Assignment 1

CS677

Alison Pedraza

**Question 1:**

1. load the ”diamonds” csv file as a list of lines using Python and construct a sublist for you group.

My group is ‘Premium’.

2. how many entries are there?

There are 13,791 entries for Premium group.

3. compute the average weight (in carats) for your group.

The average weight in carats is:

0.8919548981219524

4. compute the average price for diamonds in your group (round to the 4-th decimal point)

The average price for Premium is: $4584.2577

**Question 2:**

1. compute average prices per carat using both methods.

The average price per carat using method 'a' is: S4222.9054

The average price per carat using method 'b' is: $5139.5622

2. which average price is lower?

The average price is lower using method ‘a’.

3. compute the maximum price per carat: max price per carat = max(p1/w1, . . . , pN/wN)

The maximum price per carat: $17083.1776

4. compute the minimum price per carat min price per carat = min(p1/w1, . . . , pN/wN)

The minimum price per carat: $1051.1628

5. compute the median price per carat median price per carat = median(p1/w1, . . . , pN/wN)

The median price per carat is: $3763.33

**Question 3:**

For each of the two methods to compute price per carat, what combination of other parameters (color, clarity, depth, etc.) gave you

1. highest value

There are 5 color, clarity, and depth combinations that gave me the highest average price per carat of $5139 - which was from using method b , are:

[['11963', '1.01', 'Premium', 'G', 'SI1', '62.3', '59', '5139', '6.38', '6.3', '3.95'] – color G, clarity SI1, depth 62.3

['11965', '1.15', 'Premium', 'H', 'SI1', '60.9', '57', '5139', '6.8', '6.72', '4.12']- color H, clarity SI1, depth 60.9

['11966', '1', 'Premium', 'H', 'VS2', '59.7', '59', '5139', '6.57', '6.52', '3.91']- color H, clarity VS2, depth 59.7

['11969', '1.05', 'Premium', 'H', 'SI1', '60.2', '62', '5139', '6.57', '6.52', '3.94']- color H, clarity SI1, depth 60.2

['11970', '1.14', 'Premium', 'G', 'SI2', '60.9', '56', '5139', '6.8', '6.73', '4.12']]- color G, clarity SI2, depth 60.9

If using method ‘a’ to get highest value total from the list of price/carat, then the max price of price per carat from this list was $ 17083/carat.

2. lowest value

Out of the two methods, method ‘a’ gave the lower average price per carat.

The color and clarity combination using method ‘a’ which gave the lower price per carat ($4,222) is:

[['7376', '1.16', 'Premium', 'J', 'SI1', '62.5', '58', '4222', '6.73', '6.67', '4.19']] – color J, clarity SI1, depth 62.5

**Question 4:**

Using your prices per carat, compute the price you will be paying for 102 carat diamond sold recently by Sotheby.

Using average price per carat using method 'a', the approximate price I expect to pay for a 102 carat diamond would be: $430,736

Using average price per carat using method 'b', the approximate price I would expect to pay for a 102 carat diamond would be: $524,235

How close is your price to the real price? Why?

According to the article below, the 102 carat diamond was to auction off for anywhere over $10million and in fact it sold for $15million. My estimation using average price for Premium is completely off.

Even when using the highest price among the Premium group, the estimate fell short. I think this is because Premium diamonds are not the most pure and sought after diamond cut. The ‘Ideal’ cut is the most costly diamond cut and that was not what I subset from the diamond dataset.

https://www.cnbc.com/video/2020/09/10/sothebys-w ill-auction-off-a-102-carat-diamond-that-could -set-a-new-record.html