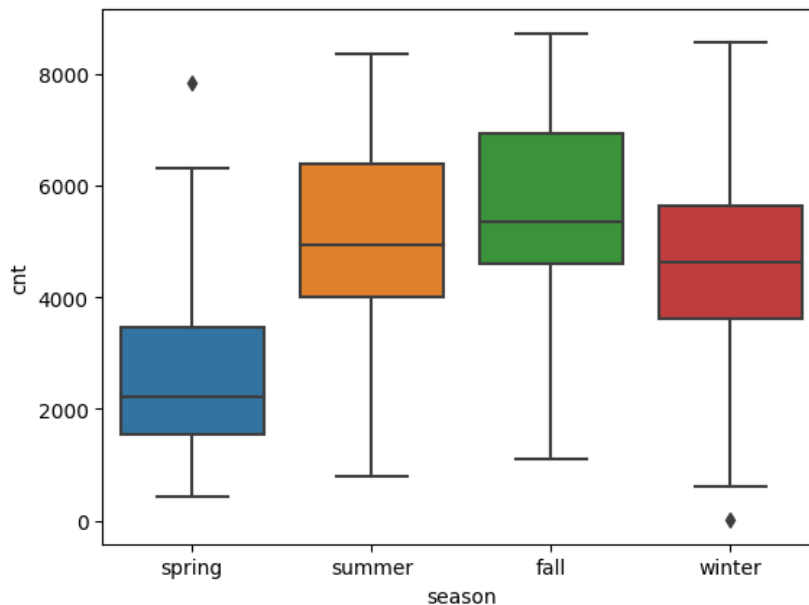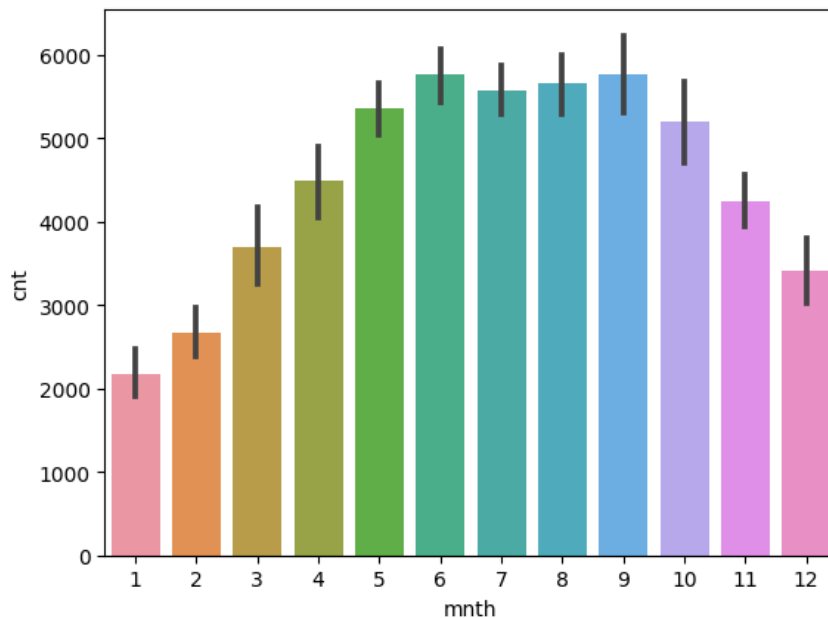# Assignment-based subjective Questions.

Answers

1. In the feature season there is a increase in the count of bike registrations

In the month of march the count for bike rental starts to increase

2. If a feature has k values then we can still describe the values by adding dummy variables by k-1

3. 'temp' variable has the highest correlation with the cnt target variable

4. After building a model we do residual analysis on the train dataset to see the error distribution after it we validate our model by using r2 score of the model on test dataset.

5. Based on the final model the top 3 features are 'thunderstorm','December','July' because these three features have very less VIF and p values compared to other features from the analysis

General Subjective Questions

Answers

1. Linear regression is a supervised algorithm the model will try to learn from the dataset and predicts the result linear regression is used for continuous features
   The linear function of a given line is

   $h_\theta(x) = \theta_0 + \theta_1 X_1$
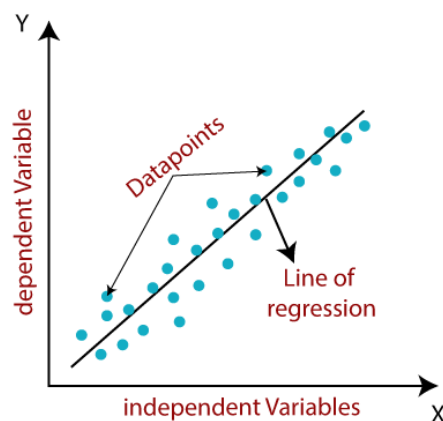
   The Line is a bestfit line or it's a prediction line

   The linear regression uses gradient descent algorithm.

