# Title: What Drives Ratings? Exploring the Impact of Selling Price

**Introduction:**

This project is centred on the processing of sales data from mobile phones in the mobile phone industry. This dataset contains data about different MOBILE PHONE brands including their models The data includes information about colors available, memory, storage, camera, the price of the models and even the discount rate for such models and customer rating. The mobile phone industry is highly competitive and the industry is undergoing tremendous growth making it a suitable area for study. The ideas for selecting this dataset is based on consumer behavior, demand, pricing models, and competitor hierarchies. In a personal capacity, the interest is focused on the object under consideration – brand, features, and price and their impact on the customer’s purchase decision.

**Business Value/Problem Statement:**

The primary business question for this analysis is: What can mobile phone retailers do in regard to sales promotion, stock, price and discount techniques to achieve the highest levels of sales and customer satisfaction? Understanding the factors that drive sales and customer ratings can help businesses:

* Improve inventory management.
* Conceive good and appealing marketing and advertisements.
* They proposed to identify high contribution product categories; and, areas where contribution could be enhanced.
* It also increases the quality of the products as well as the experience of the users through targeted solutions.

**Dataset Description:**

The dataset contains 3,114 entries and 12 columns, summarizing information about mobile phone sales:

* Brands and Models: Find out the name of the manufacturer and the particular kind of product.
* Colors: A list of models and the different color variations in stock.
* Memory and Storage: Technical requirements that determine efficiency in dealing with projects.
* Camera: Points out if the phone’s model has one or not.
* Ratings: The scores of customer satisfaction (There are some missing values in the rows).
* Selling Price and Original Price: Price at which phones are sold and the first time they were marketed in the market at a retail.
* Discount and Discount Percentage: Cases that relate to price reduction and the proportion in which it is reduced from the original price.
* Mobile: People refer to the brand and the model together forming a single unit.

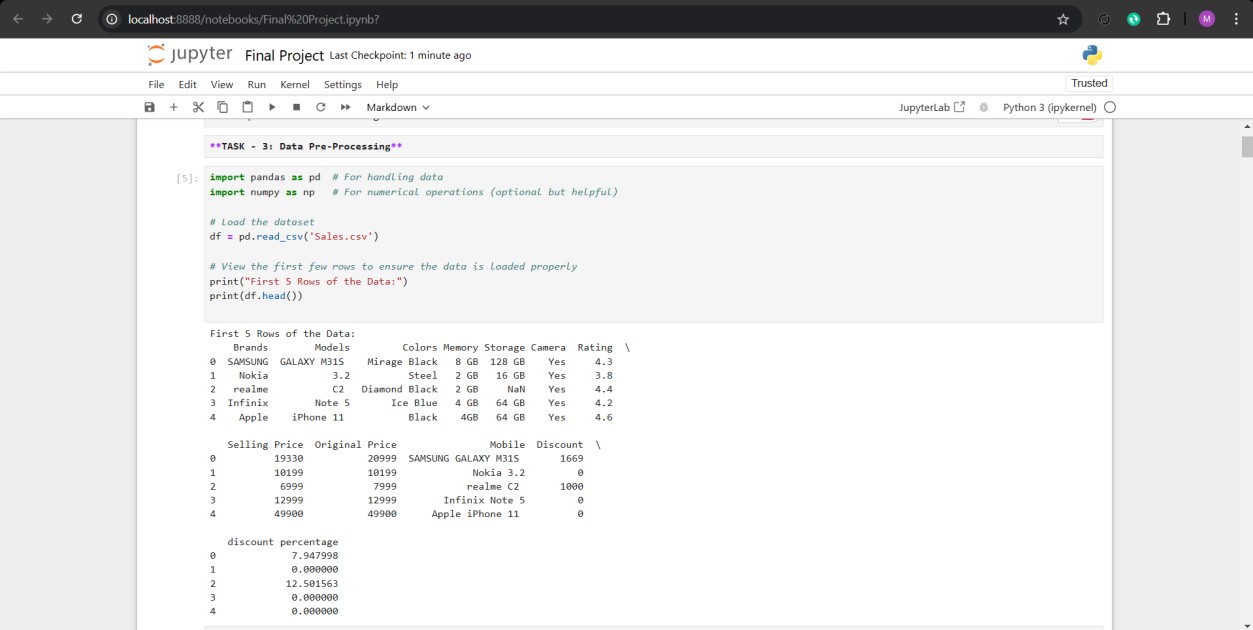
Responses from some participants are missing, mainly in the Memory, Storage, and Rating features; hence, data cleaning is necessary for analysis.

