INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Analyzing Tablet Prices Of Flipkart

ABOUT US

Name : Akshara
Qualification : BSC(Data Science)

INNOMATICS
RESEARCH LABS

Contents:

Problem Statement

- Analysing Prices Of Tablets.

- The current tabular representation of pricing data presents with its complexities and limitations for businesses.

- The tabs provides insights into customer preferences and brand satisfaction , helping to discern which brands are more commonly associated.

Introduction:

- Brief overview of the pricing of Tabs using Flipkart.
- Analysis of variation of price with respect to different features of Tabs.
- Finding the best Tabs that has most benefits.
- From this project , I am interested to provide good insights to choose better Tabs to smoothen life.

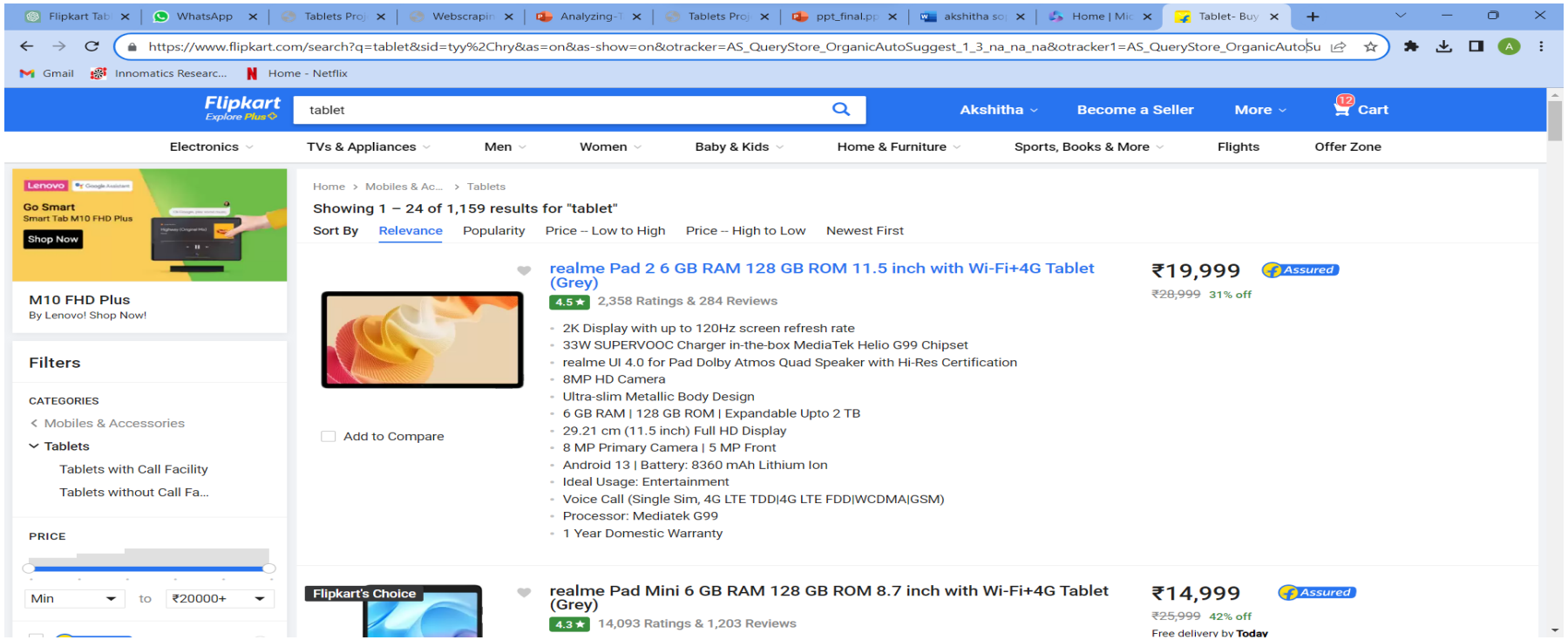
Libraries Used:

