Kaggle Master-Week 3 Quiz Total points 100/100 ?

Email address *

sevdanurgenc@gmail.com

Name-Last Name: *

Sevdanur GENÇ

✓ **Q1- Which of the following methods does not allow us to draw a**   10/10
   subplot? *

⊙ sns.relplot()

⊙ sns.catplot()

◉ sns.heatmap()   ✓

⊙ plt.subplots()

✓ Q2- Which statement is not true with the following code?   10/10
museum_data =
pd.read_csv(museum_filepath,index_col="Date",parse_dates=True) *

○ When parse_dates = True, the type of Date column in museum_data becomes datetime

○ The index value of the museum_data is the Date column in the csv file.

◉ Above code means that creating a new Date column that is not in csv and this   ✓
column is defined as the index of the museum_data.

○ If we did not use parse_dates, the type of the Date column would not change to datetime. (Assuming that the original type of the column is not datetime)

✓ Q3- Imagine that you have a company and you would like to create a   10/10
plot which shows the sales based on date. Which one of the following
plot function is less likely suitable for this job? *

○ sns.scatterplot

○ sns.barplot

○ sns.lineplot

◉ sns.heatmap   ✓

✓ Q4- When do we use bar charts? *   10/10

○ To see how the data distributed.

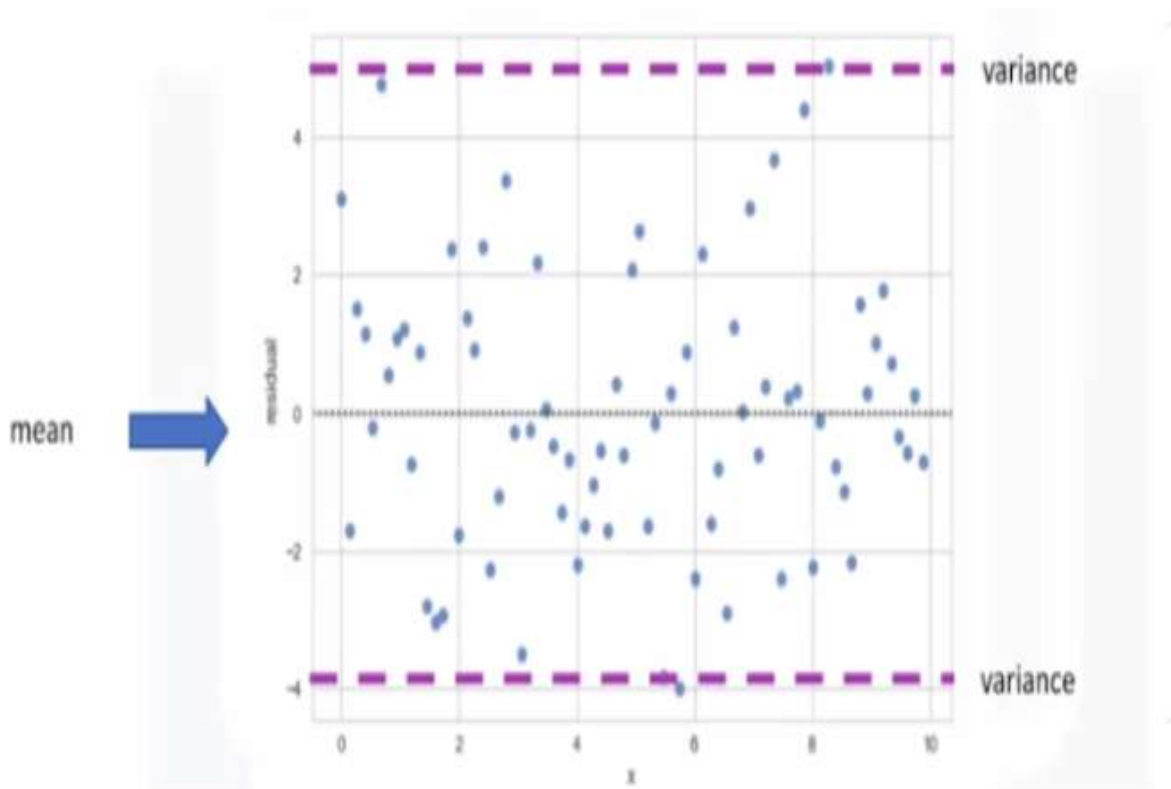○ To see if data is a normal distribution.

○ To compare categories by a feature like quantities. ✓

○ To see trend of a time series data.

✓ Q5- According to the residual plot below, which statement is false? *   10/10

mean → variance (top) ... variance (bottom)

○ It suggest that the linear model would be appropriate.

◉ Data points distributed on a curvature.   ✓

○ Data points spread around randomly.