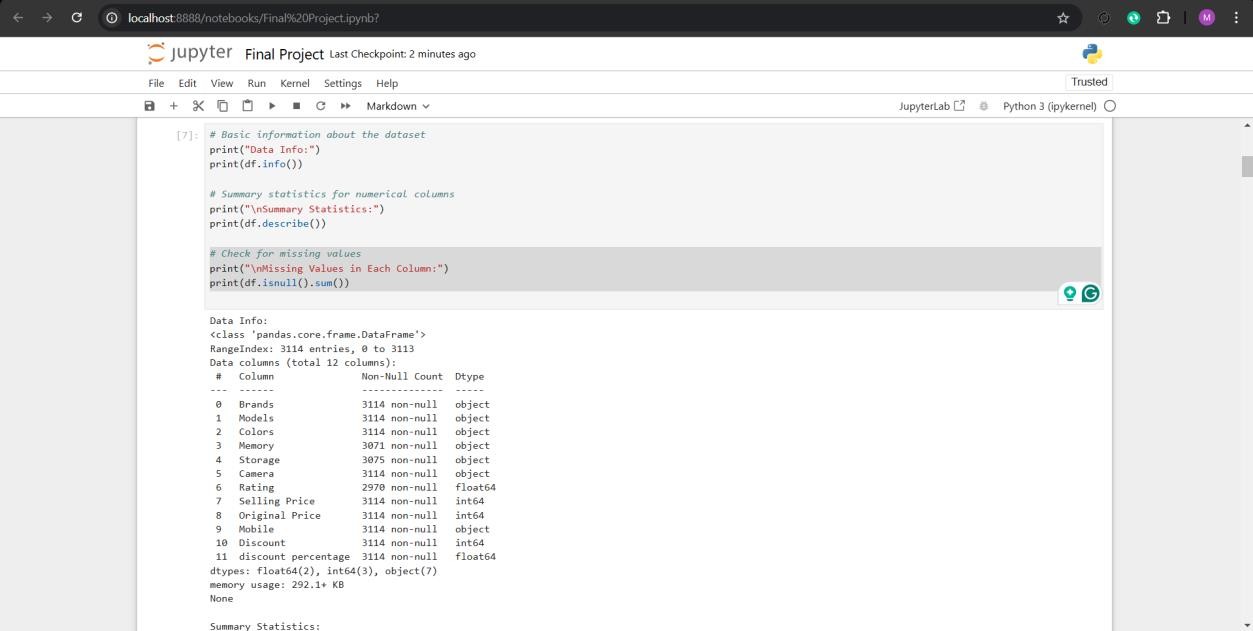
**Data Cleaning:**

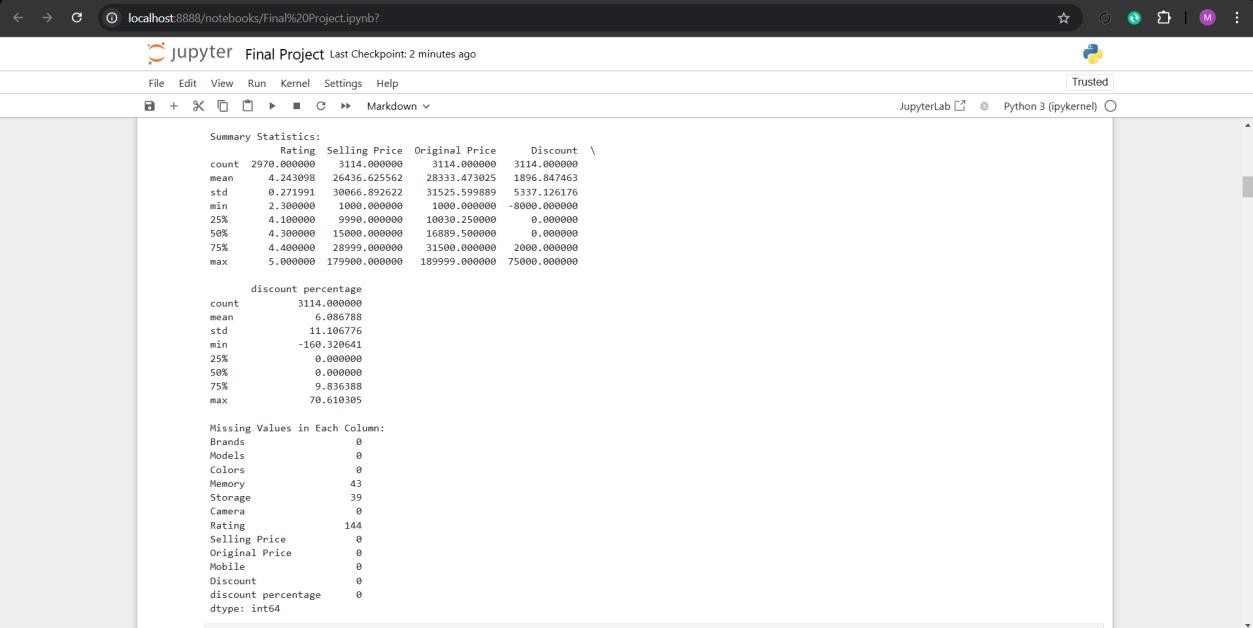
The first step is to load the dataset and examine the client’s dataset for signs of missing values. The rows with missing values are omitted, and where do not have a ‘Storage’ value, it filled with zeros. The code then strips off ‘GB’ and ‘MB’ from ‘Memory’ and ‘Storage’ respectively in order to normalize the data then adds ‘GB’ for purposes of uniformity.

