**PBD Project Proposal**

**Group Members:**

Muhammad Ali Shakoor L1S17BSCS0082

Syed Osja Hussain L1S17BSCS0109

**Section:**

G

**Project Title:**

BigMart Sales Data Regression

**Kaggle Link(Project):**

<https://www.kaggle.com/brijbhushannanda1979/bigmart-sales-data>

**Github Link(Project):**

<https://github.com/alishakoor/BigMartSales/>

**AIM**

The motive of this Project is to train a Machine Learning algorithm that is capable of predicting the sales of the product in a particular store using python.

**DATASET**

The dataset has been acquired in kaggle which consists of 2 csv files (Train.csv, Text.csv). We have Train (8523) and Test (5681) data set, Train data set has both input and output variable(s). We need to predict the sales for the Test data set.

* **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 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.