CAPASTONE PROJECT

Seoul Bike Sharing Demand Prediction

Team Members:

- kunal Gawande
- Chinmay Rojatkar
- Deepali Mahajan
- Nikhil Aggarwal
- Bipasha zade

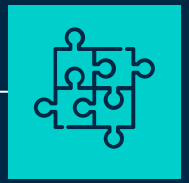
Acknowledgement

We would express our gratitude towards the entire team of "*Almabetter Team*" for acknowledging us with such important domain and providing us an opportunity to work on real life problems through Capstone Project

Problem Statement

- Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

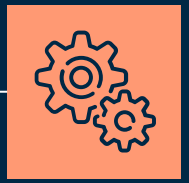
TABLE OF CONTENTS



01

PROBLEM & SOLUTION

- Importing the necessary packages and libraries
- Mounting the drive for importing the data.
- Checking for missing, NaN values,Null values.



02

OUR PROCESS

- Observing the datatypes
- Observing the correlation among independent variables.
- Exploring the data set.
- Exploring the categorical value numerical features from data

03

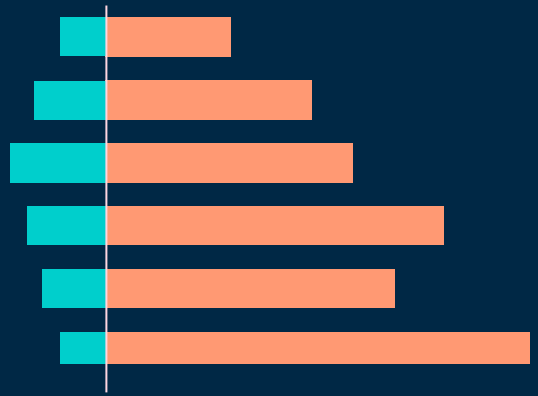
TARGET

- Exploring different target variable.
- Splitting the data and training the data.
- Observing the results.

Data Description :

Dependent variables:

Rented Bike count - Count of bikes rented at each hour



Independent variables:

Date : year-month-day
Hour - Hour of day
Temperature- Temperature in Celsius
Humidity - %
Windspeed - m/s
Visibility - 10 m
Dew point temperature - Celsius
Solar radiation - MJ/m
Rainfall - mm
Snowfall - cm
Seasons - Winter, Spring, Summer, Autumn
Holiday - Holiday/No holiday •
Functional Day –No Func, Func

8,760 Rows × 11 Columns

In the data is no missing values

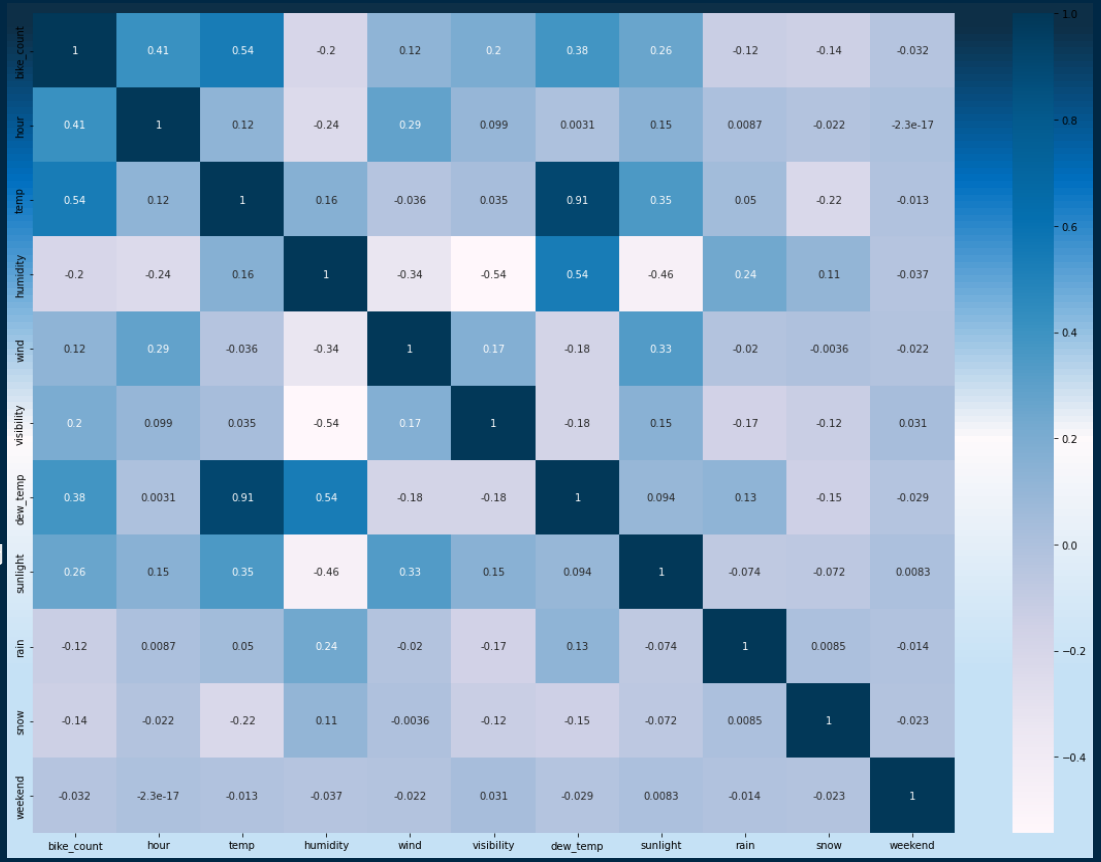
CORRELATION:

Rented Bike count and Temperature

Rented bike count are highly{0.54} correlated with temperature

Rented Bike count and weekends

Rented bike count are negativity {-0.54} correlated with temperature



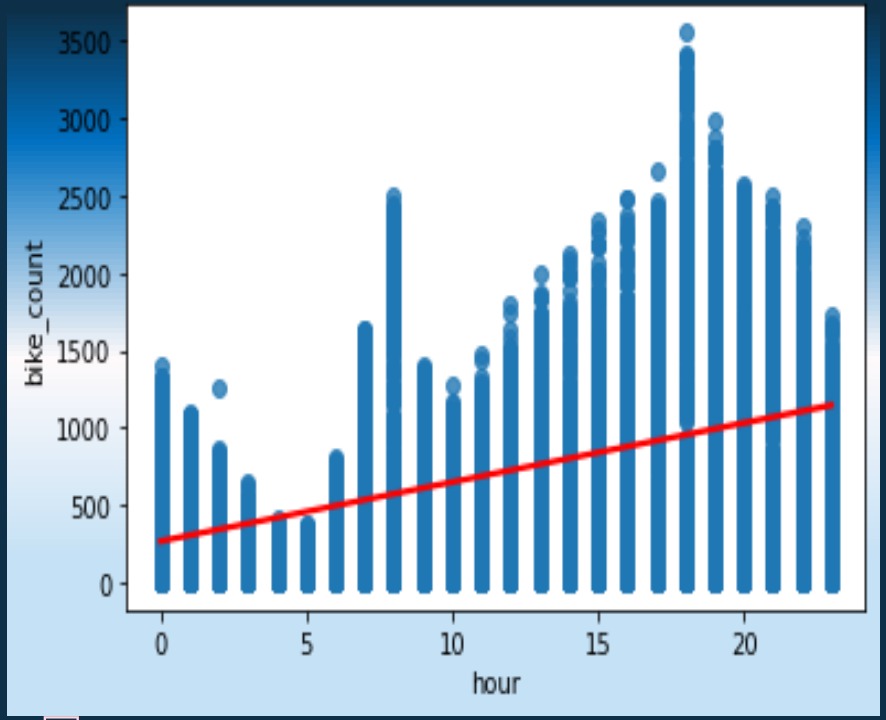
Rented Bike count and Rain

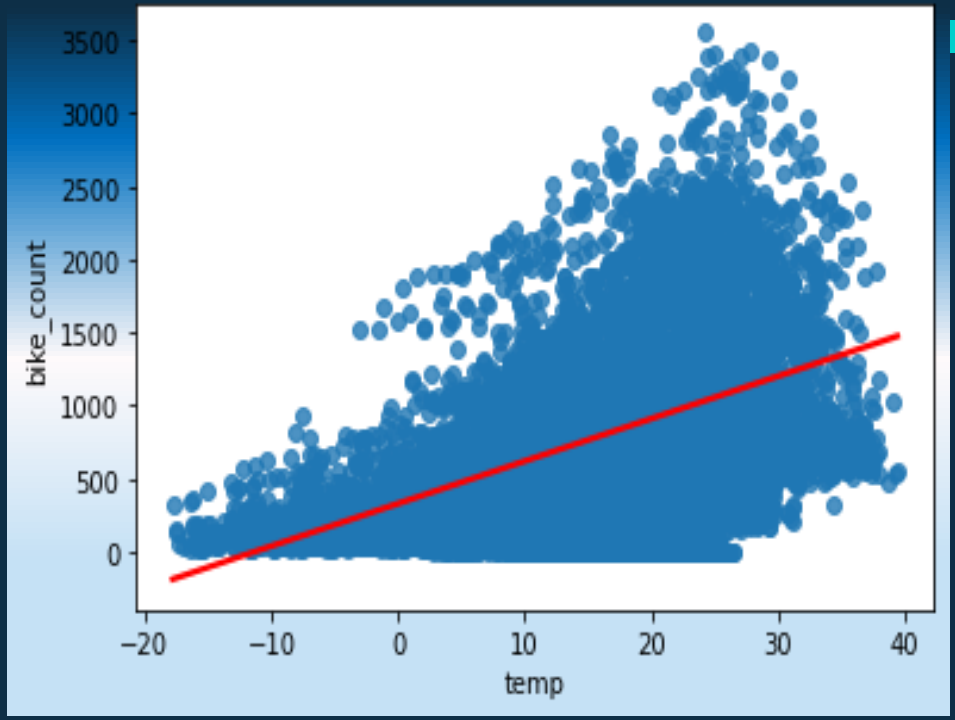
Rented bike count are negativity {-0.12} correlated with rain

