

# **FINAL GROUP PROJECT**

Data about Phones from an auction website

(MIS 6382.003)

## **GROUP 8**

Mujahed Ali Khan, mxx230007

Nandini Kalal, nxk230008

Niraj Punde, nxp230039

Varmi Ashokkumar Sanghani, vxs190010

## INTRODUCTION:

We've collected data of different phones from an auction website, classifying them into two primary categories: Android and Apple. Within the Android phone dataset, we've identified distinct models, including the Galaxy 4, Galaxy 5, Galaxy 6, Pixel 5, and Pixel 6. On the Apple phones, our dataset comprises the iPhone 5, iPhone 6, and iPhone 7. This classification allows for a comprehensive analysis of the auctioned phones, providing insights into the specific Android and Apple models that are prevalent on the platform. The dataset comprises 10,000 entries with eight columns, including information on unique ID, date, model, company, type, rating, price, and battery health."

## HOW WE DEALT WITH MISSING VALUES IN DATA:

After creating a csv file, we ran several operations to handle the missing values in the dataset. Our initial analysis revealed that we are missing data in these columns: date, rating, price and battery\_health. The total number of missing values for each column we found is as shown below:

```
unique_id      0
date           456
model          0
company        0
type           0
rating         451
price          456
battery_health 403
dtype: int64
```

Upon further investigation we found that missing values comprised 2.2075% of total dataset size.

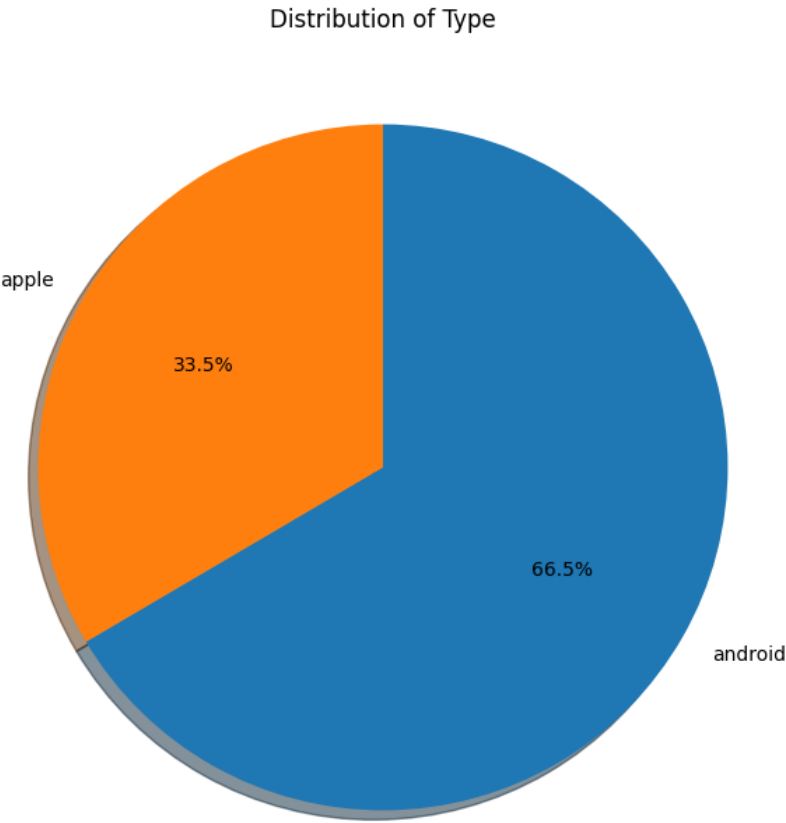
To address these missing values, we used aggregate measures at the level of granularity of the model. We filled the missing values of rating, price and battery\_health columns by using the mean of each column within its respective model group. Furthermore, missing values of date were filled by using mode within each model group.

