**BigMart Sales prediction**

**Project member: Taufique Ahmed**

**Problem Statement**:

**Nowadays, shopping malls and Big Marts keep track of individual item sales data in**

**order to forecast future client demand and adjust inventory management. In a data**

**warehouse, these data stores hold a significant amount of consumer information and**

**particular item details. By mining the data store from the data warehouse, more**

**anomalies and common patterns can be discovered.**

**Approach:**

**The classical machine learning tasks like Data Exploration, Data Cleaning,**

**Feature Engineering, Model Building and Model Testing. Try out different machine**

**learning algorithms that’s best fit for the above case.**

**Description:**

**1. Item\_Identifier: Unique product ID**

**2. Item\_Weight: Weight of product**

**3. Item\_Fat\_Content: Whether the product is low fat or not**

**4. Item\_Visibility: The % of total display area of all products in a store allocated to the particular product**

**5. tem\_Type: The category to which the product belongs**

**6. Item\_MRP: Maximum Retail Price (list price) of the product**

**7. Outlet\_Identifier: Unique store ID**

**8. Outlet\_Establishment\_Year: The year in which store was established**

**9. Outlet\_Size: The size of the store in terms of ground area covered**

**10. Outlet\_Location\_Type: The type of city in which the store is located**

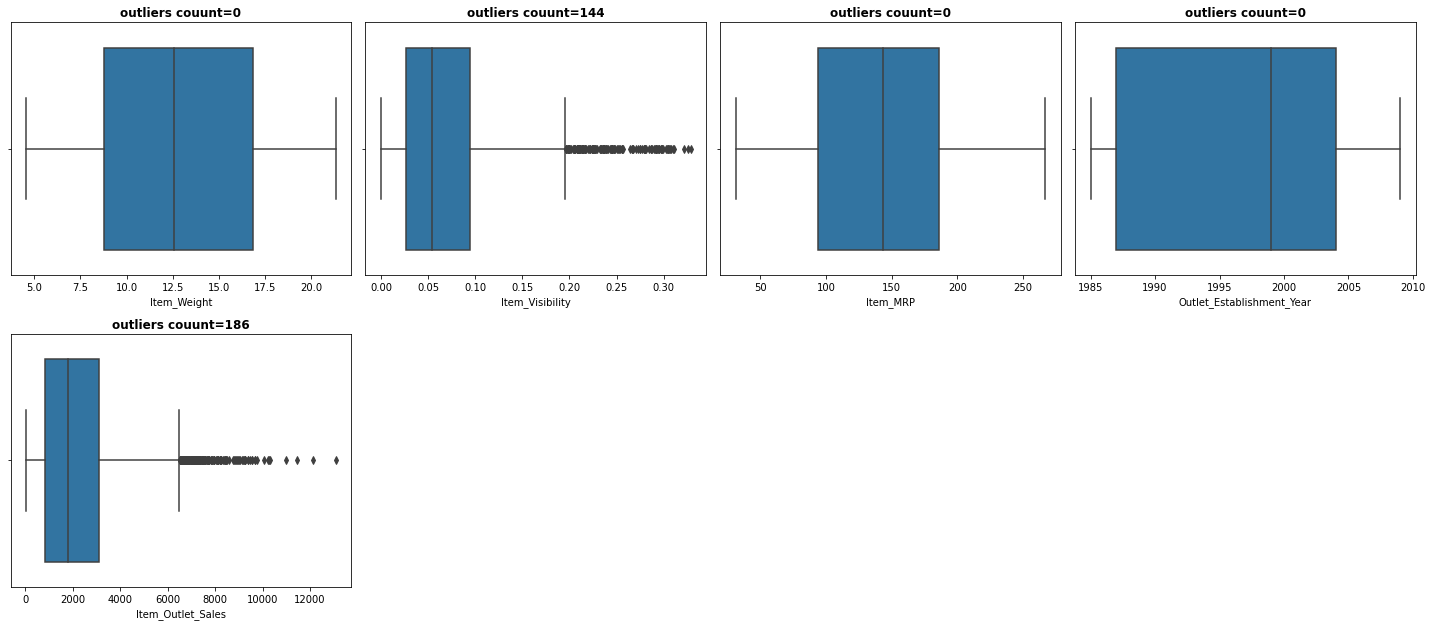
**11. Outlet\_Type: Whether the outlet is just a grocery store or some sort of supermarket**

