Capstone Project-2

Bike Sharing Demand Prediction

TEAM MEMBERS

Bikesh kumar maharana

Pratyush kumar rath

GAurav singh

CONTENT

- ❑ BUSINESS UNDERSTANDING
- ❑ DATA SUMMARY
- ❑ FEATURE ANALYSIS
- ❑ EXPLORATORY DATA ANALYSIS
- ❑ DATA PREPROCESSING
- ❑ IMPLEMENTING ALGORITHMS
- ❑ CHALLENGES
- ❑ CONCLUSIONS

Business Understanding

➢ Bike rentals have became a popular service in recent years and it seems people are using it more often. With relatively cheaper rates and ease of pick up and drop at own convenience is what making this business thrive.

➢ Mostly used by people having no personal vehicles and also to avoid congested public transport which that's why they prefer rental bikes.

➢ Therefore, the business to strive and profit more, it has to be always ready and supply no. of bikes at different locations, to fulfil the demand.

➢ Our project goal is a pre planned set of bike count values that can be a handy solution to meet all demands.

Data summary

	0	1	2	3	4	5	6	7	8	9
Date	01/12/2017	01/12/2017	01/12/2017	01/12/2017	01/12/2017	01/12/2017	01/12/2017	01/12/2017	01/12/2017	01/12/2017
Rented Bike Count	254	204	173	107	78	100	181	460	930	490
Hour	0	1	2	3	4	5	6	7	8	9
Temperature(°C)	-5.2	-5.5	-6.0	-6.2	-6.0	-6.4	-6.6	-7.4	-7.6	-6.5
Humidity(%)	37	38	39	40	36	37	35	38	37	27
Wind speed (m/s)	2.2	0.8	1.0	0.9	2.3	1.5	1.3	0.9	1.1	0.5
Visibility (10m)	2000	2000	2000	2000	2000	2000	2000	2000	2000	1928
Dew point temperature(°C)	-17.6	-17.6	-17.7	-17.6	-18.6	-18.7	-19.5	-19.3	-19.8	-22.4
Solar Radiation (MJ/m2)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.23
Rainfall(mm)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Snowfall (cm)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Seasons	Winter	Winter	Winter	Winter	Winter	Winter	Winter	Winter	Winter	Winter
Holiday	No Holiday	No Holiday	No Holiday	No Holiday	No Holiday	No Holiday	No Holiday	No Holiday	No Holiday	No Holiday
Functioning Day	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

- This Dataset contains 8760 lines and 14 columns.
- Three categorical features ‘Seasons’, ‘Holiday’, & ‘Functioning Day’.
- One Datetime features ‘Date’.
- We have some numerical type variables such as temperature, humidity, wind, visibility, dew point temp, solar radiation, rainfall, snowfall which tells the environment conditions at that particular hour of the day.

