# 1. Introduction

In today's fast-paced world, mobile technology is essential for connectivity, communication, and entertainment. This project investigates a dataset containing various characteristics of mobile phones and their prices. We intend to use statistical techniques and R programming to analyze the relationships between mobile phone features such as screen size, RAM, storage, battery capacity, camera quality, and price in order to determine how these factors affect the cost of mobile devices. This analysis provides information about consumer preferences, industry trends, and the impact of technological advancements on pricing. The goal is to create a regression model that can predict mobile phone prices based on these characteristics, thereby assisting manufacturers and consumers in understanding the key determinants of mobile pricing.

# 2. Literature Review

Understanding the relationship between technology, consumer demand, and market trends is typical of mobile phone pricing research. Previous research has shown that mobile phone features such as screen size, processing power, and camera quality have a significant impact on consumer preferences. Furthermore, factors such as battery capacity and storage influence the overall user experience, affecting the device's market value. As technology advances, manufacturers must balance cost, functionality, and market demand. This project expands on previous research by focusing on a dataset containing various mobile phone features and using statistical methods to investigate their impact on pricing. (Prof.B.Prajna)

# 3. Regression Model and Review

In this section, we will build a multiple linear regression model to investigate the relationship between mobile phone characteristics and price. Our model's dependent variable is price, and the independent variables are screen size, RAM, storage, battery capacity, and camera quality.

## (i). Formula used

The regression model was formulated as follows:

**Score = 𝛃0+ 𝛃1 . GDP\_Per\_Capita + 𝛃2 . Social\_Support + 𝛃3 .Expectancy + 𝛃4 . Freedom\_To\_Make\_Choices + 𝛃5 . Perceptions\_of\_corruption + 𝛆**

### (a). Description of variables

* + - Price: The dependent variable representing the cost of the mobile phone.
    - Screen\_Size\_inches: The diagonal measurement of the mobile phone's screen in inches.
    - RAM\_GB: The amount of Random Access Memory in gigabytes.
    - Storage\_GB: The internal storage capacity in gigabytes.
    - Battery\_Capacity\_mAh: The battery capacity in milliamp-hours.
    - Camera\_Quality\_MP: The camera quality in megapixels.

(Mobile Phone Price Prediction, n.d.)

### (b). Regression Coefficients (𝛃)

**𝛃0**: Intercept term representing the baseline price

**(𝛃1, 𝛃2,  3, 𝛃4, 𝛃5)**: Regression coefficients indicating the impact of each independent variable on the baseline price

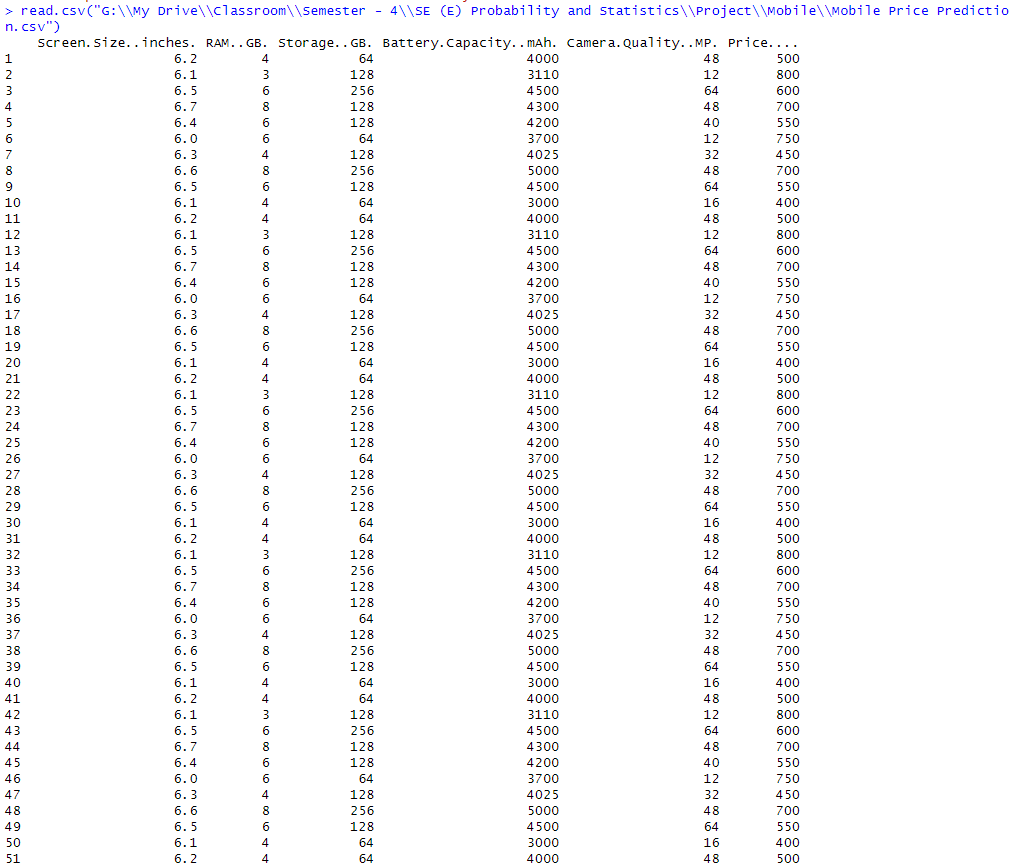
**𝛆**: Error term representing the unexplained variablity

