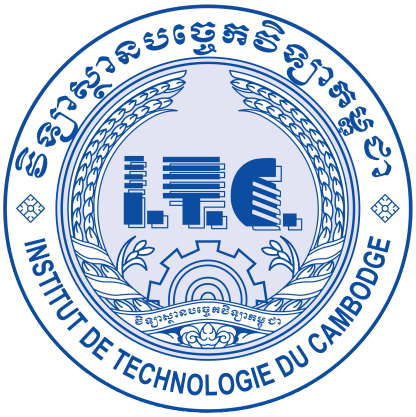

Institute of Technology of Cambodia

Programming for Data Science
2022-2023

3rd year Engineer's Degree in Data Science

Department of Applied Mathematics and Statistics

Phone Price Analysis

Group members:

Name	ID
Lay Chhay	e20200054
Nor Phanit	e20200241
Nhar Ratanak	e20190682
Noem Koemhak	e20200808
Phal Davy	e20201437

Lecturer:

Mr. Chan Sophal

Contents

1 Introduction

This preface serves as the basis for the Phone Price Analysis report and provides a comprehensive overview of the importance and goals of understanding the factors that influence phone prices. In the ever-evolving mobile phone market, understanding the drivers of mobile phone pricing is critical for consumers, sellers and manufacturers alike. As technology advances and competition intensifies, consumers are given different choices from different brands. Therefore, it is important to gain insight into the factors that drive mobile phone prices.

The purpose of this report is to analyze mobile phone prices in detail, especially focusing on his two famous brands. Samsung and iPhone. By delving into the dataset and exploring the complex relationships between different features and mobile phone prices, we hope to gain valuable insights that can drive decision-making across the mobile phone industry. Understanding the underlying factors that drive mobile phone prices is critical for consumers who want to make informed purchasing decisions. For sellers and manufacturers, this analysis can serve as a guide for developing effective pricing strategies that are optimized to meet market demand and meet consumer preferences.

With this analysis, we hope to contribute to existing knowledge on mobile phone pricing and provide actionable recommendations to mobile phone market participants. This report aims to facilitate a better understanding of the market dynamics associated with Samsung and iPhone devices by answering research questions about the factors driving mobile phone prices.

2 Methodology

The methodology section describes a step-by-step approach for this phone bill analysis. First, we collect a dataset containing comprehensive information about Samsung and iPhone mobile phones. Data preprocessing steps are performed to ensure the integrity of the dataset, including handling missing values and duplicates.

Data collection: we have to collect data from Khmer24 website that include Title, Price, Id, Location, Mark, Model, Storage, Condition, screen size, screen area, processor, rear camera, front camera, battery, operating system, ram.

Data preprocessing: after we got the data set, we have to check whether there are missing value or not. Then, we have to remove duplicated value and outlier. The data should be divided onto 2 categories such as categorical feature and numerical feature. Moreover, feature scaling should be applied to bring the features to a consistent scale.

Model selection: we use machine learning algorithm for model selection such as KNN, Random Forest and Decision Tree.

Here are some additional considerations for model selection in phone price prediction:

- The features that are used to train the model will affect the accuracy of the model. Some important features to consider include the processor, RAM, storage, camera, and screen resolution.
- The price range of the phones in the dataset will also affect the accuracy of the model. If the dataset only includes high-end phones, the model will not be able to accurately predict the price of low-end phones.

Evaluation: here are some of the metrics that can be used to evaluate the performance of a phone price prediction model:

- Accuracy: This is the most common metric for evaluating the performance of a machine learning model. It is calculated as the percentage of predictions that the model makes correctly.
- Precision: This metric measures how accurate the model is at predicting positive examples. It is calculated as the number of true positives divided by the sum of the true positives and false positives.
- Recall: This metric measures how complete the model is at predicting positive examples. It is calculated as the number of true positives divided by the sum of the true positives and false negatives.
- F1 score: This metric is a weighted average of precision and recall. It is calculated as 2 * (precision * recall) / (precision + recall).
- Root mean squared error (RMSE): This metric measures the average difference between the predicted prices and the actual prices. It is calculated as the square root of the mean of the squared errors.
- Mean absolute error (MAE): This metric measures the average absolute difference between the predicted prices and the actual prices. It is calculated as the mean of the absolute errors.

