USED SMARTPHONE PRICE Analytics

## Introduction

This project investigates used mobile phone pricing with an emphasis on the wide range of variables impacting consumer decision. You will be analyzing how phone features such as screen size, rear and front camera, internal memory, and other properties affect the price of a used phone. You start with exploratory data analysis on the features to explore the relationship between the explanatory variables and the response (target) variable, and you will also use this information to identify possible issues with multi-collinearity and identify potential outliers in the dataset. You will also explore the mathematical relationship between the explanatory variables and response variable.

## Project

The modern mobile phone market is a complex ecosystem where pricing strategies are influenced by a multitude of factors. Predictive models such as regression models are essential for predicting the prices of mobile phones, or classification models that can classify the price of a phone as either high or low. This information will help sellers of new and used phone, as well as customers, in making data driven informed decisions.

This project seeks to explore the complex correlations between fundamental phone features and their impact on new or used mobile phone prices. This study aims to uncover the crucial determinants shaping pricing strategies within the mobile phone industry.

The first task of the project requires data explanation and understanding. The second task of this project focus is on exploratory data analysis where you explore relationships, trends, patterns and correlations in and between the variables. The third task of the project preparing the data for modeling. The fourth task is where you will build predictive and/or classification models using selected explanatory variables and then perform an analysis on the results of those models. In the final section of this project, you will write conclusions about your findings and discuss in detail the further work that could be used to improve the final selected models to be more robust and more accurate and provide recommendations.

## Key Project Objectives

The machine learning process requirements of this project will be provided by instructor during the course. The following requirements are the basic requirements and are in addition to the major process as described above.

1. Detail attribute definitions and explanations.
2. Detail attribute explorations and analysis based on the machine learning requirements
3. Detail your complete data preparation process and provide the justifications as to why you did what you did.
4. Detail your selection of regression and classification models for this project and provide the justifications as to why you selected those models.
5. Detail how you prepared your model for the dataset and justifications as to why you did what you did.
6. Detail how you implemented your model and provide a detailed analysis of the model outcomes.

## Key Tasks

* Analyze data.
* Identify meaningful patterns.
* Build machine learning models.
* Evaluate model performance.
* Provide business relevant suggestions and recommendations.

## Data

The “used\_device\_data.csv” is the project dataset and its attributes are:

|  |  |
| --- | --- |
| Order | Attribute |
| 1 | device\_brand |
| 2 | OS |
| 3 | screen\_size |
| 4 | 4g |
| 5 | 5g |
| 6 | rear\_camera\_mp |
| 7 | front\_camera\_mp |
| 8 | internal\_memory |
| 9 | ram |
| 10 | battery |
| 11 | weight |
| 12 | release\_year |
| 13 | days\_used |
| 14 | normalized\_used\_price |
| 15 | normalized\_new\_price |

## Requirements

Project Name: used smartphone price analytics

Dataset: used\_device\_data.csv

Method: supervised modeling

Models: classification and prediction

## Approach:

1. Project introduction
2. Business and analytics goals
3. Data preprocessing (such as attributes definition, data exploration, checking missing value, checking zero, and more)
4. Predictor analysis and relevancy
5. Dimension reduction (if needed)
6. Data engineering and transformation (if needed)
7. Data partitioning methods (if needed)
8. Model selection
9. Model fitting, validation accuracy and test accuracy
10. Report models’ performance
11. Model evaluation (of the selected models)
12. Observation and conclusion. Include how your analytics solution and outcomes supports the accomplishment of the business goals identified.

## Delivery

Weekly delivery (Deadline: week 5)

* Week 2: part 1, 2, and 3. No program submission
* Week 3: part 4, 5, 6, 7, and 8. No program submission
* Week 4: part 9, 10, 11, and 12. No program submission
* Week 5: complete work submission (documentation in .docx, program in .R, and presentation in .pptx or .pdf )

## Deliverable Policy

Demonstrate your analytics ability as entry level data analytics professional. Your work should be delivered on time; late work will not be accepted. Do not share your work with other students. Avoid machine learning blunders and fundamental mistakes.

## Basic Grading Policy

1. **Document**
2. Your document should be in your own words, not cut and paste from the internet. (**Use of AI will result in an automatic F for the project)**
3. Should be written in a technical write up style.
4. Should include description and explanation of your work process and related to your implementation through R codes
5. Each section of your document should be in the correct sequence based on workflow process
6. Should not include irrelevant information that is not related to the project requirements work
7. Should not include repeating content and visualizations
8. Should not include any R codes or examples of R codes in your document**.**
9. Remember, you are not teaching readers. Instead, you are explaining and describing your solution to the potential clients of your company who are selling new or used phones. They are not as technical as you so keep it simple, keep it plain so they can understand.
10. **Program**
    1. Your R program should be easy to read and understand.
    2. Your R program should have comments. Comments should explain a process and not the codes.
    3. Avoid complex R codes. All code solutions should be based on what you have learned through lecture notes and the approved textbooks. Use of AI or code found on the internet is prohibited. You should be able to explain your codes completely during the in class presentations.
    4. Remember, we run the entire codes at once and won’t fix the error. A program with error will receive a zero grade.