### (c). Review and Interpretation

The regression model reveals the most significant factors influencing mobile phone prices. The positive coefficient for screen size (𝛃1) indicates that larger screens tend to increase in price, likely due to their appeal and manufacturing costs. Higher RAM (𝛃2) and storage capacity (𝛃3) lead to higher prices, indicating a demand for performance and storage space. Battery capacity (𝛃4) and camera quality (𝛃5) have a positive impact on price, indicating that consumers prioritize longer battery life and higher-quality cameras. However, it is important to consider linear regression's limitations, such as linearity assumptions and the possibility of overfitting.

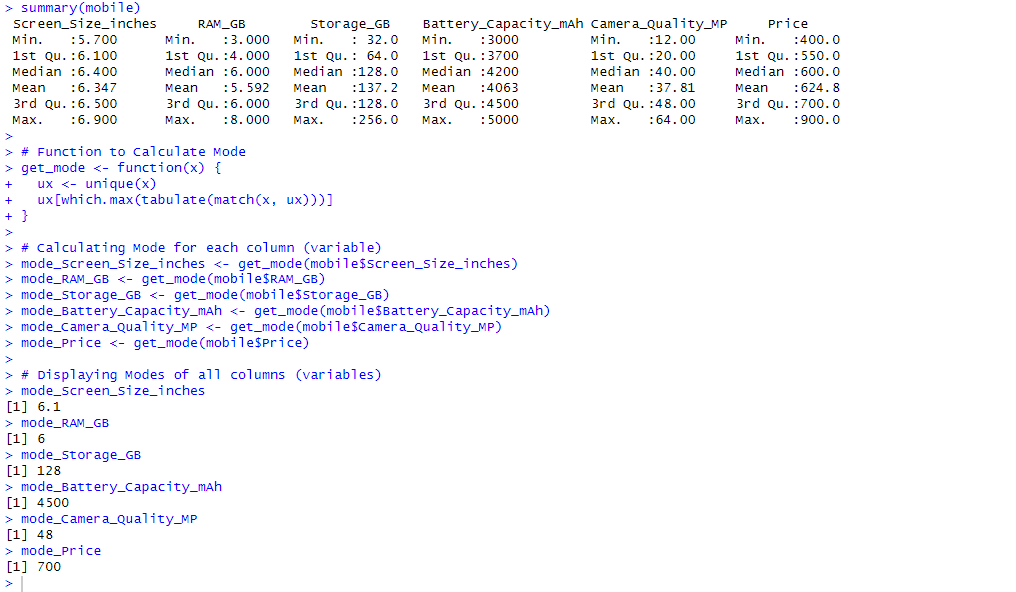
# 4. Description, Results, and Discussion

## (i). Task 1



The above dataset for mobile phone includes enough data so that a regression model can be developed for it in addition to other tests, values and plots that will be generated from the dataset like boxplots, scatterplots, mean, median, mode etc. (Linear Regression DataSets, n.d.)

## (ii). Task 2



### (a). Screen Size (inches)

* Min: 5.7 inches — This is the smallest screen size among the mobiles in the dataset.
* Max: 6.9 inches — This is the largest screen size among the mobiles.
* Median: 6.4 inches — This represents the middle value, where half of the mobiles have screens smaller than 6.4 inches, and half have screens larger.