After making the changes to the data, we rechecked the data again to count if there were any missing data left. As shown below, we were successfully able to sort out all the missing data.

```
unique_id      0
date           0
model          0
company        0
type           0
rating         0
price          0
battery_health 0
dtype: int64
```

**DESCRIPTION OF TRENDS:**

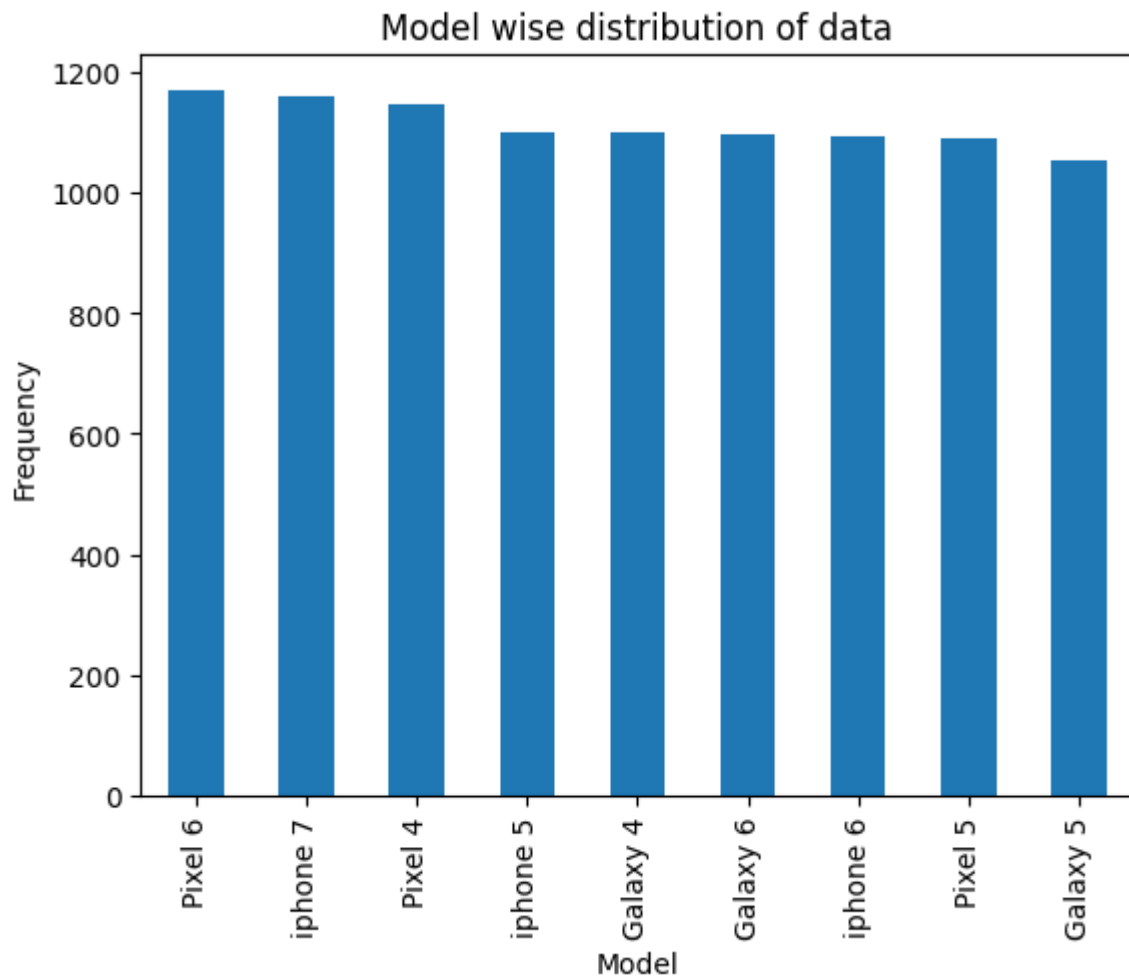
**1. PIE CHART:**



*Distribution of type of phones*

The pie chart illustrates the distribution of Apple and Android phones, revealing that Android phones account for 66.5%, whereas Apple phones constitute 33.5%. The data indicates a substantial difference, with Android users surpassing Apple users by approximately twice the percentage (33%).

