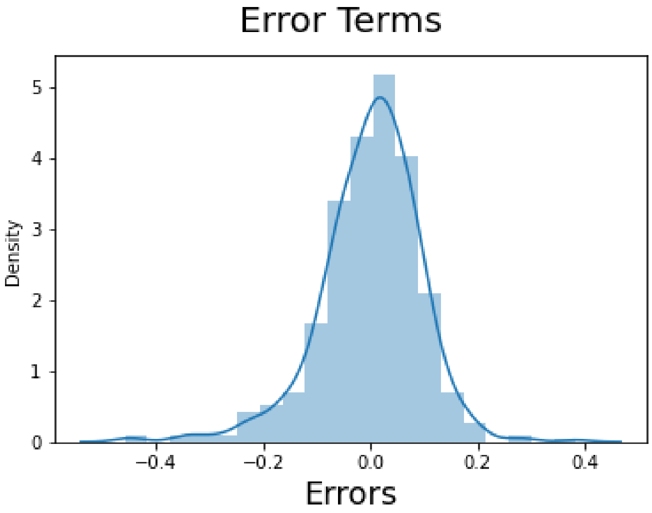
1. From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?
   - The Fall season which happens around September , october months have higher counts for bike demands.
   - Clear weather is also indicator of high demand of bikes
   - Monday and Weekends mainly Friday and Saturday show high bike demands as well

2. Why is it important to use drop_first=True during dummy variable creation?
   - For n levels in a categorical variable, the number of dummy variables required is n-1 . The n-1 dummy variables can define the dropped column easily. If the column is not dropped , it will exhibit high multicollinearity since all the dummy variables are correlated.

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

4. How did you validate the assumptions of Linear Regression after building the model on the training set?
   - One of the major assumptions of linear regression model is that the error terms are normally distributed. I created a distribution plot between the predicted count vs the actual count after model building.
     (y_train – y_train_count) gave the eror terms which was plotted shown below. The error terms are normally distributed with mean at 0 and a standard devation of 0.2.



5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?
   - Temperature , Year and Clear weather are the top three with coefficients 0.4496 , 0.2347 and 0.079 respectively.

6. Explain the linear regression algorithm in detail.
   - Linear regression is a supervised ML approach which tries to identify relations between the target and independent variables. Linear regression searches for relationships among variables and the idea is to find a function that maps all the variables well. It tries to explain the relation between target and independent variables with a best fit straight line which has the minimum error. It is defined as:

     y or f(x) = c + b1(x1) + b2(x2) + b3(x3) +b4(x4) + b5(x5) .....bn(xn) + e
     where x1, x2….xn are all independent variables with b1, b2…bn coefficients . c is the slope intercept and e is the error

     The strength of the linear regression model can be assessed using 2 metrics:
     1. R^2 or Coefficient of Determination
     2. Residual Standard Error (RSE)
     Linear regression works under certain assumptions:
     1. Linear relationship between X (independent) and Y (dependent)

2. Error terms are normally distributed (not X, Y)
3. Error terms are independent of each other.
4. Error terms have constant variance (homoscedasticity)

7. Explain the Anscombe's quartet in detail.
    - Statistician Francis Anscombe created 4 datasets to explain the importance of plotting graphs before model building or making any statistical observation. These 4 dataset have nearly similar statistical observations with same mean ,variance etc , however when plotted over a scatterplot the difference becomes very clear therefore emphasizing the import of data visualization.

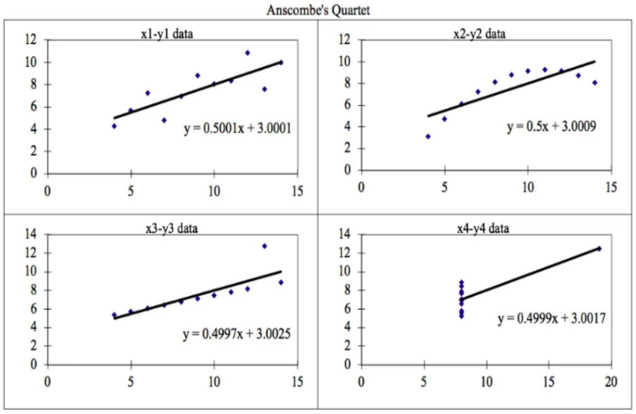


Image by Author

8. What is Pearson's R?
    - Pearson' R is a bivariate correlation which is a measure of correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviation. The values of Person's r lie between -1 to 1. 1 implies that x and y are completely linear with all data lying on the line with positive slope and vice versa. 0 means there is no linear relationship between the variables. The formula is given as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

9. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
    - Scaling is the method to bring all the variables on a similar scale for ease of interpreting the coefficients. If the variables are at different scale , comparing the coefficients becomes very difficult to comprehend. Also scaling helps in faster convergence of gradient descent methods , meaning identifying the cost of reducing errors becomes faster.
    There are two types of scaling , Standardization and Normalization (Min Max Scaling).
    Standardization – The variables are scaled in such a way that the mean becomes 0 and the SD is 1.
    Formula: x= (x- mean(x))/ sd(x).
    Normalization (min max scaling) -- The variables are scaled using min and max values so that all the values lie between 0 and 1.
    Formula: x = (x-min(x)) /max(x) – min(x)

10. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
    - When there is perfect correlation i.e when r squared becomes 1 then VIF which is : 1/ 1 – r squared gives 1/0 and hence VIF becomes infinite.
    This situation might happen if we have large set of predictors available for model fitting and some variables are able to create perfect multiple regressions on other variables or the model might remember all the data points therefore giving r squared as 1.

11. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
    - A Q-Q plot (quantile to quantile) plot is graphical representation technique to determine if two data sets come from same population with a common distribution. It also helps in identifying if the datasets have common location and scaling.It checks the data distribution to see if both the datasets have same distributional shape. It also helps in identifying if the datasets have similar tail behaviour.