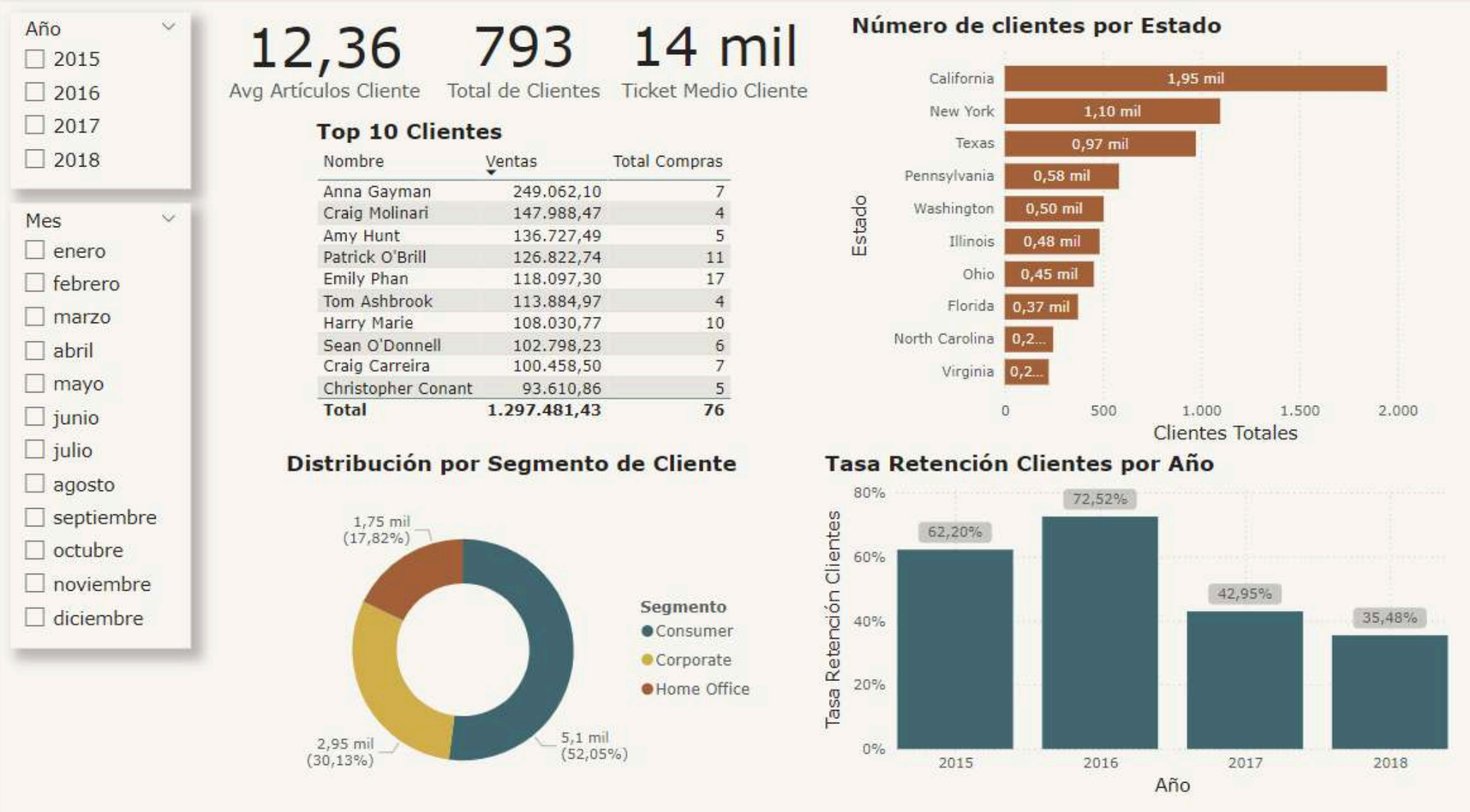
[Dashboards Exploratorio Ventas]