- Regular Expressions
- Beautiful soup
- Numpy
- Pandas
- Matplotlib
- Seaborn

INNOMATICS
RESEARCH LABS

Website URL's for Web-Scrapping :

https://www.flipkart.com/search?q=Tablets&otracker=search&otracker1=search&marketplace=FLIPKART&as-show=on&as=off



INNOMATICS RESEARCH LABS

Raw data from website :

	Brand	Color	Ram	Rom	Screen_size	Battery	Megapixel	Rating	price
0	MOTOROLA	Modernist Teal	6.0	128	11.00	7700.0	13	4.3	₹22,990
1	SAMSUNG	Silver	3.0	32	8.70	5100.0	8	4.3	₹11,999
2	SAMSUNG	Blue	4.0	128	10.40	7040.0	8	4.5	₹31,999
3	Lenovo	Abyss Blue	4.0	128	10.61	7700.0	13	4.2	₹24,490
4	SAMSUNG	Beige	12.0	256	14.60	10090.0	13	NaN	₹1,08,999
...	...	...	...	...	...	...	...	...	...
475	APPLE	Silver	NaN	128	11.00	NaN	12	4.6	₹79,900
476	Lenovo	NaN	4.0	128	10.30	5000.0	8	4.3	₹15,999
477	SAMSUNG	Silver	6.0	128	12.40	10090.0	8	4.5	₹49,999
478	APPLE	Space Grey	NaN	256	10.90	NaN	12	4.6	₹81,400
479	MOTOROLA	Frost Blue	4.0	64	10.61	7700.0	8	4.1	₹16,490

Total: 480 rows × 9 columns of Raw Data

DATA CLEANING

1) Handling Missing Values
2) Removing Duplicate Values
3)Outlier Treatment
4)Data Analysis

Cleaned Data Frame:

	Brand	Color	Ram	Rom	Screen_size	Battery	Megapixel	Rating	price
0	MOTOROLA	Modernist Teal	6	128	11	7700	13	4	22990
1	SAMSUNG	Silver	3	32	8	5100	8	4	11999
2	SAMSUNG	Blue	4	128	10	7040	8	4	31999
3	Lenovo	Abyss Blue	4	128	10	7700	13	4	24490
4	SAMSUNG	Beige	12	256	14	10090	13	4	108999
...	...	...	...	...	...	...	...	...	...
·75	APPLE	Silver	4	128	11	10090	12	4	79900
·76	Lenovo	Silver	4	128	10	5000	8	4	15999
·77	SAMSUNG	Silver	6	128	12	10090	8	4	49999
·78	APPLE	Space Grey	4	256	10	10090	12	4	81400
·79	MOTOROLA	Frost Blue	4	64	10	7700	8	4	16490

Features that Impact Tab Prices:

1 **Ram**

Higher the Ram, higher the price

2 **Battery**

Longer battery , higher the price.

3 **Color**

Higher the price, according to different colors.

INNOMATİCS RESEARCH LABS

EXPLORATORY DATA ANALYSIS(EDA)

Exploratory Data Analysis is majorly performed using the following methods:

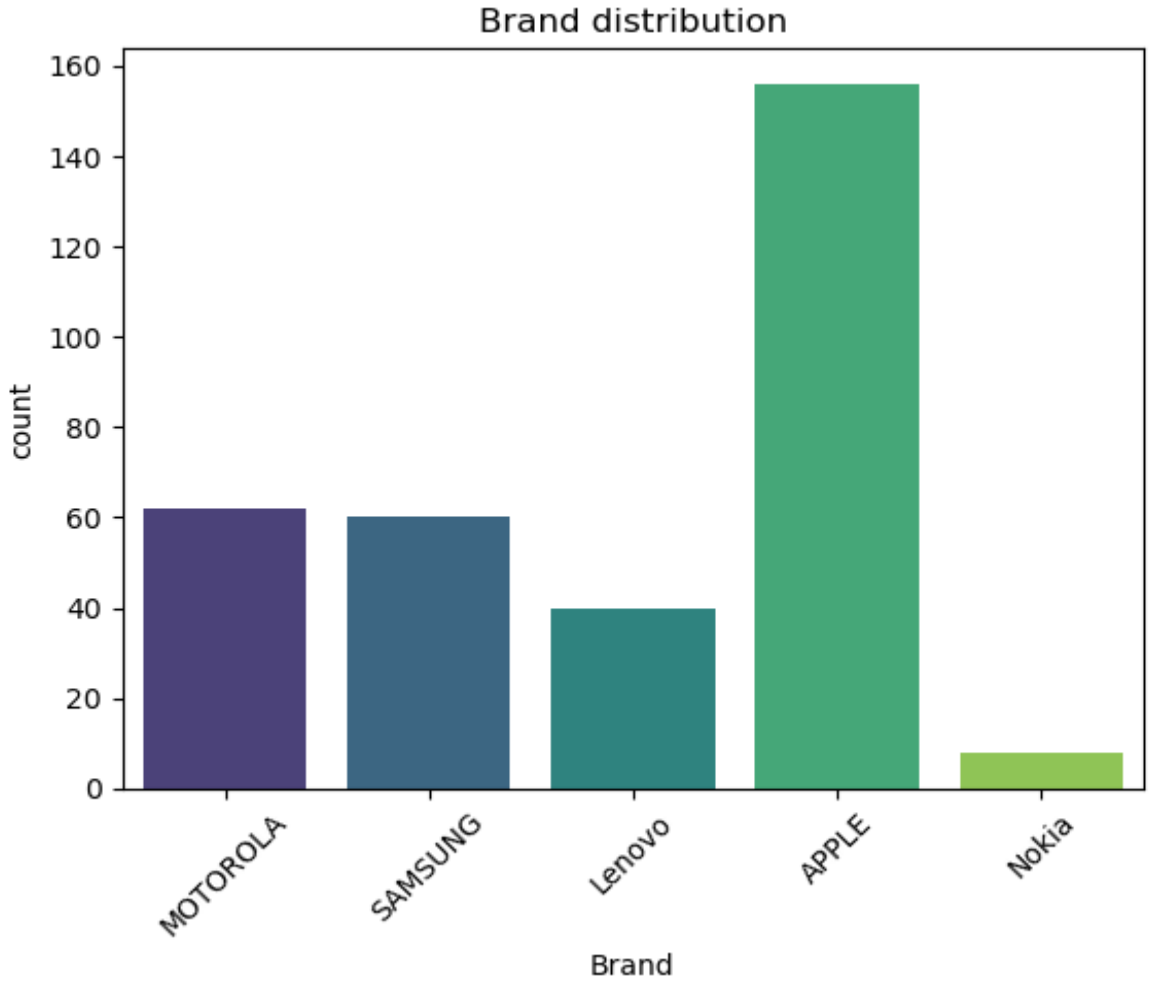
- **Univariate analysis**:- It provides summary for each field in the raw data set (or) summary only on one variable.

- **Bivariate analysis**:- It is performed to find the relationship between each variable in the dataset and the target variable of interest (or) using 2 variables and finding the relationship between them

- **Multivariate analysis**:- It is performed to understand interactions between different fields in the dataset (or) finding interactions between variables more than 2..

