# STAT394 Group Project Report2

Ken MacIver, Tom Tribe, Jundi Yang, Mei Huang

2022-08-19

## Contents

# 1  EDA summmary

```r
diamonds <- read.csv("./diamonds.csv")
df <- data.frame(diamonds)
df$cut <- as.factor(df$cut)
df$clarity <- as.factor((df$clarity))
df$color <- as.factor(df$color)


summary(df)
```

```
##        X              carat                   cut          color        clarity
##  Min.   :    1   Min.   :0.2000   Fair     : 1610   D: 6775   SI1    :13065
##  1st Qu.:13486   1st Qu.:0.4000   Good     : 4906   E: 9797   VS2    :12258
##  Median :26970   Median :0.7000   Ideal    :21551   F: 9542   SI2    : 9194
##  Mean   :26970   Mean   :0.7979   Premium  :13791   G:11292   VS1    : 8171
##  3rd Qu.:40455   3rd Qu.:1.0400   Very Good:12082   H: 8304   VVS2   : 5066
##  Max.   :53940   Max.   :5.0100                     I: 5422   VVS1   : 3655
##                                                     J: 2808   (Other): 2531
##      depth           table           price             x
##  Min.   :43.00   Min.   :43.00   Min.   :  326   Min.   : 0.000
##  1st Qu.:61.00   1st Qu.:56.00   1st Qu.:  950   1st Qu.: 4.710
##  Median :61.80   Median :57.00   Median : 2401   Median : 5.700
##  Mean   :61.75   Mean   :57.46   Mean   : 3933   Mean   : 5.731
##  3rd Qu.:62.50   3rd Qu.:59.00   3rd Qu.: 5324   3rd Qu.: 6.540
##  Max.   :79.00   Max.   :95.00   Max.   :18823   Max.   :10.740
##
##        y               z
```

```
##   Min.    : 0.000   Min.    : 0.000
##   1st Qu.: 4.720   1st Qu.: 2.910
##   Median : 5.710   Median : 3.530
##   Mean   : 5.735   Mean    : 3.539
##   3rd Qu.: 6.540   3rd Qu.: 4.040
##   Max.   :58.900   Max.    :31.800
##
```