Proposal for final project (MDSA Winter 2023)

Rudy Brown, Jose Palacios, Stuart Finley, Andrii Voitkiv, Yongpeng Fu

27/01/2023

Chapter 1: 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 features, like brand, memory, storage, camera, resolution, just to name a few. And as you can imagine, with all this new technology and features jam packed in one little device costs money, costs a lot of money. A 2020 reviewof 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 our data from curated MobilePhone's dataset [https://www.kaggle.com/datasets/sudhanshuy17/mobilephone] from Kaggle and apply a set of statistical analysis hoping to answer some guiding questions:

1. Can we estimate the true 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?

Chapter 2: Dataset

We chose a dataset weighting by *simplicity*. That is, we would like to maximize the learning experience applying class content to a toy/stylized modelthat may or may not have any practical use.

Columns

The dataset consists of 8 columns and 28'036 rows and no missing values.

- *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.
- *Company*: categorical variable. Name of the phone's manufacturer.
- *Price*: continuous variable. Units in Indian Rupees.
- *Rating*: continuous variable. Units in Indian Rupees.
- *Number of ratings*: discrete variable: a simple count.
- *Total reviews*: discrete variable: a simple count.
- *RAM size*: categorical variable. RAM specification of the phone.
- *ROM size*: categorical variable. Storage (non-volatile memory) capacity of the phone.

Scope

TODO: Hash out modelling. I recommend a diagram using a tool like lucidcharts.com. See the templates section. For instance (see Figure 1 below).

References

TODO

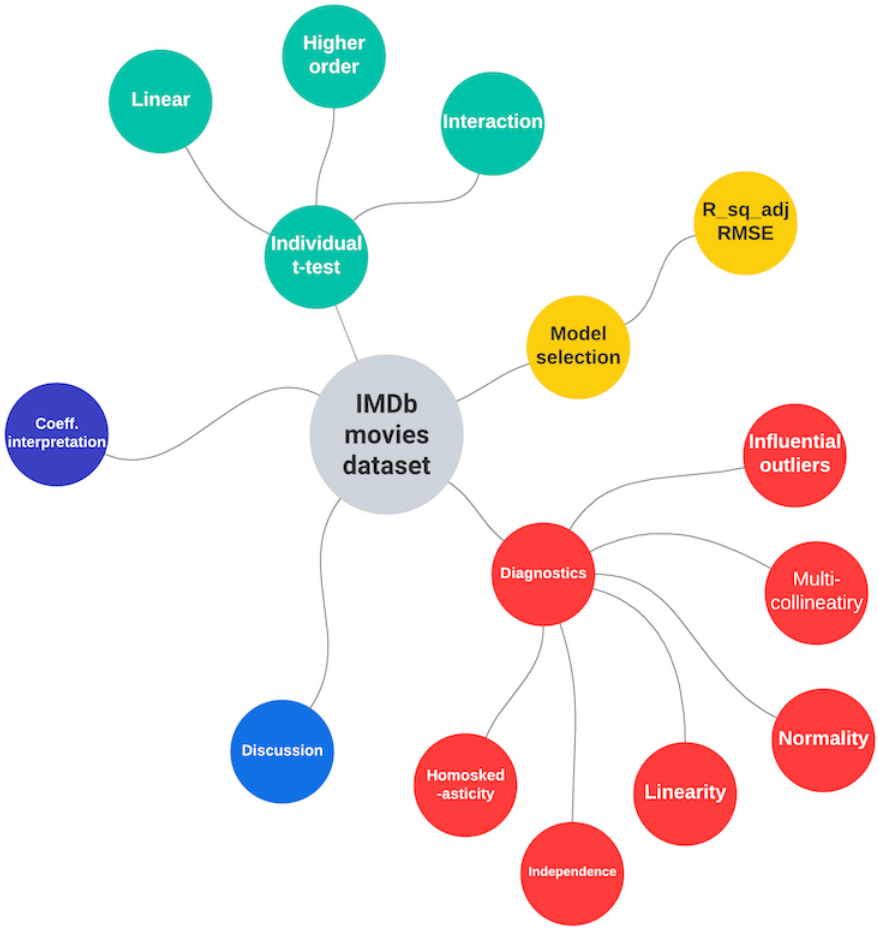

Figure 1: Example diagram