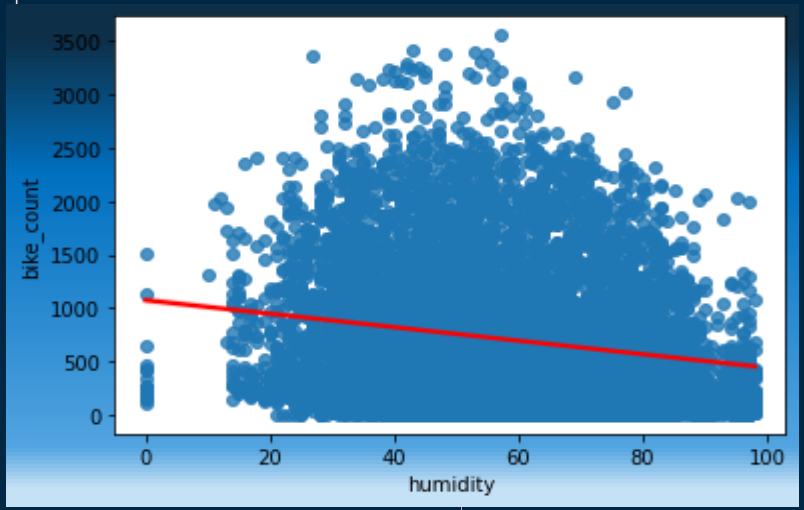
Rented Bike count and visibility

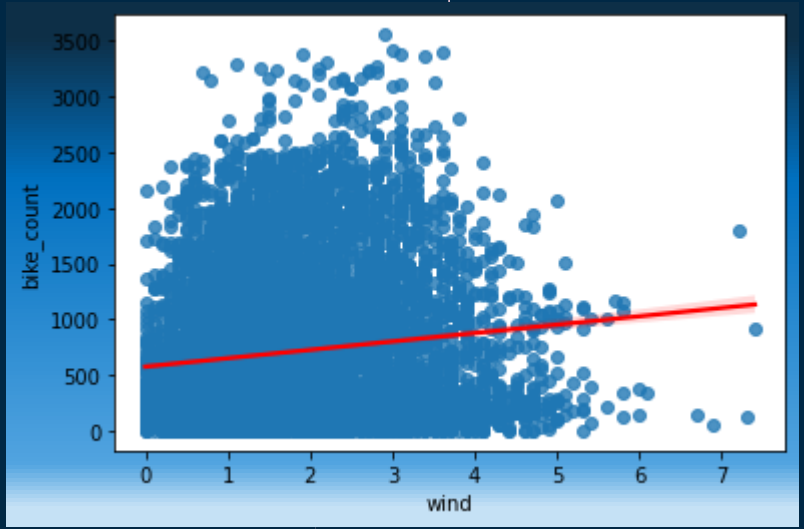
Rented bike count are positively{0.26} correlated with temperature

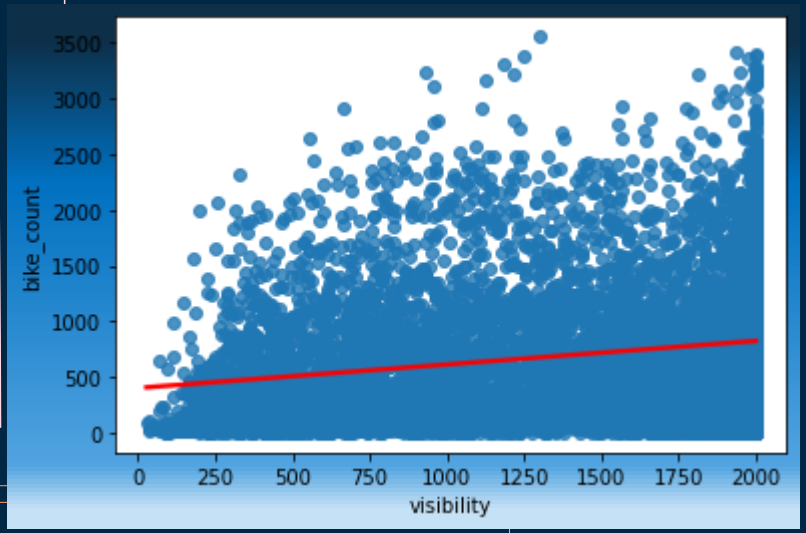
Plotting Regression plot of each columns of dataset v/s bike count columns

