



CODING DOJO

DATA SCIENTIST & MACHINE LEARNING AT PYTHON

Rafael Johannes Proaño Rodríguez

Ing. Software_UDLA

ANÁLISIS DE DATOS

- Presentación de dataset: “Predicciones de ventas”

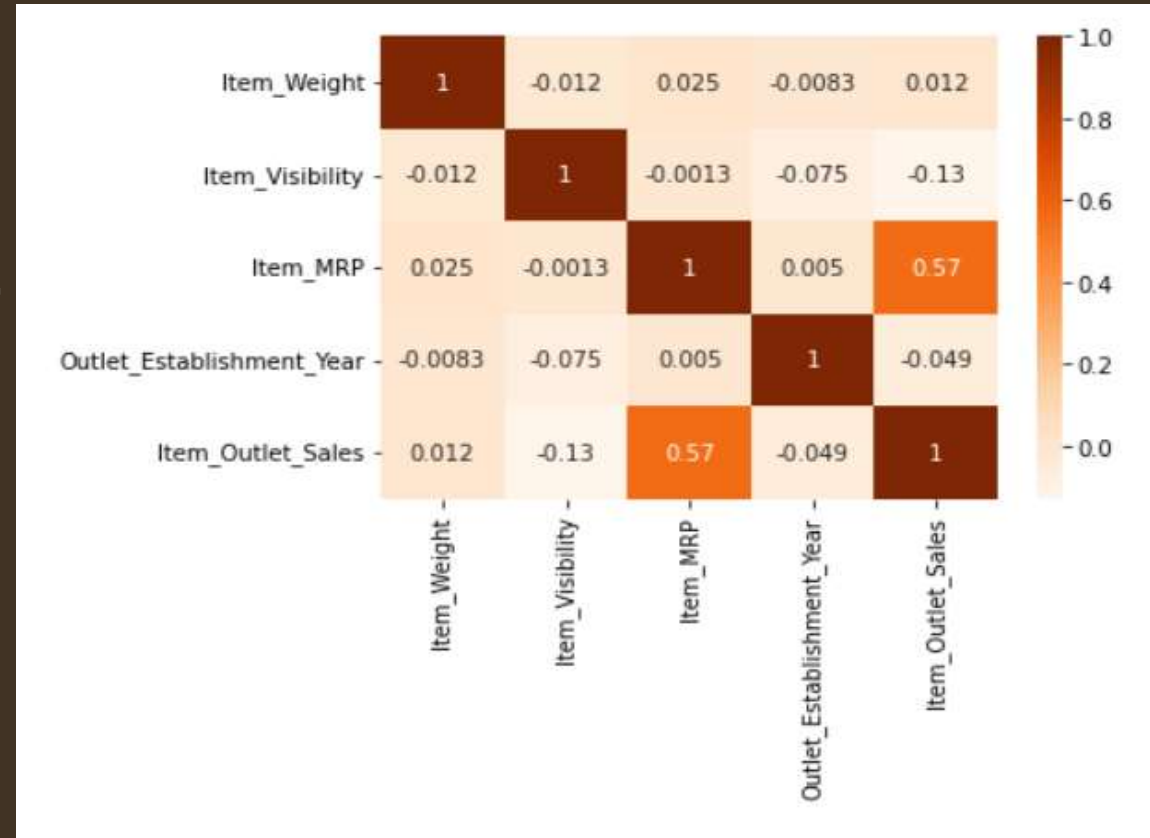
	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 1
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	Tier 1
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	Tier 1

- 8523 rows & 12 columns

https://docs.google.com/spreadsheets/d/1L7DUb__bCuJziKviXRQIUur2TXb56khysBaN9oTZgBs/edit#gid=431704500

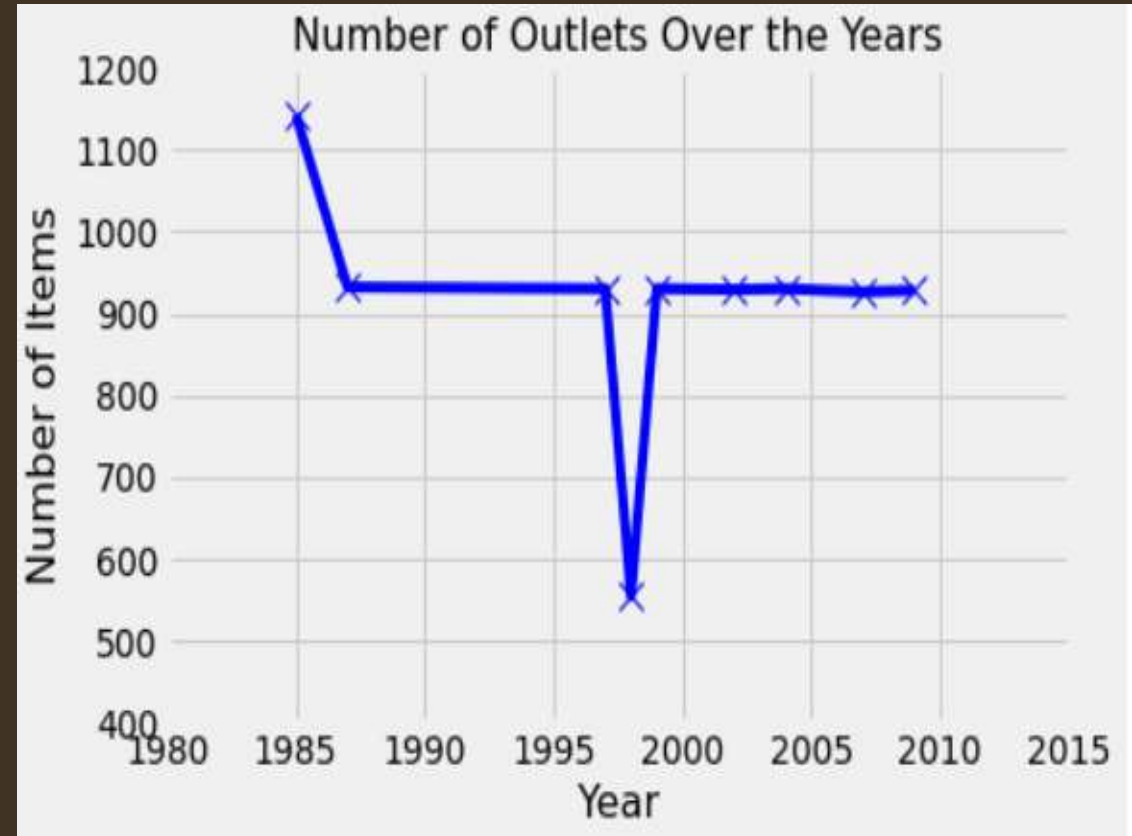
ANÁLISIS DE DATOS

- **Presentación de informes**
- Heatmap. - las variables "Item_MRP" "Item_Outlet_Sales" se correlacionan positivamente, por lo que podremos obtener una satisfactoria predicción de ventas a través de esta ventana, canal.



ANÁLISIS DE DATOS

- **Presentación de informes**
- Scatter plot. – La intersección de las variables “Outlet_Establishment_Year” e “Item_Identifier” refleja la precisión de ventas con las que se puede contrastar.



ANÁLISIS DE DATOS

- **Recomendaciones**

- El modelo de Árbol de regression. – Describe datos aproximados a la realidad: conjunto de prueba, evaluando en R^2 .

