

# Capstone Project Weekly Progress Report

Project Title	Big Mart Data Visualization and Analysis
Group Name	Group D
Student names/Student IDs	Avik Kundal(744823),Jasmeet Kaur(744215),Kirandeep Kaur(742276),Savreet Kaur(742785),Sukhjinder Singh(743143)
Reporting Week	30 sept 2019 to 6 oct 2019
Faculty Supervisor	William Pourmajidi

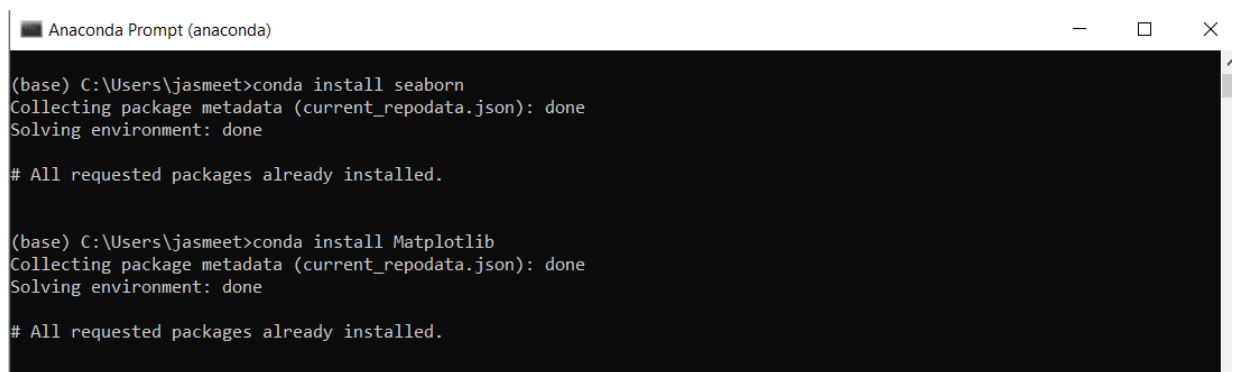
## 1. Tasks Outlined in Previous Weekly Progress Report

To know the data types present in the Big\_Mart Data Set and to differentiate the categorical and numerical features.

## 2. Progress Made in Reporting Week

Libraries used for Visualizing the Big\_Mart Data Set

- **Seaborn**
- **Matplotlib**



```

Anaconda Prompt (anaconda)

(base) C:\Users\jasmeet>conda install seaborn
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

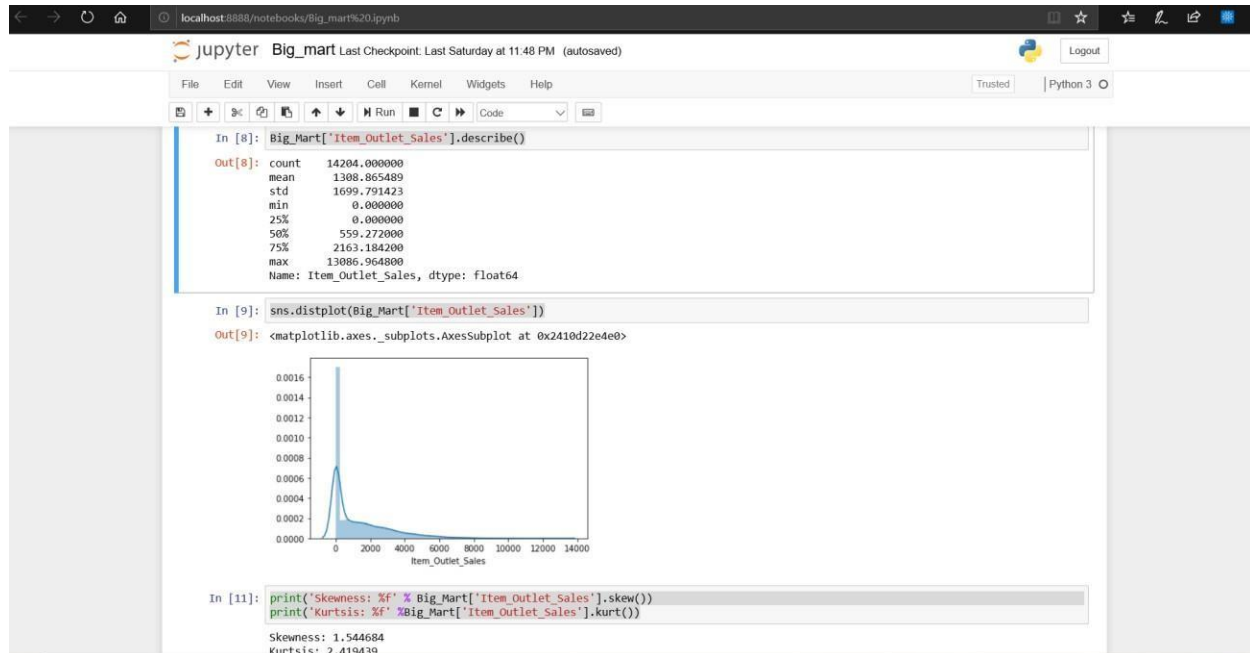
(base) C:\Users\jasmeet>conda install Matplotlib
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