Resultados

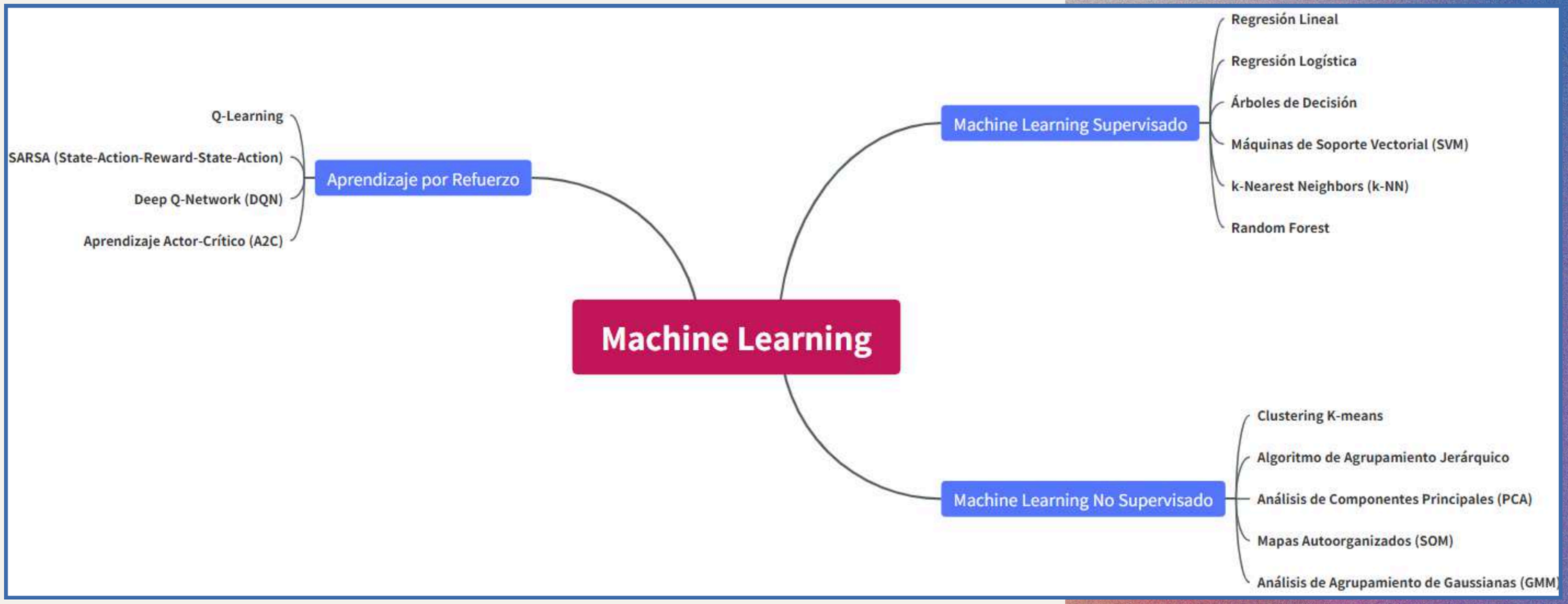


[Dashboard Exploratorio Productos]

Resultados



[Dashboards Exploratorio Consumidores]



IT ACADEMY

Proyecto Final **Sara Gutierrez**

Machine Learning

FAILS

```
X = sales_predicciones[['Month', 'Year']] #nuestras variables
y = sales_predicciones['Sales'] #la predicción
```

- Resultados excesivamente dispares

```
   Year         Sales
0  2015  354108.5921
1  2016  372897.2024
2  2017  448276.5274
3  2018  552617.0552
   Year  Predicted Sales
0  2019      2022.395636
```

```
X = copia_sales[['Ship Mode', 'Segment', 'Category', 'Region', 'Year', 'Month']]
y = copia_sales['Sales'] #Nuestra predicción
```

```
Error medio cuadrático (MSE): 66145.71688879425
Coeficiente de determinación (R²): 0.13297240534610433
```

