Capstone Project Weekly Progress Report

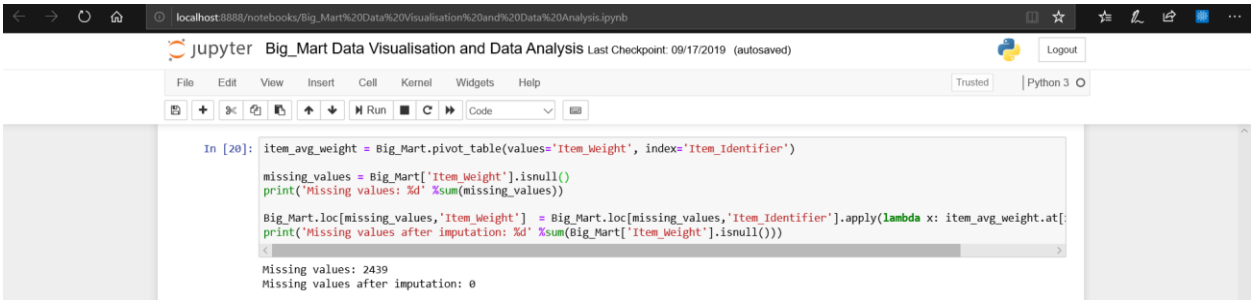
Project Title	Big_Mart Data Visualization and Analysis
Group Name	Group D
Student names/Student IDs	Avik Kundal (744823), Jasmeet Kaur (744215), Kirandeep Kaur (742276), Savreet Kaur (742785), Sukhjinder Singh (743143)
Reporting Week	21 Oct 2019 to 27 Oct 2019
Faculty Supervisor	William Pourmajidi

1. Tasks Outlined in Previous Weekly Progress Report

- **To work on handling missing values in Big_Mart Data set**
- **To impute the null values with their Average value such as in case of Item_weight and to use the mode function in case of Item_Outlet size**

2. Progress Made in Reporting Week

- **#Handling the missing values of Item_weight Feature**
- **missing_values = Big_Mart['Item_Weight'].isnull()**
- **print('Missing values: %d' %sum(missing_values))**
- **Big_Mart.loc[missing_values,'Item_Weight'] = Big_Mart.loc[missing_values,'Item_Identifier'].apply(lambda x: item_avg_weight.at[x,'Item_Weight'])**
- **print('Missing values after imputation: %d' %sum(Big_Mart['Item_Weight'].isnull()))**



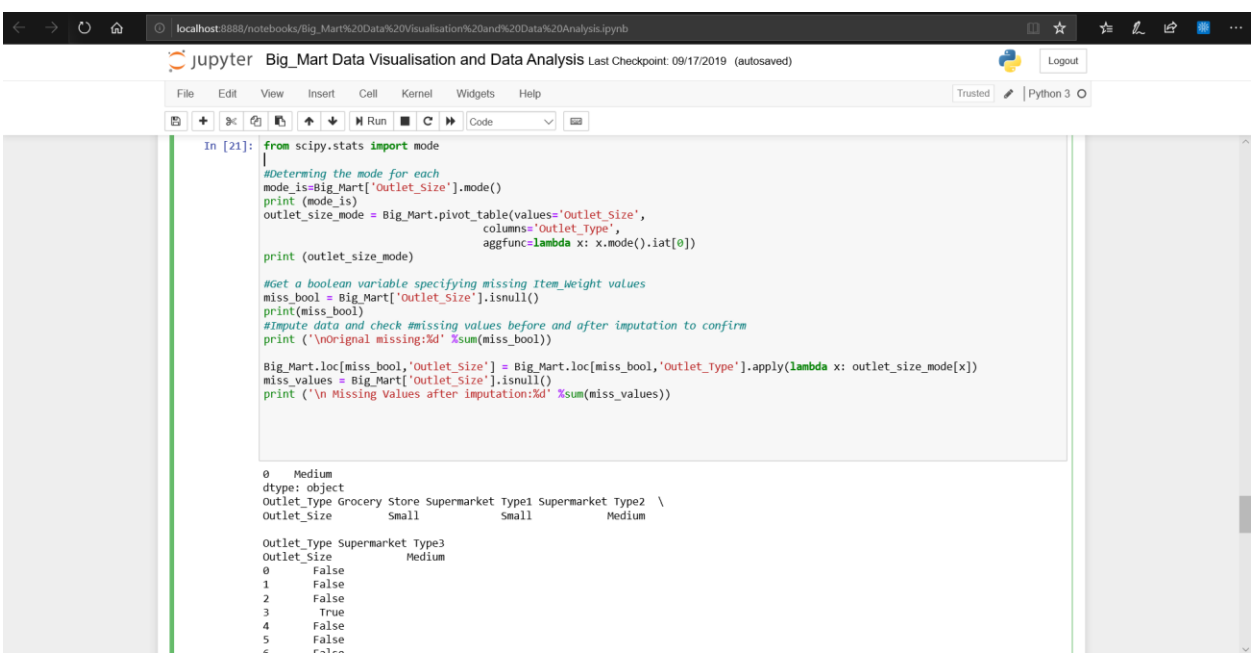
- **Imputing the missing values in Outlet_size Feature with their mode function**

