

ADSC1000
Statistical Data Analysis

Thompson Rivers University
Fall-2023



Mobile Price Classification

Submitted To:
Prof. Sean Hellingman

Submitted By:
Akansha Bhargavi-T00736533
Solomon Maccarthy- T00734513
Viswateja Adothi-T00736529

Table of Contents

| | |
|--|----------|
| 1.Introduction | 3 |
| 1.1 Background | 3 |
| 1.2 Objective | 3 |
| 1.3 Scope of Project | 3 |
| 1.4 Significance and Motivation..... | 3 |
| 2. Methodology..... | 3 |
| 2.1 Data Source | 3 |
| 2.2 Variables and Measures..... | 4 |
| 2.3 Data Analysis Techniques..... | 5 |
| 3.Exploratory Data Analysis (EDA) | 6 |
| 3.1 Data Visualization | 6 |
| 3.1.1 Categorical Variable Visuals..... | 6 |
| 3.1.2 A pictorial view of the Price Range in percentage (Pie Chart) | 6 |
| 3.1.3 A pictorial view of four_g (Bar Chart) | 6 |
| 3.2 Continues Variable Visuals..... | 7 |
| 3.2.1 A pictorial view of battery power, RAM, px_width and px_height (Histogram) | 7 |
| 3.2.2 A pictorial view of battery power, RAM, px_width and px_height (box plot) | 8 |
| 4. Hypothesis Tests | 8 |
| 4.1 | 8 |
| 4.1.2 Chi-Square Test | 9 |
| 4.1.3 Anova hypothesis test..... | 10 |
| 4.1.4. Shapiro-Wilk test and Levene's test..... | 10 |
| 4.1.5 Kruskal-Wallis Test | 10 |
| 5. Results and Findings..... | 11 |
| 6.Conclusion..... | 11 |
| 7.Appendix | 11 |
| 7.1 References | 11 |
| 7.2 Code | 11 |

1.Introduction

1.1 Background

Starting one's own mobile company and giving a tough fight to big companies like Apple, Samsung etc. can be one of a great deal of an adventure. Bob an entrepreneur is keen on penetrating in the mobile phone creating and selling space. A major factor to this quest is Pricing of his new device. Mobile phones come in various models, each with different features and specifications. Our analysis helps in understanding the diverse factors influencing mobile phone prices.

1.2 Objective

Sales data collected from various mobile phone companies to specifically find out relation between features of a mobile phone and its selling price. We as data science students have been tasked to use our knowledge to help relate mobile phone features and that of setting price of mobile phones.

1.3 Scope of Project

The project's scope encompasses in-depth feature analysis, statistical modeling in R, and leveraging a substantial dataset of 2,000 observations to classify mobile phones into distinct price ranges. The goal is to provide consumers with valuable insights for informed decision-making and contribute a systematic approach relevant to the competitive mobile device industry.

1.4 Significance and Motivation

The significance and motivation of this project is to exhibit our knowledge acquired in class on real-life situations and to prove our ability on what we have understood so far in the course under study.

2. Methodology

2.1 Data Source

Mobile price dataset is collected from Kaggle to find out relation between features of a mobile phone and its association with pricing.

Target Population: Our target population is all available mobile phones in the market.

Sample size: Mobile phones released from 2012 to 2016.

2.2 Variables and Measures

We will use the Mobile Price Classification data set which contains various variables both continues and categorical.

| Feature | Description | Variable type |
|---------------|--|---------------|
| Battery Power | Total energy a battery can store in one time measured in mAh | Continuous |
| Blue | Has Bluetooth or not | Categorical |
| Clock speed | Speed at which microprocessor executes instructions | Continuous |
| Dual_sim | Has dual sim support or not | Categorical |
| Fc | Front Camera mega pixels | Continuous |
| four_g | Has 4G or not | Categorical |
| Int_memory | Internal Memory in Gigabytes | Continuous |
| m_dep | Mobile Depth in cm | Continuous |
| mobile_wt | Weight of mobile phone | Continuous |
| n_cores | Number of cores of processor | Continuous |
| Pc | Primary Camera mega pixels | Continuous |
| px_height | Pixel Resolution Height | Continuous |
| px_width | Pixel Resolution Width | Continuous |
| Ram | Random Access Memory in Megabytes | Continuous |

| | | |
|--------------|--|-------------|
| sc_h | Screen Height of mobile in cm | Continuous |
| sc_w | Screen Width of mobile in cm | Continuous |
| talk_time | longest time that a single battery charge will last when you are | Continuous |
| three_g | Has 3G or not | Categorical |
| touch_screen | Has touch screen or not | Categorical |
| Wifi | Has wifi or not | Categorical |

But for the purpose of our analysis, we will be focusing on these features – Battery Power ,Ram,Internal Memory,Pixel Width,Pixel Height,Four_g,Three_g

