Curious IQ

Standard Metropolitan Areas Dataset

Test in Python

About the Dataset

It contains data of 99 standard metropolitan areas in the US. The data set provides information on 10 variables for each area for the period 1976-1977. The areas have been divided into 4 geographic regions: 1=North-East, 2=North-Central, 3=South, 4=West.

Data Description

land_area:

size in square miles

percent_city:

percent of population in central city/cities

percent_senior:

percent of population ≤ 65 years

physicians:

number of professionally active physicians

hospital_beds:

total number of hospital beds

graduates:

percent of adults that finished high school

work_force:

number of persons in work force in thousands

income:

total income in 1976 in millions of dollars

crime_rate:

Ratio of number of serious crimes by total population

region:

geographic region according to US Census

Curious IQ

We can see that the **regions** have 4 values, where:

1 = North-East

2 = North-Central

3 = South

4 = West

In the attached you will find the dataset (data.csv). You can import the dataset with the following code:

data = pd.read_csv("data.csv")

Now, since you are done with the above, here is the first question for you:

Which of the following information is correct about the data? (0.25) Provide the code that you found the correct answer (1.25).

a) There are 4 variables of 'float' dtype, 5 variables of 'int' dtype
b) There are 3 variables of 'float' dtype, 7 variables of 'int' dtype
c) There are 5 variabels of 'float' dtype, 5 variables of 'int' dtype
d) There are 6 variables of 'float' dtype, 4 variables of 'int' dtype

data.dtypes data.info()

What is the line of code that provides the descriptive statistics for all the numerical variables/features of the dataset?

data.describe()

How can you choose the first five elements of a dataframe?

a) data[0:5]
b) data[1,2,3,4,5]
c) data.head(5)
d) data.first_rows(5)
e) Answers c and d
f) Answers b and a

data.head(5)

data[0:5]

g) Answers a and c
h) All of the above (a,b,c,d)

Which of the following is true about bar chart and histogram?

a) Histogram indicates distribution of non-discrete variables
b) Histogram presents quantitative data while bar chart presents categorical data/discrete data
c) Bar chart indicates comparison of discrete variables/categorical variables
d) In histogram bars cannot be re-ordered while in bar charts bars can be re-ordered
e) All of the above

If you want to express in percentages the distribution of people around each region in the Metropolitan areas, what chart would you use?

a) Bar Chart
b) Scatter Plot
c) Pie Chart
d) Histogram

Which of the following information is correct about the data?

data['land_area'].median() data['crime_rate'].min()

a) Median land areas of these Metropolitan areas = 1951 and Minimum crime rate among all the Metropolitan areas = 23.32
b) Maximum income in million dollars of all the metropolitan areas = 72100
c) Total count of null entries in 'hospital beds' = 0 and the least of the Metropolitan areas lie in region 4 data. isna(). sum(). sum()
d) Only 25 Metropolitan areas lie in region 2 and Average crime rate region 4 = 69.455
e) All of the above data[data.region == 4].mean()

Which of the following is the top three land areas of region 4 based on income? (0.5) Provide the code that you found the correct answer (1).

a) 2480, 782, 4226,
b) 27293, 782, 9155
c) 2480, 2793, 9155
d) 27923, 2480, 782

data.sort_values(['income'], ascending = False)[data.region == 4][:3]

Curious IQ

Which of the following information is correct about Standard Metropolitan Data? (0.25) Provide the code that you found the correct answer (1.25).

a) There are 16 Metropolitan areas that their income is greater than 10000
b) There are 40 Metropolitan areas where their percent city is above the average percent city
c) The land area 9155 has the maximum crime rate
d) There is no Metropolitan area in region 1 where graduates are less than 35
e) Only a,b,d answers are correct
f) Only a,c,d answers are correct
g) All of the above

```
a) data[data['income'] > 10000]
b) data[data['percent_city'] < 44.270588]
c) data[data['crime_rate'] == 85.62]

data[[data['region'] ==1],
[data['graduates']< 35]]
```