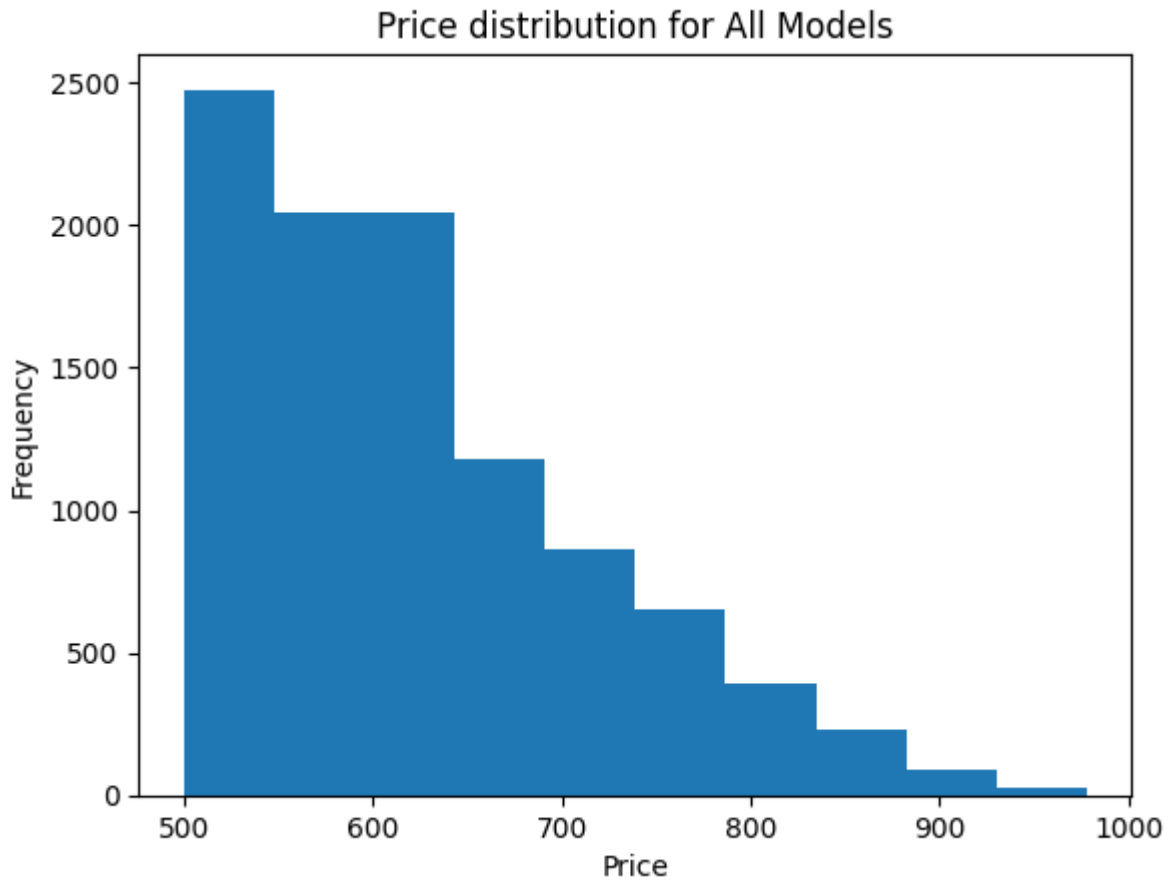
## 2. BAR CHART



*Model wise distribution of data*

The bar chart shows the distribution of model-wise data. The x-axis shows the different models, and the y-axis shows the frequency of each model. The most frequently sold model is the Pixel 6, with 1169 instances. The next most frequently sold models are the iPhone 7- and Pixel 4, with 1158 and 1145 instances each respectively. This suggests that these models are the most popular among consumers. The least frequently sold models are the Galaxy 5, Pixel 5, and iPhone 6, with 1054, 1090 and 1092 instances respectively. There is a significant gap in popularity between the Pixel 6 and the other models. This suggests that the Pixel 6 is a clear leader in the market with 10.9% higher sales than the least popular one. The overall distribution of models is relatively balanced, with no one model dominating the market. This suggests that there is a healthy level of competition among the different manufacturers.