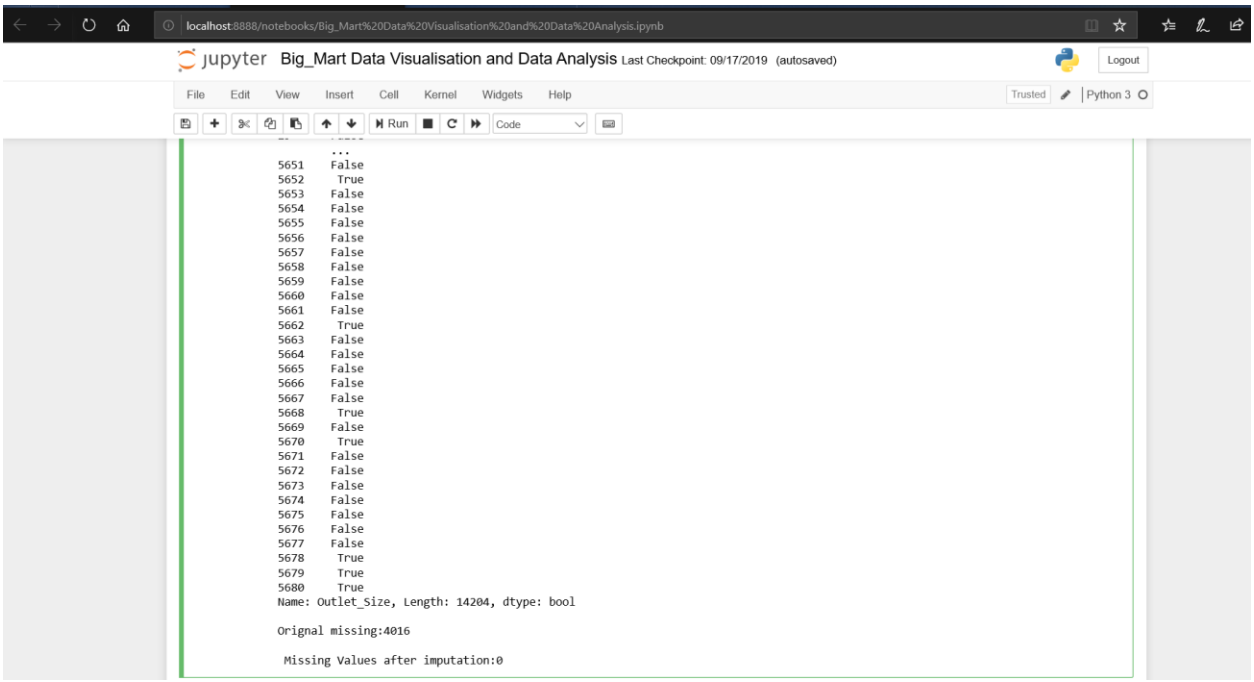
- #Determing the mode for each
- mode_is=Big_Mart['Outlet_Size'].mode()
- print (mode_is)
- outlet_size_mode = Big_Mart.pivot_table(values='Outlet_Size',

  columns='Outlet_Type',

  aggfunc=lambda x: x.mode().iat[0])

- print (outlet_size_mode)
- #Get a boolean variable specifying missing Item_Weight values
- miss_bool = Big_Mart['Outlet_Size'].isnull()
- print(miss_bool)
- #Impute data and check missing values before and after imputation to confirm
- print ('\nOrignal missing:%d' %sum(miss_bool))
- Big_Mart.loc[miss_bool,'Outlet_Size'] = Big_Mart.loc[miss_bool,'Outlet_Type'].apply(lambda x: outlet_size_mode[x])
- miss_values = Big_Mart['Outlet_Size'].isnull()
- print ('\n Missing Values after imputation:%d' %sum(miss_values))





3. **Difficulties Encountered in Reporting Week**
   - **To begin with handling missing values, we found difficulty in observing whether mean or mode function is good for imputing values into the field of null values.**
   - **For Numerical Features such as Item_weight ,we decided to use average of Item_weight and to put this value of Avg weight into the records of null values. It was difficult for us to use the lambda and apply fuction with the .loc fuction. The error we got while using the functions were like "expected an indented block" etc.**
   - **For Categorical variables such as Outlet_size, we choose the mode function (aggfunc=lambda x: x.mode().iat[0]))as suitable function to replace null values of Outlet_size.The pivot table and .iat[] was big hurdle in imputing the mode value.**

4. **Tasks to Be Completed in Next Week**
   - **To observe the output of handling missing values with the aid of heat map**
   - **Determine average visibility of a product**
   - **Replace Values with suitable name as different names were used to represent single item.**
   - **For instance, Low Fat and LF in dataset have same meaning. So, replace 'LF' with 'Low Fat' and 'reg' with 'Regular''.**