

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?
All the range of date, month & year are in the valid range, hence we can proceed ahead with this column without dropping / changing any records. On analysing date column we can see that all the attributes of date except for day are comfortably covered under month & year column.
2. Why is it important to use drop_first=True during dummy variable creation?
3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?
4. How did you validate the assumptions of Linear Regression after building the model on the training set?
5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?

General Subjective Questions

1. Explain the linear regression algorithm in detail.
Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task.
2. Explain the Anscombe's quartet in detail.
Anscombe's Quartet can be defined as a group of four data sets which are nearly identical in simple descriptive statistic
3. What is Pearson's R?
The Pearson correlation coefficient or as it denoted by r is a measure of any linear trend between two variables
4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.