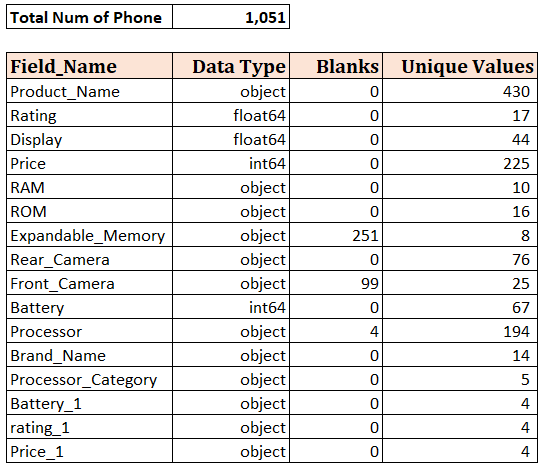
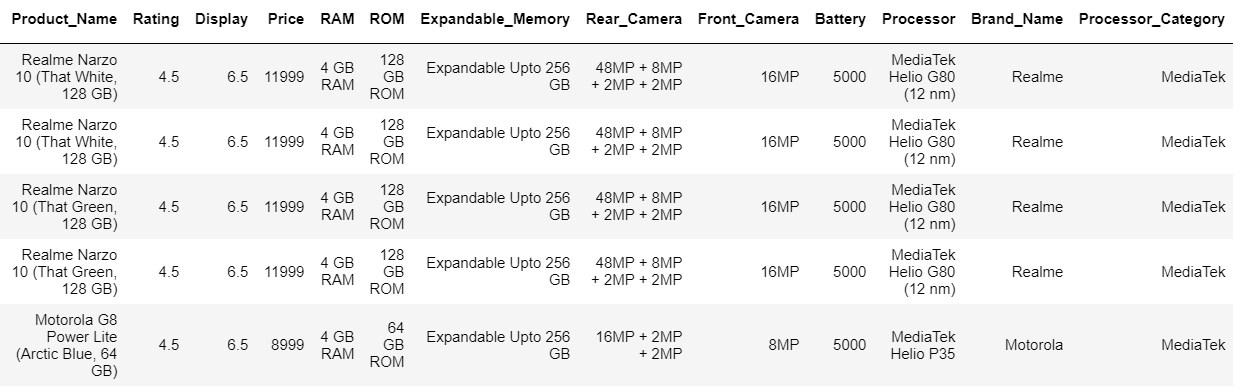
**Exploratory Data Analysis**

*Overview of the dataset:*

* In this project, 50 pages from Flipkart – smart phones data has been scrapped
* Population dataset contains 1,051 rows/observations and 16 columns/fields