Price range was categorized into the below:

- 0 (low cost)
- 1 (medium cost)
- 2 (high cost)
- 3 (very high cost)

2.3 Data Analysis Techniques

Our data analysis techniques used were test ranging from a one-sample hypothesis test to determine the means of one major variable used in our study battery_power. A Two sample hypothesis test was also employed to verify the significant difference between battery_power and four_g networks and the difference in RAM and dual_sim. An Anova test was conducted to test the different means of our preferred variables and to check the normality of our data set a Shapiro test was conducted which proved that our data set was not normally distributed. Due to our non-normality of our data set a levene test was conducted to check how equal our homogeneity variance and our sample are. Due to violation of Anova assumptions, we performed Kruskal Wallis test. To verify the independence of the variables of our project a Chi-square test was conducted to prove whether our categorical variables were dependent or not to our Price range.

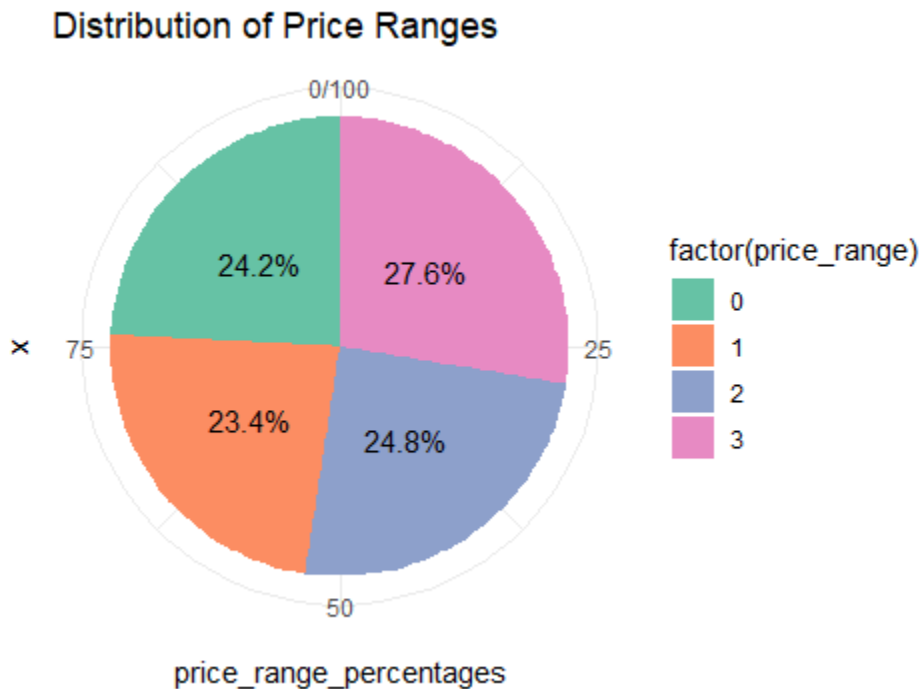
3.Exploratory Data Analysis (EDA)

3.1 Data Visualization

Visualizations, including Pie chart, Bar chart, histogram, boxplot were employed to discover trends and associations within the data.

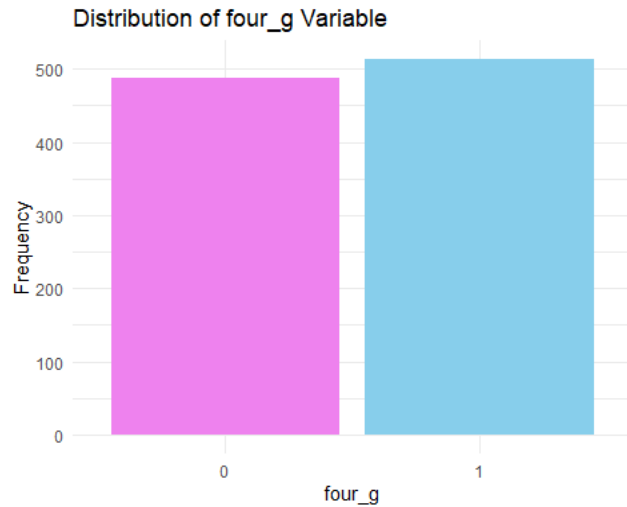
3.1.1 Categorical Variable Visuals

3.1.2 A pictorial view of the Price Range in percentage (Pie Chart)



From the visualization we see that category 3 has the most phones priced in that range. Category 0 and 2 are almost equal in terms of price range distribution in the mobile phone market. Category 1 has less pricing range terms of percentage.