### (b). RAM (GB)

* Min: 3.0 GB — The minimum amount of RAM among the mobiles in the dataset.
* Max: 8.0 GB — The maximum RAM available in the dataset.
* Median: 6.0 GB — The median RAM value, indicating that half of the mobiles have less than or equal to 6 GB of RAM, and half have more.

### (c). Storage (GB)

* Min: 32.0 GB — This is the smallest internal storage capacity among the mobiles.
* Max: 256.0 GB — This is the largest internal storage capacity.
* Median: 128.0 GB — The median storage capacity, suggesting that half of the mobiles have 128 GB or less, and half have more.

### (d). Battery Capacity (mAh)

* Min: 3,000 mAh — The smallest battery capacity in the dataset.
* Max: 5,000 mAh — The largest battery capacity.
* Median: 4,200 mAh — The median value for battery capacity, indicating that half of the mobiles have 4,200 mAh or less, and half have more.

### (e). Camera Quality (MP)

* Min: 12.0 MP — The lowest camera quality (in megapixels) among the mobiles.
* Max: 64.0 MP — The highest camera quality.
* Median: 40.0 MP — This value indicates that half of the mobiles have 40 MP or less for their cameras, and half have more.

### (f). Price

* Min: 400.0 — The lowest price among the mobiles in the dataset.
* Max: 900.0 — The highest price.
* Median: 600.0 — The median price, indicating that half of the mobiles cost 600 or less, and half cost more.

### (g). Mode of Screen Size (inches)

* Mode: 6.1 inches
* The most frequently occurring screen size among the mobiles in the dataset is 6.1 inches. This indicates that many mobiles share this specific screen size, suggesting it might be a common or popular dimension in the market.

### (h). Mode of RAM (GB)

* Mode: 6 GB
* The mode for RAM indicates that the most common RAM capacity among the mobiles is 6 GB. This could reflect a typical configuration used by manufacturers to balance performance and cost, appealing to a broad range of users.

### (i). Mode of Storage (GB)

* Mode: 128 GB
* A mode of 128 GB for storage implies that this is the most frequently encountered storage capacity in the dataset. This capacity might be considered a standard or middle-ground option, offering sufficient space for typical usage without reaching premium levels.

### (j). Mode of Battery Capacity (mAh)

* Mode: 4,500 mAh
* The most common battery capacity in the dataset is 4,500 mAh, indicating a typical range for modern smartphones. It reflects a balance between battery life and physical size/weight constraints.