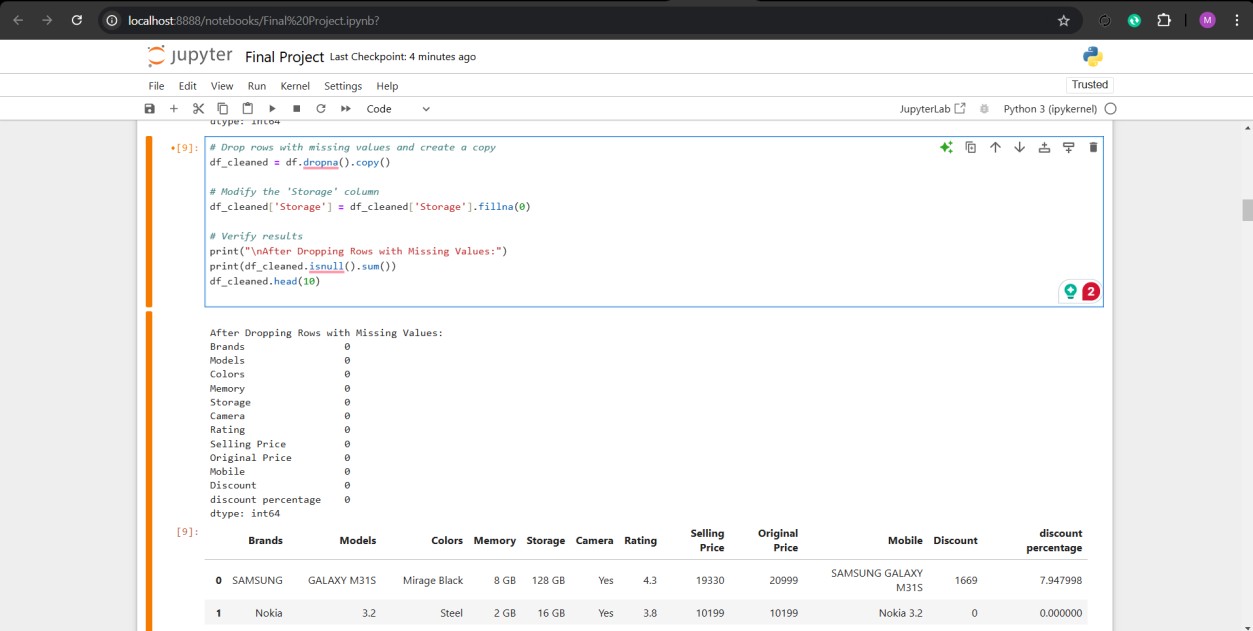
Then, using the Z-score method and setting the limit for rows containing more than 3, outliers are deleted. Since the dataset has two similar columns ‘Camera’ and ‘Mobile’, these are removed to enhance the dataset. Last of all, the names of columns are modified for better understanding and to have a clean look of a DataFrame to proceed with analysis.