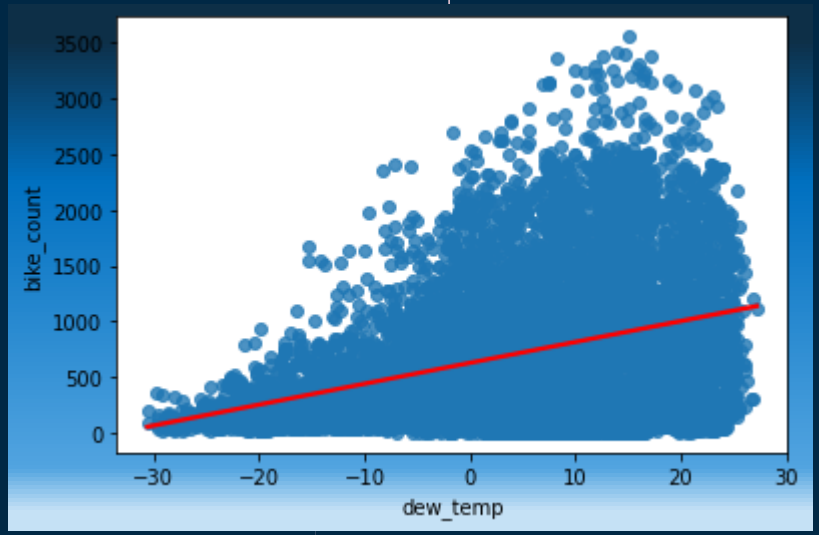


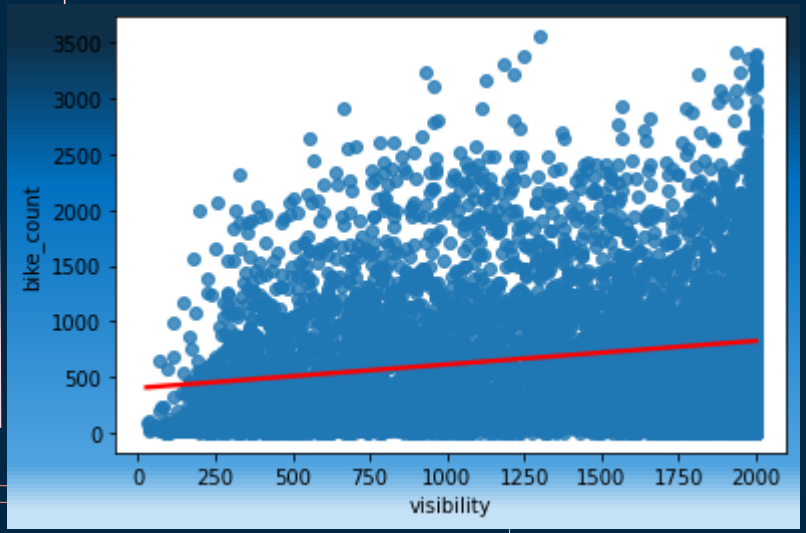


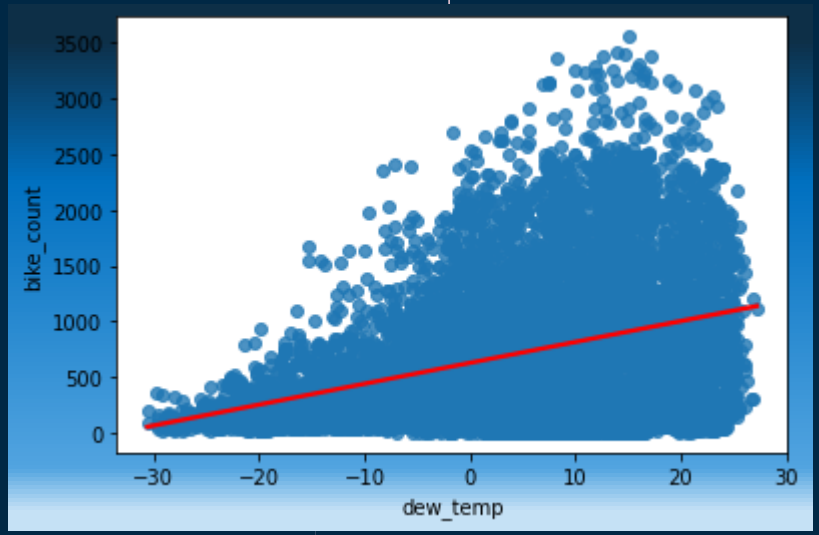

bike_count
humidity
0
500
1000
1500
2000
2500
3000
3500
0
20
40
60
80
100

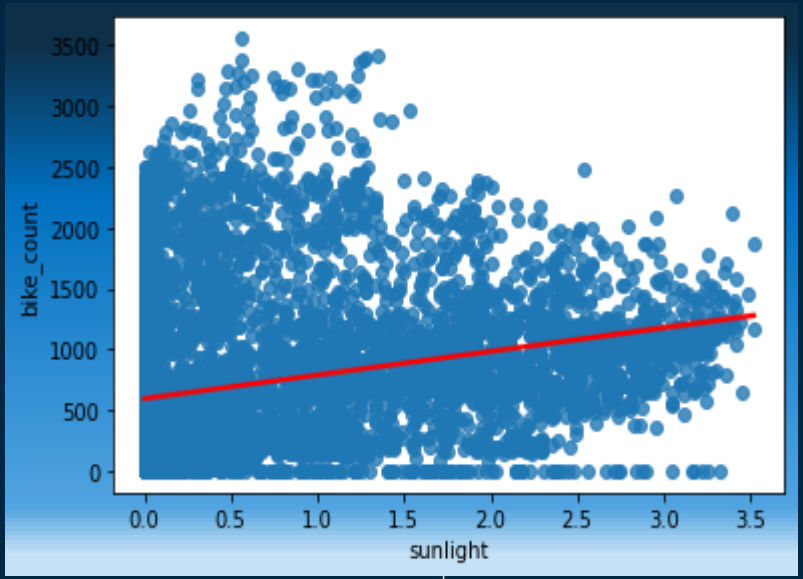

bike_count
wind
0
500
1000
1500
2000
2500
3000
3500
0
1
2
3
4
5
6
7

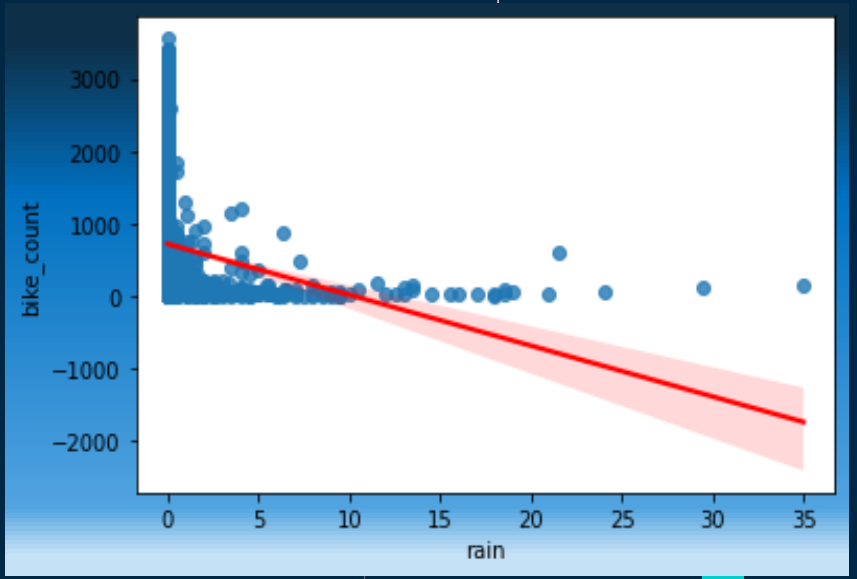

bike_count
visibility
0
250
500
750
1000
1250
1500
1750
2000
0
500
1000
1500
2000
2500
3000
3500

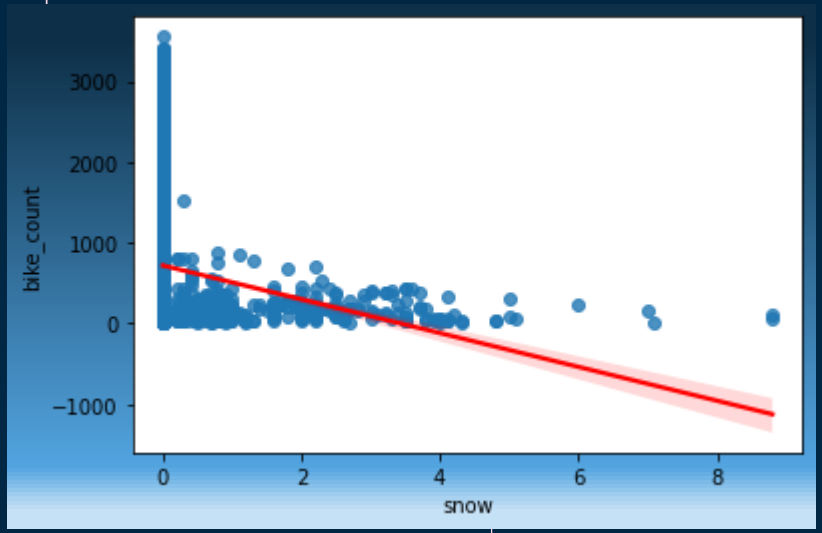

bike_count
dew_temp
−30
−20
−10
0
10
20
30
0
500
1000
1500
2000
2500
3000
3500

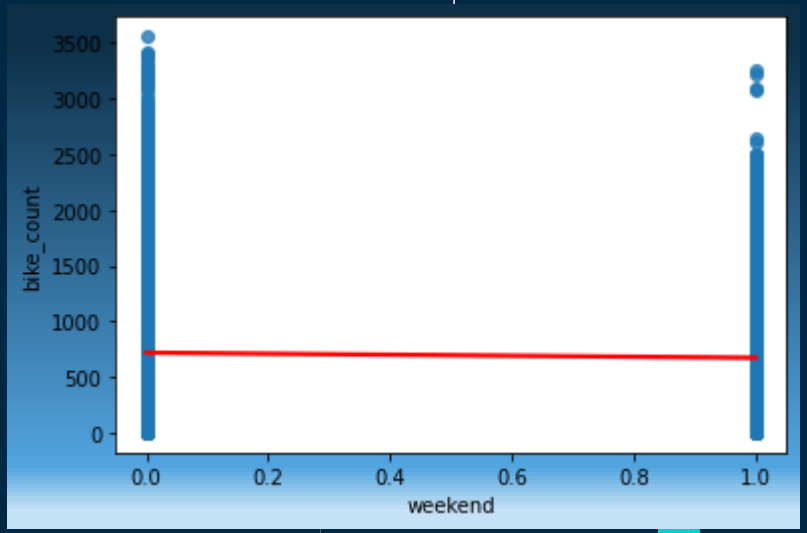

bike_count
visibility
0
250
500
750
1000
1250
1500
1750
2000
0
500
1000
1500
2000
2500
3000
3500


bike_count
dew_temp
−30
−20
−10
0
10
20
30
0
500
1000
1500
2000
2500
3000
3500


bike_count
sunlight
3500
3000
2500
2000
1500
1000
500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5


bike_count
rain
3000
2000
1000
0
−1000
−2000
0
5
10
15
20
25
30
35


bike_count
3000
2000
1000
0
−1000
0
2
4
6
8
snow

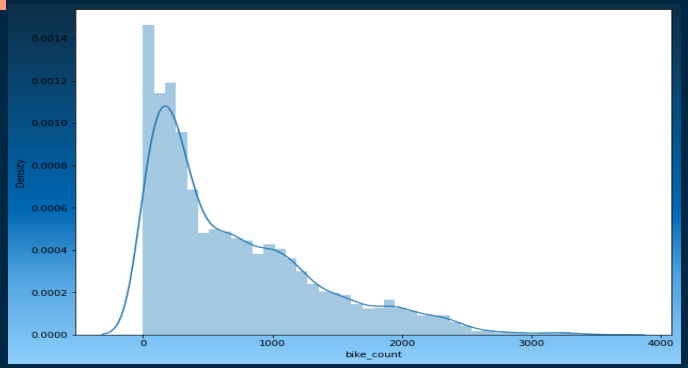

bike_count
3500
3000
2500
2000
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
weekend

CONCLUSION

01

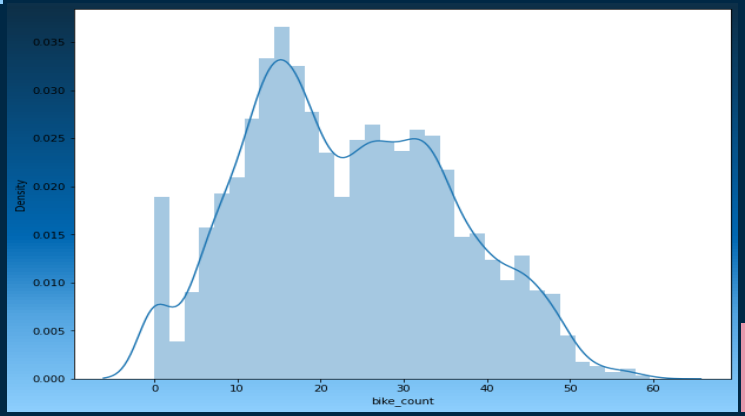
- From above 'Regression Plots' we observe **'Temperature'**, **'Wind speed'** , **'Visibility'** ,**'Solar Radiation**' this features are positively related with our dep endent variable.
- **'Rainfall'** , **'Snowfall'** , **'Humidity'** these features are nega tively related with the dependent variable or Target variabl e.
- We can see the Left or Right skewness in data