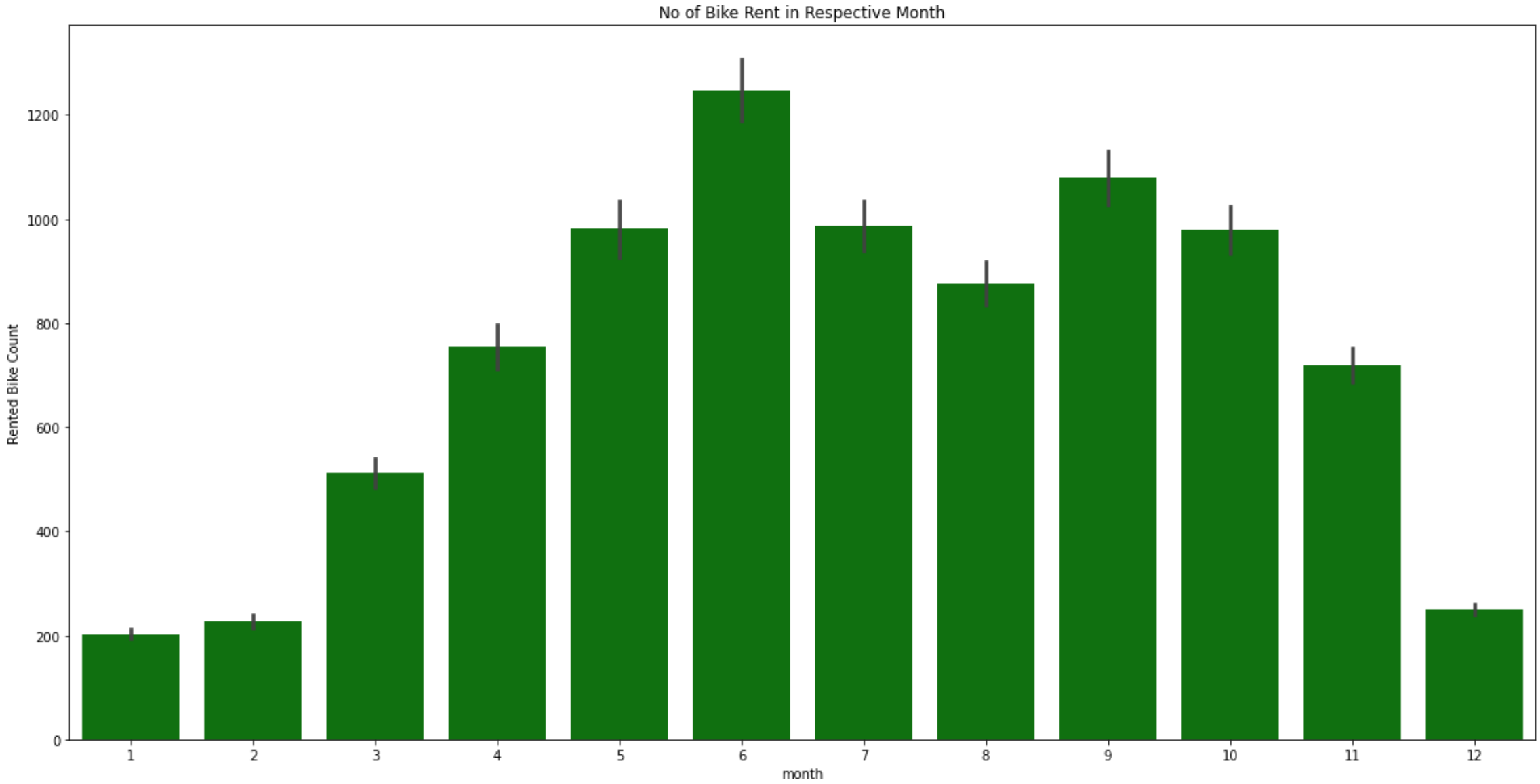
Feature summary

- Date : Year-Month-Day
- Rented Bike Count - Count of bikes rented at each hour
- Hour - Hour of the day
- Temperature - Temperature in Celsius
- Humidity - %
- Wind Speed - m/s
- Visibility - 10m
- Dew point temperature -Celsius
- Solar radiation -MJ/m2
- Rainfall -mm
- Snowfall -cm
- Seasons -Winter, Spring, Summer, Autumn
- Holiday -Holiday/No Holiday
- Functioning Day

Insights from our dataset

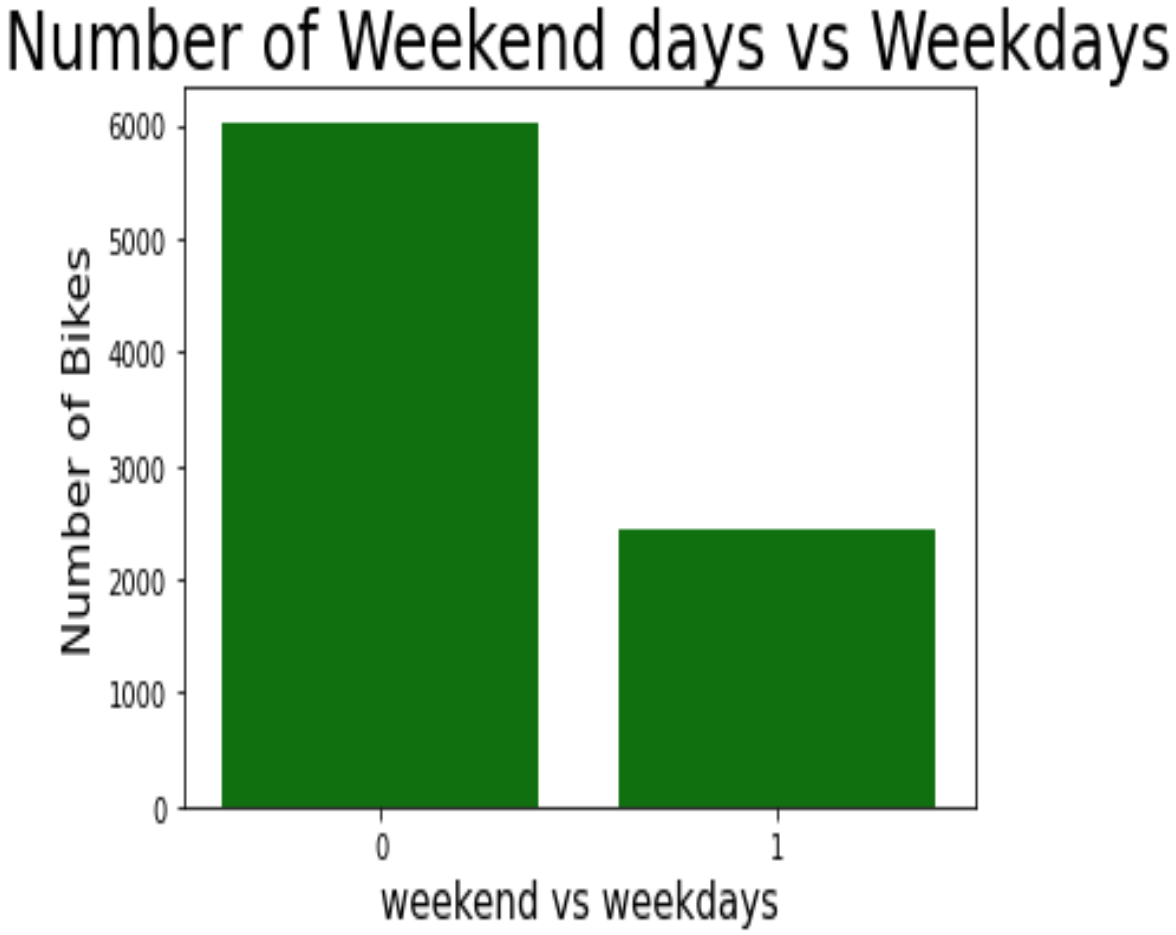
- There are No Missing Values present .
- There are No Duplicate values present.
- There are No null values.
- And finally we have 'rented bike count' variable which we need to predict for new observations .
- The dataset shows hourly rental data for one year (1 December 2017 to 31 November(2018)(365 days).we consider this as a single year data
- So we convert the "date" column into 3 different column i.e "year","month","day".
- We change the name of some features for our convenience , they are as below 'Rented_Bike_Count', 'Hour', 'Temperature', 'Humidity', 'Wind_speed', 'Visibility', ‘Dew_point_temperature', 'Solar_Radiation', 'Rainfall', 'Snowfall', 'Seasons', 'Holiday', 'Functioning_Day', 'month','weekdays_weekend'

