**Sales Prediction for Big Mart Outlets:**

The data scientists at BigMart have collected 2013 sales data for 1559 products across 10 stores in different cities. Also, certain attributes of each product and store have been defined. The aim is to build a predictive model and predict the sales of each product at a particular outlet.

Using this model, BigMart will try to understand the properties of products and outlets which play a key role in increasing sales.

Please note that the data may have missing values as some stores might not report all the data due to technical glitches. Hence, it will be required to treat them accordingly.

**About the Dataset:**

We have a train (8523) data set, the train data set has both input and output variable(s).

**Variable 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 the 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 the 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 t particular store. This is the outcome variable to be predicted.

**structured workflow:**

Step 1: Understanding problem and data

**• Objective**: Predict sales of each product at a particular outlet.

• Input and the target data.

* Understand the features (Item\_Identifier, Outlet\_Identifier, etc.).
* Check for categorical vs. numerical columns.

Step 2: Data Cleaning

• Check for Missing Values.

• Fix Data Quality Issues

• Remove Outliers

### **Step 3: Exploratory Data Analysis (EDA)**

• Visualize Target Distribution

• Correlation Analysis.

• Categorical Data Analysis

• Understand Relationships

Step 4: Feature Engineering

• Encoding Categorical Variables

• Create New Features

• Normalize/Scale Numerical Features

### **Step 5: Data Splitting and model building**

• Split the data into training and test sets

• Model Selection

Step 6: Hyperparameter Tuning

• Optimize model parameters.