EDA

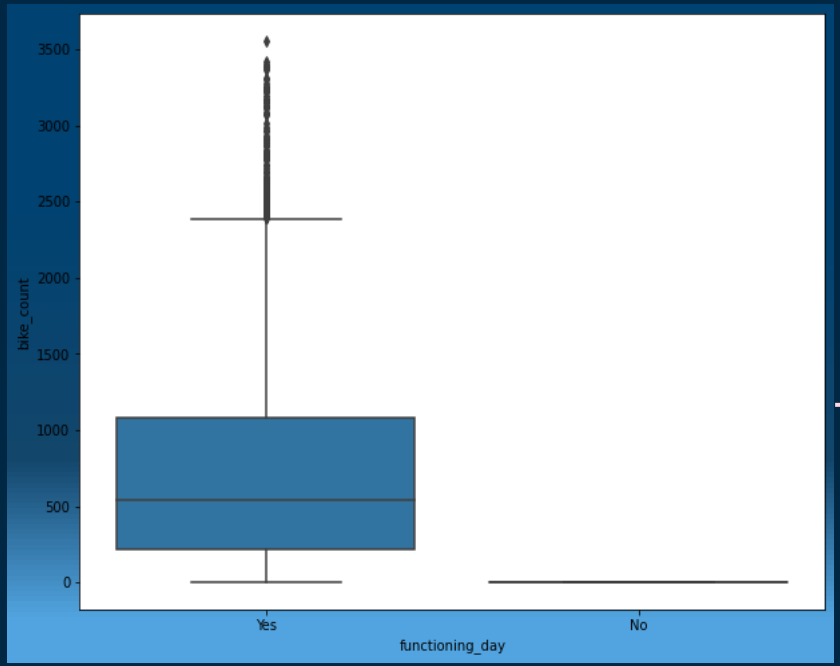


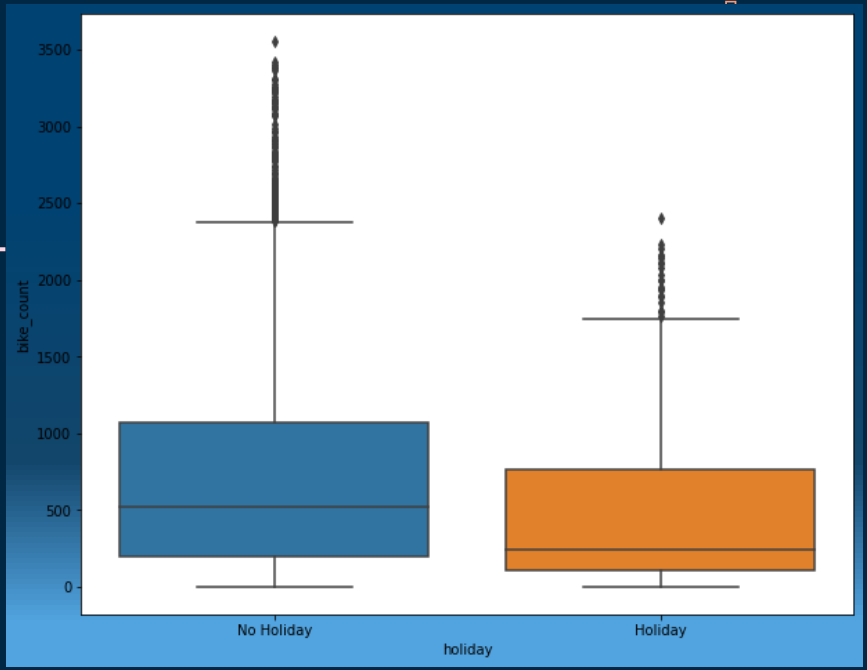
Distribution of rented bike count

Square root transformation of rented bike count

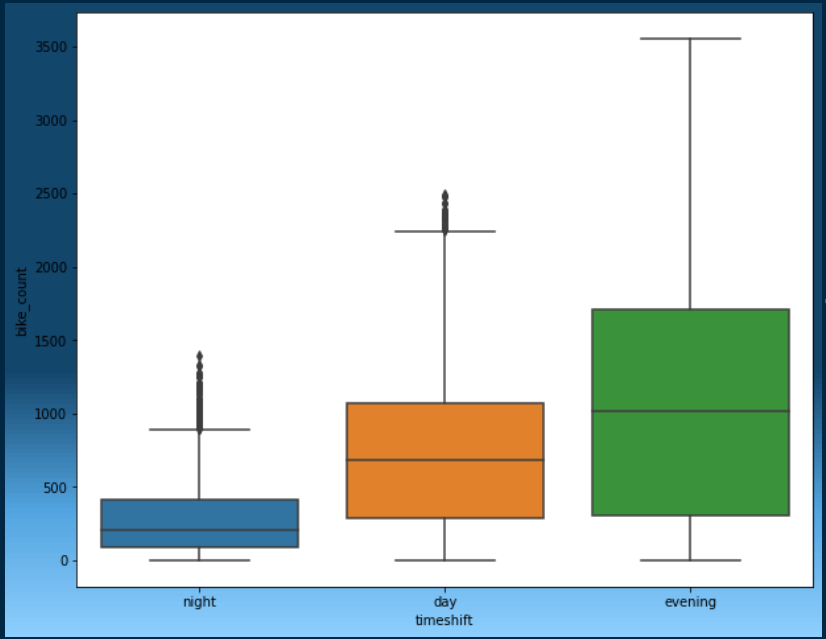


OUR PROCESS


3500
3000
2500
2000
1500
1000
500
0
bike_count
Yes
No
functioning_day

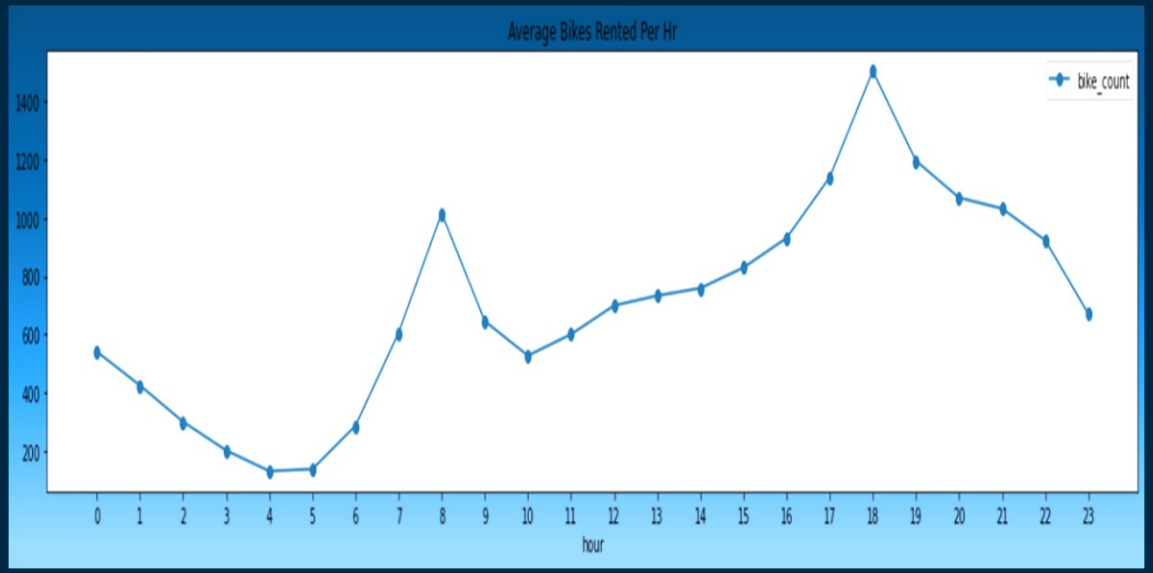

3500
3000
2500
2000
1500
1000
500
0
bike_count
No Holiday
Holiday
holiday

OUR PROCESS



- Less demand on winter seasons
- Slightly Higher demand during Non holidays
- Almost no demand on non functioning day
- Weekday or weekend does not affect the rented bike count , we will try to see on the basis of hours how it affects.

