Big Mart Sales Prediction

Abstract:

Big Mart is a large retail chain that operates multiple stores across different locations. The data scientists at Big Mart have collected 2013 sales data for various products across 10 stores in different cities. The aim is to build a predictive model and find out the sales of each product at a particular store. They want to leverage this data to build a predictive model that can accurately predict the sales of their products in order to optimize their inventory management and improve overall profitability.

Problem Statement:

The goal of this project is to develop a machine learning model that can predict the sales of each product at different stores based on historical sales data and other relevant features. You are tasked with developing a machine learning model that takes into account historical sales data, product attributes, store information and any other relevant features available in the dataset. The model should be able to accurately predict the sales of each product at different stores.

Variable Description:

Column	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.

Scope:

- Understand data by performing exploratory data analysis
- Training and building regression algorithms to predict the sales of a product in a particular store
- Understand feature importance and improve the model
- Understand various model performance metrics and measure the performance of each model