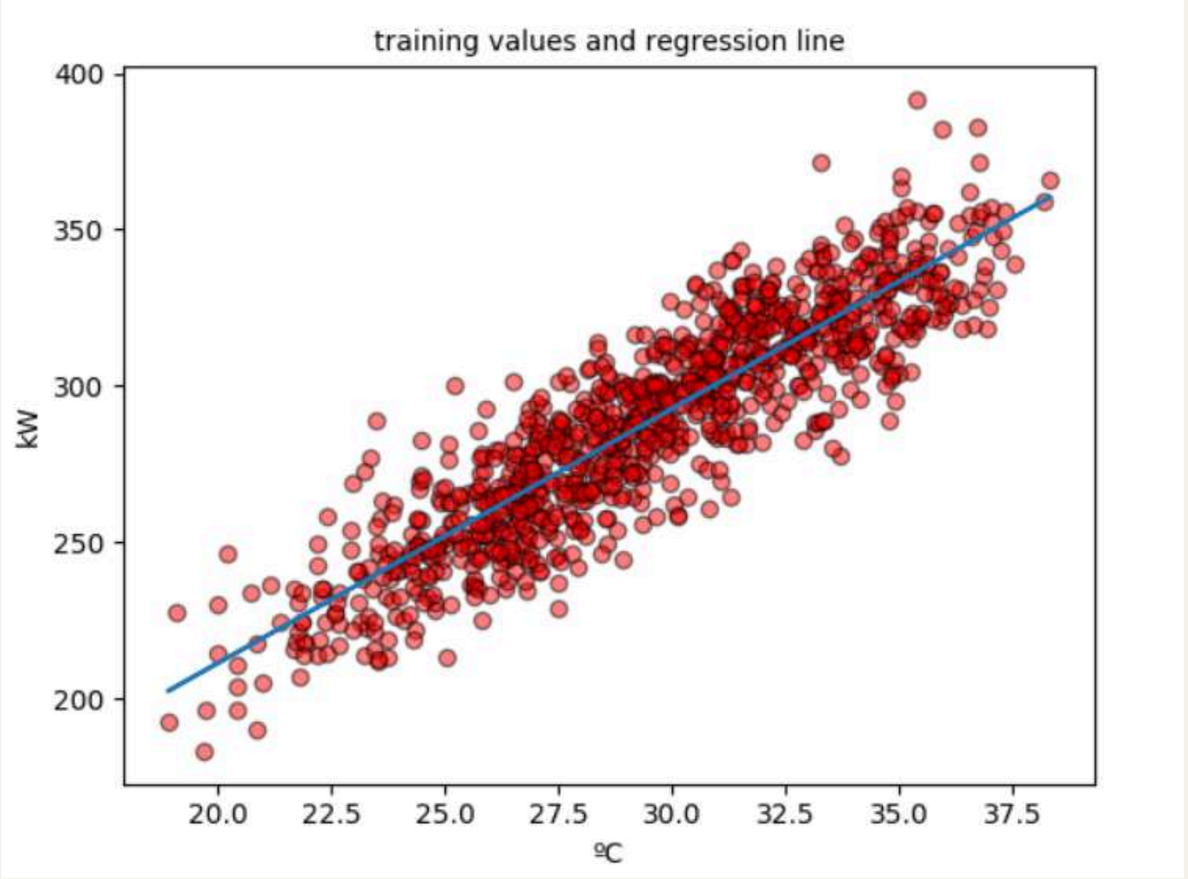
- Métricas que reflejan un rendimiento muy bajo

IT ACADEMY

Proyecto Final **Sara Gutierrez**

Regresión Lineal

¿Por qué no fue posible?