```

Function used in order to get proper plot

- **Sns.distplot()**-function is used to get plot of Item\_Outlet\_Sales using Matplotlib



**Big\_Mart.datatypes**-Get the Data Types of Variables present in the Big\_Mart data set

### #Filter categorical variables

- `categorical_features = Big_Mart.select_dtypes(include=[np.object])`

### #Filter numerical variables

- `numerical_features = Big_Mart.select_dtypes(include=[np.number])`

```

In [12]: Big_Mart.dtypes
Out[12]: Item_Identifier      object
Item_Weight      float64
Item_Fat_Content  object
Item_Visibility   float64
Item_Type        object
Item_MRP         float64
Outlet_Identifier  object
Outlet_Establishment_Year  int64
Outlet_Size      object
Outlet_Location_Type  object
Outlet_Type      object
Item_Outlet_Sales  float64
source          object
dtype: object

In [13]: categorical_features = Big_Mart.select_dtypes(include=[np.object])
categorical_features.head(2)
Out[13]:
  Item_Identifier Item_Fat_Content Item_Type Outlet_Identifier Outlet_Size Outlet_Location_Type Outlet_Type source
0      FDA15      Low Fat    Dairy      OUT049      Medium      Tier 1 Supermarket Type1      train
1      DRC01      Regular  Soft Drinks      OUT018      Medium      Tier 3 Supermarket Type2      train

In [14]: numerical_features = Big_Mart.select_dtypes(include=[np.number])
numerical_features.head(2)
Out[14]:
  Item_Weight Item_Visibility Item_MRP Outlet_Establishment_Year Item_Outlet_Sales
0         9.30      0.016047    249.8092              1999          3735.1380
1         5.92      0.019278    48.2692              2009          443.4228

```

**Big\_Mart['Outlet\_Establishment\_Year'].value\_counts()-** to count the number of outlets of different supermarkets that were established in particular year

```

In [16]: Big_Mart['Outlet_Establishment_Year'].value_counts()
Out[16]: 1985    2439
1987    1553
1999    1550
1997    1550
2004    1550
2002    1548
2009    1546
2007    1543
1998     925
Name: Outlet_Establishment_Year, dtype: int64

```

### 3. Difficulties Encountered in Reporting Week

Choose proper function to visualize the output\_sales using seaborn and to differentiate the categorical and numerical variables in Big\_Mart data set.

### 4. Tasks to Be Completed in Next Week

**Data Wrangling :** Work on missing values or null values present in the data set.

**Tableau :** Learning how to work with tableau dashboard.