**Boom Bike – The Bike Sharing Service**

**Exploratory Data Analysis- (EDA)**

A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" which is usually computer-controlled wherein the user enters the payment information, and the system unlocks it.

**1) Business Scenario:**

Here, A US bike-sharing provider ‘Boom Bikes’ has recently suffered considerable dips in its revenues due to the ongoing Corona pandemic. The company is finding it exceedingly difficult to sustain itself in the current market scenario. So, it has decided to produce a mindful business plan to be able to accelerate its revenue as soon as the ongoing lockdown comes to an end, and the economy restores to a healthy state.

**2) Problem Statement:**

* Which variables are significant in predicting the demand for shared bikes?
* How well do those variables describe the bike demands?

**3) Strategies:**

We required a model which gives the demand for shared bikes with available independent variables. It will be used by the management to understand how exactly the demands vary with distinctive features. So, they can accordingly manipulate the business strategy to meet the demand levels and meet the customer’s expectations. We will use “The Linear Regression Model” to solve the problems and minimize the error, this model will be an effective way for management to understand the demand of a new market.

**4) Sources:**

This dataset has 730 rows and 16 columns.

Data columns (total 16 columns):

dtypes: float64(4), int64(11), object(1)

instant 730 non-null int64

dteday 730 non-null object

season 730 non-null int64

yr 730 non-null int64

mnth 730 non-null int64

holiday 730 non-null int64

weekday 730 non-null int64

workingday 730 non-null int64

weathersit 730 non-null int64

temp 730 non-null float64

atemp 730 non-null float64

hum 730 non-null float64

windspeed 730 non-null float64

casual 730 non-null int64

registered 730 non-null int64

cnt 730 non-null int64

**5) Exploratory Data Analysis (EDA):**

**Questions:**

1) Visualising Numeric Variables. (Pair plot)

2) Visualising Categorical Variables. (boxplot)

3) What is the distribution of count? use Univariate Analysis:

1. Relation between season and count. (Bar plot)
2. Relation between Weather and count. (Bar plot)
3. Relation between Year and count. (Bar plot)
4. Relation between Month and count. (Bar plot)
5. Analysis between temperature and count. (Scatter plot)
6. Analysis between Humidity and count. (Scatter plot)

4) Finding Correlation between variables. (Heatmap)

5) Data Preparation.

6) Splitting the Data into Training and testing Sets.

1. Split the data into Training sets and Testing sets. (Ratio should be for Train\_size = 0.7, and Test\_size=0.30)
2. Train data:

rows 510, columns 29.

1. Test Data:

rows 220,

columns 29.

7) Rescaling the Features:

All the numeric variables should be mapped between 0 to 1.

8) Building a linear model. (using ‘Statsmodel’).

a) Check VIF: gives a basic quantitative idea about how much the feature variable are correlated with each other.

b) If a variable is insignificant, can be dropped. In this data ‘January’, ‘humidity’, ‘holiday’, ‘windspeed’, and ‘July’ variables are insignificant. So, we can drop it off and rebuild the new model.

c) The VIFs and p-values both are in an acceptable range. We can build our linear model.

d) Here, R-squared value is 82.6% means 0.826. which is between 0 to 1.

