Great Learning

← Go Back to Python Foundations

Course Content

Weekly Quiz 4 - Exploratory Data Analysis

Type	:	Graded Quiz
Attempts	:	1/1
Questions	:	10
Time	:	45m
Due Date	:	May 06, 3:30 AM EDT
Your Marks	:	9/10

Instructions

Attempt History

Attempt #1
May 05, 1:47 PM

Marks: 9

Q No: 1 Correct Answer

Marks: 1/1

Following is the summary of all the variables that play an important role in deciding the price of diamond.

	count	mean	std	min	25%	50%	75%	max
carat	53940.0	0.797940	0.474011	0.2	0.40	0.70	1.04	5.01
depth	53940.0	61.749405	1.432621	43.0	61.00	61.80	62.50	79.00
table	53940.0	57.457184	2.234491	43.0	56.00	57.00	59.00	95.00
price	53940.0	3932.799722	3989.439738	326.0	950.00	2401.00	5324.25	18823.00
x	53940.0	5.731157	1.121761	0.0	4.71	5.70	6.54	10.74
y	53940.0	5.734526	1.142135	0.0	4.72	5.71	6.54	58.90
z	53940.0	3.538734	0.705699	0.0	2.91	3.53	4.04	31.80

Based on the above data description, which of the following statement is correct?

A. The average price of diamonds is 3932.8 dollars.

B. The average carat value of diamonds is greater than the carat value of 50% of the diamonds in the data.

- ○ Only A
- ○ Only B
- ○ Neither A nor B
- ◉ Both A & B — You Selected

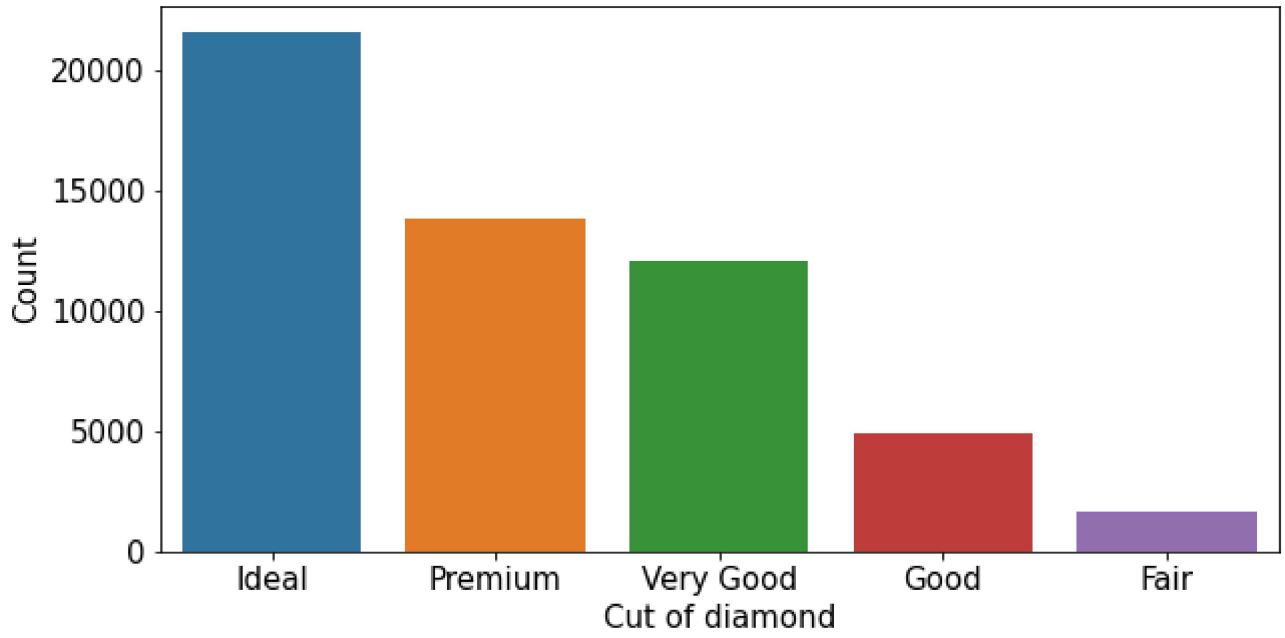
From the statistical summary of the data, we can observe that

1. The average price of diamonds is 3932.8 dollars.
2. Since the average carat value of diamonds is 0.798 and the median value is 0.7, we can say that the average carat value of diamonds is greater than the carat value of 50% of the diamonds in the data.