AVERAGE BIKES RENTED PER HOUT :



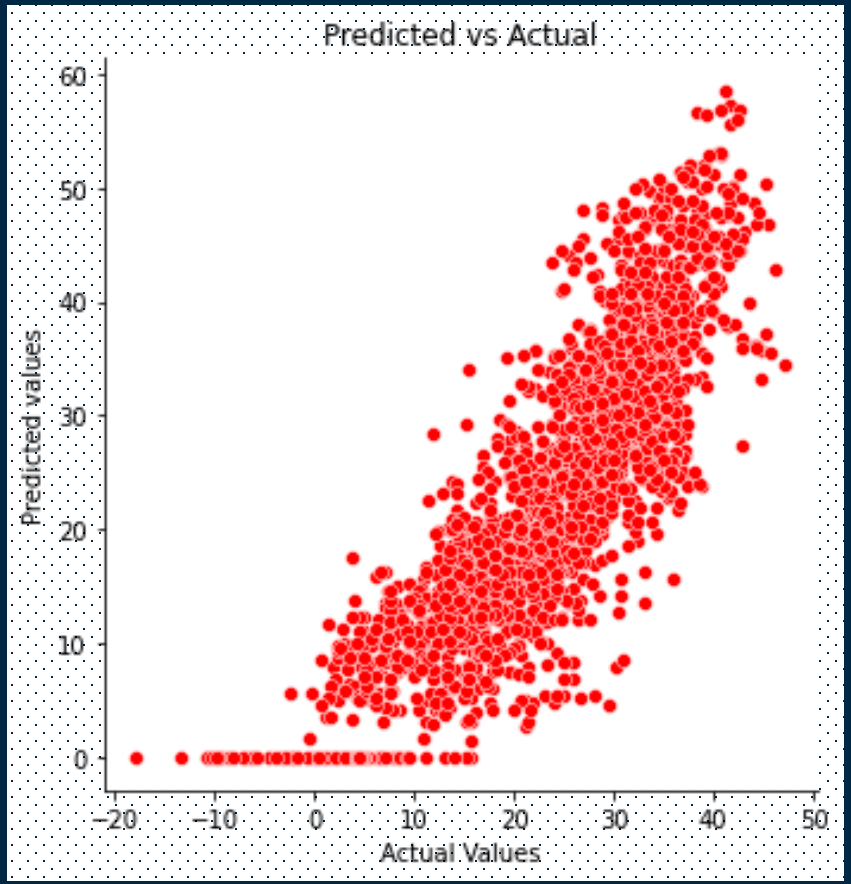
- High rise of Rented Bikes from 8:00 a.m to 9:00 p.m means people prefer rented bike during rush hour.
- we can clearly see that demand rises most at 8 a.m and 6:00 p.m so we can say that that during office opening and closing time there is much high demand

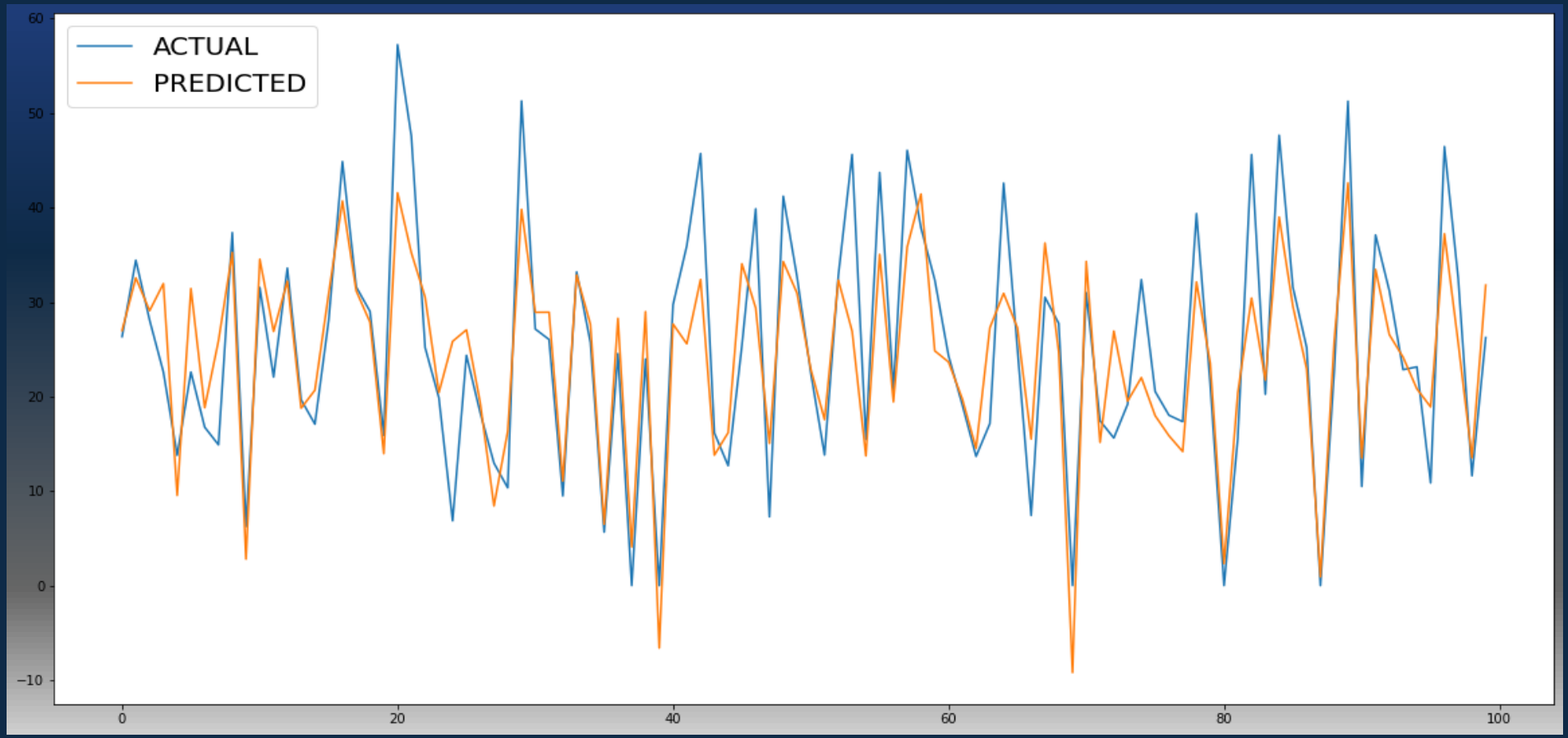
Model's Performed

Linear Regression

Actual value

Predicted value




ACTUAL
PREDICTED
60
50
40
30
20
10
0
−10
0
20
40
60
80
100

RESULT

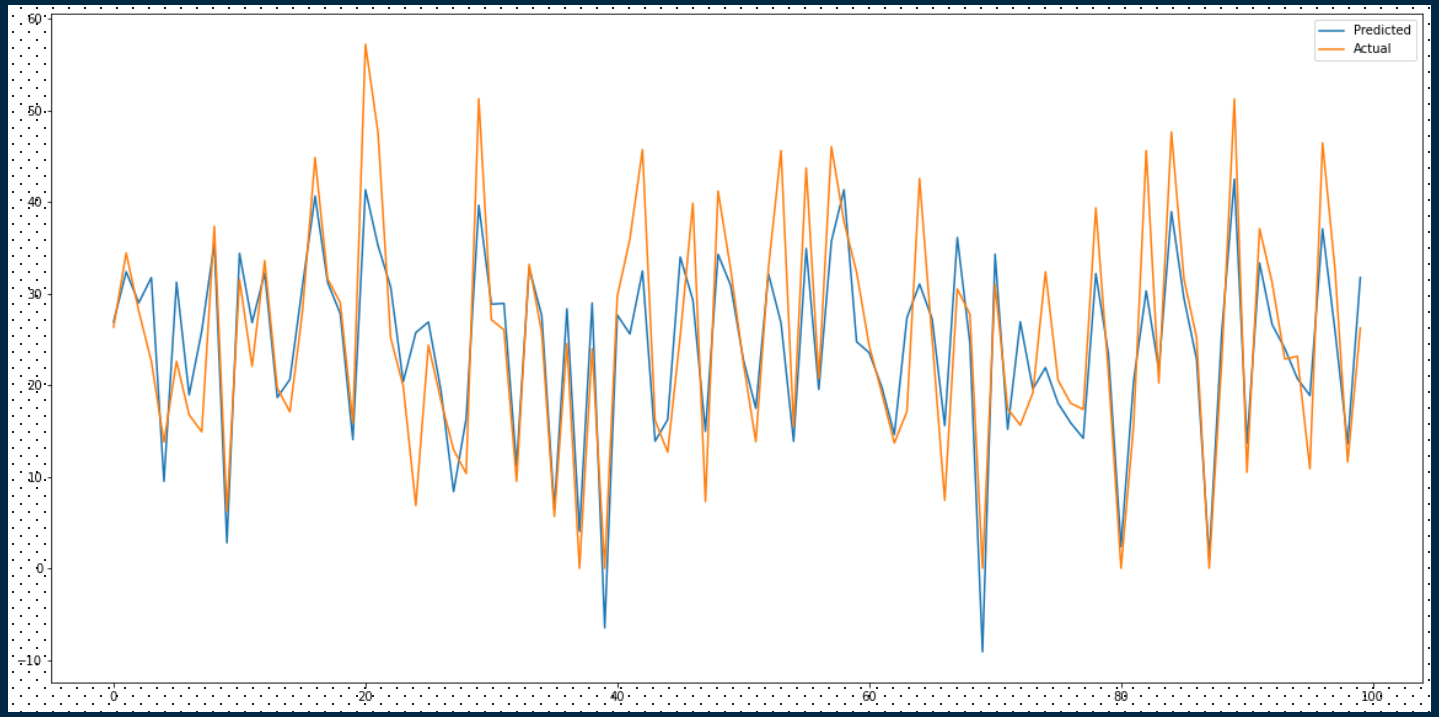
Model		MSE	RMSE	R2_score	Adjusted R2
Linear regression		143456.60	378.75	0.651	0.648