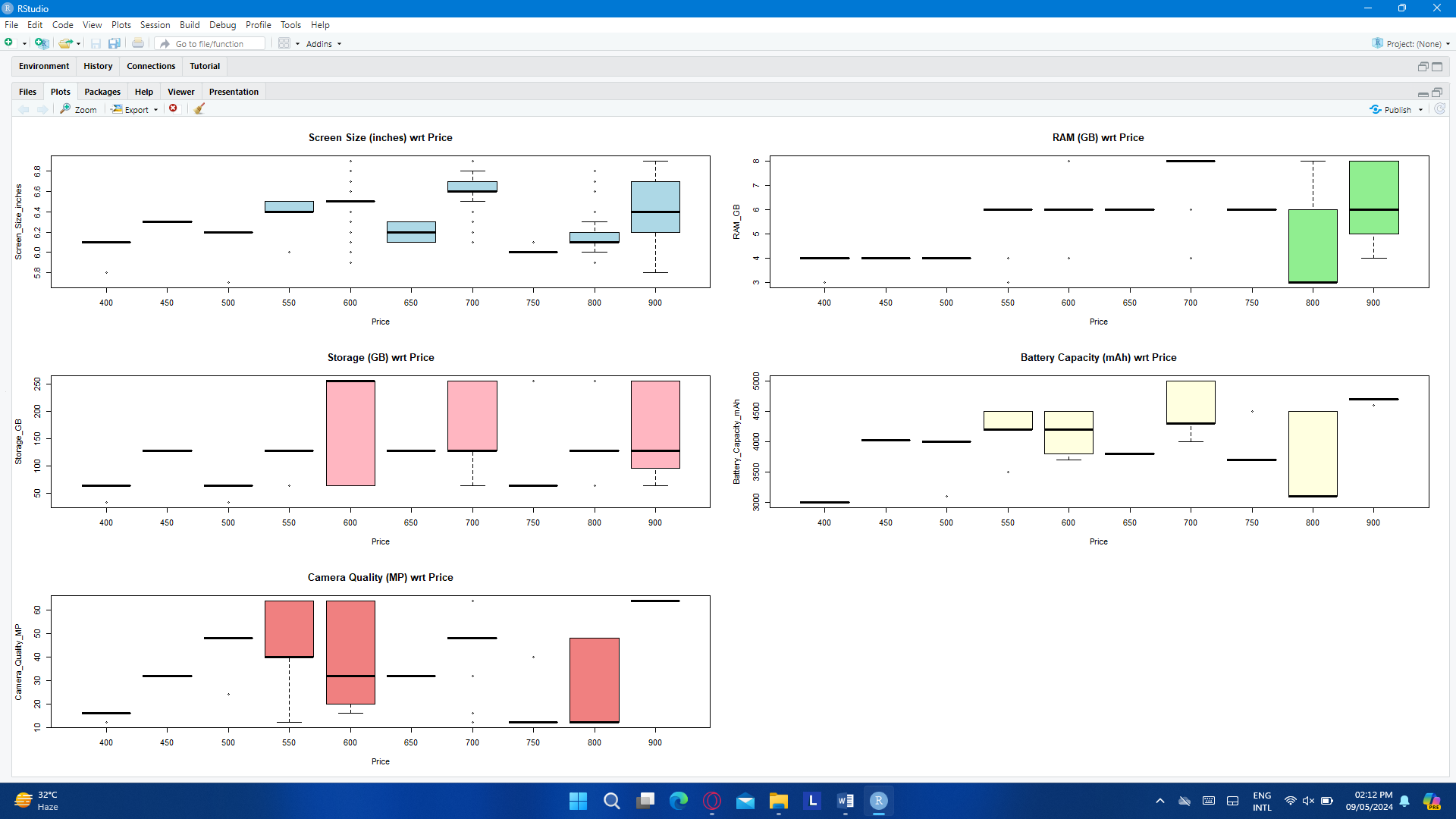
### (k). Mode of Camera Quality (MP)

* Mode: 48 MP
* A mode of 48 MP for camera quality suggests that this is a frequently chosen configuration for smartphone cameras. It may represent a popular choice among manufacturers to offer higher-resolution cameras for improved photo quality and consumer appeal.

### (l). Mode of Price

* Mode: 700
* The mode for price indicates that 700 is the most frequently encountered price point among the mobiles in the dataset. This could be a common price range for mid-range smartphones, balancing features and affordability.

## (iii). Task 3



The above plots represent the box and whisker plot which include the presence or absence of outliers for different variables, which in the case of our chosen dataset are Screen Size (inches), RAM (GB), Storage (GB), Battery Capacity (mAh), Camera Quality (MP) and Price. An outlier is a data point that significantly differs from other observations in the dataset

### (a). Screen Size (inches) Outliers

* There are plenty of mobile phones that are outliers in terms of screen size (inches), e.g a mobile phone with price 400$ and screen size (inches) of 5.8 is an outlier, a mobile phone with price 600$ and screen sizes (inches) of 5.9, 6.0, 6.1 etc. are outliers, meaning that theses values deviate from the typical range of the dataset
* Some Mobile phones don’t have outliers, e.g a mobile phone with price 450$ and 900$ don’t have any outliers, meaning there are no extreme values that significantly deviate from the typical range of the dataset

### (b). RAM (GB) Outliers

* There are few mobile phones that are outliers in terms of RA, e.g a mobile (GB) phone with a price 550$ and RAM (GB)of 3, 4 are outliers, a mobile phone with a price 700$ and RAM (GB) of 4, 7 etc. are outliers, meaning that theses values deviate from the typical range of the dataset
* Most Mobile phones don’t have outliers, e.g a mobile phone with price 450$, 500$ and 750$ don’t have any outliers, meaning there are no extreme values that significantly deviate from the typical range of the dataset

