Proposal for final project (MDSA606 Winter 23)

Group 10

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

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 model that 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).
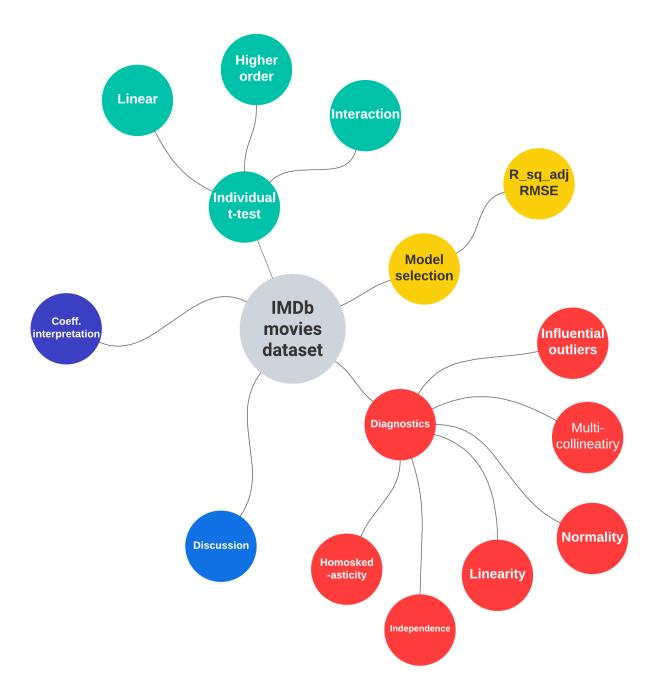
References

TODO

Figure 1: Example diagram