Q No: 2 Correct Answer

Marks: 1/1

Following is a graph of the different types of cuts of diamonds and how many diamonds of each type of cut have been sold.



Based on the above graph, which of the following statements is true?

- Good Cut and Fair Cut alone comprise 50% of the overall sales.
- Premium Cut has the highest sales.
- Fair and Premium Cut combined have more sales compared to that of Ideal Cut.
- Very Good and Premium Cut combined have more sales compared to that of Ideal Cut. You Selected

Based on the above graph, we can say that:

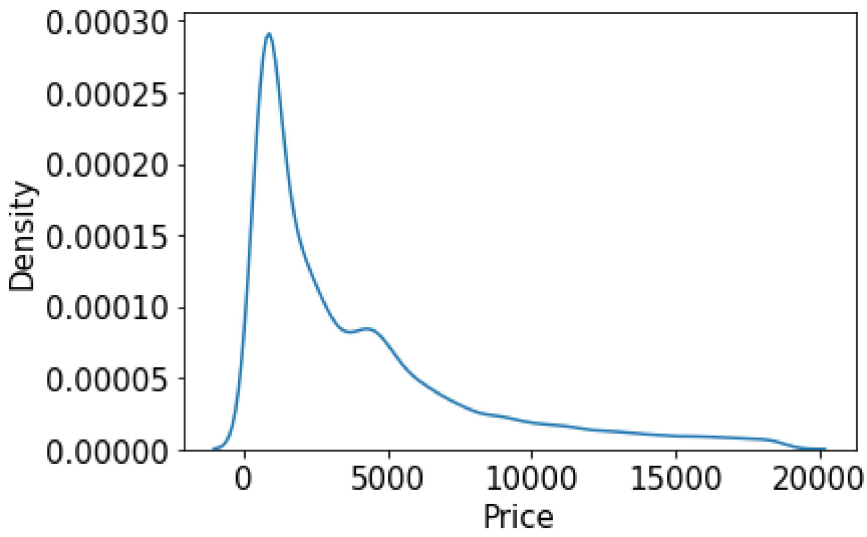
- Good Cut and Fair Cut alone do not comprise 50% of overall sales.
- Ideal cut has the highest Sales. (approximately 21000 diamonds of this cut were sold).

- Very Good(around 13000) and Premium Cut(around 14000) combined(around 27000) have more sales compared to Ideal Cut (around 21000).

Q No: 3 Correct Answer

Marks: 1/1

Following is a density curve of price of the diamonds:



Based on the above graph, which of the following statement is correct?