### (c). Storage (GB) Outliers

* There are few mobile phones that are outliers in terms of storage (GB), e.g a mobile phone with price 750$ and storage (GB) of 250 is an outlier, a mobile phone with price 800$ and storage (GB) of 75 are outliers, meaning that theses values deviate from the typical range of the dataset
* Most Mobile phones don’t have outliers, e.g a mobile phone with price 450$, 600$ and 700$ don’t have any outliers, meaning there are no extreme values that significantly deviate from the typical range of the dataset

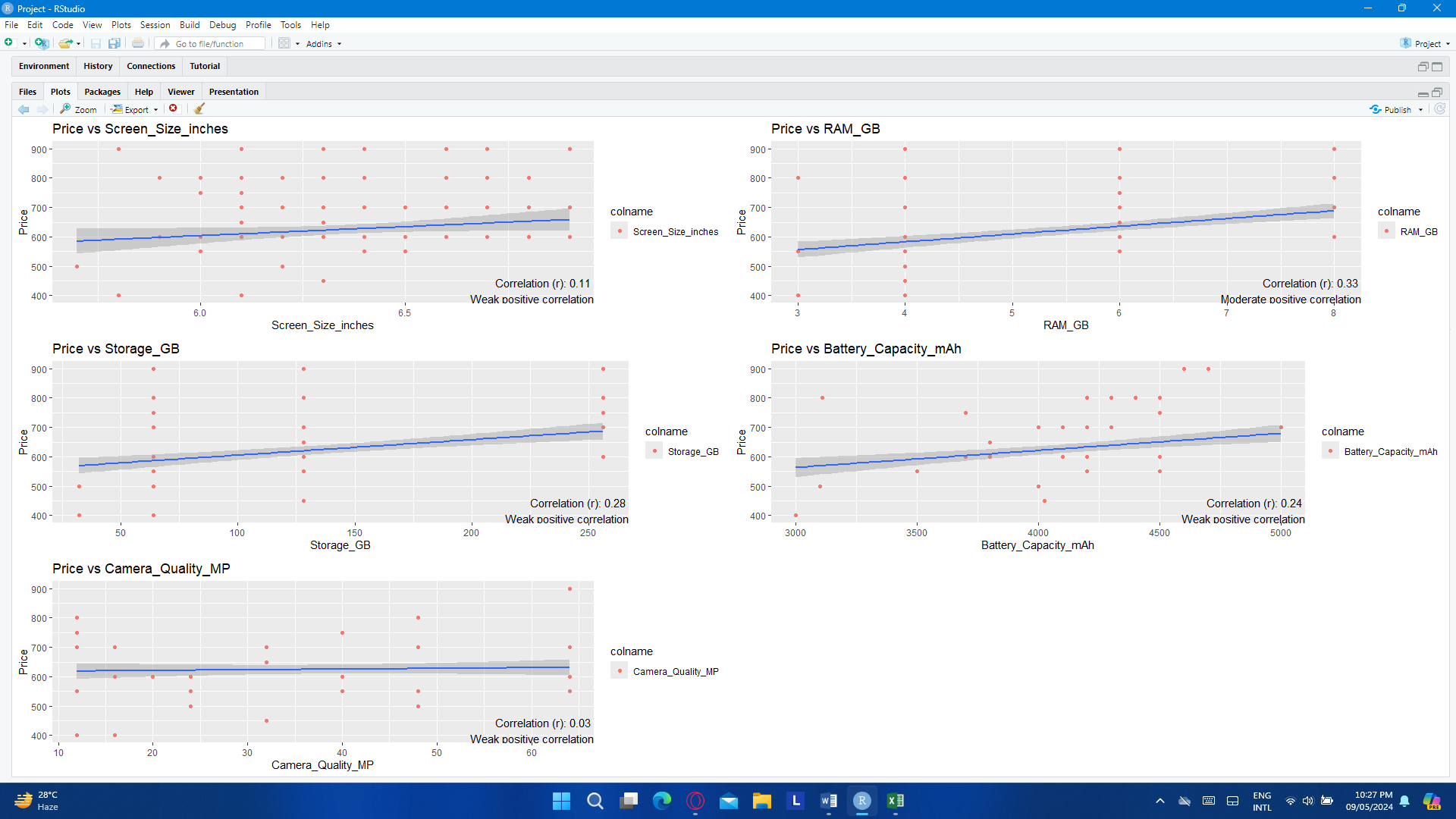
### (d). Battery Capacity (mAh) Outliers

