

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?

Ans: 1. Users tend not to rent and use bikes during Snow or Rain.
2. Users generally prefer riding bikes when the temperature is higher.
3. Bike rentals on Holidays are lesser than working days.
4. Windspeed has a significant effect on bikes being rented.
5. Spring season generally sees lesser bike rentals

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

Ans: **drop_first=True** is **important to use**, as it helps in reducing the extra column created **during dummy variable creation**. Hence it reduces the correlations created among **dummy variables**

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

Ans: The numerical variable 'registered' has the highest correlation with the target variable 'cnt', if we consider all the features. But after data preparation, when we drop registered due to multicollinearity the numerical variable 'atemp' has the highest correlation with the target variable 'cnt'.

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

Ans: Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task to compute the regression coefficients. Regression models a target prediction based on independent variables. Linear Regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So this regression technique finds out a linear relationship between x(input) and y(output). Hence it has got the name Linear Regression.

Assumptions of Linear Regression. There are 5 basic assumptions of Linear Regression Algorithm:

- 1) **Linear Relationship between the features and target**
- 2) **Little or no Multicollinearity between the features**
- 3) **Homoscedasticity Assumption**
- 4) **Normal distribution of error terms**
- 5) **Little or No autocorrelation in the residuals**

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

Ans: i) Users tend not to rent and use bikes during Snow or Rain.

We can reduce the price of bike rentals during times of snow or rain by implementing dynamic pricing

We can also implement hand warmers into the bike handlebars

At the bike docking station, we can also have thin disposable rain coats

ii)Users generally prefer riding bikes when the temperature is higher.

We can have different offers and coupons sent to riders when we can see a dip in the temperature to increase interest

We can have dynamic pricing to increase rates by 5% for every 5°C change in temp above the mean of 20°C

iii)Bike rentals on Holidays are lesser than working days.

Like in the above suggestion, we can send coupons and offers to riders to increase interest on holidays

We can see that bikes are predominantly rented as a means of workplace commuting during workdays. We can have additional marketing to pose bike rides as a lifestyle activity and host events during holidays so that riders can rent bikes and participate in these events.