Model's Performed

Lasso Regression

Actual value

Predicted value



RESULT

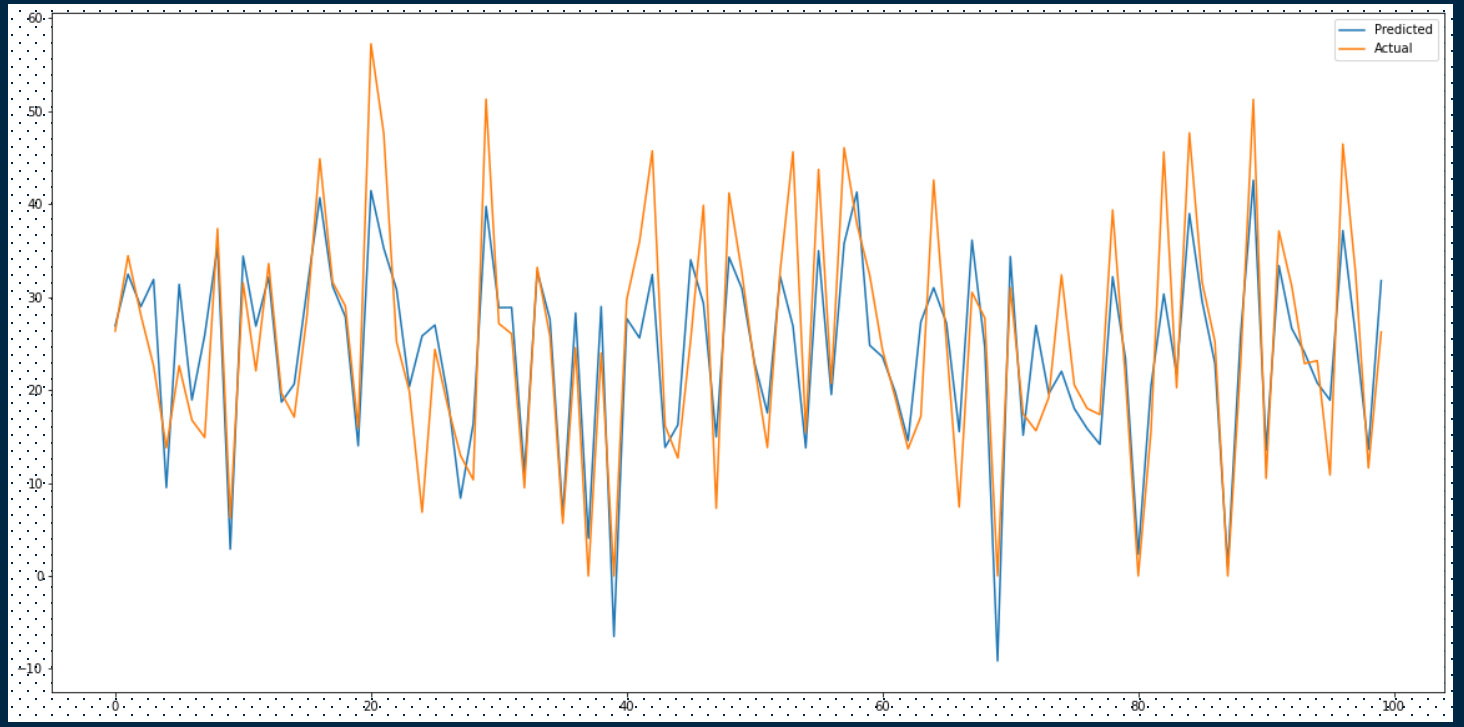
Model		MSE	RMSE	R2_score	Adjusted R2
Linear regression		143543.00	378.87	0.650	0.648

Model's Performed

Ridge Regression

Actual value

Predicted value



RESULT

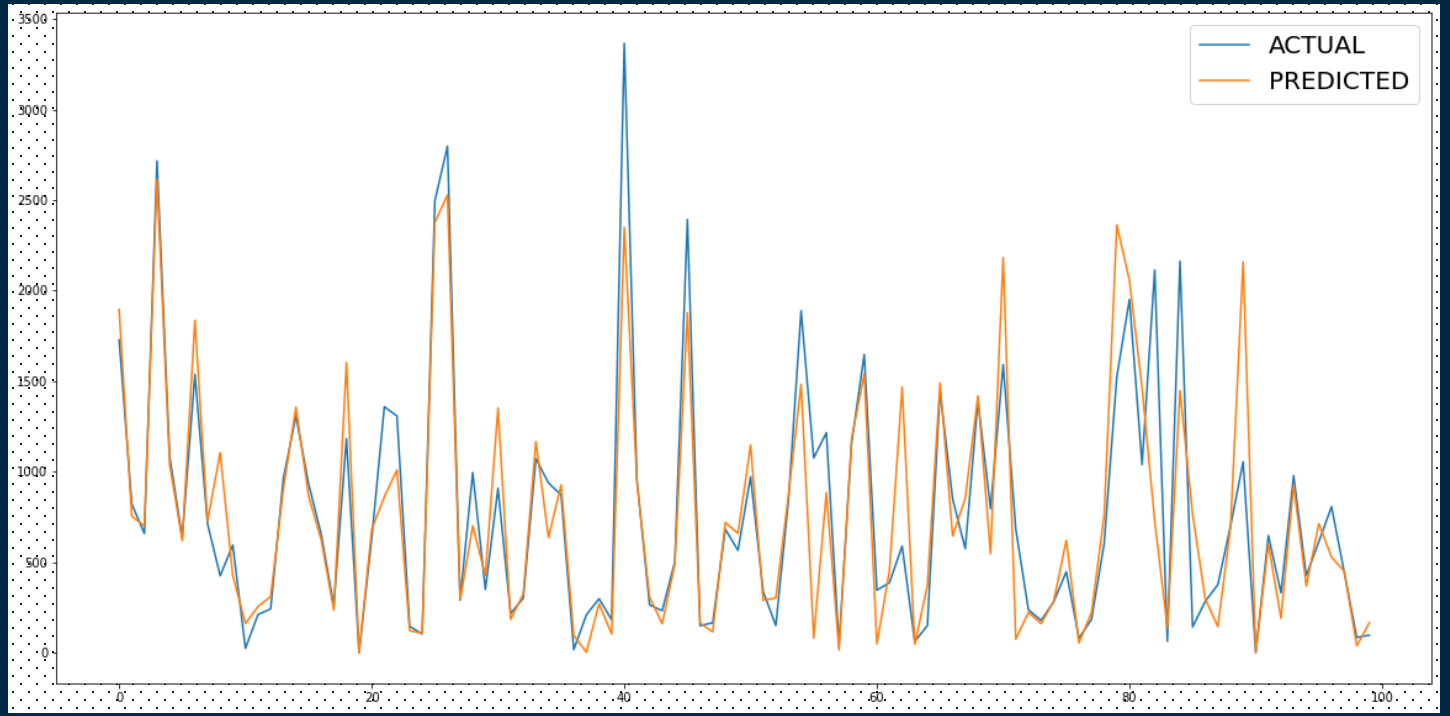
Model		MSE	RMSE	R2_score	Adjusted R2
Linear regression		143543.03	378.65	0.651	0.648

