Introduction to EDA

Problem Statement: BigMart

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 find out the sales of each product at a particular store. Using this model, BigMart will try to understand the properties of products and stores 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.

Dataset Link: [BigMart dataset](#)

Variable: Description

- **Item_Identifier**: Unique product ID
- **Item_Weight**: Weight of the 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 the particular store. This is the outcome variable to be predicted.