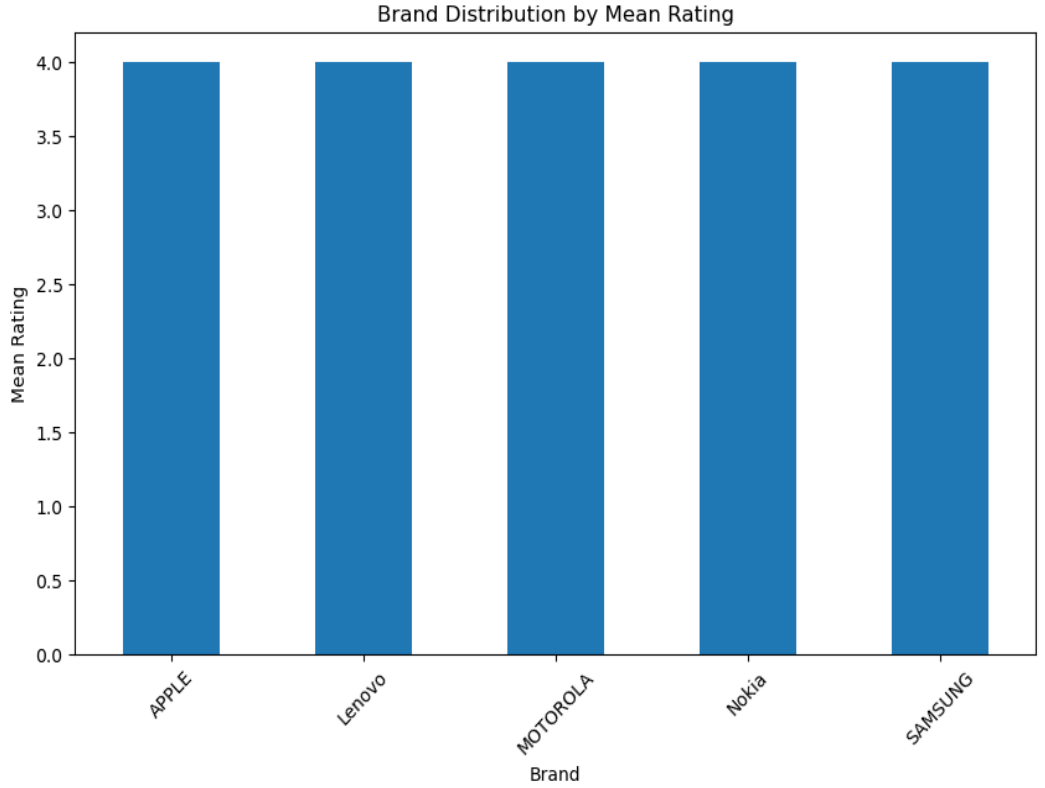
Univariate: Brand count plot

- The count plot is useful for understanding the frequency or count of each category in the dataset, providing a quick overview of the distribution.
- Brand Count Plot represent the no of brands in the tablets.
- No .of count of the tabs available in each brand.
- Apple and Motorola is the leading company having huge variety of products.



