**Report**

1. Data Loading and Exploration

The goal of this project is to predict the price of a product based on its features such as brand, color material, size and etc. We will be using Random Forest Regressor for our project since we need high accuracy.

2. Data Preprocessing

Data Cleaning

• autoclean() and klib.datacleaning() were not used

• Removed unnecessary columns which are non-predictive such as id

• Handled missing values by filling them as appropriate (mode, median).

• encoding was done as there are multiple categorical columns which prevents our machine learning model to predict

• scaling was done only on features assigning them to x variable with target value included to increase the performance

3. Data Visualisation is also shown with barplot based on value counts of each categorical columns

4. Mutual information between columns and target value is also calculated to see the relativeness

5. Model Training

Splitted the dataset into training and test sets using an 80-20 split

Model Selection

one regression machine learning model was considered since the target output is continous value:  
• Random Forest Regressor

6. Evaluation

The following metrics were to evaluate the model

• MSE

• R2