Model's Performed

Decision Tree

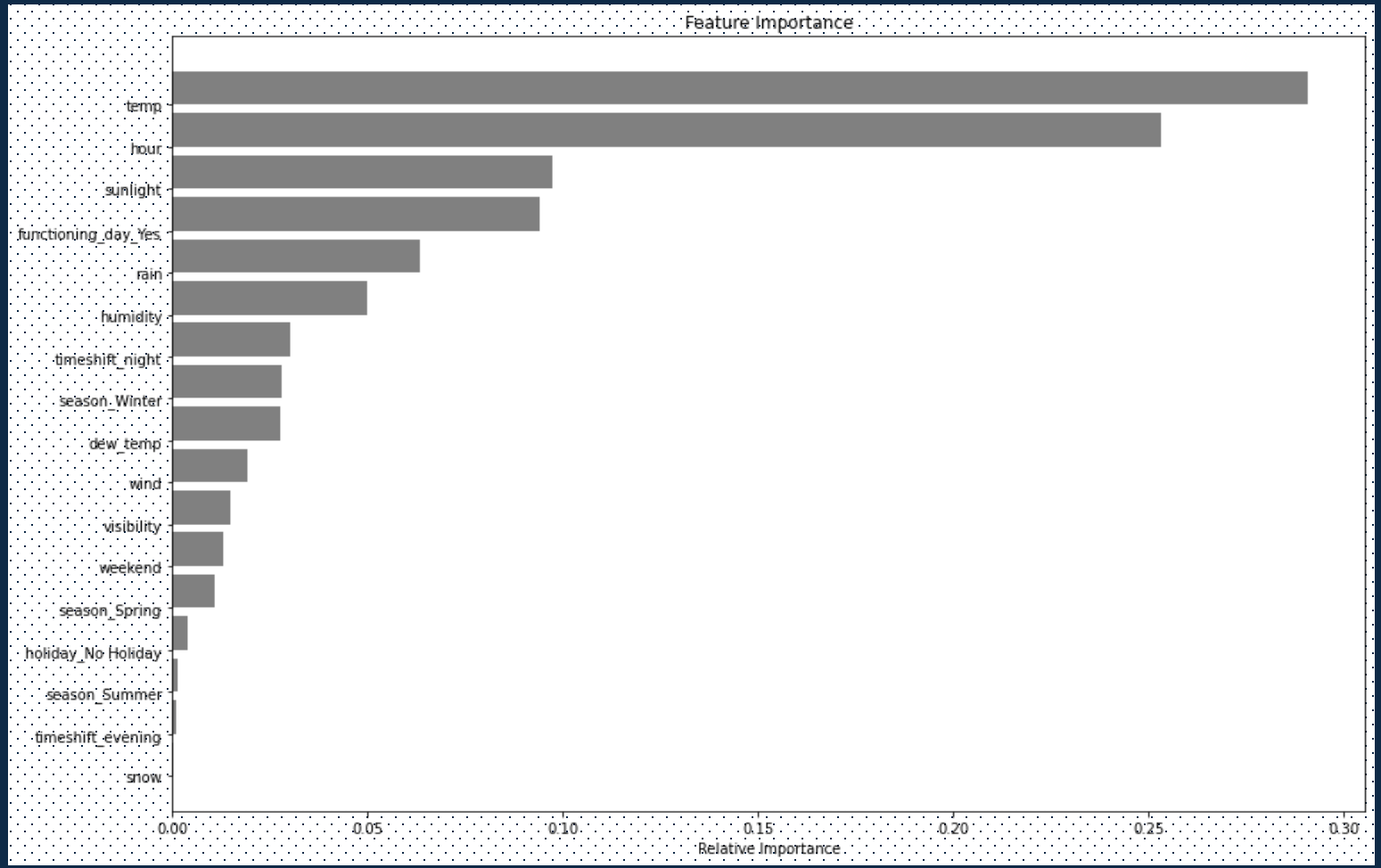
Actual value

Predicted value



RESULT

Model		MSE	RMSE	R2_score	Adjusted R2
Linear regression		98626.36	314.04	0.758	0.757

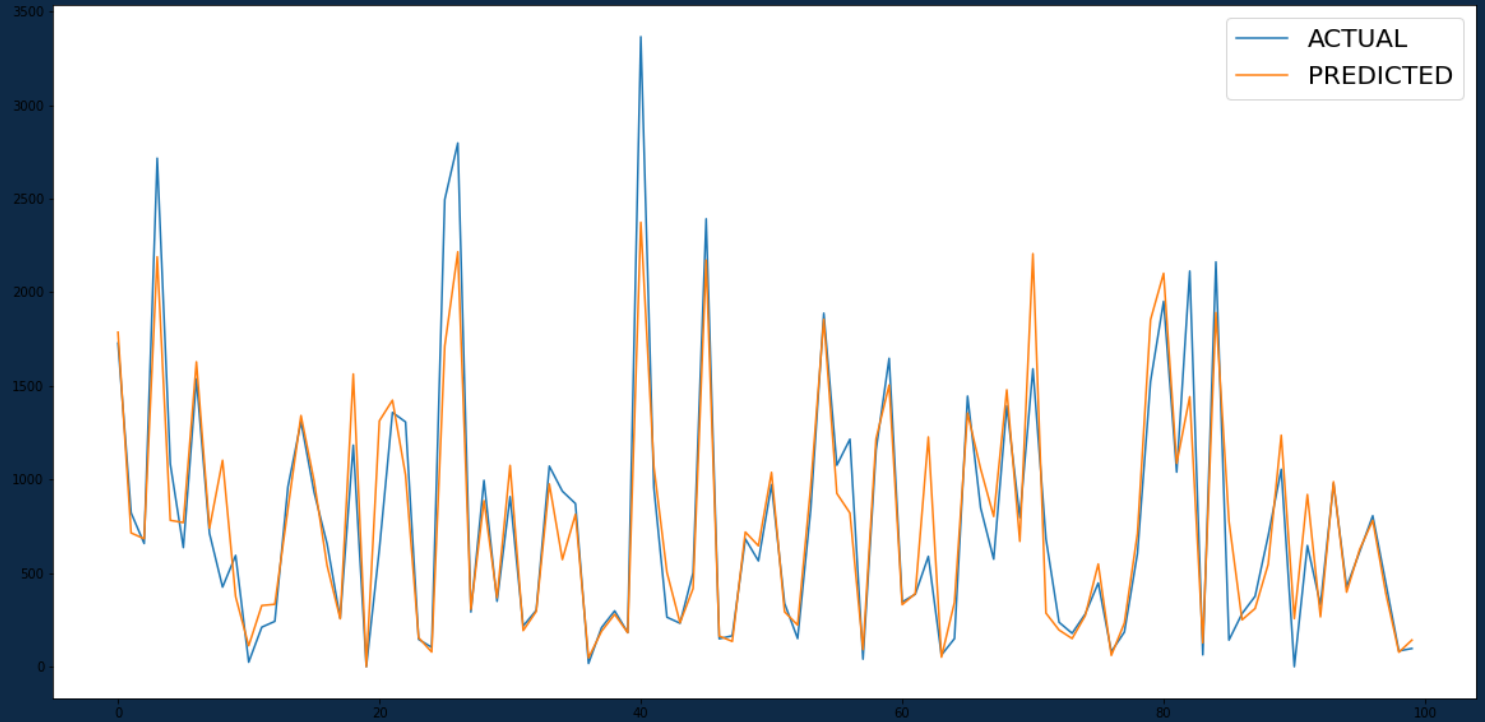

Feature Importance
temp
hour
sunlight
functioning_day_Yes
rain
humidity
timeshift_night
season_Winter
dew_temp
wind
visibility
weekend
season_Spring
holiday_No Holiday
season_Summer
timeshift_evening
snow
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Relative Importance

Model's Performed

Random Forest

Actual value

Predicted value



RESULT

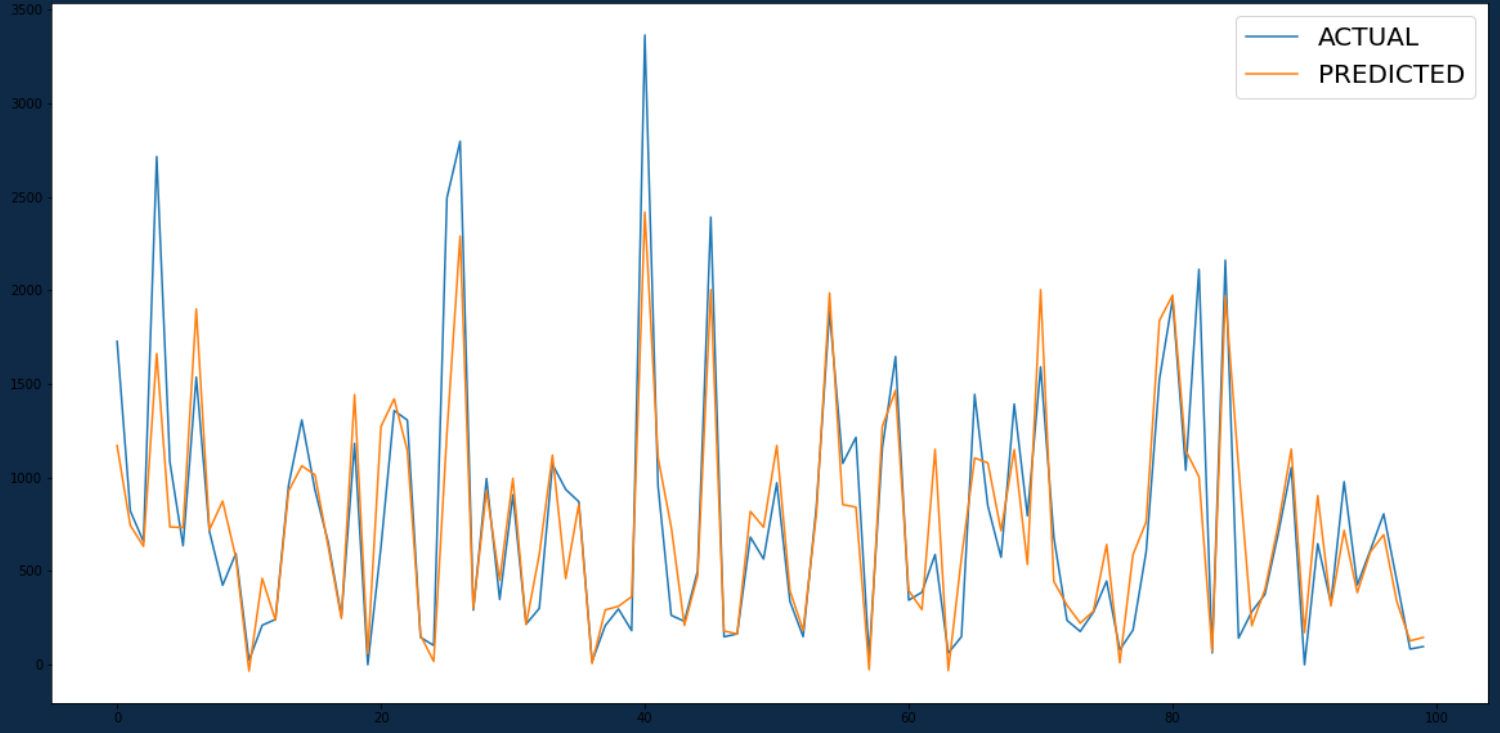
Model		MSE	RMSE	R2_score	Adjusted R2
Random Forest		47887.70	218.83	0.8829	0.8820

