

Data classification

En esta práctica se usó la columna 'Sales'.

Además, el parámetro 'Category' se transformó a valores numéricos para poder hacer el modelo y se utilizaron media de 0 y desviación estándar 1, lo que es importante para KNN.

Se creó un clasificador KNN con 5 vecinos y lo entrenamos y se calculó la precisión, reporte de clasificación.

```
[Running] python -u "e:\AAAAAAAAAAAAAAAAAAAAAAAAAAAA\UNIVERS
Precision del modelo: 0.6479591836734694
```

Reporte de clasificacion:					
		precision	recall	f1-score	support
	Furniture	0.44	0.44	0.44	629
Office Supplies		0.75	0.81	0.78	1760
	Technology	0.51	0.38	0.43	551
	accuracy			0.65	2940
	macro avg	0.57	0.54	0.55	2940
	weighted avg	0.64	0.65	0.64	2940

A partir del reporte de clasificación se puede deducir que **"Office Supplies"** es la clase con mejor rendimiento, con una precisión del **75%**, un recall del **81%** y un F1-score del **78%**. Esto indica que el modelo identifica bien esta categoría ya que tiene el mayor soporte (1760 instancias) y el mejor rendimiento del modelo (F1-score: **0.78**). Esto sugiere que es la categoría más frecuente en los datos, posiblemente porque es la más vendida o la más registrada y que el modelo la identifica bien, lo que podría indicar que sus patrones de venta son más predecibles.

"Furniture" y **"Technology"** tienen un rendimiento inferior, con F1-scores de **0.44** y **0.43** respectivamente. En particular, **"Technology"** tiene una precisión moderada (51%) pero un recall bajo (38%), lo que significa que el modelo no logra identificar muchos casos reales de esta clase y esto podría deberse a un menor volumen de ventas (apoyado por su bajo "support": 629 y 551 respectivamente).