○ The plot is created using seaborn.

```
I-      sns.lineplot(data=fifa_data) --> Plot the data
II-     fifa_data = pd.read_csv(fifa_filepath, index_col="Date", parse_dates=True) --> Load the data
III-    fifa_data.head() --> Examine the data
```
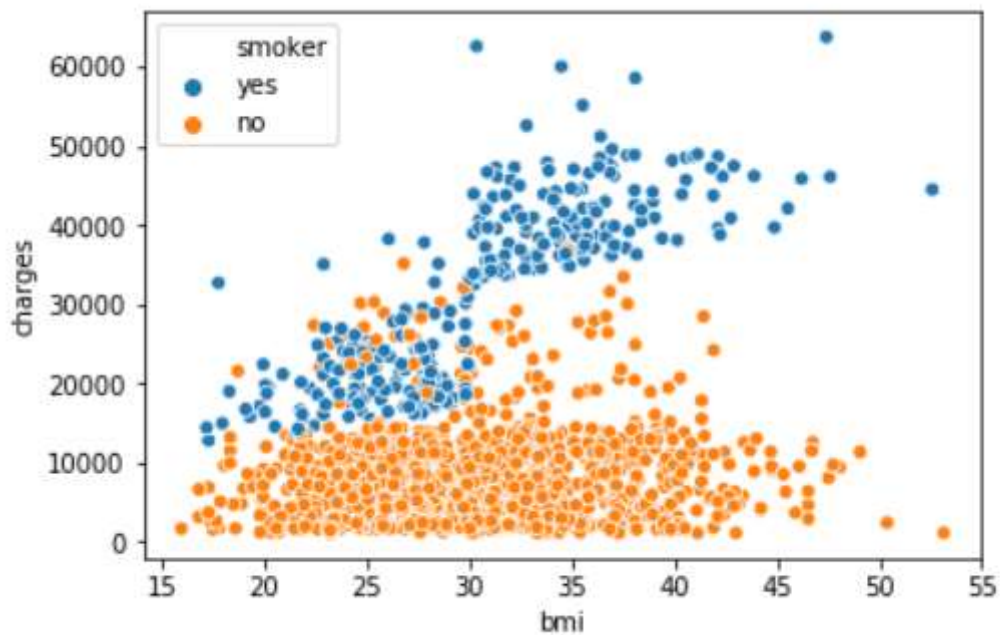
○ I-II-III

○ II-I-III

○ III-II-I

◉ II-III-I   ✓

✓ Q7- Which one is the correct option for the below visualization? *   10/10

○ sns.scatterplot(x=insurance_data['bmi'], y=insurance_data['charges'])

○ sns.lmplot(x="bmi", y="charges", hue="smoker", data=insurance_data)

● sns.scatterplot(x=insurance_data['bmi'], y=insurance_data['charges'], hue=insurance_data['smoker'])   ✓

○ sns.regplot(x=insurance_data['bmi'], y=insurance_data['charges'])

✓ Q8- Which of the below statement is false? *   10/10

○ sns.lmplot - Useful for drawing multiple regression lines, if the scatter plot contains multiple, color-coded groups.

● sns.swarmplot - Useful for comparing quantities corresponding to different groups   ✓

○ sns.heatmap - Used to find color-coded patterns in tables of numbers

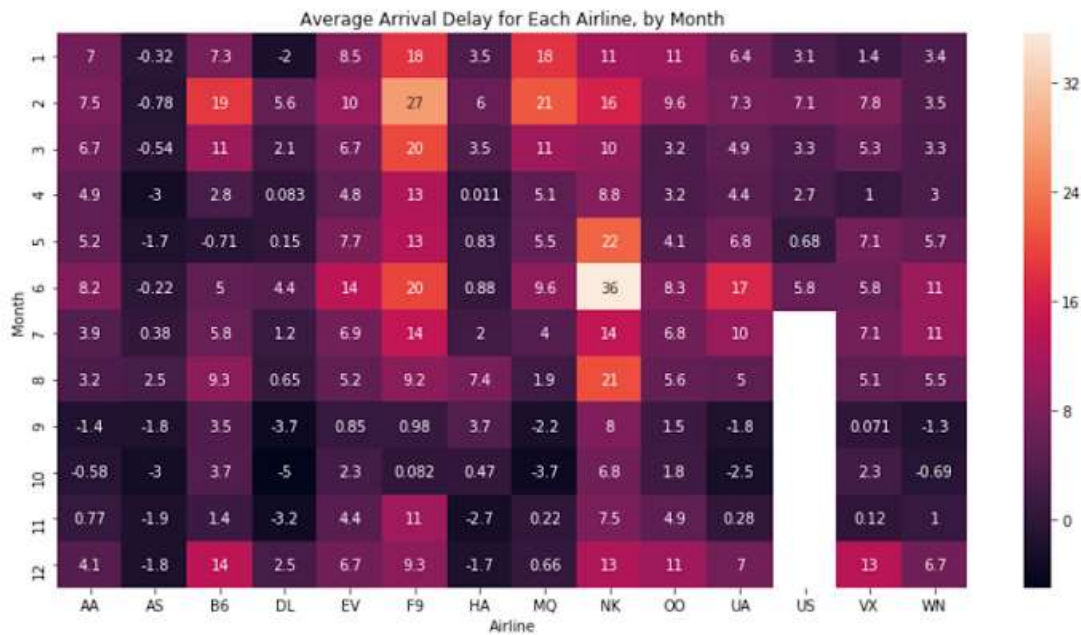○ sns.distplot - Show the distribution of a single numerical variable

---

✓ Q9- There are many different chart types that you can use to understand relationships between variables in your data. Which one is not used for showing a relationship? *   10/10

○ sns.regplot

○ sns.lmplot

○ sns.scatterplot

● sns.jointplot   ✓

Average Arrival Delay for Each Airline, by Month

○ This shows a bar chart.

◉ The months 9-11 are the best schedules in that year. ✓

○ AS Airline has the most delayed flights.

○ The light boxes are the desired relations if we want to define the least delayed flights.