```
In [1]: import pandas as pd
           import numpy as np
           import matplotlib.pyplot as plt
           %matplotlib inline
          Predictive Diagram
           Predictive Diagram
           Price Equation
           Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity
 In [2]: diamonds = pd.read csv('data/diamonds.csv')
           new_diamonds = pd.read_csv('data/new-diamonds.csv')
           Preview the data used to build the regression model
 In [3]: diamonds
 Out[3]:
                  Unnamed: 0 carat
                                          cut | cut_ord | color | clarity | clarity_ord |
                                                                                price
                                    Premium
                                                                               1749
                  2
                                                       G
                                                             11
                              2.25
                                    Fair
                                                                               7069
                 13
           2
                              0.70
                                    Very Good 3
                                                       Е
                                                             VS2
                                                                               2757
                              0.47
                                    Good
                                                             VS1
                                                                               1243
                                                       G
                  5
                              0.30
                                    Ideal
                                                             VVS1
                                                                               789
                  6
                                                       D
           5
                              0.33
                                    Ideal
                                                             SI1
                                                                   3
                                                                               728
                                    Very Good
                              2.01
                                                       G
                                                             SI1
                                                                               18398
                  8
                              0.51
                                    Ideal
                                                             VVS2
                                                                   6
                                                                               2203
                                                       D
                              1.70
                                    Premium
                                                                               15100
                  10
                                                       D
           9
                              0.53
                                    Premium
                                                             VS2
                                                                               1857
                  11
                              0.39
                                    Premium
                                                       Н
                                                             SI1
                                                                               834
           10
                  12
                                                       Н
                                                                               7708
                                                             SI1
           11
                                    Very Good
                  13
                              1.00
                                    Premium
                                                             VS2
                                                                               6272
           12
                                              5
                  14
           13
                               1.29
                                    Ideal
                                                             VS1
                                                                               5676
                  15
                              2.01
                                    Good
                                                       D
                                                                               16776
           14
                                                             SI2
                                                       G
                  16
                                                             VS1
                                                                               7404
           15
                              1.13
                                    Ideal
                  17
                                                                               1702
           16
                              0.70
                                    Ideal
                                                             SI2
           17
                  18
                              0.38
                                    Very Good 3
                                                             VS1
                                                                               606
           18
                  19
                              1.17
                                    Ideal
                                                       Н
                                                             SI2
                                                                               5423
                  20
                                    Premium
                                                             SI1
                                                                               8033
                              1.51
                                                       D
           20
                  21
                              0.40
                                    Ideal
                                                             VVS1
                                                                               1279
                  22
                                    Very Good 3
                                                             VS2
                                                                               863
           21
                              0.41
                  23
                                                             VVS1
                                                                               1893
                              0.51
                                    Ideal
                                              4
                                                                               3584
                  24
                                                             SI2
                               1.00
                                    Premium
                  25
                              1.09
                                    Ideal
                                                             VVS2
                                                                               10196
                  26
                                                       Е
                                                             VS1
                                                                               1082
           25
                              0.39
                                    Good
                  27
                                                       Е
                                                                               3024
                              0.72
                                    Premium
                                                             VS2
           26
                  28
                                                       Е
                                                                               5593
           27
                                    Very Good 3
                                                             SI2
                              1.14
           28
                  29
                              0.30
                                    Ideal
                                                       D
                                                             VS2
                                                                               710
                  30
                              0.30
                                                       Н
                                                             SI1
                                                                               465
           29
                                    Ideal
                 49971
                                    Premium
                              1.83
                                                       Н
                                                             SI1
                                                                               10162
           49970
                  49972