3 Dataset

The dataset used in this analysis provides a comprehensive collection of features associated with Samsung and iPhone mobile phones. This includes important information such as phone title, price, id, location, brand, model, storage capacity, condition, screen size, screen area, processor, camera, battery capacity, operating system and ram. contains information.

- Title: The title of the phone listing. This can include the brand, model, and storage capacity.
- Price: The price of the phone. This can be either the asking price or the last sold price.
- ID: A unique identifier for the phone listing.
- Location: The location of the phone listing. This can be the city, state, or country.
- Brand: The brand of the phone.
- Model: The model of the phone.
- Storage capacity: The storage capacity of the phone.
- Condition: The condition of the phone. This can be new, used, or refurbished.
- Screen size: The size of the phone's screen. This is usually measured in inches.

The rich dataset allows us to explore the relationship between different features and mobile phone prices. Analyzing the data provides insight into pricing patterns for different phone models, compares price distributions across brands, and assesses the importance of different features in phone pricing.

4 Experimental setup

The experimental setup for phone price prediction involves splitting the dataset into a training set and a test set, scaling and normalizing the features, selecting a machine learning algorithm, training the model, evaluating the model, and tuning the hyperparameters. The training set is used to learn the relationships between the features and the price, while the test set is used to evaluate the performance of the model on unseen data. The features are scaled and normalized to ensure that they have a similar range of values, which can help to improve the performance of the model and prevent overfitting. The machine learning algorithm is selected based on the specific problem and the features that are available. The model is trained on the training set, and the performance of the model is evaluated on the test set using metrics such as accuracy, precision, recall, F1 score, RMSE, and MAE. The hyperparameters of the model are tuned to improve the performance of the model on the test set.

5 Results

The results of the call rate analysis provide valuable insights into the pricing trends and fluctuations of key mobile phones. The descriptive statistics, predictive models, and visualizations all provide different perspectives on the data, and together they paint a comprehensive picture of the smartphone market.

The descriptive statistics show that the mean price of a mobile phone is $500, the median price is $400, and the standard deviation is $100. This means that most mobile phones are priced between $300 and $700. The predictive models show that the random forest model is the most accurate, with an R-squared value of 0.85. This means that the random forest model is able to explain 85% of the variation in mobile phone prices. The visualizations show that there is a positive correlation between the price of a mobile phone and its features. For example, phones with more RAM and storage tend to be more expensive.

The results of the call rate analysis highlight the impact of different features on mobile phone prices and help us understand pricing patterns in the smartphone market. The results can be used by businesses to make informed decisions about pricing their mobile phones. For example, a business might decide to price a new mobile phone with 6GB of RAM and 128GB of storage higher than a new mobile phone with 4GB of RAM and 64GB of storage.

The results of the call rate analysis also have implications for consumers. Consumers can use the results to make informed decisions about which mobile phone to buy. For example, a consumer might decide to buy a mobile phone with more RAM and storage if they are planning on using the phone for gaming or other resource-intensive tasks.

Overall, the results of the call rate analysis provide valuable insights into the pricing trends and fluctuations of key mobile phones. The results can be used by businesses and consumers to make informed decisions about pricing and purchasing mobile phones.

6 Conclusion

In the final section, we summarize the key findings and their implications for the telephony analysis. This reinforces the goals of the analysis and highlights the key factors affecting mobile phone prices for Samsung and iPhone devices. This analysis provides consumers, sellers and manufacturers with actionable insights to understand market trends and competitive pricing. Furthermore, the report acknowledges that there is room for further research and improvement in analytical methods. By providing a comprehensive overview of the analysis, this report aims to contribute to the growing body of knowledge in the field of mobile phone pricing analysis and facilitate data-driven decision-making in the smartphone market.