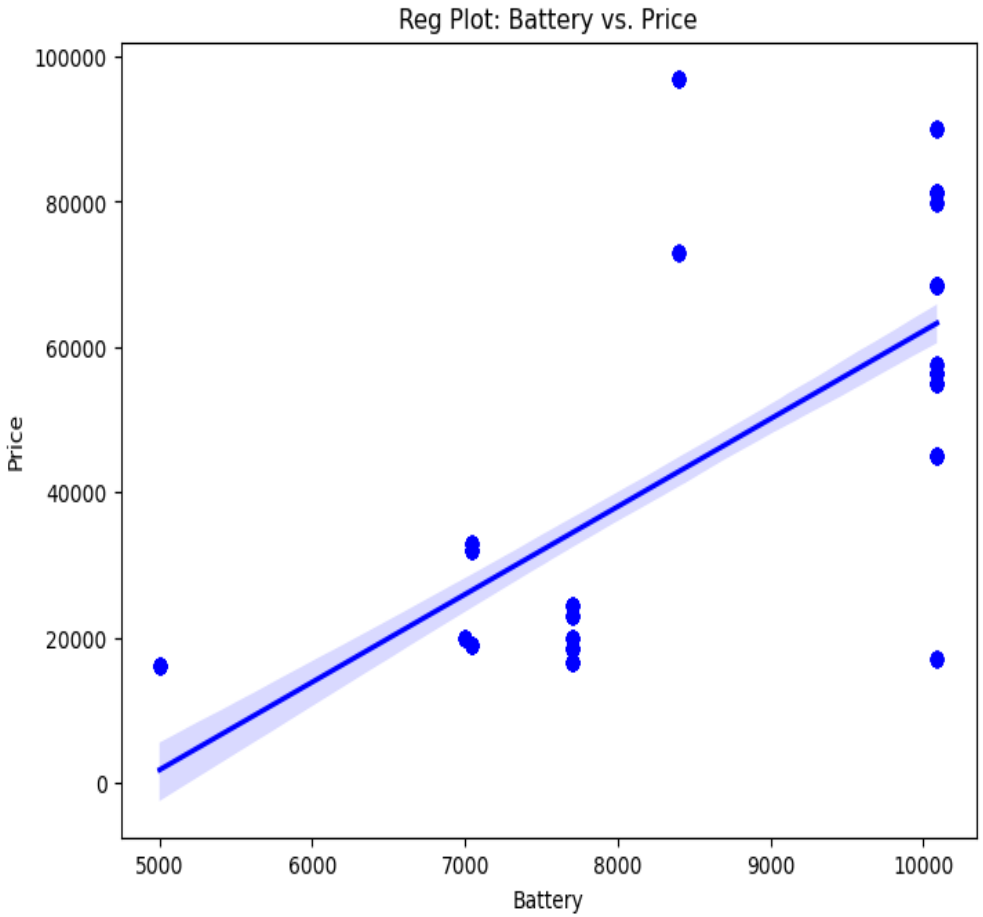
Bivariate: Brands with 5-star rating

- The plot provides insights into which brands have a higher frequency of top ratings in the dataset.
- The code aims to visualize the distribution of 5-star ratings among different brands using a count plot.
- Getting brands of 5 star rating.
- 5 Star Tabs is more suggestible because it consumes more count compared to other ones.
- Apple and Motorola products are having more rating.



Bivariate : Battery vs Price

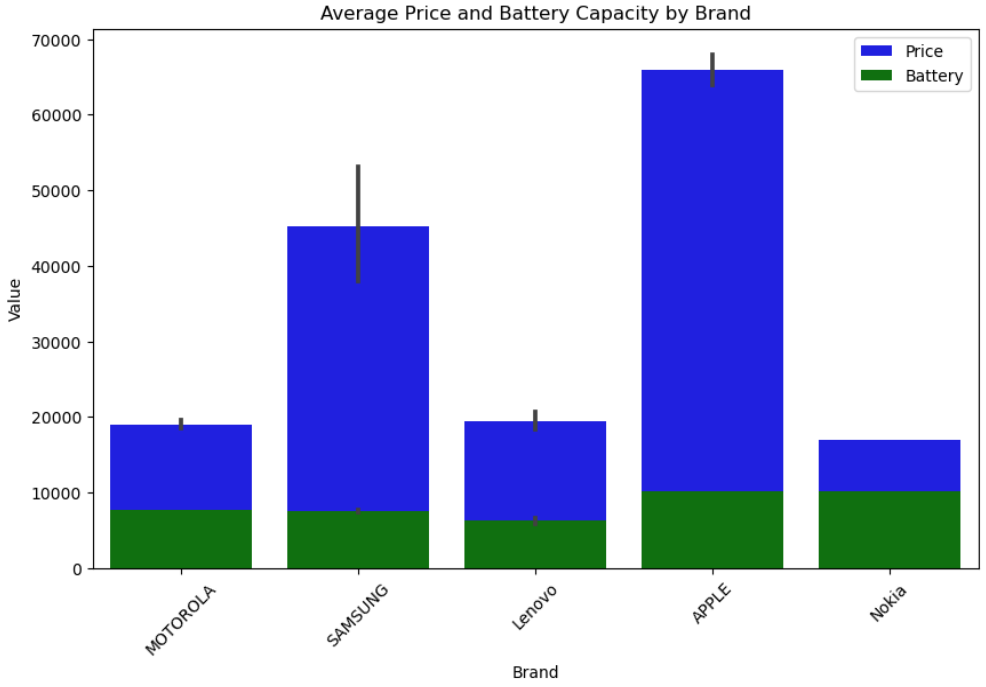
- It is a continuous and continuous bivariate plot
- It visually represent the linear relationship between the 'Battery' and 'Price' columns using a regression plot.
- The plot provides insights into how changes in the battery capacity may be associated with changes in the price of tablets.
- The distribution of batteries are more around 20000.