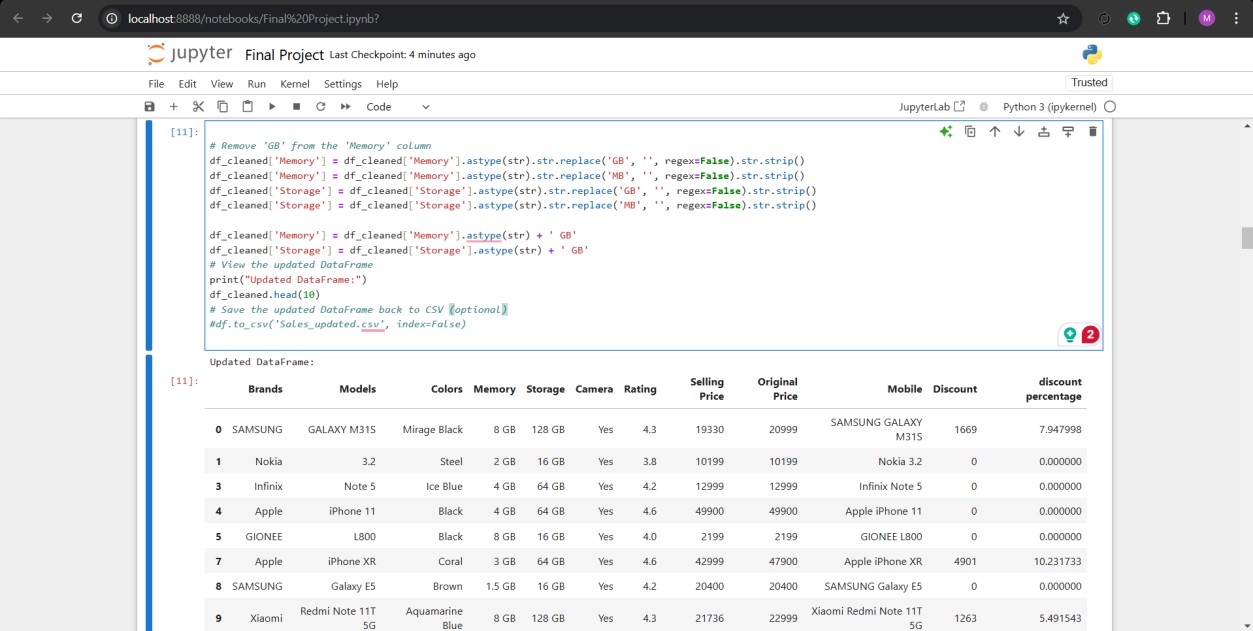
**Screenshots of the Data Cleaning with Python code and Output:**