                                                                               5506
           49971
                              1.11
                                    Ideal
                                                             VS2
                                    Very Good 3
           49972
                  49973
                              1.00
                                                             SI2
                                                                               3763
                  49974
                                                       D
                                                             VVS1
           49973
                              0.61
                                    Ideal
                                                                               3625
                                                       G
           49974
                  49975
                              0.80
                                    Very Good 3
                                                                               2451
                                                       Ε
           49975
                 49976
                              0.53
                                    Ideal
                                                             SI1
                                                                               1564
           49976 49977
                                                       Е
                                                                               659
                              0.34
                                    Very Good 3
                                                             VS2
           49977
                  49978
                              1.22
                                                             VS2
                                                                               7584
                                    Ideal
                  49979
                                                             VS1
           49978
                              0.57
                                    Very Good 3
                                                                               1270
           49979
                  49980
                              0.58
                                    Very Good 3
                                                             VVS1
                                                                               1790
                  49981
                              0.32
                                                             SI2
                                                                               371
           49980
                                    Ideal
                                                       D
                                                                               1040
           49981
                  49982
                              0.42
                                    Premium
                                                             SI1
           49982
                  49983
                              0.35
                                    Very Good
                                                             SI1
                                                                               491
                                                       Е
                  49984
                                                                               870
           49983
                              0.46
                                    Ideal
                                                             SI2
                  49985
                              0.70
                                                             VS1
                                                                               2398
           49984
                                    Ideal
                                                                               1129
           49985
                  49986
                              0.43
                                                       G
                                                             VVS2
                                    Ideal
                                                       G
                                                                               6932
           49986
                  49987
                              1.01
                                                             VS1
                                    Premium
           49987
                  49988
                              1.03
                                    Premium
                                                             VS1
                                                                               7328
                  49989
                                                       Η
                                                             VS2
                                                                               1058
           49988
                              0.47
                                    Ideal
                  49990
                                                             VVS1
           49989
                              1.01
                                    Premium
                                                                               4989
                                                             VS2
           49990
                  49991
                              1.21
                                    Good
                                                                               7786
           49991
                  49992
                              0.43
                                    Ideal
                                                             VVS1
                                                                               848
           49992
                  49993
                              0.70
                                    Premium
                                                       G
                                                             VVS1
                                                                               3105
                  49994
                                                             VS2
                                                                               1449
           49993
                              0.50
                                    Ideal
                                                       G
                  49995
                                                       G
                                                             VVS2
           49994
                              0.33
                                    Very Good 3
                                                                               692
                                                       Н
           49995
                  49996
                              0.71
                                    Ideal
                                                             VVS1
                                                                               2918
                                                       G
                                                             VVS2
           49996
                  49997
                              0.43
                                    Ideal
                                                                               1056
           49997
                  49998
                              1.14
                                    Premium
                                                       G
                                                             VS2
                                                                               6619
                                                       Е
                                                             VS2
                                                                               6787
           49998
                 49999
                              1.01
                                    Premium
           49999 50000
                                                             VS2
                              1.77
                                    Premium
                                                                               9428
           50000 rows × 8 columns
            • Carat represents the weight of the diamond, and is a numerical variable.
            • Cut represents the quality of the cut of the diamond, and falls into 5 categories: fair, good, very good, ideal, and
              premium. Each of these categories are represented by a number, 1-5, in the Cut_Ord variable.
            • Clarity represents the internal purity of the diamond, and falls into 8 categories: I1, SI2, SI1, VS1, VS2, VVS2, VVS1,
              and IF. Each of these categories are represented by a number, 1-8, in the Clarity_Ord variable
          Preview the data of the diamonds the company would like to purchase
 In [4]: new_diamonds
 Out[4]:
                                                            clarity clarity_ord
                                         cut cut_ord color
                 Unnamed: 0 carat
                                   Premium
                             1.22
                                                            SI1
                                                                   5
                 2
                             1.01
                                   Good
                                                      G
