**Machine Learning Project Proposal: Diamond Pricing**

**Team**

Ziad Abouchadi

Susann Almasi

Aditya Bindal

Jinesh Ramani

**Project Overview**

Diamond e-commerce retail is a highly price-competitive industry. Diamonds are commodity products, with a variety of categorical and numerical variables influencing their price. Companies in this space often use competitor prices as a benchmark when pricing diamonds.

We will assume that we are working with a new diamond e-commerce retailer on improving diamond purchasing and pricing decisions. We will do so by models to:

1. Predict how a competitor would price certain diamonds
2. Identify large “clusters” in inventory data of a competitor – i.e. more popular types of diamonds that are stocked by the competitor

**Data Set**

We have scraped the online diamond inventory of Brilliant Earth, a large diamond e-commerce retailer. This dataset contains over 45,000 distinct diamonds, each with 43 populated variables (including country of origin, size, cut, color, clarity, fluorescence, girdle width, dimensions, shape, and many other characteristics).

**Methodology**

We will predict the diamond’s price given its features using KNN, Random Forests, and Boosting. We will also attempt to use unsupervised models (e.g., neural nets) to identify clusters within the data without training the model.

We will divide the test data into train, validate, and test. We will pick the model that has the best out-of-sample performance.