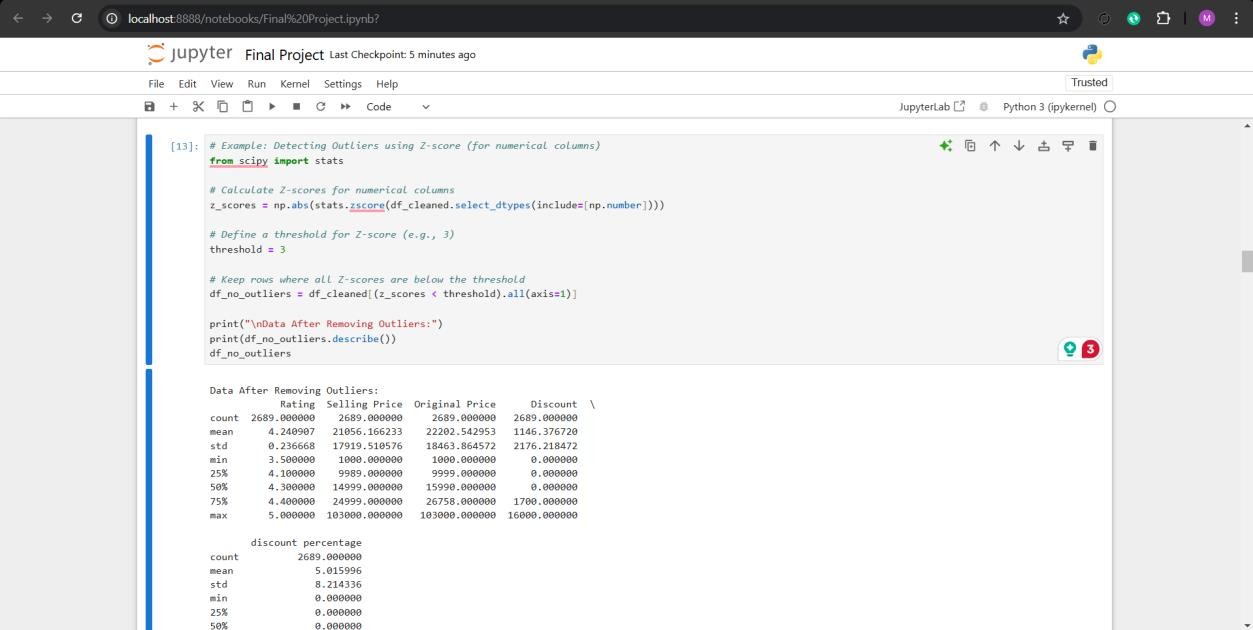


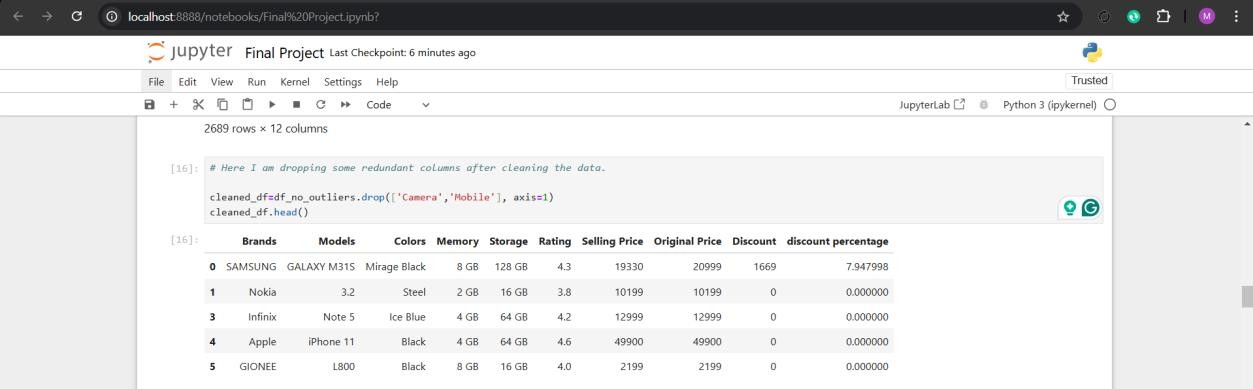


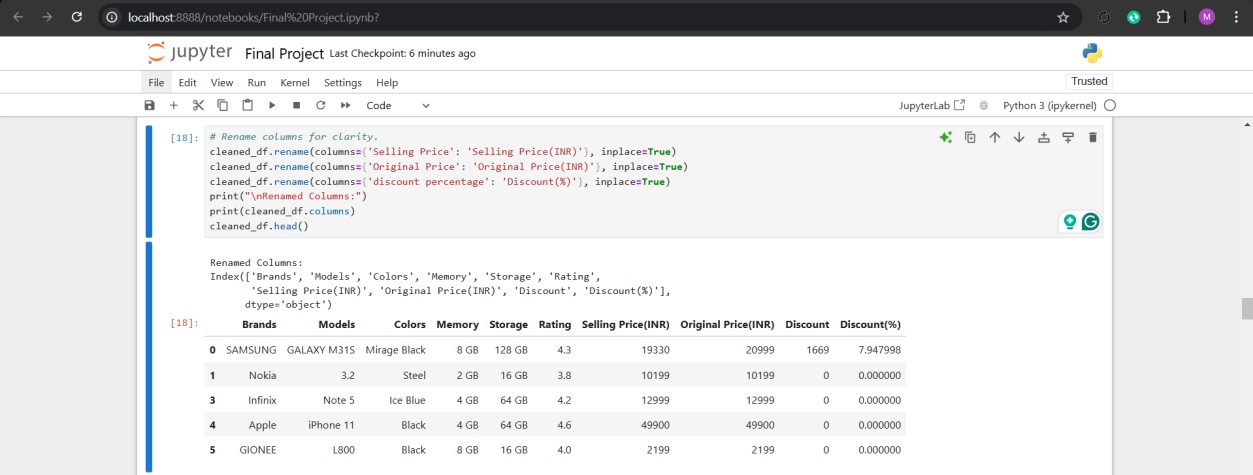








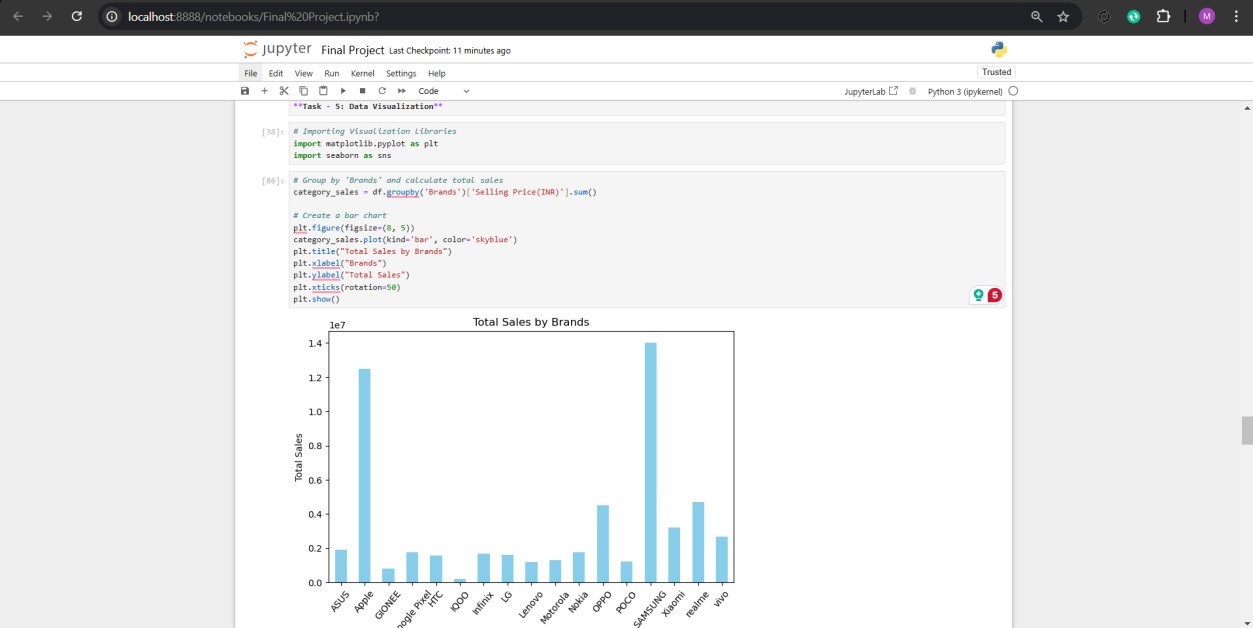


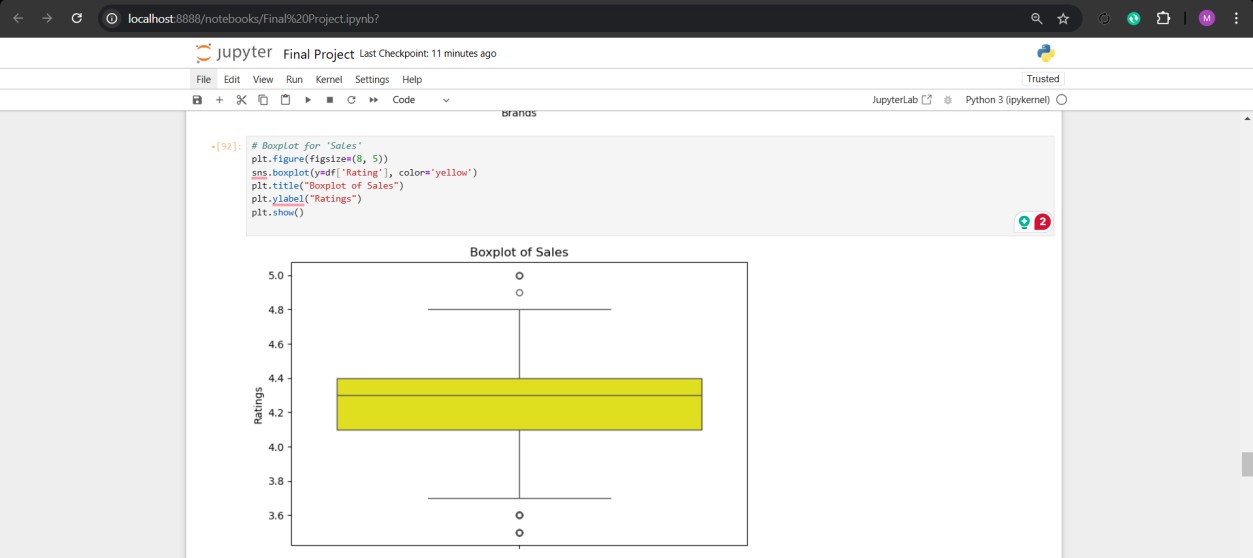


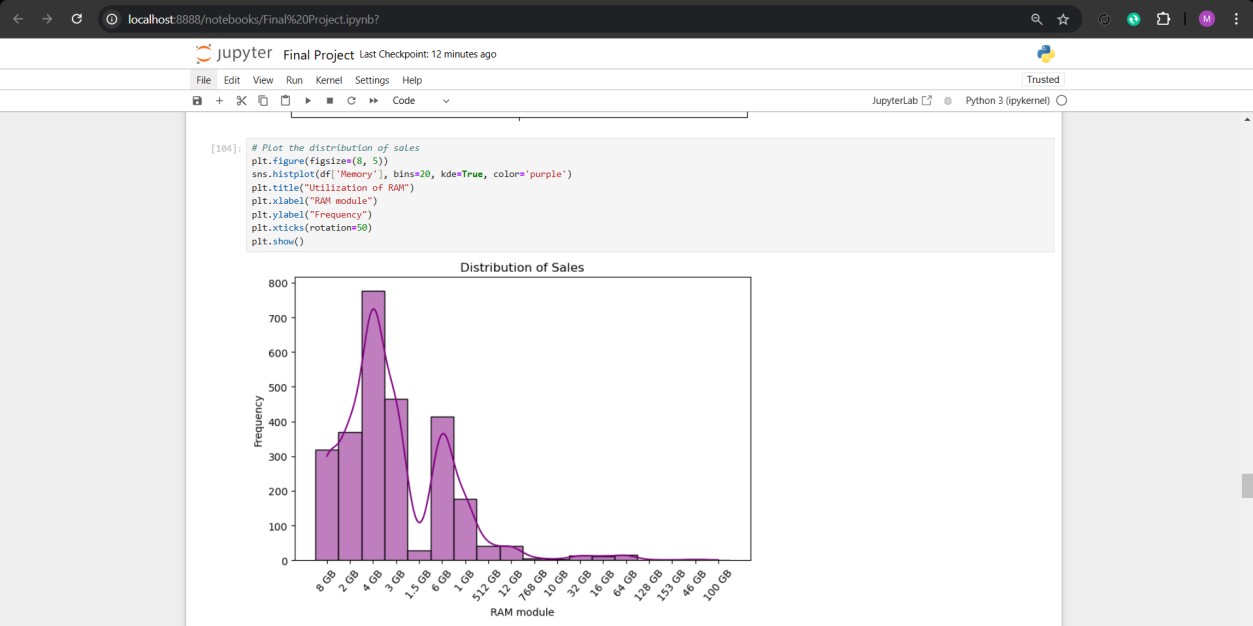
**Data Exploration (EDA):**

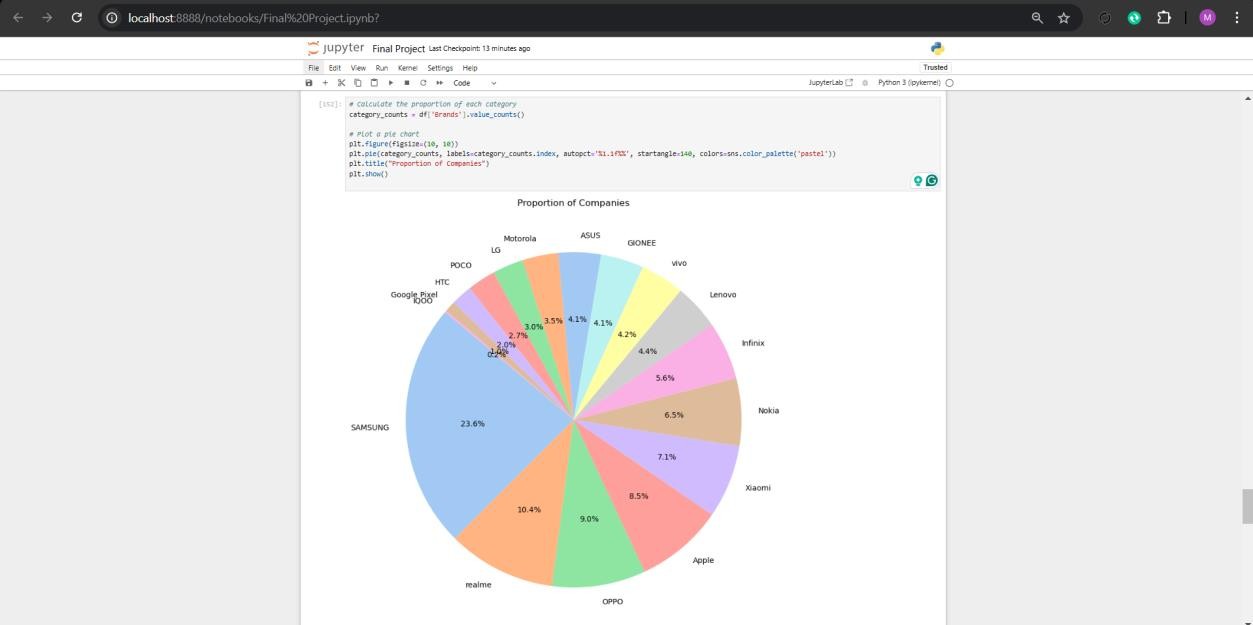
Common Python methods for EDA were applied while exploring the cleaned dataset. Descriptive analytics in form of mean median mode were done on numerical columns to get statistical analysis. The given dataset was then analyzed by categories to understand certain patterns, for example average value of given Brands by Selling Price. Pearson’s coefficient was used in order to determine the correlation of one column with another. Moreover, outliers were found and data distribution was depicted in order to get clearer view about the data and thus facilitate realization of the structure and variability of the given data set.