                                                            VS2
                                                                   5
           2
                             0.71
                                   Very Good 3
                                                            VS2
                                                                   2
                             1.01
                                   Ideal
                                                            SI2
                                                                   6
                 5
                             0.27
                                                      Н
                                                            VVS2
                                   Ideal
           5
                             0.52
                                                      G
                                                            VS1
                                   Premium
                                                                   3
           6
                             1.01
                                   Premium
                                                            SI1
                                                                   3
                             0.59
                                                            SI1
                                   Ideal
                                             2
                                                                   3
                 9
                             1.01
           8
                                   Good
                                                            SI1
                                                                   2
           9
                 10
                             2.03
                                   Ideal
                                                            SI2
                                                                   5
                 11
                                   Premium
                                                            VS2
                             1.35
           10
                 12
                             0.74
                                                                   3
           11
                                   Ideal
                                                      G
                                                            SI1
                                                                   3
                 13
           12
                             0.90
                                   Premium
                                                      D
                                                            SI1
                 14
                                                                   5
           13
                             0.30
                                   Good
                                                            VS2
                                             2
                 15
                                                                   5
                             1.01
                                   Good
                                                            VS2
           14
                                                                   2
                 16
                             1.02
                                                            SI2
           15
                                   Good
                 17
                             2.05
                                                      G
                                                            SI1
                                                                   3
           16
                                   Premium
                                                                   3
           17
                 18
                             0.54
                                   Ideal
                                                            SI1
                 19
                                                            VS2
           18
                             0.72
                                   Ideal
                                                                   2
                 20
           19
                             2.00
                                   Premium
                                                            SI2
                                                                   2
                 21
           20
                             1.57
                                   Premium
                                                      G
                                                            SI2
           21
                 22
                                                      G
                                                            SI1
                                                                   3
                             0.89
                                   Premium
                 23
                             0.33
                                   Premium
                                                            VVS2
                                                                   6
           22
                 24
                                                                   8
                                   Very Good 3
           23
                             0.30
                                                                   5
                 25
                             1.79
                                   Very Good 3
           24
                                                            VS2
                                                                   2
                 26
                             1.11
                                   Ideal
                                                      Ε
                                                            SI2
           25
                 27
                             0.79
                                                                   3
           26
                                   Premium
                                                            SI1
                                                                   3
           27
                 28
                             0.71
                                   Very Good 3
                                                            SI1
                 29
                                                                   3
           28
                             0.73
                                   Ideal
                                                      G
                                                            SI1
                                                                   5
                 30
                             0.90
                                   Very Good 3
                                                            VS2
           29
           2970 2971
                                   Premium
                                                                   5
                             1.00
                                                            VS2
                                                                   2
           2971 2972
                             1.07
                                   Premium
                                                      Ε
                                                            SI2
                                                                   4
           2972 2973
                                   Very Good 3
                             1.21
                                                            VS1
                                  Fair
                                                            VS1
           2973 2974
           2974 2975
                             0.79
                                   Good
           2975 2976
                                                            VVS2
                                                                   6
                             0.32
                                   Ideal
           2976 2977
                                                            VS2
                             0.40
                                   Ideal
           2977 | 2978
                             0.51
                                   Fair
                                                            VS2
                                                                   5
           2978 | 2979
                             1.05
                                   Premium
                                                            VS1
                                                                   3
           2979 2980
                                   Very Good 3
                                                            SI1
                             0.43
           2980 2981
                                                      G
                                                            VVS2
                                                                   6
                             0.30
                                   Ideal
                                                      Ε
           2981
                 2982
                             1.51
                                   Premium
                                                            SI1
           2982 | 2983
                             1.54
                                   Ideal