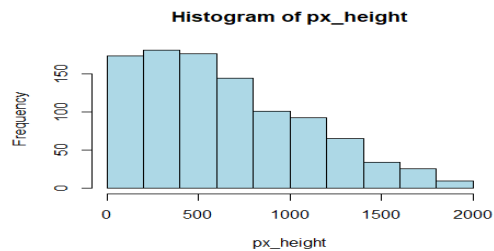
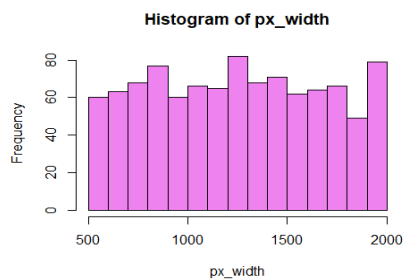
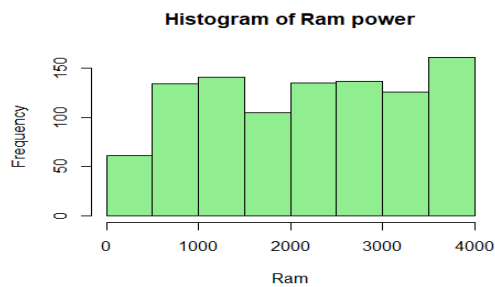
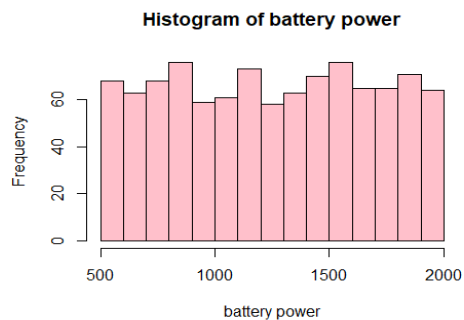
3.1.3 A pictorial view of four_g (Bar Chart)



From the visualization we can state that mobiles with four_g are high in number compared to mobiles without four_g. The purple bar represents a non four_g mobile phone while the blue bar represents that of four_g.

3.2 Continues Variable Visuals

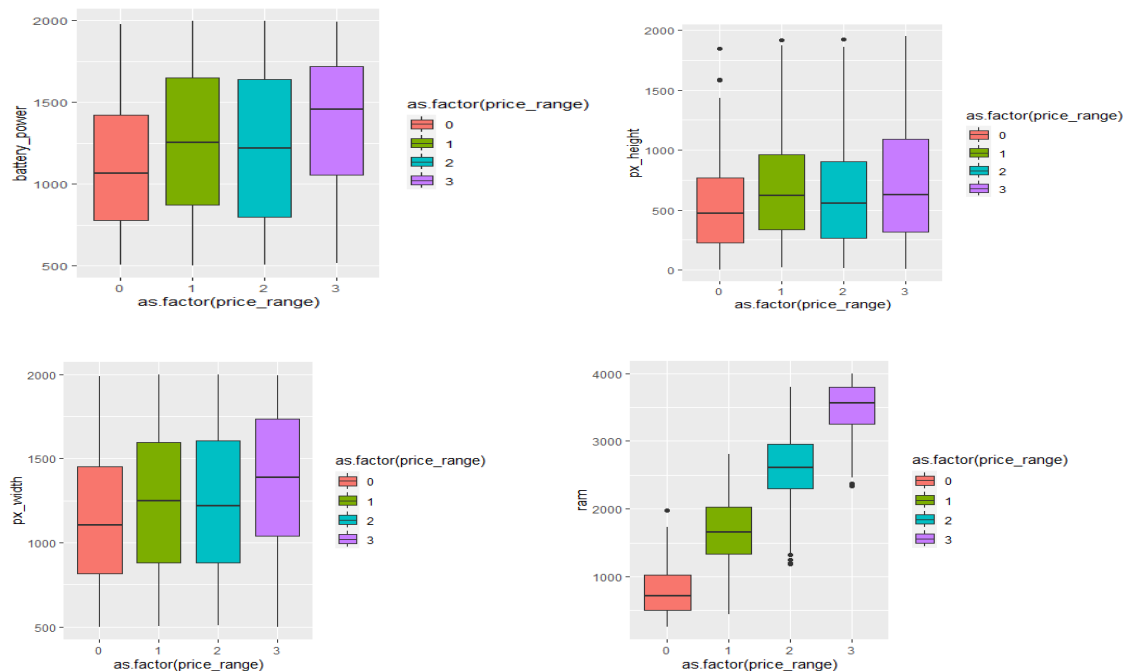
3.2.1 A pictorial view of battery power, RAM, px_width and px_height (Histogram)



From the visualization we can see the various distributions of the variables being shown via a histogram.

PX_Height histogram is right skewed and all the other three histograms are mostly uniformly distributed

3.2.2 A pictorial view of battery power, RAM, px_width and px_height (box plot)



From the visualization we can see that the various medians are very close and also no major significant difference in the first three visuals. But for that of the fourth visual (RAM) the means are significantly different.

