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 predict the sales per product for each product at a particular store

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).

Data description:

Big mart dataset contains 8523 entries and 12 features.

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 people bought more low fat or regular fat?
- Which product have more profitable the low fat or regular fat?
- What kind of product people bought more?
- What product have most profitable?
- What is the most outlets size of big mart in our data?
- What is the most profitable outlet size?
- Which type of city (tire1, tire2, tire3) have most profitable?
- Dose the display area of products effects on the sales?
- Dose items with higher MRP (Maximum Retail Price) sold more?
- Which outlet have the top sales?

Model:

1- Predict the sales of the product in a particular store.

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.