**Task: Data Visualization Results**









**Task: Predictive Modelling:**

It begins with the import of the necessary libraries then the loading of a cleaned dataset from a CSV file. The dependent variable is given the name ‘Rating’, whereas the independent variable used to predict ‘Selling Price(INR)’. The obtained data is then partitioned across two sets namely training and testing data. Owing to this, a linear regression model is created and fit into the training data. The estimators which include the intercept and the slope are displayed. The model then goes on to the task of rating generation on the test set and then the output is compared with the actual ratings. This is performed by assessing the model where tools like Mean Squared Error (MSE) and R-squared (R²) are obtained.

Performance and Evaluation of the Predictive Model:

Mean Squared Error (MSE): This means it measures the average of the squared differences between the exact and the predicted values. It also emerged that for the current study, a lower mean square error signifies a better fit of the model against the data.

R-squared (R²): Essentially this one measures the extent to which the dependent variable can be explained by the one or many independent variables. A closer observation to 1 has a higher levels of explanatory variables.

**Analysis:**

Based on an initial exploration of the data, the following insights and recommendations could be derived:

Pricing Strategy: This signifies that models that are cheaper in the market and whose discounts are considerably higher tend to lead in performance. Quantitative examination can reveal the best low price band that will help to increase sales volume while preserving profitability.

Consumer Preferences: There initial assessments show that factors like having more storage, improved memory, and high ratings from customers makes an impact. Thereby, retailers should display models that reflect these preferences.

Brand Performance: With the help of sales and rating data collected, a business might target popular brands and enhance their sales or explore problematic brands and find out possible problems.

Promotional Campaigns: Topping up when there are many clients or during off peak hours are ways to time promotional offers well.

Stores managers and other stakeholders could utilise the finding from this analysis in developing the right strategies that would help them select and price their products right and better market to their customers hence driving profitability and customers’ satisfaction.