* There are few mobile phones that are outliers in terms of Battery Capacity (mAh), e.g a mobile phone with price 900$ and battery capacity (mAh) of 4750 is an outlier, a mobile phone with price 750$ and battery capacity of 4500 are outliers, meaning that theses values deviate from the typical range of the dataset
* Most Mobile phones don’t have outliers, e.g a mobile phone with prices 400$, 450$ and 600$ don’t have any outliers, meaning there are no extreme values that significantly deviate from the typical range of the dataset

### (e). Camera Quality (MP) Outliers

* There are few mobile phones that are outliers in terms of Camera Quality (MP), e.g a mobile phone with price 400$ and camera quality (MP) of 10 is an outlier, a mobile phone with price 700$ and camera quality (MP) of 30 are outliers, meaning that theses values deviate from the typical range of the dataset
* Most Mobile phones don’t have outliers, e.g a mobile phone with price 450$, 550$ and 650$ don’t have any outliers, meaning there are no extreme values that significantly deviate from the typical range of the dataset

## (iv). Task 4



### (a). Price VS Screen\_Size\_inches

The scatter plot of Price vs. Screen Size (inches) shows a weak positive correlation (r = 0.11). This suggests that as screen size increases, the price rises slightly, but the relationship is not strong. Larger screen sizes may be perceived as premium features, but other factors influence the price. This weak correlation could also be attributed to the availability of phones with varying screen sizes and price ranges, including both budget and high-end models.

### (b). Price VS RAM\_GB

The scatter plot of Price versus RAM reveals a moderate positive correlation (r = 0.33). This correlation suggests that phones with more RAM are more expensive. Higher RAM is typically associated with improved performance, which may explain why these devices are more expensive. However, the correlation's moderate strength suggests that other factors influence price.

### (c). Price VS Storage\_GB

The scatter plot of Price versus Storage shows a weak positive correlation (r = 0.28). Although there is a positive relationship, it is not as strong as one would expect. This could be due to varying pricing strategies across the mobile industry, where some devices with more storage are competitively priced, while others may have additional features that raise the price. It indicates that storage capacity influences the price, but not as significantly as other factors.

### (d). Price VS Battery\_Capacity\_mAh

The scatter plot of Price versus Battery Capacity shows a weak positive correlation (r = 0.24). This suggests that phones with larger battery capacities are more expensive, but the correlation is not particularly strong. This weak correlation could be the result of other design and feature considerations that influence pricing. While a larger battery typically means a longer battery life, this does not always translate into higher prices due to advances in battery technology and different pricing strategies.

### (e). Price VS Camera\_Quality\_MP

The scatter plot of Price versus Camera Quality shows a weak positive correlation (r = 0.03). This suggests that the relationship between price and camera quality is negligible. Although high-end smartphones with advanced cameras are more expensive, the weak correlation could be attributed to other factors such as brand, design, or additional features. It suggests that camera quality alone may not be a significant factor in mobile phone pricing.

### (f). Strength of Correlation

The mobile dataset's scatter plots show a positive correlation between price and other features such as screen size, RAM, storage, battery capacity, and camera quality. The correlation coefficient between price and RAM (r = 0.33) is moderate, indicating that phones with more RAM are generally more expensive. However, other variables, such as screen size, storage, battery capacity, and camera quality, have weaker correlations, with coefficients ranging from 0.03 to 0.28. These weaker correlations indicate that, while prices tend to rise with these characteristics, the relationships are not as strong or consistent.

