Assignment based subjective question

2) drop_first=True is used during dummy creation to eliminate the one of the dummy variable created. This is because to represent a categorical data with ‘n’ variables can be represented by “n-1” dummy variables. Hence the excess dummy variable is not used for the analysis.

3) Temperature has the highest correlation with the target variable **Count** among all the numerical variable.

4) The model is valuated using the following three steps.

a) Plot the error distribution. The error distribution should have mean at zero.
b) Plot y test data with the predicted y_pred data. The scatter plot should show a linear relationship.
c) R2 Score- R2 score of the model with the y-test data should be arrived and the R2 score should be significant and comparable with that of the train data.

5) The three most significant contributing features are –

a) Atemp or temp- both are highly correlated and hence can be considered as one and same
b) Year- the business seems to increase its demand Y-o_Y
c) Humidity and Windspeed seems to have a negative correlation with the demand

General Subjective Questions

1.

2. Ascombes quartet represents 4 data sets. These data sets contain data which has similar characteristics like mean, variance etc. But while plotting a distribution chart we get 4 charts which some of which shows different patterns. Ascombes quartet is a proof that inferences cannot be made just on characteristics like mean or median and that graphical representation of the data is required to get complete and correct insights.

3. Pearsons R- is a measure of correlation between two sets of data. It is calculated by arriving the ratio between covariance between two sets of data and the product of their standard deviations.

4.