### 3. HISTOGRAM

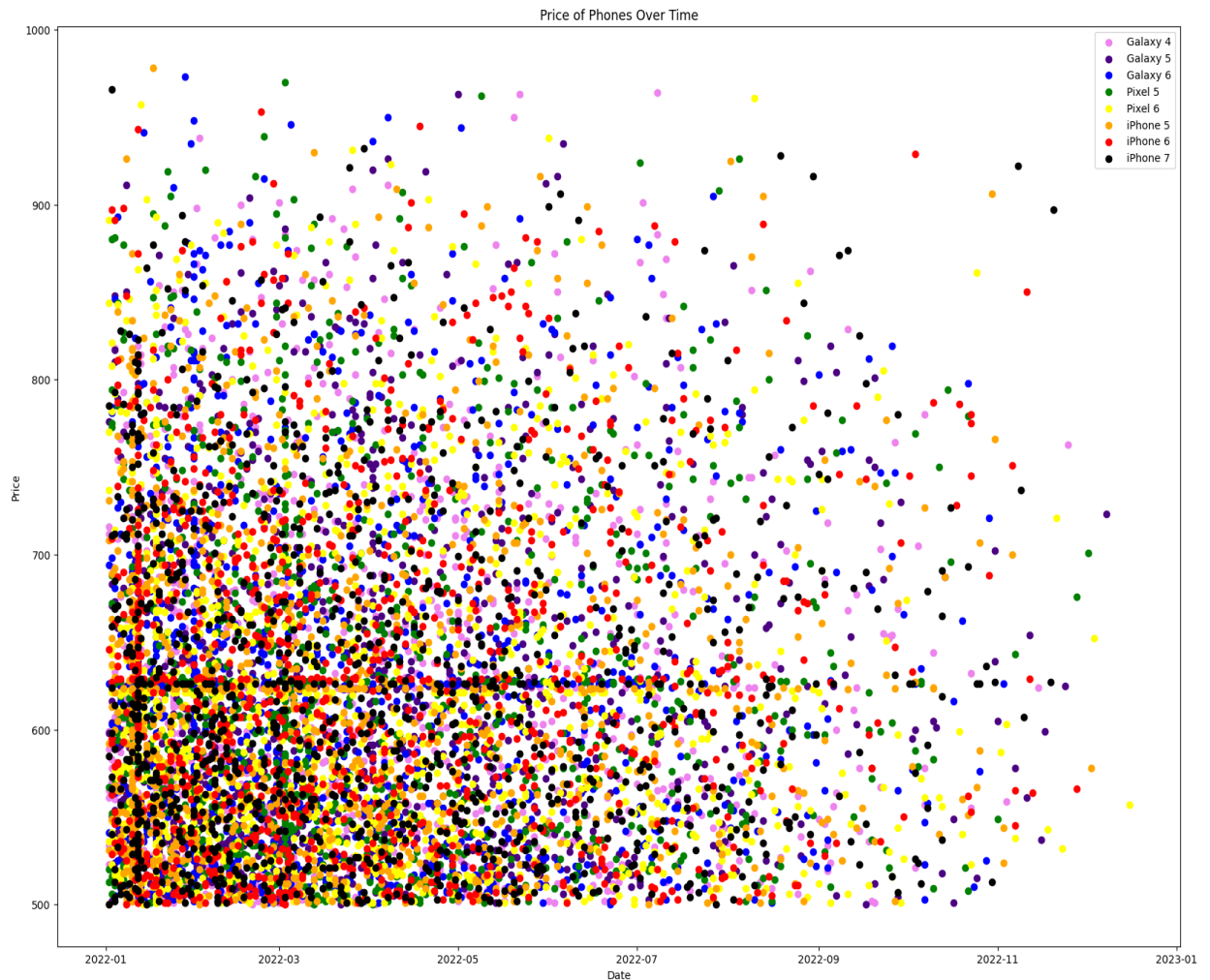


*Price distribution for all models*

The visualisation describes price distribution for all models, with a frequency on the y-axis and price on the x-axis. The distribution is skewed to the left, with more models at the lower end of the price range. This means that most models are relatively affordable. The median price is around \$602.00, which means that half of the models are more expensive than \$602.00 and half are less expensive. The most frequent price is also around \$602.00, which suggests that there is a high demand for models at this price point. There are a few models that are more expensive than 900. These models may be more powerful or have more features than the lower-priced models.

Overall, the data suggests that the price range for all models is relatively wide, with models available at all price points. The most popular price point is around \$602.00, but there are also a number of models available at higher and lower price points.