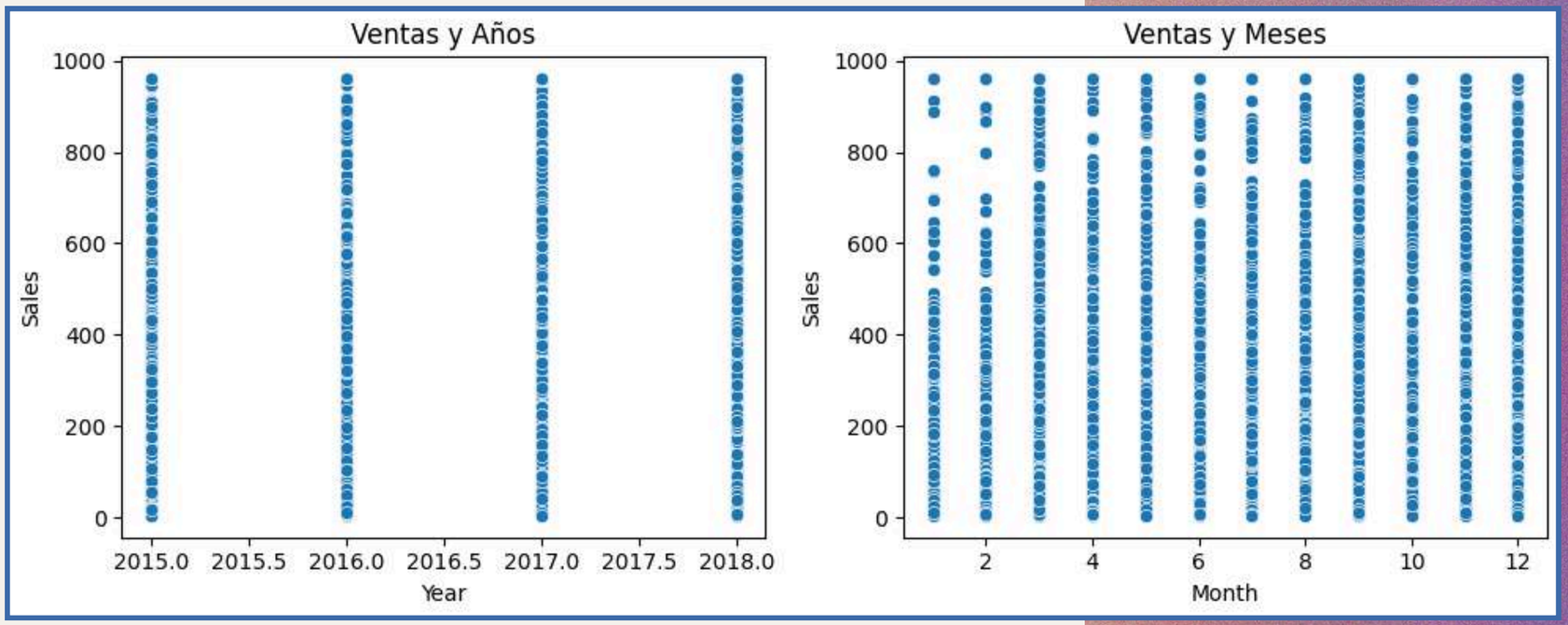


IT ACADEMY

Proyecto Final **Sara Gutierrez**

Regresión Lineal

No hay relación lineal y menos con variables categóricas



IT ACADEMY

Proyecto Final **Sara Gutierrez**

Regresión Logística

Predecir categorías

RESULTADOS

```
X = copia_sales[['Sales', 'Category']] #Variables que van a predecir
y = copia_sales['Ship Mode'] #Variable a predecir
```

