

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	28 Oct 2019 to 2 Nov 2019
Faculty Supervisor	William Pourmajidi

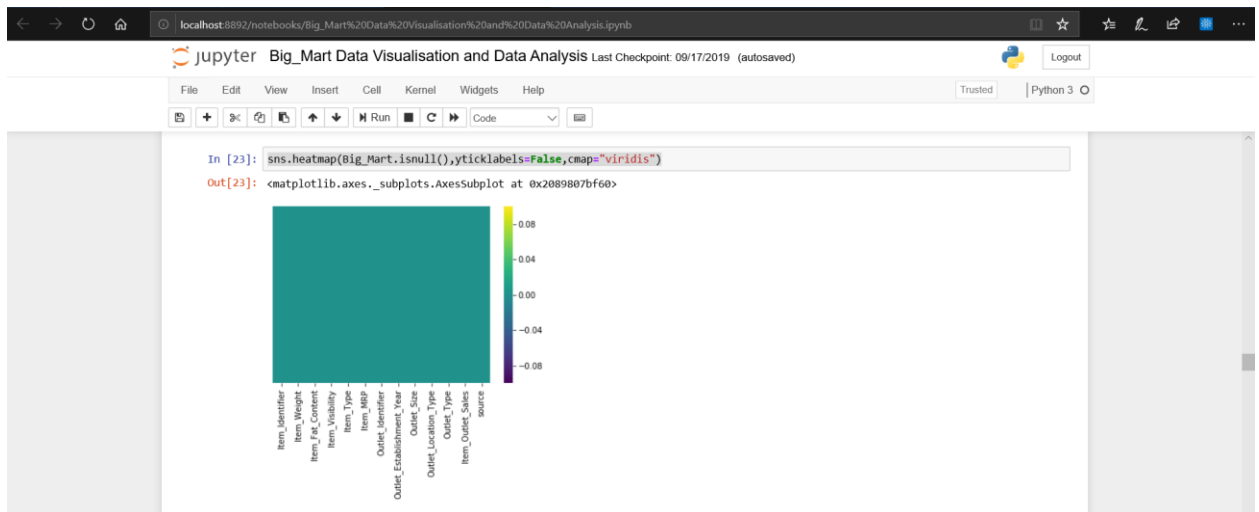
1. Tasks Outlined in Previous Weekly Progress Report

- To observe the output of imputing missing values with the aid of heat map
- To observe the correlation between dependent and independent features of Big_Mart Data Set by using heat map

2. Progress Made in Reporting Week

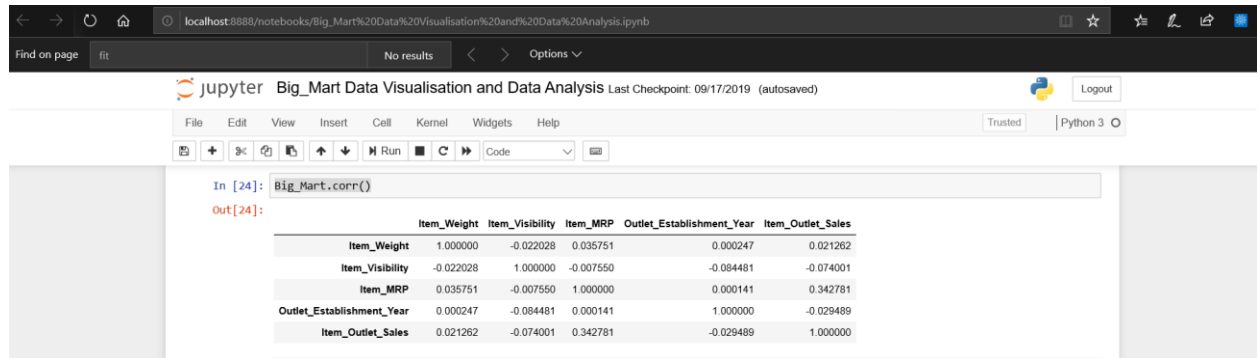
A heat map (or heatmap) is a graphical representation of data where the individual values contained in a matrix are represented as colors.