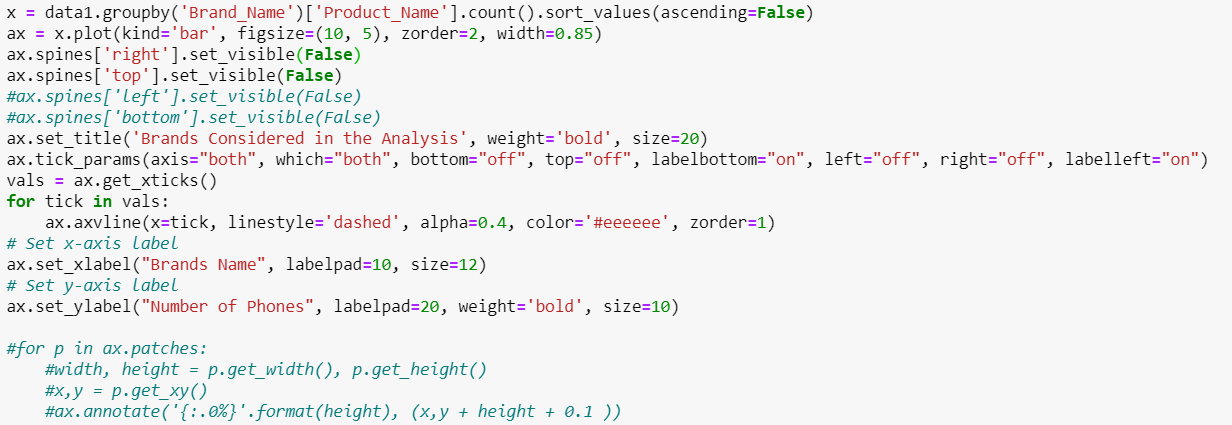


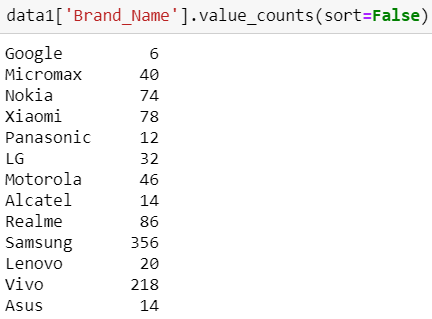
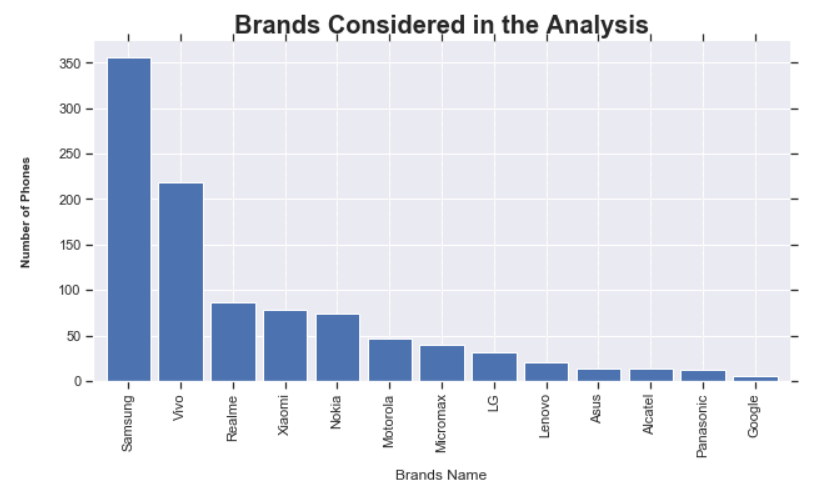
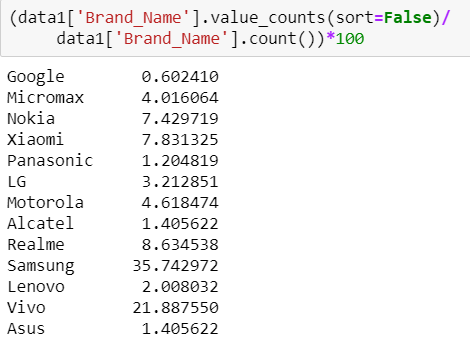
* Input and data cleansing files:
  + Input - Flipkart scrapped file
  + Data cleansing:
    - Processor category – to tag all processors from specific brand into one group
    - Brand Name – Based on product name bunch of phones from same brand are grouped into one
* Below is the sample dataset:



*Exploratory Data Analysis:*

Which brand phone do you use?



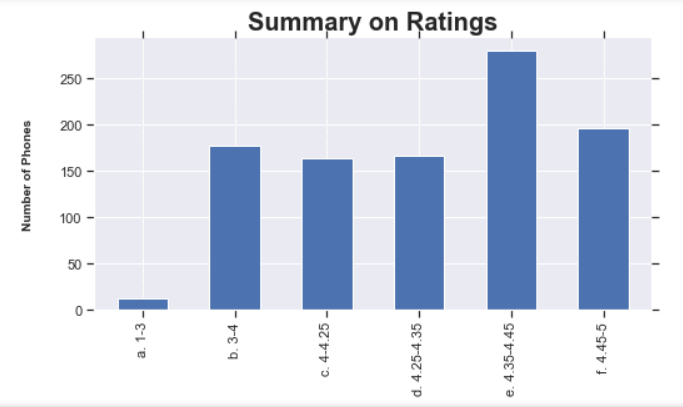
 

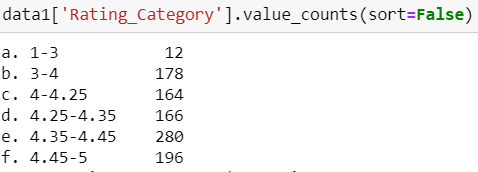
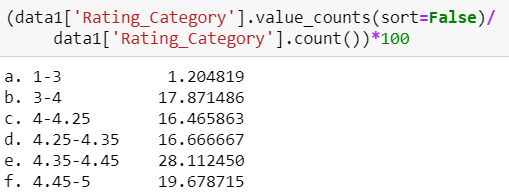
Insights:

* There are 13 companies considered in this analysis
* Maximum number of records scrapped are from Samsung followed by Vivo, Realme and Xiaomi, top 4 brands constitute to 74% of the total data scrapped
* Google has least number of phones with 0.6% of the data

How much do you rate your phone?