4. Hypothesis Tests

4.1 A one & two sample hypothesis test

| Test | Null Hypothesis (H0): | Alternative Hypothesis (H1): | Result |
|-------------------|--|--|--------|
| One Sample t-test | The average battery power of mobile phones in the population is equal to or less than 1238 mAh | The average battery power of mobile phones in the population is greater than 1238 mAh. | 0.1466 |

| | | | |
|-------------------------|---|--|---------|
| Welch Two Sample t-test | There is no difference in the average battery power between phones with and without 4G. | There is a significant difference in the average battery power between phones with and without 4G. | 0.02489 |
| Welch Two Sample t-test | There is no difference in the average RAM between phones with and without dual SIM | There is a significant difference in the average RAM between phones with and without dual SIM. | 0.03928 |

From the results of one-sample t-test we can say that we don't have enough evidence to reject the null hypothesis which means the average battery power of mobile phones in the population is equal to or less than 1238 mAh

As p-values are less than 0.05 for both two sample t-test we can say that there is a significant difference in the average battery power between phones with and without 4G and there is a significant difference in the average RAM between phones with and without dual SIM.

4.1.2 Chi-Square Test

A chi-square test results showing the independence or dependence of four_g and three_g against price range.

| Null Hypothesis (H ₀): | Alternative Hypothesis (H _a): | P-value |
|---|--|---------|
| There is no significant association between the 'Three_G' variable and 'Price_Range' in the population. | There is a significant association between the 'Three_G' variable and 'Price_Range' in the population. | 0.2994 |
| There is no significant association between the 'Four_G' variable and 'Price_Range' in the population. | There is a significant association between the 'Four_G' variable and 'Price_Range' in the population. | 0.1713 |

From the results of one-sample t-test we can say that we don't have enough evidence to reject the null hypothesis which states that there is no significant association between the 'Three_G' variable and 'Price_Range' in the population and there is no significant association between the 'Four_G' variable and 'Price_Range' in the population.

4.1.3 Anova hypothesis test

Higher F-Value suggests larger difference between group means and is more significant. Lower p-value indicates higher level of significance.

| Feature | F-Value | P-Value |
|-----------------|---------|----------|
| RAM | 5568 | <2e-16 |
| Battery Power | 48.42 | 6.23e-12 |
| Pixel width | 25.01 | 6.74e-07 |
| Pixel height | 16.2 | 6.12e-05 |
| Internal memory | 3.268 | 0.071 |
| Primary Camera | 2.24 | 0.135 |

Ram, Battery power, Pixel width, Pixel Height have average mean difference across the price range, where as Internal memory with p-value of 0.071 is marginally significant. Primary camera has no significance as p-value is greater than 0.05.

Here Ram has Higher F-Value and Lower P-Value which says there is strong association between independent variable (price range) and the dependent variable (RAM).

4.1.4. Shapiro-Wilk test and Levene's test

We conducted Shapiro-Wilk test to assess the normality of a distribution. As the p-value is less than 0.005, we concluded that the data is not normally distributed. We also performed Levene's test to assess the homogeneity of variances across different groups. The p-value from Levene's test was less than the 0.005, suggesting that there are significant differences in variances across groups.

4.1.5 Kruskal-Wallis Test

As our mobile data is not normally distributed and the variances are significantly different across groups, we consider using a non-parametric alternative to ANOVA, such as the Kruskal-Wallis test.

| Feature | P-Value |
|-----------------|-----------|
| RAM | 2.2e-16 |
| Battery Power | 1.949e-11 |
| Pixel width | 1.808e-06 |
| Pixel height | 0.0002366 |
| Internal memory | 0.1356 |
| Primary Camera | 0.4012 |

There are significant differences in battery power, pixel height, RAM, and pixel width across different price ranges.

There are no significant differences in internal memory and PC across different price ranges. There are no significant differences in internal memory and PC across.

There are significant differences in battery power, pixel height, RAM, and pixel width across different price ranges. There are no significant differences in internal memory and PC across different price ranges.

5. Results and Findings

Higher RAM, larger battery power, and better camera specifications could contribute to higher price ranges. Whereas primary camera and internal memory may not be strong indicators of differences in mobile phone prices. The network type alone may not be a strong predictor of price range. The dual SIM phones may have different RAM specifications compared to single SIM phones. 4G capability may be a factor influencing the battery power specifications of mobile phones.

6. Conclusion

This project provides a useful information for both consumers and the mobile industry by helping people make informed decisions about buying phones. It systematically analyzes key features to understand how factors like RAM and camera quality influence the price, making it easier for everyone involved to navigate the dynamic mobile market.

7. Appendix

7.1 References

<https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

Team Work:

We collaborated together to interpret the tests that we can perform using Mobile Price dataset where we divided the work of analysis required to draw the statistics and documentation equally.

7.2 Code