2. Anscombe's qurter comprises four datasest that have nearly identical simple descriptive statistics each dataset consists of 11 (x,y) points the importance of the graphing data before data analyzing it and the effect of the outliers in the statistical properties
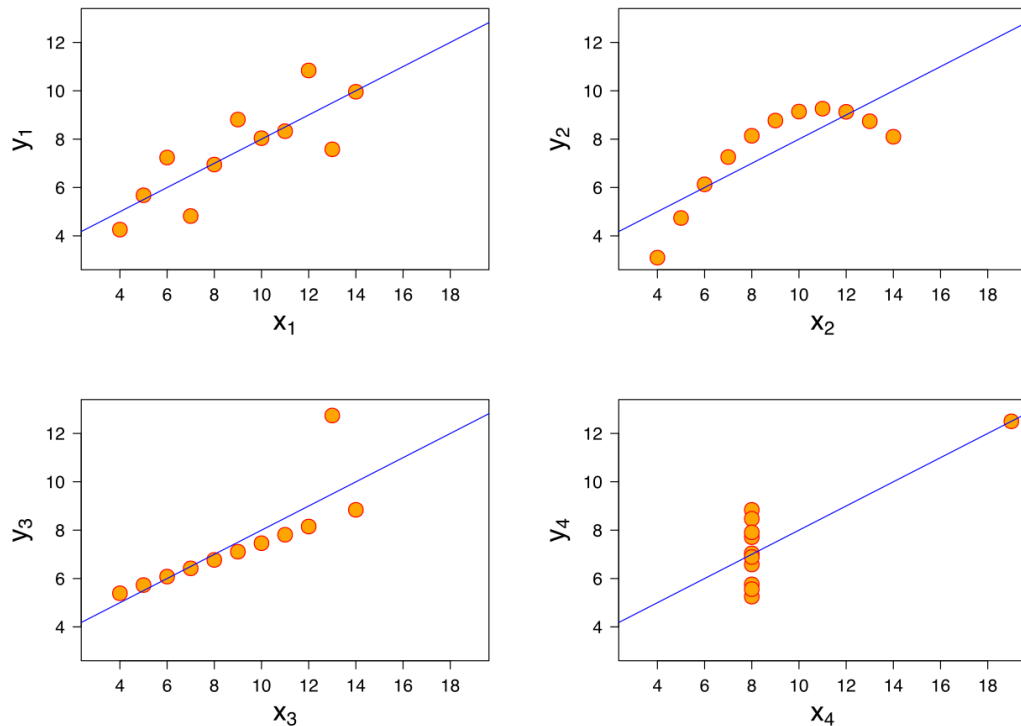
3. The **Pearson correlation coefficient (*r*)** is the most common way of measuring a linear correlation. It is a number between −1 and 1 that measures the strength and direction of the relationship between two variables.
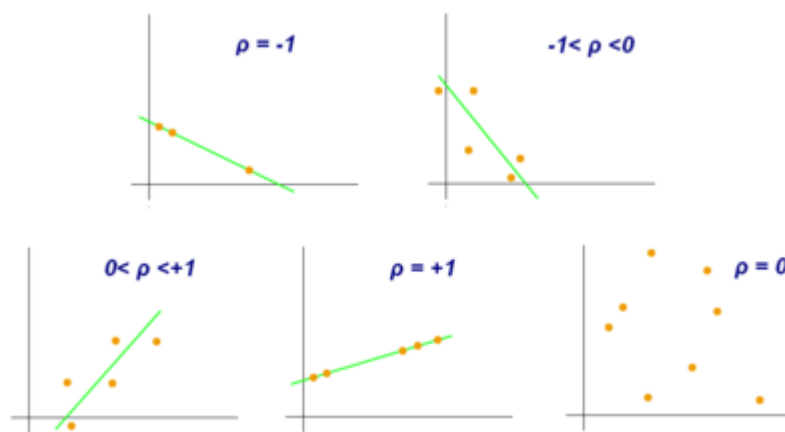
4. Scaling is performed to scale down the features to the same magnitude because in the dataset each features has different units and magnitude to run gradient descent algorithm for the faster convergence we scale down the features to the same magnitude.
Normalized scaling converts the features between 0 to 1

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Standardized scaling uses.

$$X' = \frac{X - \mu}{\sigma}$$

5. If there is perfect correlation, then VIF = infinity. This shows a perfect correlation between two independent variables. In the case of perfect correlation, we get R2 =1, which lead to 1/(1-R2) infinity.

6. Q-Q Plots (Quantile-Quantile plots) are plots of two quantiles against each other. A quantile is a fraction where certain values fall below that quantile. For example, the median is a quantile where 50% of the data fall below that point and 50% lie above it. The purpose of Q Q plots is to find out if two sets of data come from the same distribution. A 45 degree angle is plotted on the Q Q plot; if the two data sets come from a common distribution, the points will fall on that reference line.