                                                            VS2
                                                                   5
                                                                   5
           2983 | 2984
                             1.16
                                                            VS2
                                   Ideal
           2984 2985
                             1.01
                                   Premium
                                                            SI1
                                                                   5
           2985 2986
                             1.01
                                   Very Good 3
                                                            VS2
           2986
                 2987
                             0.83
                                   Premium
                                                            SI1
                                                                   5
           2987 | 2988
                             0.93
                                   Very Good 3
                                                            VS2
           2988 | 2989
                                                                   2
                             1.20
                                   Very Good 3
                                                      G
                                                            SI2
                                                            VVS2
                                                                   6
           2989 2990
                             0.80
                                   Ideal
           2990 | 2991
                             1.11
                                   Ideal
                                                            VVS2
           2991 | 2992
                             1.31
                                   Premium
                                                            SI2
                                                                   2
           2992 | 2993
                             1.11
                                                            SI1
                                                                   3
                                   Good
                                             2
                                                                   2
           2993 | 2994
                             1.10
                                   Good
                                                      D
                                                            SI2
                                                                   2
                             1.22
           2994
                 2995
                                   Premium
                                                            SI2
           2995 2996
                                                                   2
                             0.72
                                                            SI2
                                   Ideal
           2996 2997
                             1.09
                                                            VS2
                                                                   5
                                   Premium
                 2998
                             1.05
                                   Very Good 3
                                                            SI1
           2997
                                                                   3
           2998 2999
                             0.70
                                   Fair
                                                      G
                                                            SI1
           2999 3000
                                                                   3
                             1.01
                                   Very Good 3
                                                            SI1
           3000 rows × 7 columns
          Understanding the Model
           1. According to the model, if a diamond is 1 carat heavier than another with the same cut, how much more should I expect to
           pay? Why?
 In [5]: def calculate_diamond_price(carat, cut, clarity):
               diamond_price = -5269 + (8413 * carat) + (158.1 * cut) + (454 * clarity)
               return diamond_price
           def format currency(amount):
               return '${:,.2f}'.format(amount)
 In [6]: diamond_one_carat = calculate_diamond_price(1,1,1)
           diamond_two_carats = calculate_diamond_price(2,1,1)
           price_difference = diamond_two_carats - diamond_one_carat
           print("Diamond 1 carat price:\t ",diamond_one_carat,"$\nDiamond 2 carats price:\t",diamond_two_
                  "$\nPrice difference:\t ",price_difference,"$")
          Diamond 1 carat price: 3756.1 $
           Diamond 2 carats price: 12169.1 $
           Price difference:
                                        8413.0 $
           According to our equation: Price = -5,269 + 8,413 x Carat + 158.1 x Cut + 454 x Clarity
           1 carat will increase the price in 8,413 $
           2. If you were interested in a 1.5 carat diamond with a Very Good cut (represented by a 3 in the model) and a VS2 clarity rating
           (represented by a 5 in the model), how much would the model predict you should pay for it?
 In [7]: diamond one half carat = calculate diamond price(1.5,3,5)
           print("Diamond Carat: 1.5 Cut: Very Good Clarity: VS2 price:\t ",diamond_one_half_carat,"$")
           Diamond Carat: 1.5 Cut: Very Good Clarity: VS2 price: 10094.8 $
          The model predicts a price of 10,094.8 $
          Visualize the Data
          1. Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.
 In [8]: x = diamonds['carat']
           y = diamonds['price']
           plt.figure(figsize=(14,8))
           plt.ylabel('Price')
           plt.xlabel('Carat')
          plt.title('Price vs Carat')
           plt.scatter(x, y, alpha=0.4, color='green')
 Out[8]: <matplotlib.collections.PathCollection at 0x13498454c50>
             17500
             15000
             12500
           Pice 10000
              7500
              5000
              2500
 In [9]: new_diamonds["predicted_price"] = calculate_diamond_price(new_diamonds['carat'], new_diamonds['c
           ut_ord'], new_diamonds['clarity_ord'])
In [10]: new_diamonds
Out[10]:
                                                            clarity clarity_ord predicted_price
                 Unnamed: 0 carat
                                         cut | cut_ord | color |
                             1.22
                                   Premium
                                                                   3
                                                                              6989.26
                                                      G
                                                            SI1
                                                            VS2
                             1.01
                                                      G
                                                                              5814.33
                                   Good
                             0.71
                                   Very Good 3
                                                            VS2
                                                                              3448.53
                                                                   2
           3
                             1.01
                                   Ideal
                                                            SI2
                                                                              4926.63
                             0.27
                                                      Н
                                                            VVS2
                                                                   6
                                   Ideal
                                                                              517.01
                             0.52
                                                            VS1
                                                                              1554.16
                 6
                                   Premium
                                                      G
                                                                   4
           6
                                                                   3
                                                                              5222.53
                             1.01
                                   Premium
                                                            SI1
                 8
                             0.59
                                   Ideal
                                             5
                                                      D
                                                            SI1
                                                                   3
                                                                              1847.17
                 9
                                             2
                                                                   3
                                                                              4906.33
           8
                             1.01
                                   Good
                                                      Ε
                                                            SI1
