***Big mart sales prediction***

***Content:***

Big Mart is a big supermarket, with stores all over the country. I'm trying to help the management of Big Mart by building a predictive model to find out the sales of each product at a particular store and the sales of the different stores of Big Mart.

The data scientists at Big Mart have collected 2013 sales data for 1559 products with attributes such as: (Weight of product, fat content, etc.) across 10 stores with attributes such as (store size, the year established,etc) in different cities

This dataset was taken from Kaggle website ([BigMart Sales Data | Kaggle](https://www.kaggle.com/brijbhushannanda1979/bigmart-sales-data)).

***Data description:***

Big mart dataset contains 8523 entries and 12 features in **train** dataset and 5681 in test dataset.

|  |  |
| --- | --- |
| **variables** | **Description** |
| Item\_Identifier | Unique product ID |
| Item\_Weight | Weight of product |
| Item\_Fat\_Content | Whether the product is low fat or not |
| Item\_Visibility | The % of total display area of all products in a store allocated to the particular product |
| Item\_Type | The category to which the product belongs |
| Item\_MRP | Maximum Retail Price (list price) of the product |
| Outlet\_Identifier | Unique store ID |
| Outlet\_Establishment\_Year | The year in which store was established |
| Outlet\_Size | The size of the store in terms of ground area covered |
| Outlet\_Location\_Type | The type of city in which the store is located |
| Outlet\_Type | Whether the outlet is just a grocery store or some sort of supermarket |
| Item\_Outlet\_Sales | sales of the product in a particular store. This is the outcome variable to be predicted. |

***Objective:***

* Analyze the data available on Big Mart.
* Explore various visualization techniques like charts, graphs.

***Questions:***

From this data I will find out:

* Which product have higher sales the low fat or regular fat?
* Is the display area of products effects on the sales?
* Which product type have higher sales?
* Which store have higher sales than others?

***Model:***

1. Predict the sales of the product in a particular store.
2. Predict the sales of the different stores of Big Mart.

***Tools:***

* **Environment:** Jupyter notebook.
* **Programming Language:** python.
* **Libraries:** numpy, pandas, matplotlib and seaborn.

***MVP:***

* Import needed libraries.
* Load the Big Mart dataset.
* Data preparation.
* Exploratory Data Analysis (EDA).
* Model training and selection.