INNOMATİCS
RESEARCH LABS

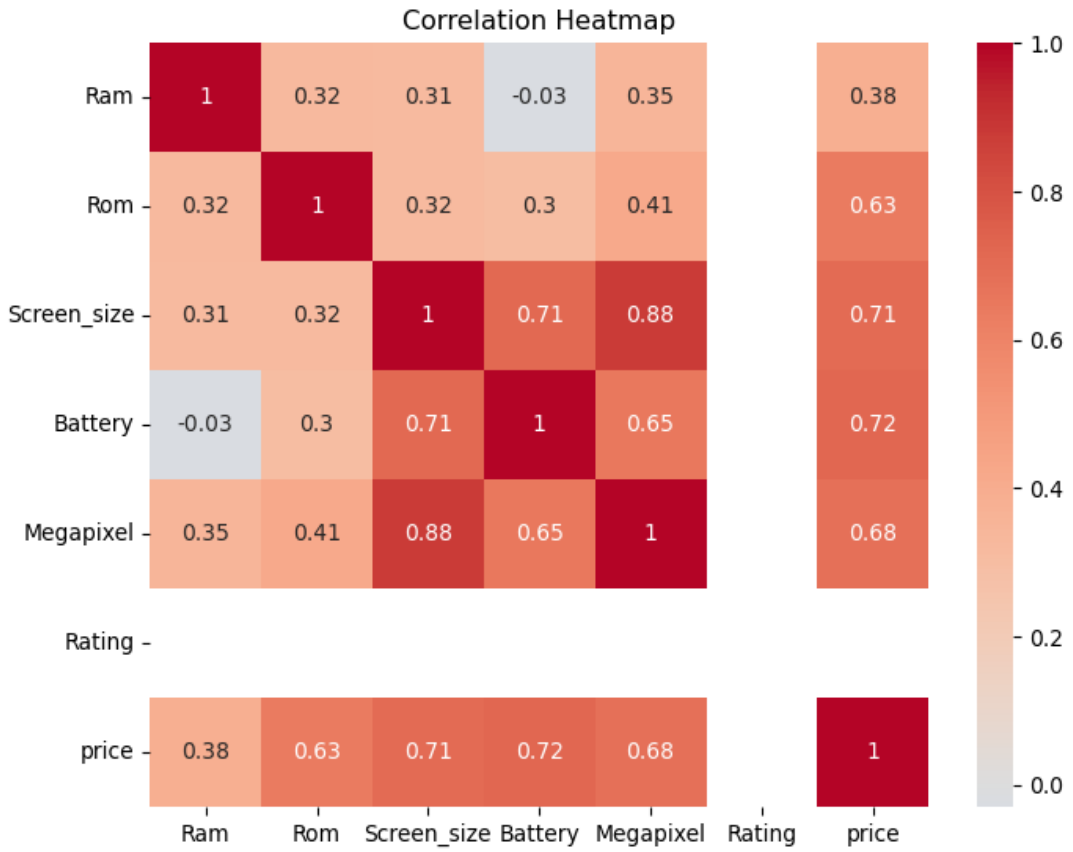
Multivariate : Price and Battery capacity

- It visualizing to compare the average price and battery capacity of different brands using a grouped bar plot. This type of plot allows for easy comparison of two numerical values across different categories.
- Average Price and Battery by Brand.
- Most of the products are in range between 10K to 65K.
- Apple is the most suitable one to buy.



Relationship

- The plot provides a visual representation of the correlation structure among numerical columns in the dataset.
- The correlation heatmap helps identify patterns and strengths of relationships between variables.
- The Battery and Ram shows there is no strong relationship between them.
- The Rating has no correlation.
- There is strong relationship between itself features.



INNOMATİCS RESEARCH LABS

Conclusion:

•Most selling products are in 10K to 65K.
•Most preferable battery brands are Apple and Motorola.
•5-Star products are more recommended for new users.
•Most trusted company is Apple and Motorola Product.

- We can infer the most suitable brand is Apple and the second brand is Motorola, according to price distribution of the most common features Ram ,Battery and Ratings.

THANK YOU!

INNOMATICS
RESEARCH LABS