

Assignment-based Subjective Questions:

1. From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?

- ✚ Categorical stuff, like types of weather or days of the week, can affect how many bikes get rented.
- ✚ We can figure out how much each type of weather, for example, affects bike rentals by looking at how they change together.

2. Why is it important to use `drop_first=True` during dummy variable creation?

- ✚ When we turn categories into numbers for the computer, we want to avoid making extra categories that we can already figure out.
- ✚ Using `drop_first=True` helps us avoid confusion and keeps things clear.

3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?

- ✚ When we look at pairs of things, like temperature and bike rentals, we can see which one seems to change the most with bike rentals.
- ✚ The one that moves the most with bike rentals is the most correlated.

4. How did you validate the assumptions of Linear Regression after building the model on the training set?

- ✚ After we make a guess about bike rentals with our math, we need to make sure our guess makes sense.
- ✚ We check our guess using graphs and tests to see if our math is okay.

5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?

- ✚ After all the math, we can see which things, like temperature or weather, are the most important for figuring out bike rentals.

General Subjective Questions

1. Explain the linear regression algorithm in detail.

- ✚ Linear regression is like trying to draw a straight line through a bunch of points on a graph. It's trying to find the best-fitting line that predicts something based on other factors. In our bike project, it's figuring out how to predict bike rentals based on things like weather.

2. Explain the Anscombe's quartet in detail.

- ✚ Anscombe's quartet is basically four sets of data that look totally different when you just see the numbers. But when you plot them out, they surprisingly make the same kind of pattern. It's a reminder that sometimes numbers can trick you, and you need to look at the big picture.

3.What is Pearson's R?

- ✚ Pearson's R is like a measure of how closely two things are related. If it's close to 1, it means they go up and down together. If it's close to -1, it means when one goes up, the other goes down. And if it's close to 0, there's not much of a connection.

4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?

- ✚ Scaling is just making sure all your numbers are playing by the same rules. In our project, it's about making sure that numbers that are really big or small don't dominate everything else. It's like making sure everyone's voice is heard equally.

5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?

- ✚ VIF is like a detective that checks if our data is too predictable. Sometimes, if one thing can be totally guessed by other things, VIF freaks out and gives us a big number, or even infinity, to say "Hey, this isn't working!"

6.What is a Q-Q Plot and Its Importance in Linear Regression:

- ✚ A Q-Q plot is like checking if your data follows the rules. It's comparing your data to what's considered normal. In our project, it's important because it helps us see if our data behaves the way we'd expect, which is crucial for making good predictions.