                                                                              13507.89
                 10
                             2.03
                                   Ideal
                                                            SI2
                                                                   5
                 11
                             1.35
                                                            VS2
           10
                                   Premium
                                                      Н
                                                                              8990.95
                 12
                             0.74
                                             5
                                                                   3
                                                                              3109.12
           11
                                   Ideal
                                                      G
                                                            SI1
           12
                 13
                             0.90
                                                      D
                                                            SI1
                                                                   3
                                                                              4297.10
                                   Premium
           13
                 14
                             0.30
                                   Good
                                             2
                                                      G
                                                            VS2
                                                                   5
                                                                              -158.90
                                             2
                                                                   5
           14
                 15
                             1.01
                                                                              5814.33
                                   Good
                                                            VS2
                 16
                             1.02
                                             2
                                                            SI2
                                                                   2
                                                                              4536.46
           15
                                   Good
                                                      Н
                 17
                                                                   3
                             2.05
                                                                              13972.05
           16
                                   Premium
                                                      G
                                                            SI1
                 18
                                                                   3
                                                                              1426.52
           17
                             0.54
                                                            SI1
                                   Ideal
                             0.72
                                                                   5
           18
                 19
                                   Ideal
                                                            VS2
                                                                              3848.86
                                                                   2
                 20
                             2.00
                                                            SI2
                                                                              13097.40
           19
                                   Premium
                                                                   2
                 21
                             1.57
                                                      G
                                                            SI2
                                                                              9479.81
           20
                                   Premium
                 22
                                                                   3
                             0.89
                                   Premium
                                                      G
                                                            SI1
                                                                              4212.97
           21
                 23
                                                                  6
           22
                             0.33
                                   Premium
                                                            VVS2
                                                                              863.69
           23
                 24
                             0.30
                                   Very Good 3
                                                            ΙF
                                                                   8
                                                                              1361.20
                 25
                             1.79
                                   Very Good 3
                                                            VS2
                                                                   5
                                                                              12534.57
           24
                                                                              5767.93
           25
                 26
                             1.11
                                   Ideal
                                                            SI2
                 27
                                                                   3
           26
                             0.79
                                   Premium
                                                            SI1
                                                                              3371.67
                 28
                             0.71
                                   Very Good 3
                                                            SI1
                                                                   3
                                                                              2540.53
           27
                                                                              3024.99
           28
                 29
                             0.73
                                   Ideal
                                                            SI1
                                                                   3
                                   Very Good 3
                                                                   5
                 30
                             0.90
                                                      G
                                                            VS2
                                                                              5047.00
           29
                                                                   5
           2970 2971
                             1.00
                                                            VS2
                                                                              6046.40
                                   Premium
           2971 | 2972
                             1.07
                                   Premium
                                                      Ε
                                                                   2
                                                                              5273.31
                                                            SI2
           2972 2973
                             1.21
                                   Very Good 3
                                                            VS1
                                                                              7201.03
           2973 2974
                             0.46
                                   Fair
                                                            VS1
                                                                              575.08
                                                                   5
           2974 2975
                             0.79
                                             2
                                                            VS2
                                                                              3963.47
                                   Good
           2975 2976
                             0.32
                                                            VVS2
                                                                  6
                                                                              937.66
                                   Ideal
           2976 2977
                             0.40
                                   Ideal
                                                            VS2
                                                                              1156.70
                                  Fair
                             0.51
                                                            VS2
           2977 2978
                                                                              1449.73