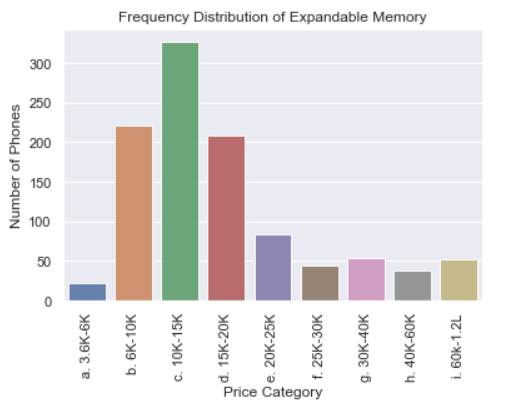
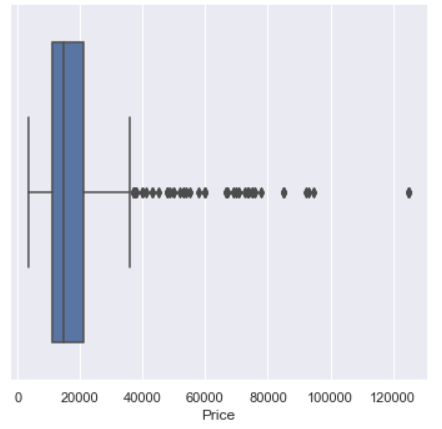


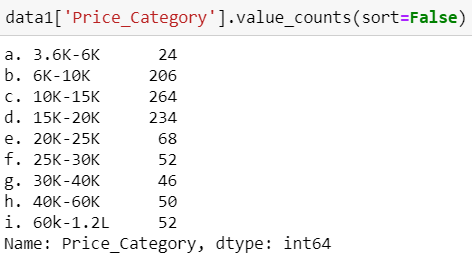
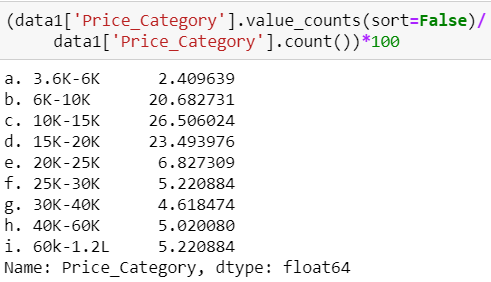
Insights:

* There are 190 records with rating <3, 19% of the population
* Between rating 4-5 we have 18% of the data i.e. 806 records.Rating 4-5 is further split to better understand the data.
* 61 % data is present within the range 4-4.45

What is the cost range of your phone?

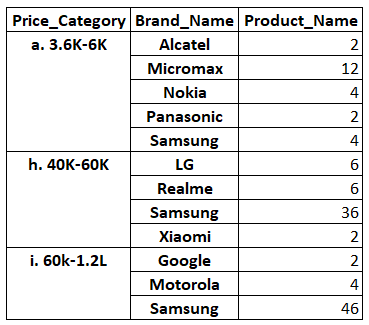
 

**To be repeated**

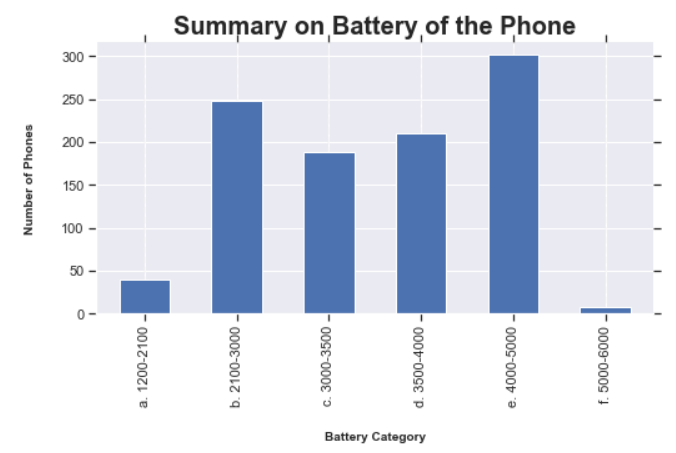
 

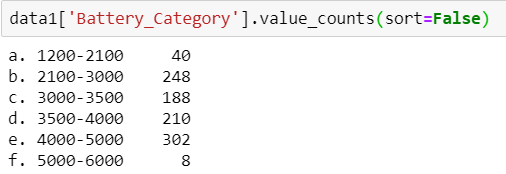
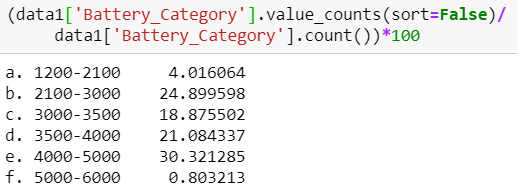
In sights:

* In our analysis 71% of the data has price range 6,000 to 20,000
* Using box plot we observe below Q1 and above Q3 line is near 40,000; after which we can consider it as outliers
* From the summary, records below 6K is 24 records and above 40K is 102 records. Together after excluding outliers we will be loosing 126 records which is 13% of the population
* Out of curiosity want to know which brands these 126 records belongs to:



Is your display and battery dependent on each other?