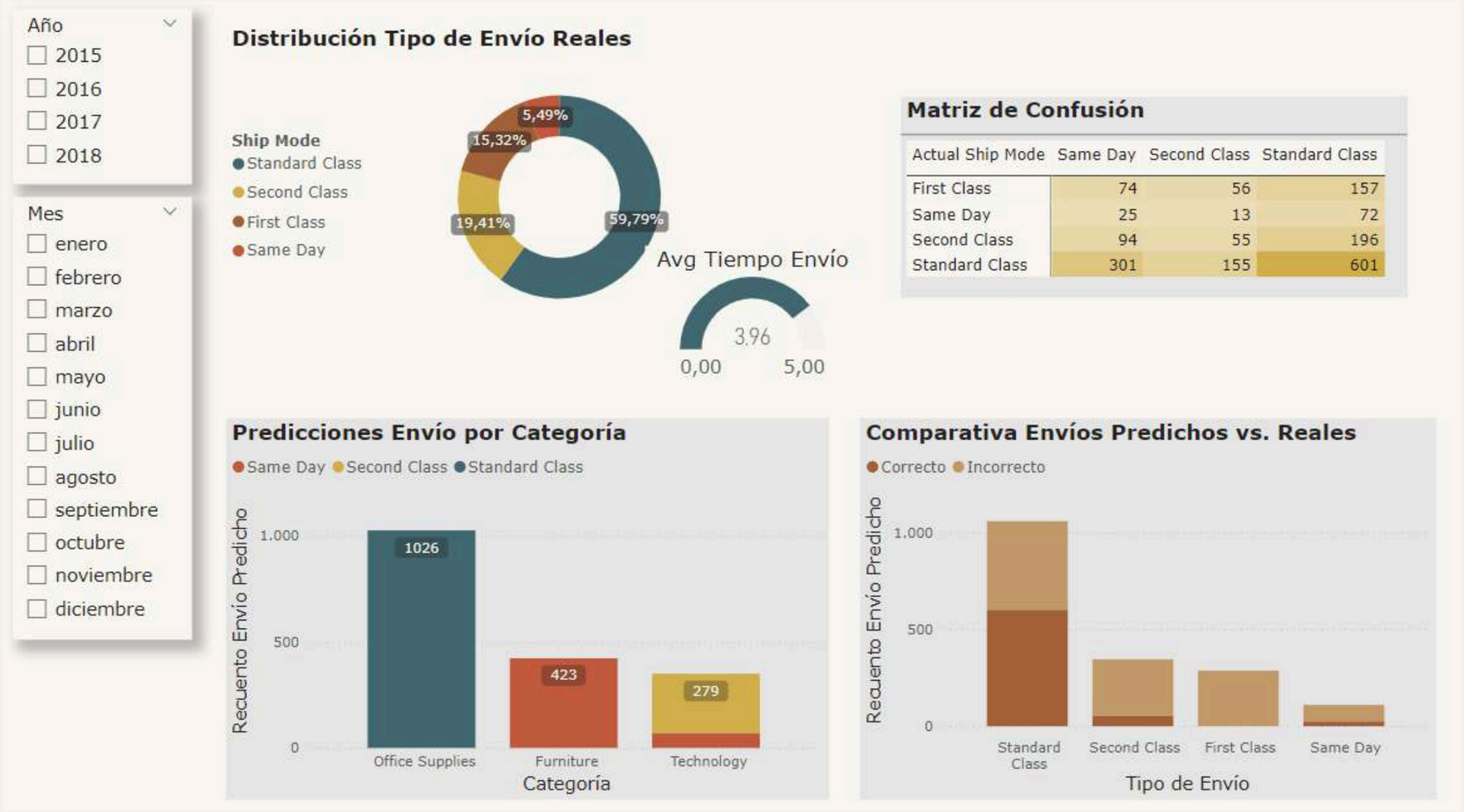
```
Precisión del modelo: 0.37854363535297386
Informe de clasificación:
                precision    recall  f1-score   support

   First Class       0.00      0.00      0.00       287
      Same Day       0.05      0.23      0.08       110
  Second Class       0.20      0.16      0.18       345
Standard Class       0.59      0.57      0.58      1057
```

- Precisión del 37%
- Predice 3 de los 4 envíos

IT ACADEMY

Proyecto Final **Sara Gutierrez**

Resultados Modelo Predictivo



[Dashboards Exploratorio Predicciones]

Conclusiones

de Machine Learning

- **Hay que estudiar estadística.**
- El procesamiento de datos y exploración es el 90% del trabajo.
- Control de la frustración: Ensayo y Error

IT ACADEMY

Proyecto Final **Sara Gutierrez**

Muchas Gracias

GITHUB
LINKEDING