### (g). Causation VS Associations

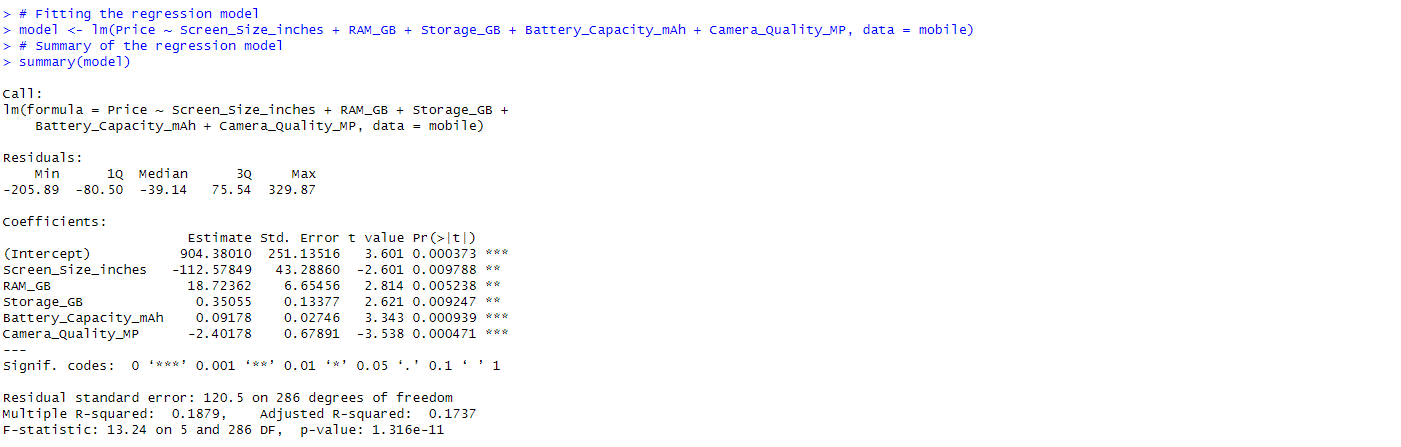
It is critical to understand that a positive correlation does not imply causation. In the case of mobile phones, a number of factors could contribute to the observed relationships. For example, a phone with a larger screen may be more expensive due to other premium features rather than the screen size itself. Similarly, phones with more RAM or storage may have additional performance benefits or brand-related costs. Thus, a positive correlation indicates a relationship between two variables, but it does not imply that one variable causes changes in the other. Confounding variables or underlying technological and consumer preferences may also influence the observed correlations.

### (h). Implications

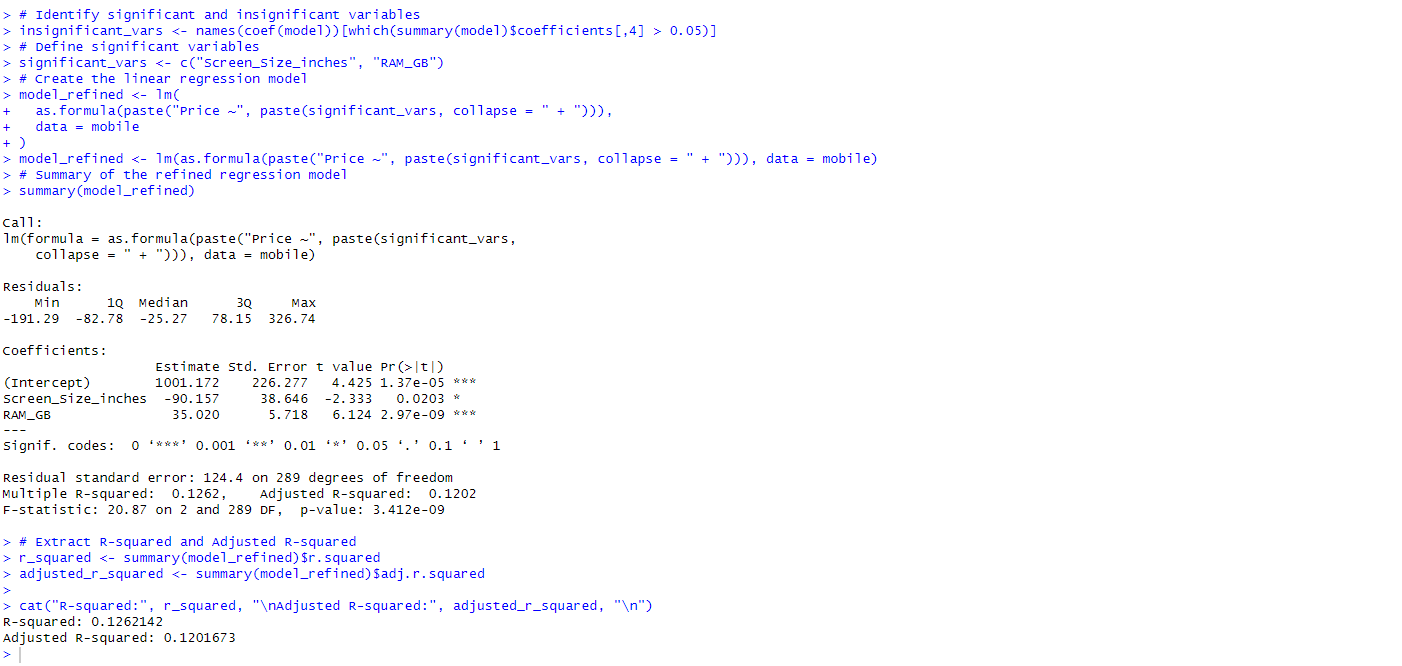
The scatter plots show several positive correlations, which have implications for the mobile phone industry and consumer behavior. For example, the moderate correlation between Price and RAM implies that better performance leads to higher pricing. This insight could help manufacturers with pricing strategies and product differentiation. The weak positive correlations between Price and other variables such as Screen Size and Storage indicate that consumers are willing to pay a premium for specific features, but these are not the only price determinants. Overall, understanding these correlations assists businesses in designing products that meet market demands while remaining cost-effective. It also informs customers about what to expect when they buy mobile phones with various specifications.