Model's Performed

Gradient Boosting

Actual value

Predicted value



RESULT

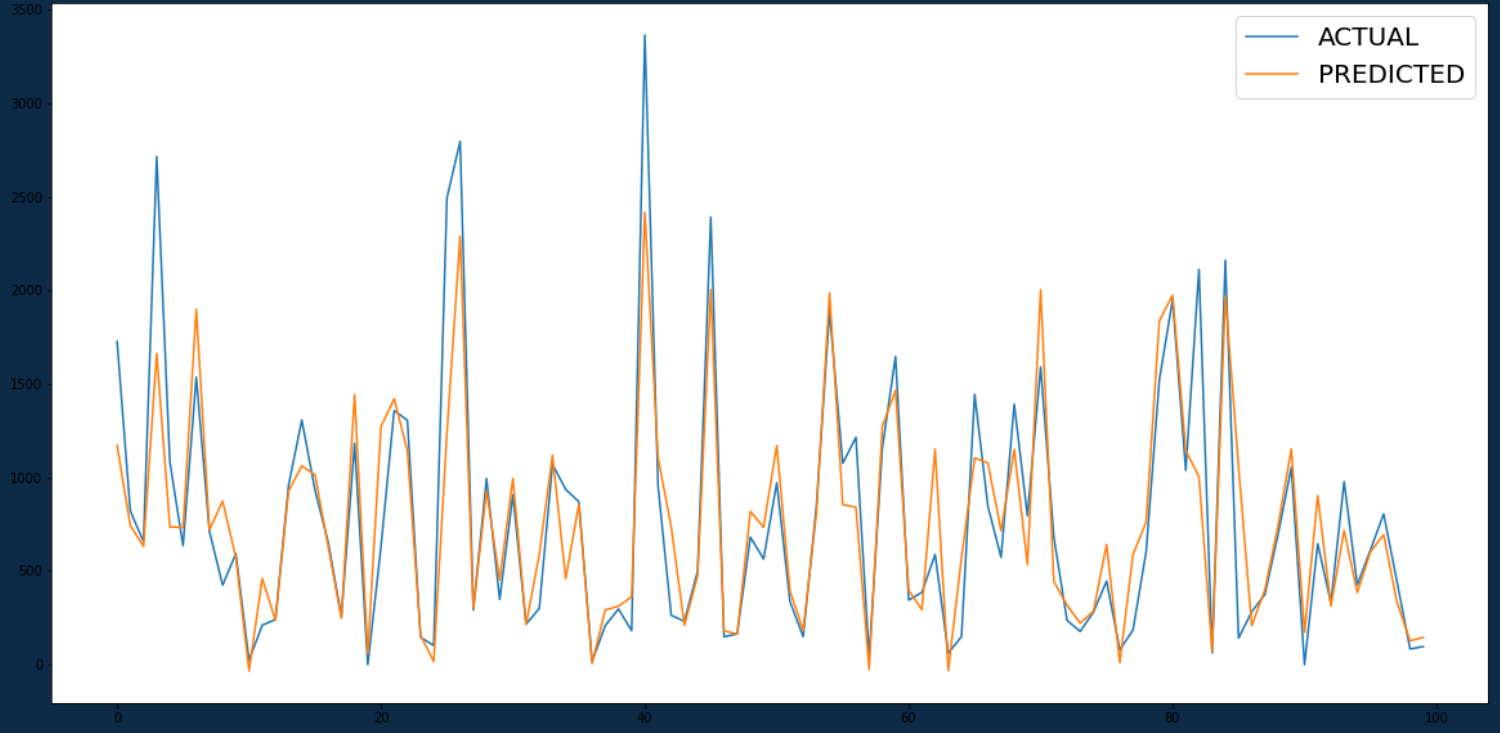
Model		MSE	RMSE	R2_score	Adjusted R2
Gradient Boosting		69123.22	262.91	0.8310	0.8297

Model's Performed

Xtreme Gradient Boosting

Actual value

Predicted value



RESULT

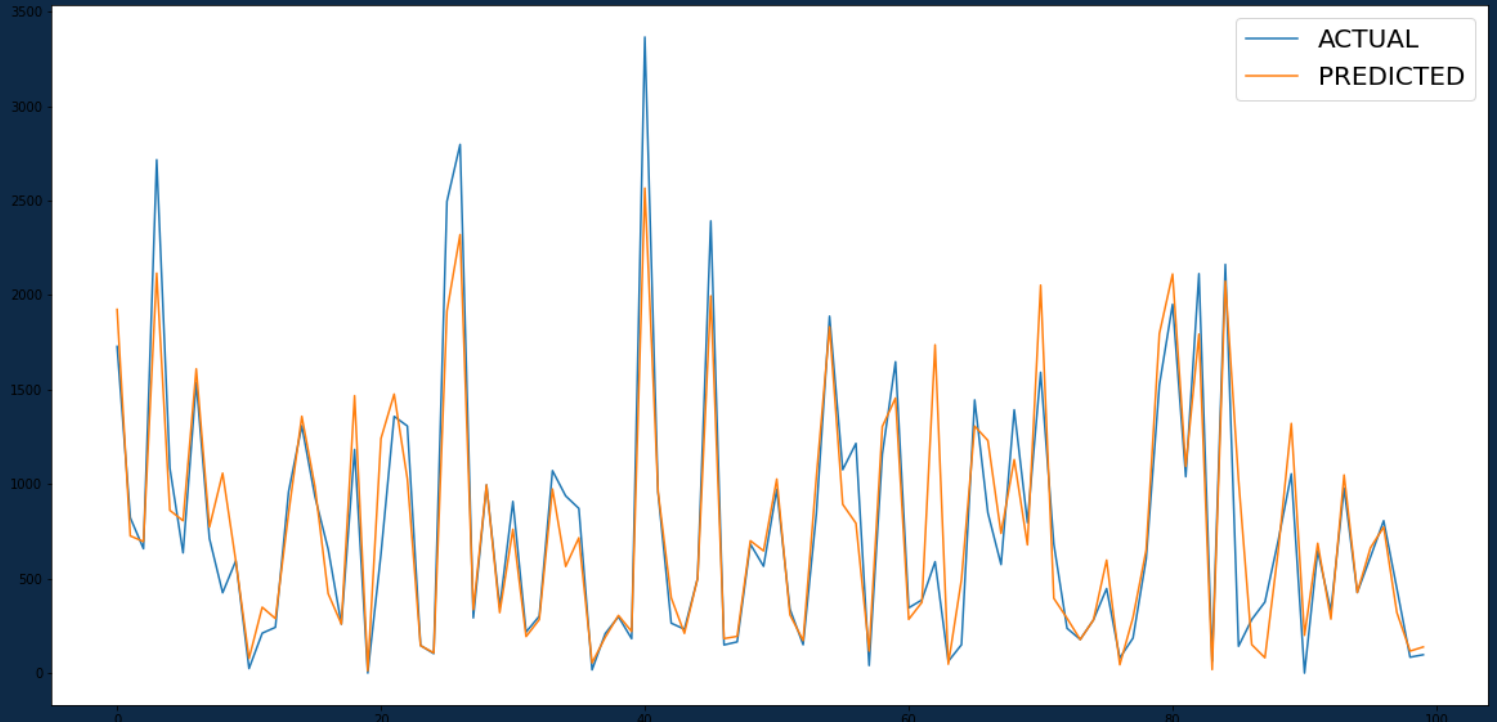
Model		MSE	RMSE	R2_score	Adjusted R2
Xtreme Gradient Boosting		46458.60	215.54	0.8864	0.8855

Model's Performed

light GBM

Actual value

Predicted value



RESULT

Model		MSE	RMSE	R2_score	Adjusted R2
Light GBM		40840.23	202.08	0.9001	0.8994

CONCLUSION

- No overfitting is seen, as we can see the models are performing well with the test data with good results.
- After performing the various models, the lightGBM and Xtreme Gradient Boosting found to be th e best model that can be used for the Bike Sharing Demand Prediction since the performance m etrics (mse,rmse) shows lower and (r2,adjusted_r2) shows a higher value for the lightGBM and X treme Gradient Boosting models !
- In holiday or non-working days there is demands in rented bikes.
- People preferred more rented bikes in the morning than the evening.
- When the rainfall was less, people have booked more bikes except some few cases.
- The Temperature, Hour & Humidity are the most important features that positively drive the total rented bikes count.
- We can use either lightGBM or Xtreme Gradient Boosting model for the bike rental stations.

THANKS