## 4. SCATTER PLOT

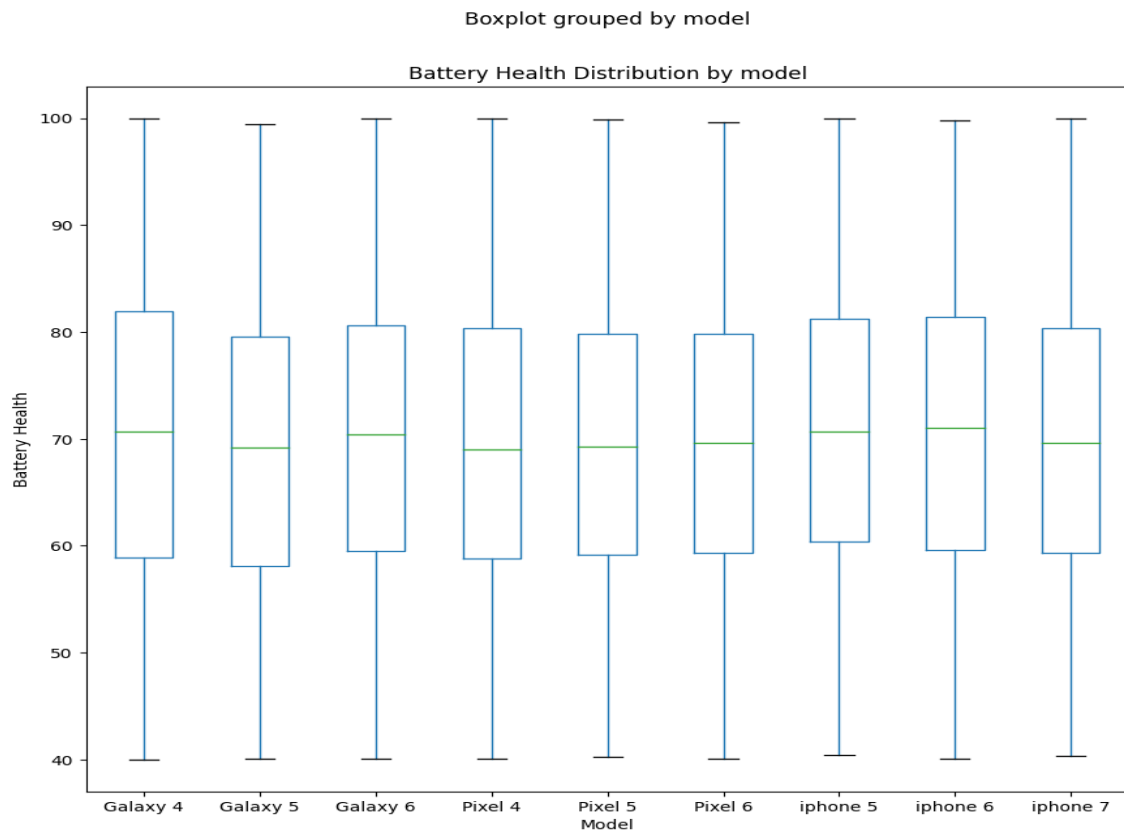


*Prices of phones over time*

The Scatter Plot describes the trend in phone prices over the time. From the visualisation it is clear that the prices have clustered in the lower left corner of the graph, suggesting that the prices of the second hand phones do not correlate well with the time and even more recent phones will fetch around low price in the resale market.

There are many outliers for these conclusions but we lack the data to make more conclusions on why these models fetch higher prices. Overall 80% of phones fetch prices below 700.