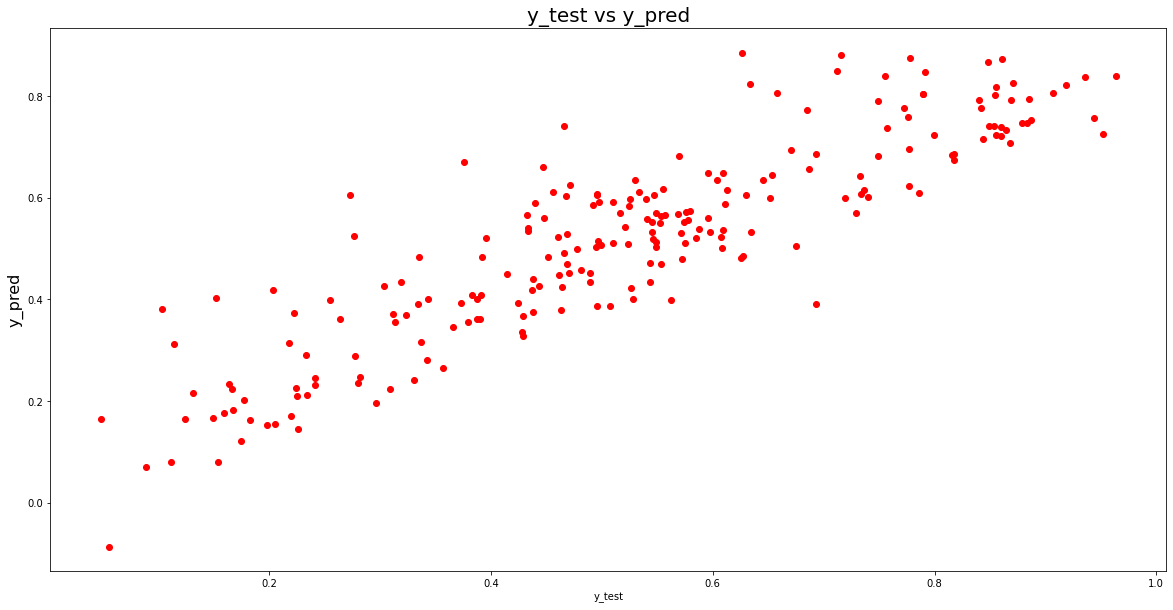
9) Residual Analysis of the Train Data for the error term. (Distribution plot)

10) Making Predictions.

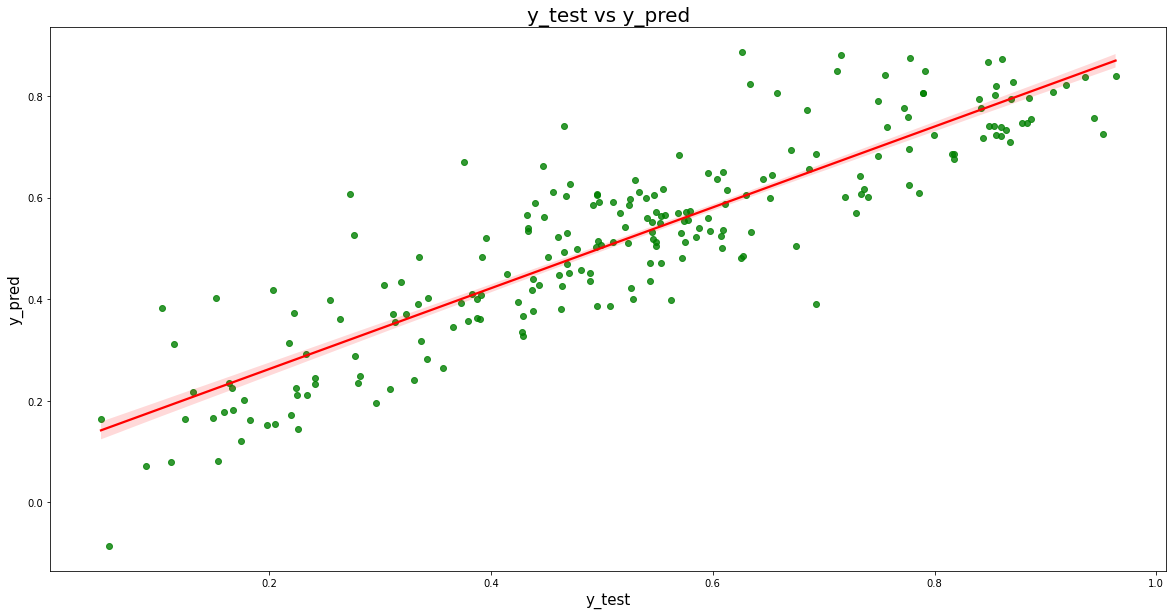
a) R2 squared of prediction: 0.82

b) And adjusted R2 squared prediction: 0.80

11) Model Evaluation.



12) Final linear regression Model, which gives an appropriate idea of the dataset. (Linear regression plot)



**6) Insights:**

a) here, the temperature variable has the highest coefficient of 0.3680, which means if the temperature increases by one unit, the number of bike rentals increases by 0.3680.

b) there are some variables with negative coefficients. A negative coefficient suggests that the dependent variable tends to decrease as the independent variable increases.

c) Mist + Cloudy: -0.0799, light\_snow: -0.3135, spring: -0.1462 has a negative coefficient, which suggests as the independent variable changes have given a one-unit shift in the independent variable while holding other variables in the model constant.