A. The majority of diamonds are priced below 5000 dollars.

B. The price of diamonds has a highly right-skewed distribution.

- Only A
- Only B
- Neither A nor B
- Both A & B — You Selected

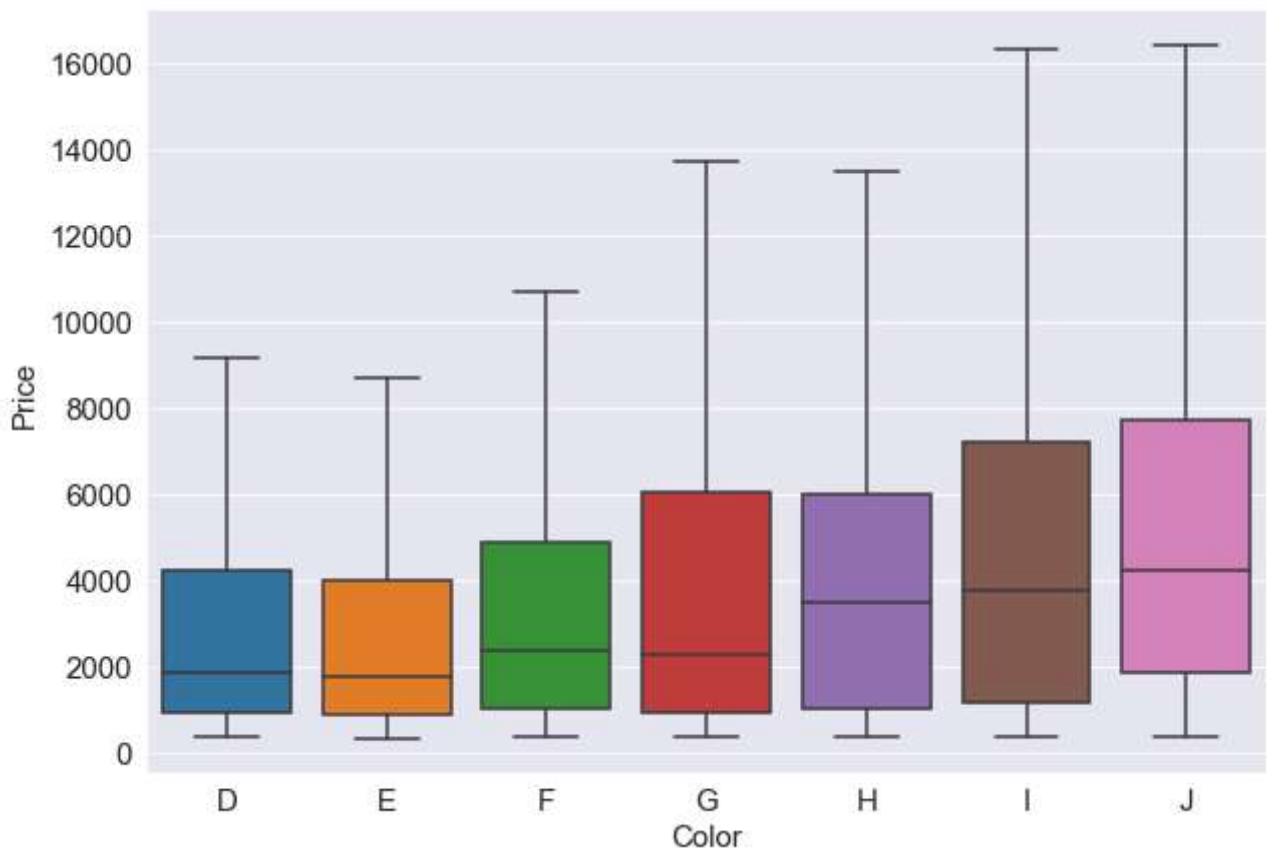
Based on the graph above,

- As we can observe from the graph, the density of the distribution is high between the range of 0 and 5000, so we can conclude that the majority of diamonds are priced below 5000 dollars.
- The price distribution is highly right-skewed.

Q No: 4 Correct Answer

Marks: 1/1

Following is a plot of price of diamonds and their colors (colors encoded):



Based on the above graph, which of the following statement is correct?

- ○ Approximately 75% of the diamonds of color F cost more than 4000 dollars
- ◉ Approximately 75% of diamonds of color G and H cost less than 6500 dollars — You Selected
- ○ 50% of diamonds of all the colors are priced less than 3000 dollars
- ○ Only 50% of diamonds of color E are priced below 4000.

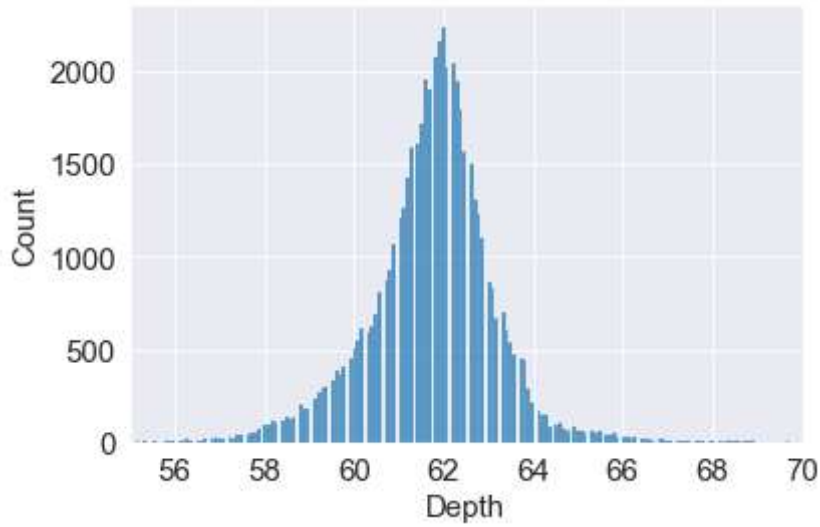
Based on the above graph,

- Here in the graph we can see 50%(median) of diamonds of color F are priced approximately below 2500, so based on this we can conclude that the statement is incorrect.
- Approximately 75% of diamonds of color G and H cost less than 6500 dollars.
- Since the median price of H, I, and J color diamonds is above 3000 dollars, the statement "50% of diamonds of all the colors are priced less than 3000 dollars" is False.
- Approximately 75% of diamonds of color E are priced below 4000.

Q No: 5 Correct Answer

Marks: 1/1

Following is a histogram of the depth of diamond (height of the diamond):



Select True/False for the following statement:

The distribution of depth of diamond is highly left-skewed.