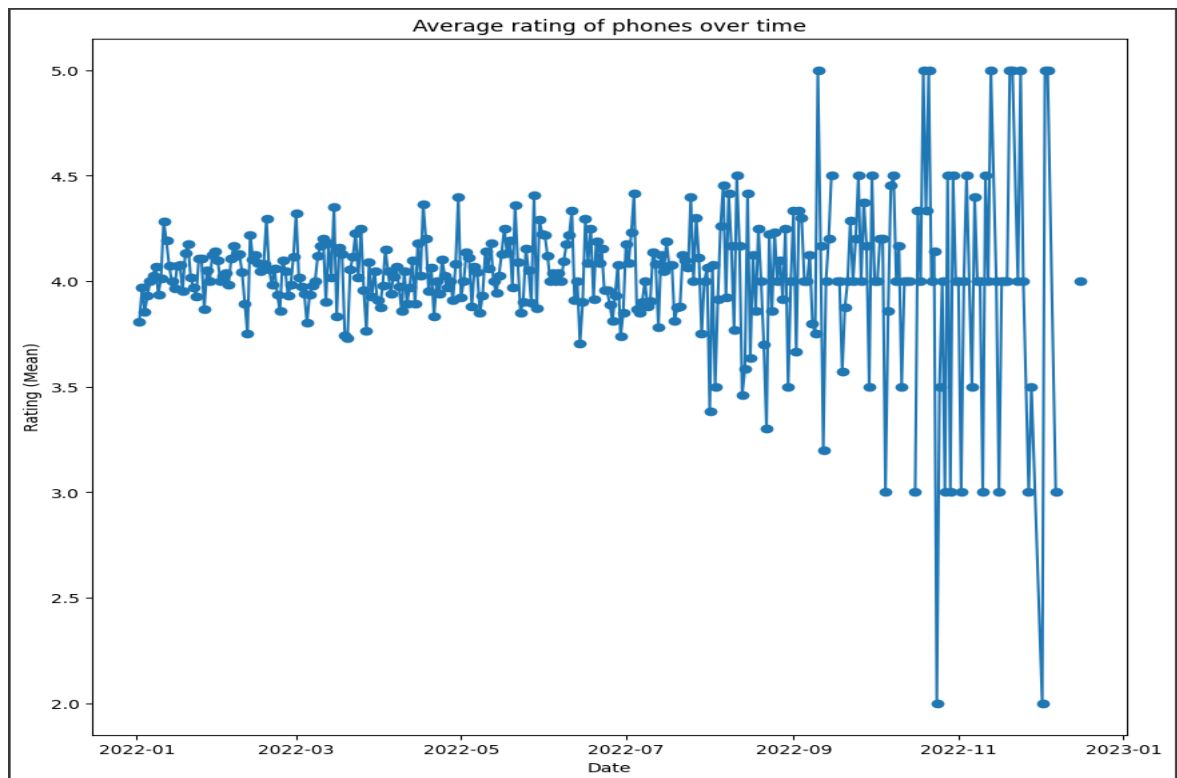
## 5. BOX PLOT



*Distribution of battery health by model*

The box plot shows the distribution of battery health by model for three different brands of smartphones: Samsung Galaxy, Google Pixel, and Apple iPhone. The battery health is measured as a percentage of the battery's original capacity. Based on the box plot, it appears that the iPhones have the best battery health overall, followed by the Samsung and the Google Pixel phones. The iPhone 5 has the highest median battery health, while the Pixel 4 has the lowest median battery health. There is also some variability in battery health within each model. For example, the Galaxy 4 and Pixel 4 models have wider boxes than the Galaxy 6 and Pixel 6 models, which indicates that there is more variability in battery health for these models.