`sns.heatmap (Big_Mart.isnull(), yticklabels=False, cmap="viridis")` – The same color of all columns of heatmap indicates that after the data cleaning of Big_Mart Data set there is no null values present in the data set.



Big_Mart.corr() –used to find correlation coefficient which is a numerical measure of statistical relationship between two variables, or two columns of a data set.

The Correlation coefficient assume values in the range from -1 to $+1$, where ± 1 indicates the strongest possible agreement and 0 the strongest possible disagreement.



```
corr = Big_Mart.corr()

ax = sns.heatmap(

    corr,

    vmin=-1, vmax=1, center=0,

    cmap=sns.diverging_palette(20, 220, n=200),

    square=True

)

ax.set_xticklabels(

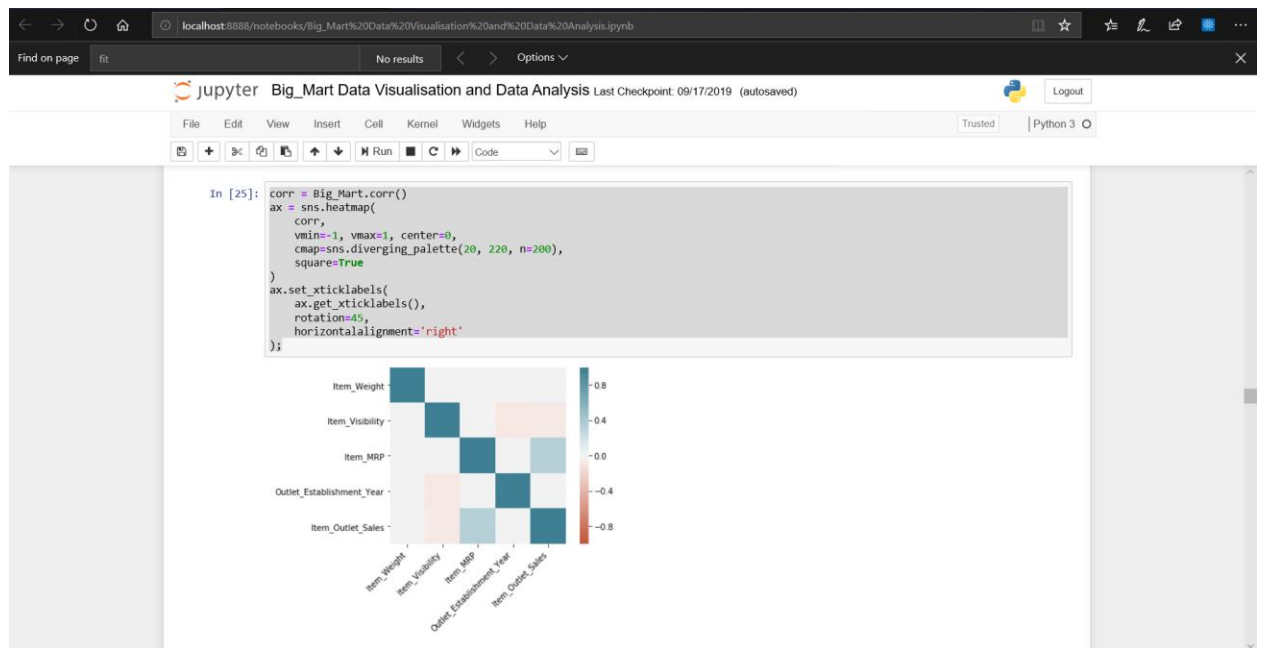
    ax.get_xticklabels(),

    rotation=45,

    horizontalalignment='right'

);
```

- Used heatmap to see the visual representation of correlation between the different features of Big_Mart Data Set.
- The brightest colour in following Heat map is used to represent the ± 1 value of correlation coefficient which indicates the strongest possible agreement
- The light color of heatmap is used to represent the 0 the strongest possible disagreement.



3. Difficulties Encountered in Reporting Week

To use the correlation coefficient function `corr()` with heat map along with `diverging_palette` `cmap=sns.diverging_palette(20, 220, n=200)` and to choose the variables with positive correlation to represent them with the use of various plots

4. Tasks to Be Completed in Next Week

To use pair plot and joint plot to find more relationships with output sales and to predict the customers' behaviours prevailing in market.