Project Proposal - Predicting Price for Diamonds

Franscisco Arrieta, Emily Schmidt, Lucia Camenisch

November 25, 2022

1 Overview

The project's business case is to help a jewelry company to determine if it is possible to forecast potential income by predicting prices for assets in their inventory. This will help in their pricing strategy for the diamonds and compare their position with regards to market competition.

2 Goal

The objective is to predict price of diamonds based on several characteristics and try to identify the most influential explanatory variables.

3 Structure and Analysis Steps

1. Introduction

2. Data Exploration

   - Dimension Summary
   - Variable Types
   - Missing Values
   - Variable Visualization

3. Dimension Reduction Analysis

4. Variable Prediction and Model Performance Evaluation

   - Linear Regression
   - k-NN
   - Regression Tree
   - Neural Network
   - Ensembles

5. Model Performance Summary

6. Conclusion

4 Dataset Details

This dataset was found `diamonds.csv` on kaggle. It is sourced from Tiffany & Co. snapshot price list in 2017. There are 53'940 rows and 10 variables, listed below:

Variable	Description
price	price in US dollars **(outcome variable)**
carat	weight of the diamond (1 ct = 200 mg)
cut	quality of the cut
color	diamond color
clarity	a measurement of how clear the diamond is
x	length in mm
y	width in mm
z	depth in mm
depth	total depth percentage
table	width of top of diamond relative to widest point