

Mobile Phones Selling Price Report

Proposal for final project (MDSA Winter 2023)

Yongpeng Fu (10182778), Rudy Brown (10057171), Jose Palacios (30190988),
Stuart Finley(30191070), Andrii Voitekiv(30199373)

2023-01-31

Contents

Introduction	2
Dataset	2
Scope of Analysis	3
References	4

Introduction

Mobile phones are everywhere, so are the prices. Despite still having the word “phone” in the name, a typical modern smartphone has much more features than just to make and receive calls. They are boasting a staggering range of applications like brand, memory, storage, camera, resolution, just to name a few. All these cutting edge technology and features packed in one little device does not come without a cost. A [2020 review](#) of premium mobile phones shows a staggering 490% rise in the last two decades.

With so many mobile phones on the market, it can be difficult to decide which one you want to buy. As a customer, we are particularly interested in finding some relation between all these features and its selling price. To this purpose, we collected the [MobilePhone's dataset](#) from Kaggle and apply a set of statistical analysis hoping to answer some guiding questions:

1. Can we estimate the average price for mobile phones?
2. What is the impact of each mobile phone's feature on the selling price?
3. Can a classification model to distinguish the selling price range?
4. Can we build a decent model to predict the selling price for a mobile phone?

Dataset

The initial dataset consists of 8 columns and 28,036 rows and no missing values. These 8 columns are:

- **Model:** categorical variables with sub-classes. These names include the color of the unit and its storage capacity. The latter being also listed as a separate column. - **Independent Variable**
- **Company:** categorical variable. Name of the phone's manufacturer. - **Independent Variable**
- **Price:** continuous variable. Units in Indian Rupees. - **Dependent Variable**
- **Rating:** continuous variable. Units in Indian Rupees. - **Independent Variable**
- **Number of ratings:** discrete variable: a simple count. - **Independent Variable**
- **Total reviews:** discrete variable: a simple count. - **Independent Variable**
- **RAM size:** categorical variable. RAM specification of the phone. - **Independent Variable**
- **ROM size:** categorical variable. Storage (non-volatile memory) capacity of the phone. - **Independent Variable**

Some initial steps can be completed to clean the dataset and create new variables which can be used in our analysis. The initial steps for cleaning the dataset are as follows:

1. Remove any duplicates in the dataset;
2. Because **Model** column contains sub-class of a mobile phone, we decide to further break it down to *Model* and *Color*;
3. Add additional column to segment the **Price** into 4 different levels.

After cleaning and breaking down columns, the dataset now consists of 11 columns and 736 rows and no missing values. These 11 columns are:

- **Model:** categorical variables with sub-classes. These names include the color of the unit and its storage capacity. The latter being also listed as a separate column. - **Independent Variable**
- **Company:** categorical variable. Name of the phone's manufacturer. - **Independent Variable**
- **Price:** continuous variable. Units in Indian Rupees. - **Dependent Variable**
- **Rating:** continuous variable. Units in Indian Rupees. - **Independent Variable**
- **Number of ratings:** discrete variable: a simple count. - **Independent Variable**
- **Total reviews:** discrete variable: a simple count. - **Independent Variable**
- **RAM size:** categorical variable. RAM specification of the phone. - **Independent Variable**
- **ROM size:** categorical variable. Storage (non-volatile memory) capacity of the phone. - **Independent Variable**

- **aug_model**: categorical variable. It only contains the Model information for a mobile phone. - **Independent Variable**
- **aug_color**: categorical variable. The color of a mobile phone. - **Independent Variable**
- **aug_price**: categorical variable. The price level of a mobile phone, with levels of “Low”, “Medium”, “High”, “Very High” - **Independent Variable**

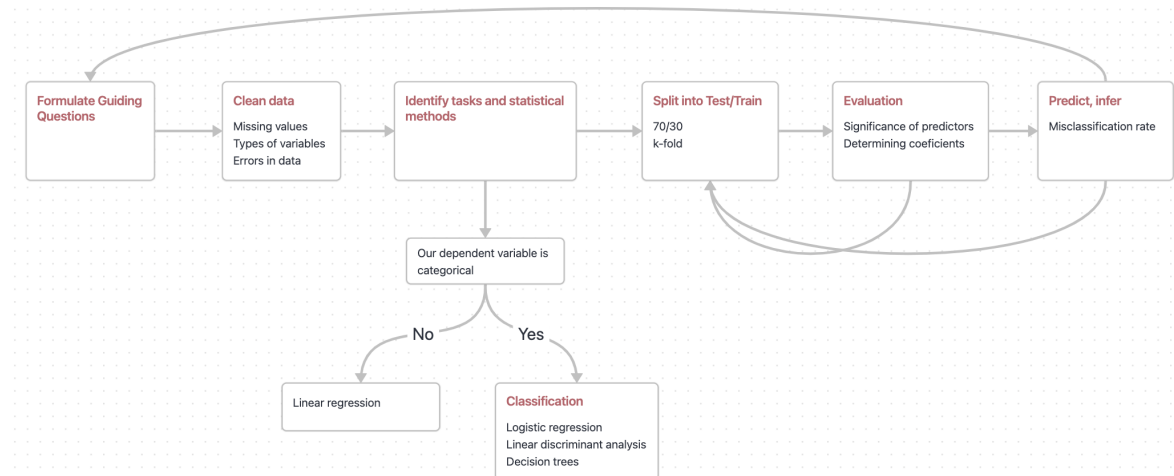
```
mobile_dataset <- as_tibble(read.csv("./augmented_dataset.csv"))
mobile_dataset %>% head(4)
```

```
## # A tibble: 4 x 11
##   Model      Company Price Rating No_of~1 Total~2 RamSize RomSize aug_m~3 aug_c~4
##   <chr>      <chr>   <int> <dbl> <int> <int> <chr> <chr> <chr> <chr>
## 1 Infinix ~ Infinix  8199   4.3   505    52 4 GB  64 GB  "HOT 2~ Luna B~
## 2 MOTOROLA~ Motoro~  7999   4.1  56085  5600 4 GB  64 GB  "e40 " Carbon~
## 3 MOTOROLA~ Motoro~  7999   4.1  56085  5600 4 GB  64 GB  "e40 " Pink C~
## 4 POCO C31~ Poco    7499   4.3  183688  11185 4 GB  64 GB  "C31 " Shadow~
## # ... with 1 more variable: aug_price <chr>, and abbreviated variable names
## #   1: No_of_ratings, 2: TotalReviews, 3: aug_model, 4: aug_color
```

The dataset and detailed analysis can be found at this [repository](#).

Scope of Analysis

Our team is finalizing what the full analysis of the dataset will look like, but a preliminary template and breakdown of work by team member has been included below. The different colors represent which components of the project different team members would take on. It is anticipated that all members will assist in the finalization of the report.



References

2020 review. “High-End Mobile Phones Price Have Soared 490% in 20 Years | This Is Money.” This Is Money, This Is Money, 23 July 2020, <https://www.thisismoney.co.uk/money/bills/article-8548235/High-end-mobile-phones-price-soared-490-20-years.html>.

MobilePhone’s dataset. “MobilePhone’s Dataset | Kaggle.” Kaggle: Your Machine Learning and Data Science Community, Kaggle, 20 Dec. 2022, <https://www.kaggle.com/datasets/sudhanshuy17/mobilephone>.