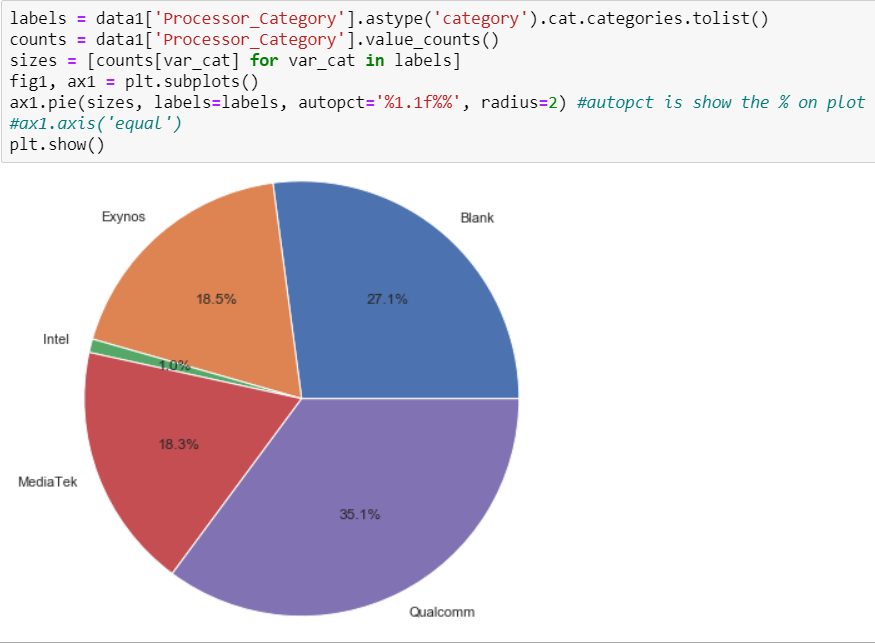
 

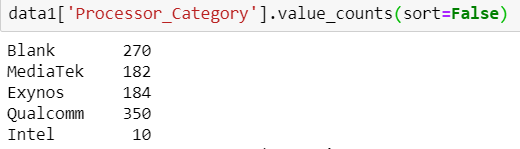
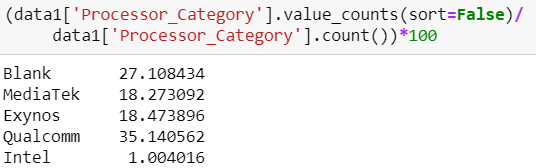
Insights:

* 95% has battery ranging from 2100 – 5000 mAh
* Looking at the summary table we can observe the categories with battery range 1200-2100 and 5000-6000 together constitutes to 4.8% of the population which is 48 records
* Is battery dependent on other features of the phone?

We will look into it in bivariate analysis!

Processor Category:



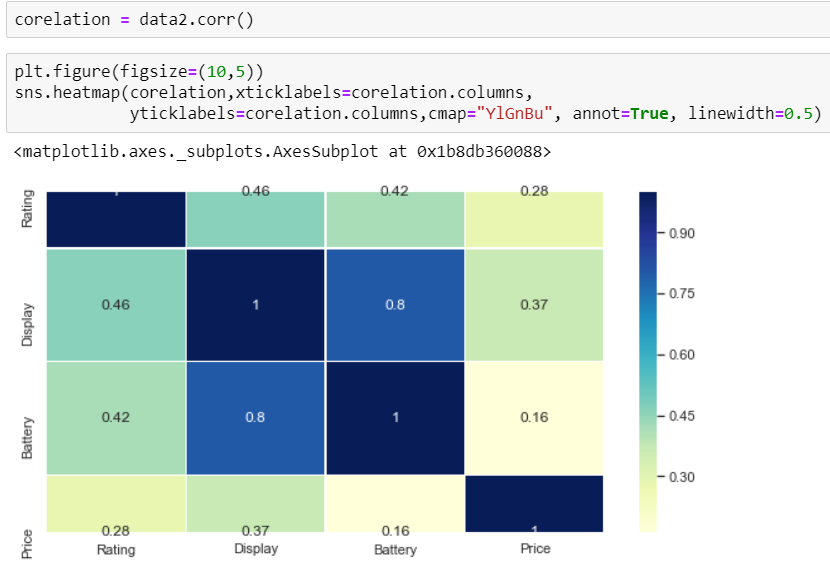
 

Insights:

* Multiple versions from Processor are grouped into each of the above values
* Qualcomm takes a major share of 35% in the entire population; however, 27% is blank.
* If we consider in the data available; then Qualcomm share increases to 48%
* There is only 1% of data available with Intel processor, which is 10 record and can be removed as outlier

**Bi-Variate Analysis**

**Correlation**

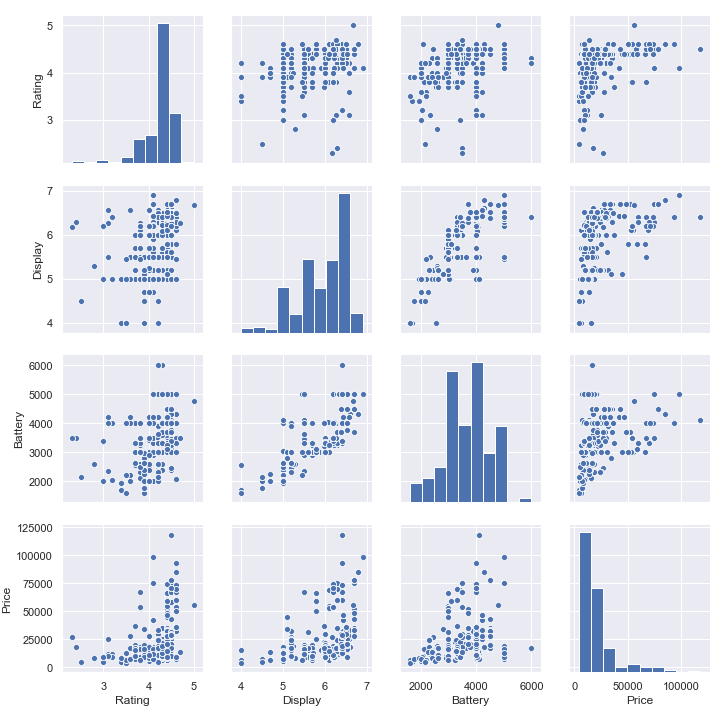


**Insights:**

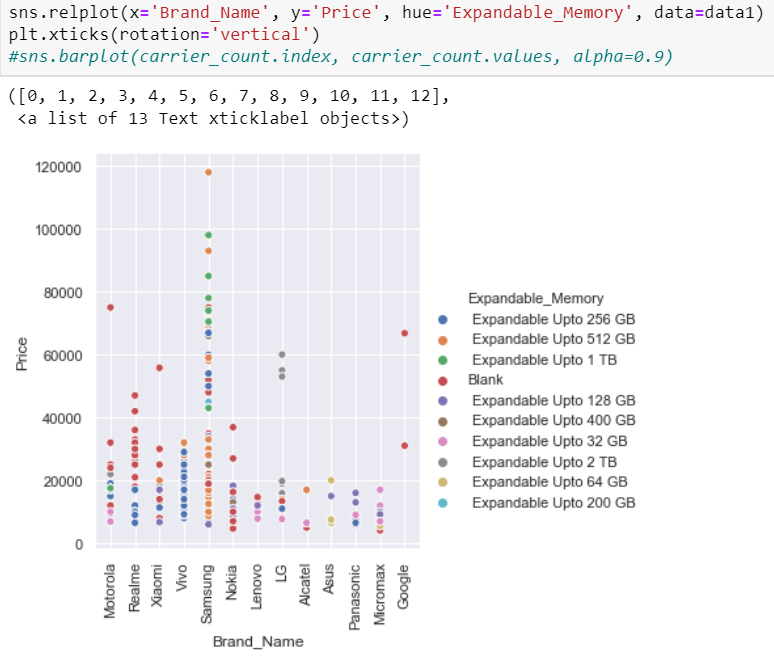
Display and battery are strongly dependent on each other

In the next category we have Rating dependent on Display and Battery

**Pairpolt:**



**Relation between Brand Name Price and Rating:**



**Insights:**

Very strong positive correlation would be between Display and Battery

Very weak positive correlation would be between Battery and Price

**Relation between RAM, Price and Rating:**