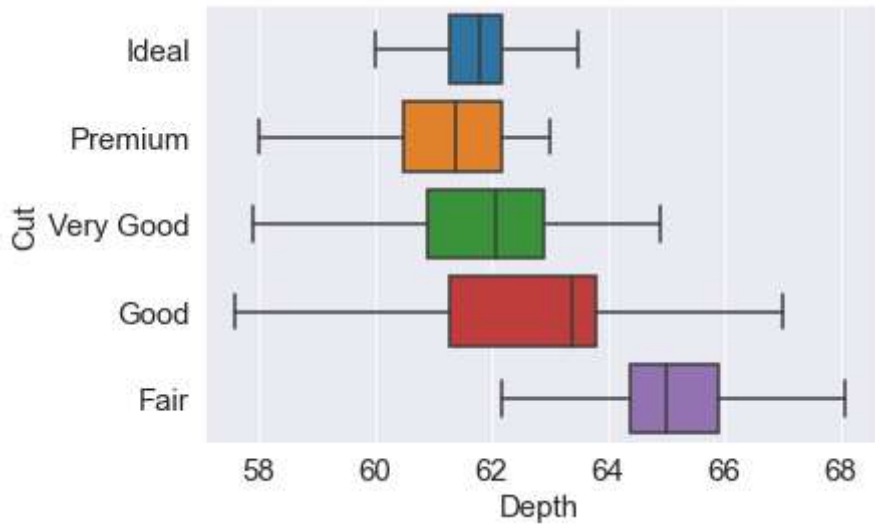
- True
- False — You Selected

Based on the above graph, we can observe that the depth of diamond has approximately a normal distribution.

Q No: 6 Incorrect Answer

Marks: 0/1

Following is a graph of relationship between cut of diamond and its depth:



Based on the above graph, which of the following statement is correct?

A. Atleast 75% of diamonds have depth of diamond greater than 60 mm.

B. Atleast 50% of diamonds of type 'Ideal' have a depth of less than 62mm.

- ○ Only A
- ◉ Only B — You Selected
- ○ Neither A nor B
- ○ Both A & B — Correct Option

Based on the above graph,

- As the Q1 of depth of various cuts of diamonds is greater than 60mm, we can conclude that approximately 75% of diamonds have a depth of diamond greater than 60 mm.
- As the median depth of diamonds of the 'Ideal' cut is less than 61mm, we can conclude that 50% of diamonds of type 'Ideal' have a depth of less than 62mm.

Q No: 7 Correct Answer

Marks: 1/1

The Diamond's table is the flat surface that you can see when you look at the diamond from above. The table in the diamond data ranges from 43 to 95. We want to convert this numerical column to a categorical column (categorize in size bins).

The rules for categorization are:

- If the table size is less than 56, categorize it as small.
- If the table size is greater than 59, categorize it as large.
- If the table size is ranging between 56 and 59, categorize them as medium.

Which of the following will correctly complete the code below to categorize the diamonds based on the given rules?

```
def table_cat(x):
    if '_______':      #A
        x = 'small'
        return x
    elif '______':     #B
        x = 'medium'
        return x
    else:
        x = 'large'
        return x

df['table'] = df['table'].apply(table_cat)
```

A. x<56

B. x>59

You Selected

A. x<56

B. x<=59

A. x>56

B. x>59

A. x>56

B. x<=59

The code to create the categorical bins is:

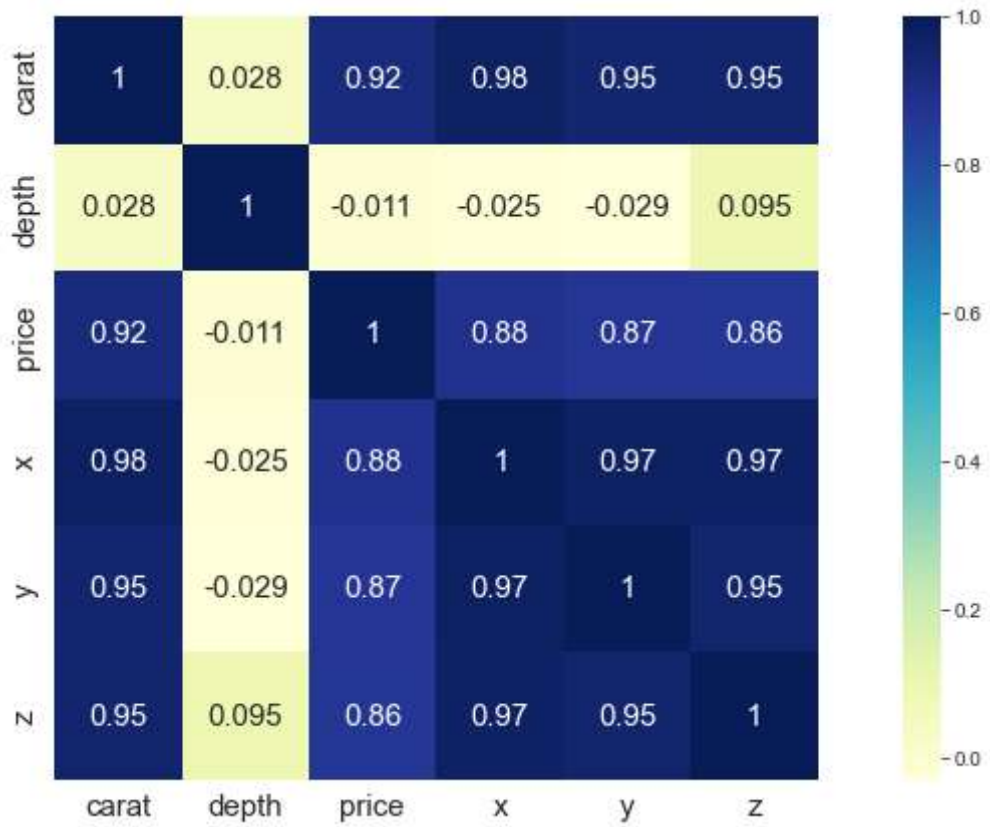
```
def table_cat(x):
    if x<56:      #A
        x = 'small'
        return x
    elif x<=59:     #B
        x = 'medium'
        return x
    else:
        x = 'large'
        return x
```

```
df['table'] = df['table'].apply(table_cat)
```

Q No: 8 Correct Answer

Marks: 1/1

Following is a correlation plot of all the variables that play important role in deciding the price of diamond.



Which among the following variables is having the highest correlation with price?

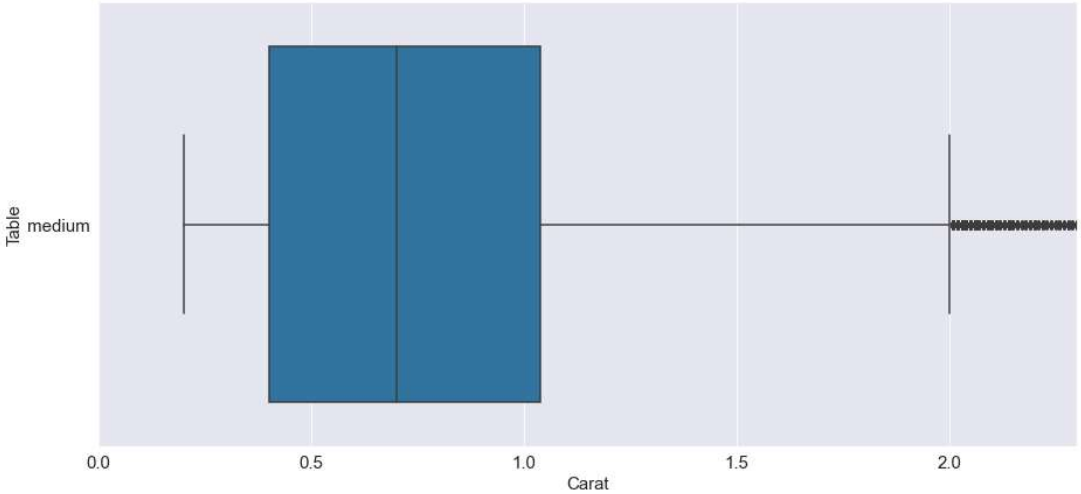
- depth
- carat — You Selected
- x
- y

Based on the above graph, carat has the highest correlation with price (0.92) among all the other columns. This indicates that higher the carat, higher will be the selling price and vice versa.

Q No: 9 Correct Answer

Marks: 1/1

The following graph shows the distribution for the carat value of diamonds having a medium table size.



Based on the above graph, which of the following statement is correct?

A. The distribution of carat is left-skewed.

B. There are diamonds which have unusually high carat value for medium table size.

Comments:

Previous

+ Add comments

Next

Proprietary content.©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

© 2024 All rights reserved

Privacy Terms of service Help