Assignment based subjective Questions

1. The categorical vars have high VIF ad thus playing a significant role in predicting the target var
2. If we do not use drop_first = True, then n dummy variables will be created, and these predictors(n dummy variables) are themselves correlated which is known as multicollinearity and it, in turn, leads to Dummy Variable Trap.
3. Windspeed
4. We can validate them using R square and p value
5. Yr, weathersit and mnth

General Subjective Questions

1. Linear Regression is an algorithm that belongs to supervised Machine Learning. It tries to apply relations that will predict the outcome of an event based on the independent variable data points. The relation is usually a straight line that best fits the different data points as close as possible. The output is of a continuous form, i.e., numerical value. For example, the output could be revenue or sales in currency, the number of products sold, etc. In the above example, the independent variable can be single or multiple
2. Scaling is performed to bring the range of variables into a same range so that they don't have bigger impact in prediction of target
3. VIF in inf when there exist a perfect correlation between the variables
4. In Statistics, Q-Q(quantile-quantile) plots play a very vital role to graphically analyze and compare two probability distributions by plotting their quantiles against each other. If the two distributions which we are comparing are exactly equal then the points on the Q-Q plot will perfectly lie on a straight line y = x