

Capstone Project Weekly Progress Report

Project Title	Market Based Data Visualization and Analysis		
Troject Title	Trial net based base visualization and ranarysis		
Group Name	Group D		
Student	Avik Kundal (744823), Jasmeet Kaur(744215), Kirandeep Kaur(742276), Savreet		
names/Student IDs	Kaur(742785),Sukhjinder Singh(743143)		
Reporting Week	oorting Week 16 Sept 2019 to 22 Sept 2019		
Faculty Supervisor	William Pourmajidi		

- 1. **Tasks Outlined in Previous Weekly Progress Report** (Provide detailed information on the tasks to be completed in this week)
 - We selected Big Mart Dataset from Kaggle for the purpose of analyzing patterns
 related to product sales and discovering hidden insights between products
 prevailing in Supermarkets based upon their fat content and maximum retail
 price, we selected a data set and modified it according to our requirements and
 added more data specific to the outlet sales and type of outlet.
 - We studied the Data set to know about what details they provided in different fields.
- 2. **Progress Made in Reporting Week** (Provide detailed information on the progress that you made in the reporting week. Limit your write-up to no more than two page)

The initial stage of our project of Data Gathering is accomplished in this reporting week.

- The Big Mart dataset has over 8000 rows and 14 columns.
- It a collection of data about approximately 1550 products across 10 stores in different cities of Canada.
- The data set includes columns represented below
 - Item Fat Content
 - Item Identifier
 - Item Type
 - Item Weight
 - Outlet Establishment Year
 - Outlet Size
 - Outlet location type

- Outlet Type
- City Name
- Population
- Item MRP
- Item visibility
- Item Outlet Sales.
- Outlet Location Type



Field	Field Name		Remote Field Name
#	ltem Weight ▼	Sheet1	Item_Weight
Abc	Item Fat Content	Sheet1	Item_Fat_Content
#	Item Visibility	Sheet1	Item_Visibility
Abc	Item Type	Sheet1	Item_Type
#	Item MRP	Sheet1	Item_MRP
#	Outlet Identifier	Sheet1	Outlet_Identifier
#	Outlet Establishment Year	Sheet1	Outlet_Establishment_Year
Abc	Outlet Size	Sheet1	Outlet_Size
Abc	Outlet Location Type	Sheet1	Outlet_Location_Type
Abc	Outlet Type	Sheet1	Outlet_Type
#	Item Outlet Sales	Sheet1	Item_Outlet_Sales
•	City	Sheet1	City
#	Population	Sheet1	Population

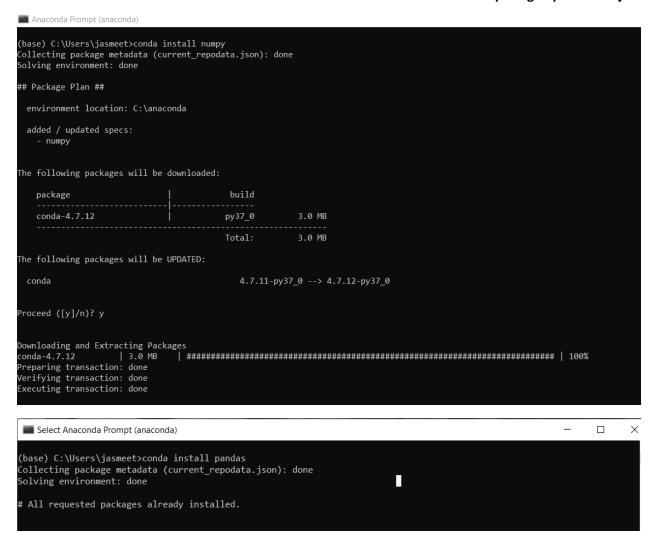
TOOLS USED

- The jupyter notebook is installed and used for imputing values and data cleaning
- Why data cleaning?
- Data might not be in its original form it may contain missing values, Incorrect categories(Low Fat, LF, low fat, regular)

DATA ANALYSIS LIBRARIES

- Pandas
- Numpy





3. **Difficulties Encountered in Reporting Week** (Provide detailed information on the difficulties and issues that you encountered in the reporting week. Limit your write-up to no more than one page)

To use Pandas library to store data set in data frame using function pd.read_csv().

To choose proper features and dimensions which are important for the retail stores to observe the hidden relationships related to their market sales and to observe the customers' behaviors about the products.

4. **Tasks to Be Completed in Next Week** (Outline the tasks to be completed in the following week)

To choose proper functions in Pandas library to know about the null values from our data set and to clean these null values.