**12. Item\_Outlet\_Sales: Sales of the product in the particular store. This is the outcome variable to be predict.**

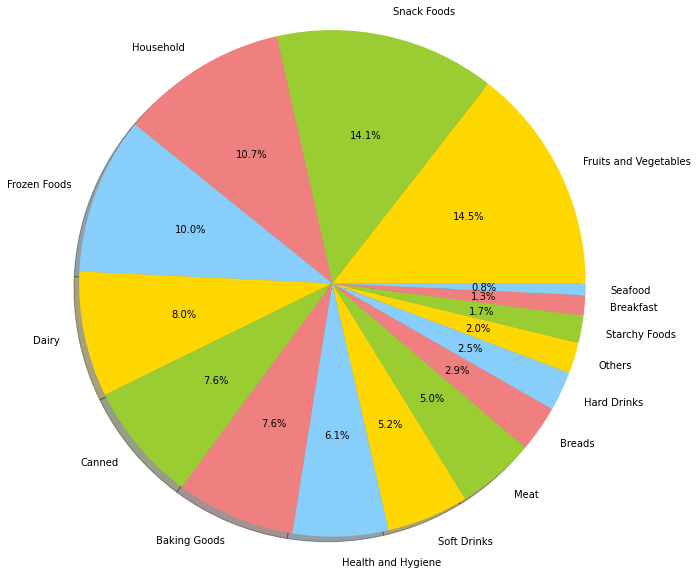
**Flow of Project:**

**Description:**

1. **In this project Mart are need to predict their sales on based on independent variable which is mentioned above. For this I have to first import train data and observed what kind of variable are and any pre-processing are required or not.**
2. **For these 2 variables are have missing's values than I had replace with their mean values. And 3 variables are having outliers than I had to deal with it.**



1. **After all of this I had perform EDA to extract insightful data. For this I have perform univariate and bivariate distribution with the help of plots.**

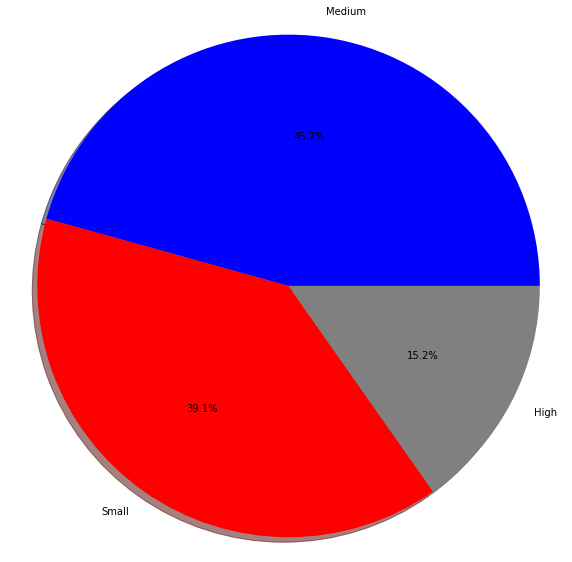


**Conclusion:**

**1. Low fat product visible more than any product. approximate 59.7% visibility.**

**2. 2nd highly visible product is kind of regular**

**3. and remaining 3 reg, LF, low fat is 1.3%, 1.4% & 3.7% visible respectively.**



**Conclusion:**

**1. total item type are 16 out these 4 categories are cover almost 50% of the total product**

**2. and remaining 50% are covered by 12 categories.**

1. **After EDA we need to go Feature Encoding. In that we introduce dummies on categorical variable.**
2. **Then we split data into train and test format. And apply regression machine learning models and try to predict.**
3. **And choose best model and proceed with to unseen data.**