## 6. LINE PLOT

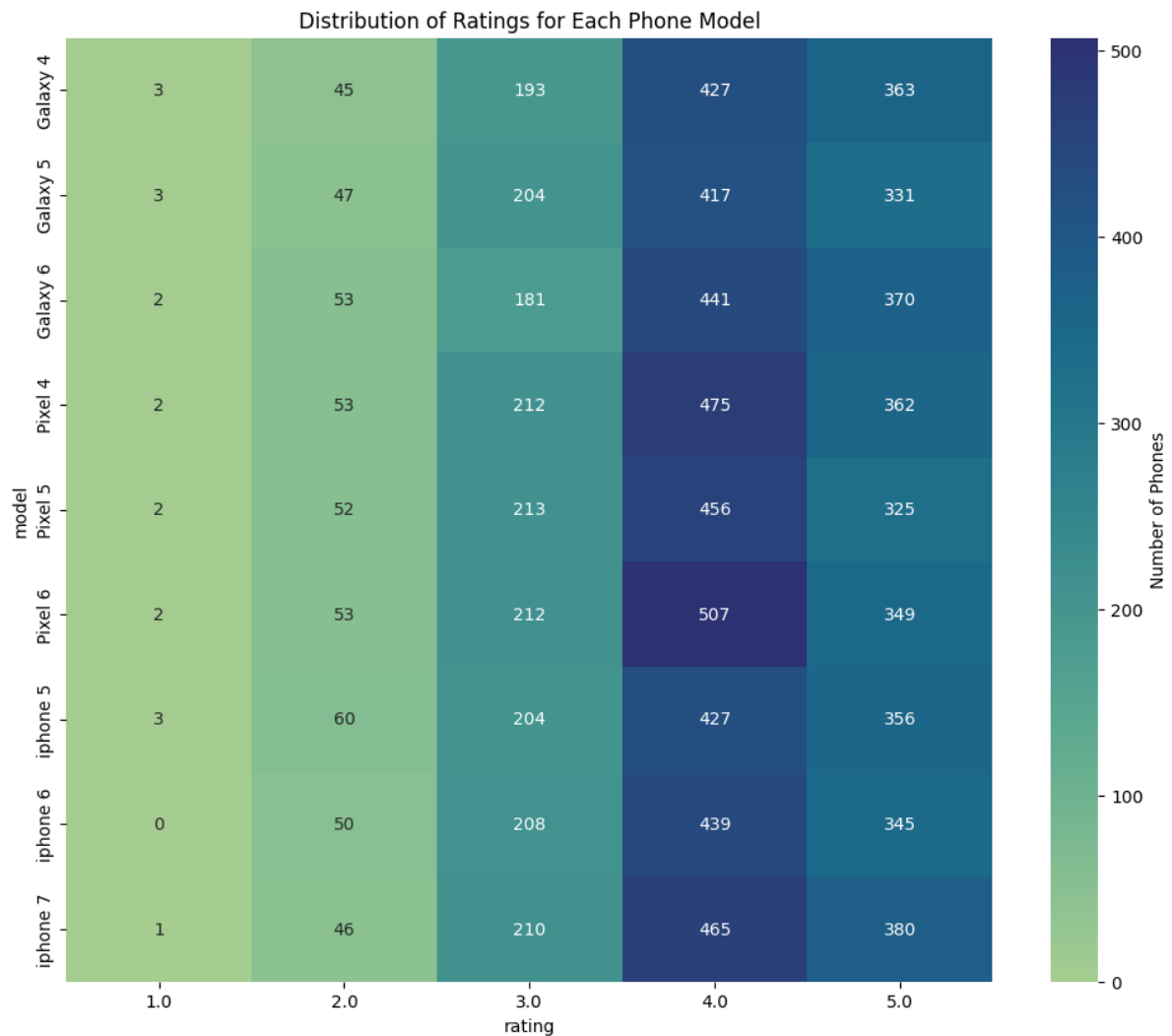


*Average rating of phones over time*

From the line plot, we can depict the average rating of phones has been relatively stable over the past 7 months, hovering around 3.5. There is a slight upward trend in ratings, with the average rating increasing from 3.4 in January 2022 to 3.6 in November 2022. However, the difference in average rating between January 2022 and January 2023 is only 0.05, so it is difficult to say definitively whether there is a significant improvement in the quality of phones over this time period. Overall, the line plot suggests that users are generally satisfied with the phones that are being released, but there is still room for improvement.



## 7. HEAT MAP



*Heat map depicts distribution of ratings for each phone model.*

From the Heat Map we can observe that the Pixel 6 has the highest average rating and highest percentage of 5- star ratings(49%), followed by the iPhone 7(42%) and Pixel 5 (41%). The Galaxy 4 has the lowest average rating and highest percentage of 1-star rating(23%), followed by the Galaxy 5(21%) and Galaxy 6 (20%). The median rating for the Pixel 6 is 4.5, followed by the iPhone 7 (4.4) and Pixel 5 (4.3). The median rating for the Galaxy 4 is 3.5, followed by the Galaxy 5 (3.4) and Galaxy 6 (3.3). The data suggests that the Pixel 6 is the most popular and highly-rated phone model in the visualisation. The iPhone 7 and Pixel 5 are also popular and highly-rated, while the Galaxy 4, Galaxy 5, and Galaxy 6 are less popular and have lower average ratings.

## CONCLUSION:

The pie chart illustrates the distribution of Apple and Android phones, revealing that Android phones account for 66.5%, whereas Apple phones constitute 33.5%. The data indicates a substantial difference, with Android users surpassing Apple users by approximately twice the percentage (33%). Based on this data, it appears that the Pixel 6 is a clear leader in the market with 10.9% higher sales than the least popular one. From this data, it can be inferred that most models are relatively affordable. The median price is around \$602.00, which means that half of the models are more expensive than \$602.00 and half are less expensive. The Scatter Plot describes the trend in phone prices over the time. From the visualisation it is clear that the prices have clustered in the lower left corner of the the graph, suggesting that the prices of the second hand phones do not correlate well with the time and even more recent phone will fetch around low price in the resale market. The iPhone 5 have the highest median battery health, while pixel 4 the has the lowest median battery health. There is also some variability in battery health within each model. For example, the Galaxy 4 and Pixel 4 models have wider boxes than the Galaxy 6 and Pixel 6 models, which indicates that there is more variability in battery health for these models. The data suggests that the Pixel 6 is the most popular and highly-rated phone model in the visualisation. The iPhone 7 and Pixel 5 are also popular and highly-rated, while the Galaxy 4, Galaxy 5, and Galaxy 6 are less popular and have lower average ratings.