           2978 2979
                                                            VS1
                                                                   4
                                                                              6013.05
                             1.05
                                   Premium
                                   Very Good 3
           2979 2980
                             0.43
                                                            SI1
                                                                              184.89
                                                      G
           2980 | 2981
                             0.30
                                                            VVS2
                                                                   6
                                                                              769.40
                                   Ideal
           2981 2982
                             1.51
                                                      Ε
                                                                   3
                                   Premium
                                                            SI1
                                                                              9429.03
           2982 | 2983
                             1.54
                                   Ideal
                                                            VS2
                                                                   5
                                                                              10747.52
           2983 | 2984
                                                            VS2
                                                                              7550.58
                             1.16
                                   Ideal
           2984 2985
                                                                   3
                             1.01
                                   Premium
                                                      D
                                                                              5222.53
                                                            SI1
           2985 2986
                                                                   5
                             1.01
                                                            VS2
                                                                              5972.43
                                   Very Good
           2986 2987
                             0.83
                                                            SI1
                                                                   3
                                                                              3708.19
                                   Premium
           2987 | 2988
                             0.93
                                   Very Good 3
                                                            VS2
                                                                   5
                                                                              5299.39
           2988 2989
                             1.20
                                   Very Good 3
                                                            SI2
                                                                              6208.90
                                                            VVS2
           2989 2990
                                                                              4975.90
                             0.80
                                   Ideal
           2990 2991
                             1.11
                                             5
                                                      D
                                                            VVS2
                                                                   6
                                                                              7583.93
                                   Ideal
           2991 2992
                             1.31
                                                      Н
                                                            SI2
                                                                   2
                                                                              7292.43
                                   Premium
                                             2
           2992 2993
                             1.11
                                   Good
                                                            SI1
                                                                   3
                                                                              5747.63
                                                                   2
                 2994
                             1.10
           2993
                                   Good
                                                            SI2
                                                                              5209.50
                                                                   2
           2994 2995
                             1.22
                                   Premium
                                                            SI2
                                                                              6535.26
                                                                   2
           2995 2996
                             0.72
                                                            SI2
                                                                              2486.86
                                   Ideal
           2996 2997
                             1.09
                                                            VS2
                                                                   5
                                                                              6803.57
                                   Premium
           2997 | 2998
                             1.05
                                   Very Good 3
                                                      G
                                                            SI1
                                                                   3
                                                                              5400.95
                                                                   3
           2998 | 2999
                             0.70
                                                      G
                                                            SI1
                                                                              2140.20
                                   Fair
           2999 3000
                                                            SI1
                                                                   3
                                                                              5064.43
                             1.01
                                   Very Good 3
           3000 rows × 8 columns
          2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-
           axis.
In [11]: plt.style.use('seaborn')
           plt.figure(figsize=(14,8))
           x new = new diamonds['carat']
           y new = new diamonds['predicted price']
           plt.ylabel('Predicted Price')
           plt.xlabel('Carat')
           plt.title('Predicted Price vs Carat')
          plt.scatter(x_new, y_new, alpha=0.4, color = 'blue')
Out[11]: <matplotlib.collections.PathCollection at 0x134987a27f0>
                                                          Predicted Price vs Carat
                                               20000
             15000
             10000
              5000
                                              1.0
          Merge both plots to compare them
In [12]: plt.figure(figsize=(14,8))
           plt.scatter(x, y, alpha=0.3, color='green')
           plt.scatter(x_new, y_new, alpha=0.3, color = 'blue')
           plt.ylabel('Price', fontsize=15)
           plt.xlabel('Carat', fontsize=15)
           plt.title('Price vs Carat', fontsize=15)
           plt.legend(loc='upper right', frameon=False, fontsize=15)
Out[12]: <function matplotlib.pyplot.show>
                                                             Price vs Carat
                                                                                                     predicted price
              20000
           Price 10000
              5000
                                                                 Carat
           3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?
           Comparing both plots we can see that the model is making a good average prediction on prices, we can conclude that the model
```

is working fine and that predicted prices are correct.

In [13]: | total\_diamonds\_cost = np.sum(new\_diamonds['predicted\_price'])

Total diamonds cost estimate: \$11,733,522.76

Total bid value recommendation: \$8,213,465.93

**\$8,213,465.93** which is exactly 70% of the retail value.

In [14]: bid\_value = (total\_diamonds\_cost \* 70) / 100

**1.** What price do you recommend the jewelry company to bid? Please explain how you arrived at that number.

print("Total bid value recommendation: ", format\_currency(bid\_value))

The final retail price the consumer will pay for all 3,000 diamonds will add up to \$11,733,522.76

Since the company generally purchases diamonds from distributors at 70% of the that price my bid recommendation is

Recommendation