## (v). Task 5

## 



## 



An Interpretation from the Results of our Regression Model is given as:

Price = **0 + 𝛃1 . Screen\_Size\_inches + 𝛃2 . RAM\_GB + 𝛃3 .Storage\_GB + 𝛃4 . Battery\_Capacity\_mAh+ 𝛃5 . Camera\_Quality\_MP + 𝛆**

### (a). Model Summary

The model aims to predict the Price of Mobile Phones (Price $) based on Screen Size (inches), RAM (GB), Storage (GB), Battery Capacity (mAh) and Camera Quality (MP)

### (b). Co-Efficients

Intercept (**β0):** The intercept represents the baseline price when all independent variables are zero. In this model, the intercept is estimated at 1001.72, indicating the expected mobile phone price when all other factors are absent or negligible.

Screen\_Size\_inches(**β1):** For every inch increase in screen size, the price is estimated to increase by approximately 90 units. This coefficient suggests a positive and significant relationship between screen size and the price of mobile phone.

RAM\_GB (**β2):** A one GB increase in RAM corresponds to an estimated increase of 35 units in the price. This coefficient suggests that higher RAM is positively associated with a higher price

Storage\_GB(**β3):** An increase of one GB of Storage is associated with an estimated increase of 0.35 units in the price. This coefficient highlights the importance of storage of a mobile phone and its link with the overall price

Battery\_Capacity\_mAh(**β4):** A one-unit increase in battery capacity leads to an estimated increase of 0.09 units in the price. This coefficient suggests that lower levels of battery capacity is linked to higher mobile prices.

Camera\_Quality(**β5):** An increase of one unit of camera quality is associated with an estimated increase of 2.4 units in the mobile price. This coefficient indicates that higher camera quality significantly adds to the mobile phone price.

### (c). Significance and Interpretation

* All coefficients of the regression model are significant at a significance level of **0.05 or lower**. This means that **Screen\_Size\_inches, RAM\_GB, Storage\_GB**, **Battery\_Capacity\_mAh,** and **Camera\_Quality\_MP** significantly influence the price of mobile phones
  + The adjusted **R-squared value of 0.1202** suggests that approximately **12.02%** of the variability in mobile phone prices can be explained by the independent variables included in the model. This indicates a moderately weak fit of the regression model to the data.

### (d). Residuals and Model-Fit

* The residual standard error of **124.4** indicates the average distance of data points from the fitted regression line, representing the model's accuracy in predicting mobile phone prices
  + An F-statistic of 20.87 with a very low p-value of 3.412e-09, we can conclude that the model is statistically significant, suggesting a meaningful relationship between the dependent variable and the independent variables. This result implies that the model has a good fit and is likely capturing key patterns in the data.

# 5. References

*Linear Regression DataSets*. (n.d.). Retrieved from Kaggle: https://www.kaggle.com/datasets/vishwajeetvijaykale/mobile-price-prediction-very-beginner-friendly

*Mobile Phone Price Prediction*. (n.d.). Retrieved from RPubs: https://rpubs.com/Nkunim\_2023/1133431

Prof.B.Prajna. (n.d.). Review of Literature. *Mobile Price Prediction using Machine Learning*.