Analysis of Month vs Rented bike



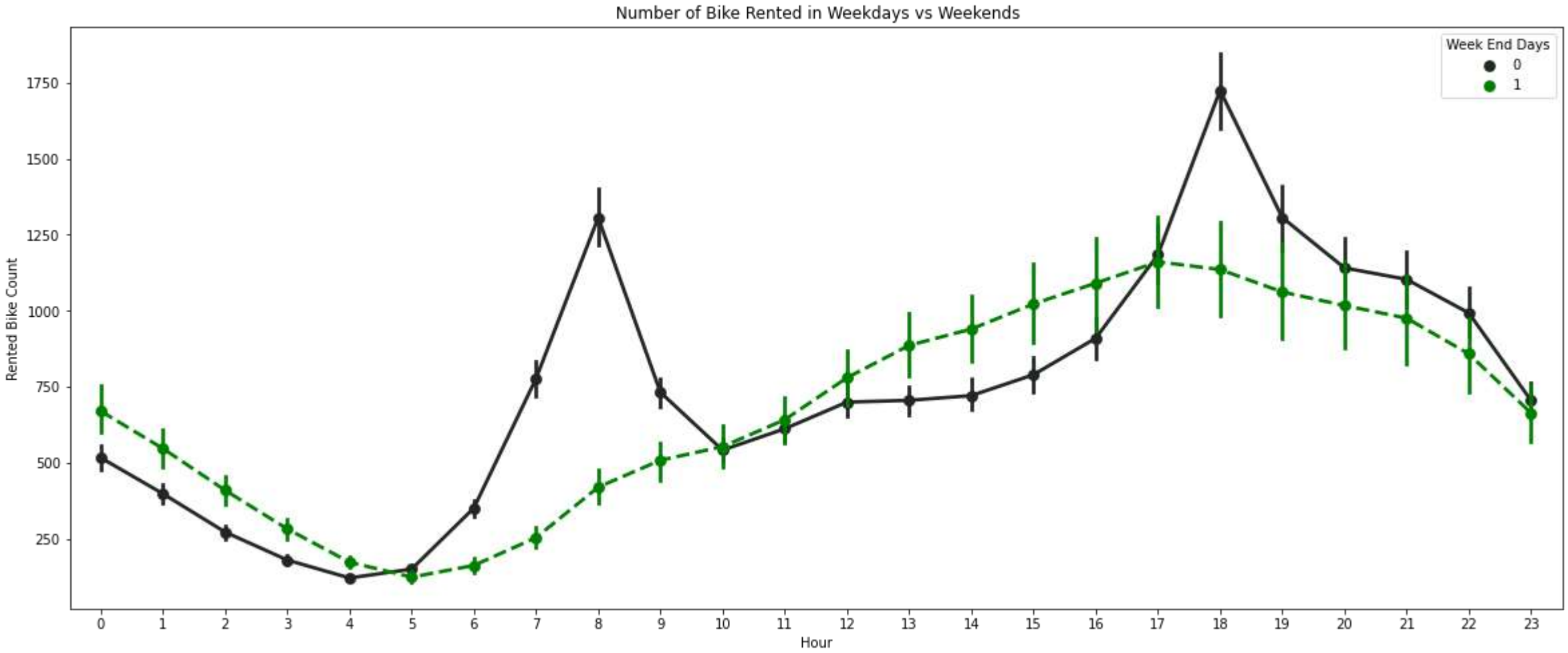
Analysis of rented bike

(Weekend vs weekdays)



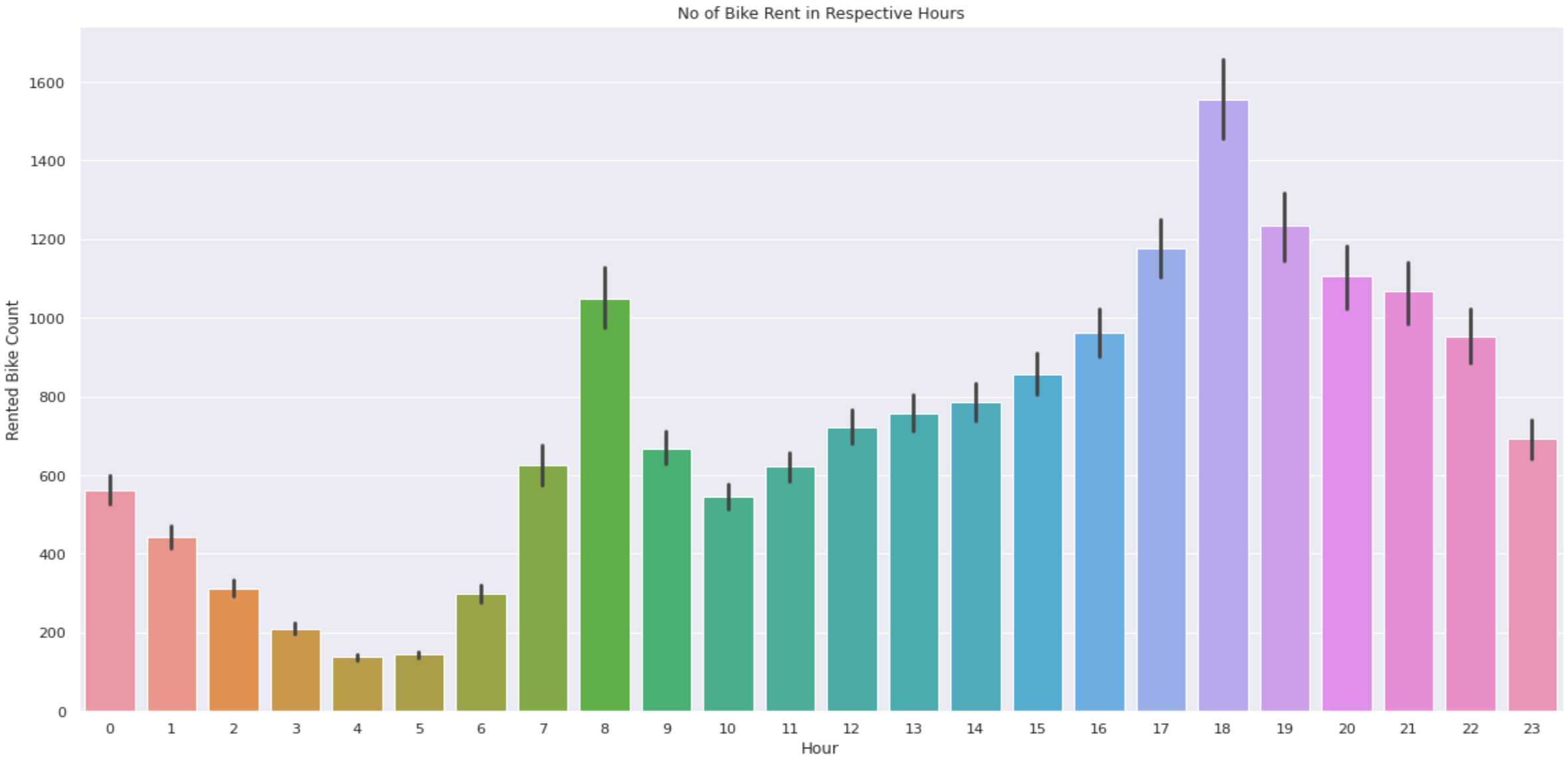
Analysis of rented bike

(HOUR)



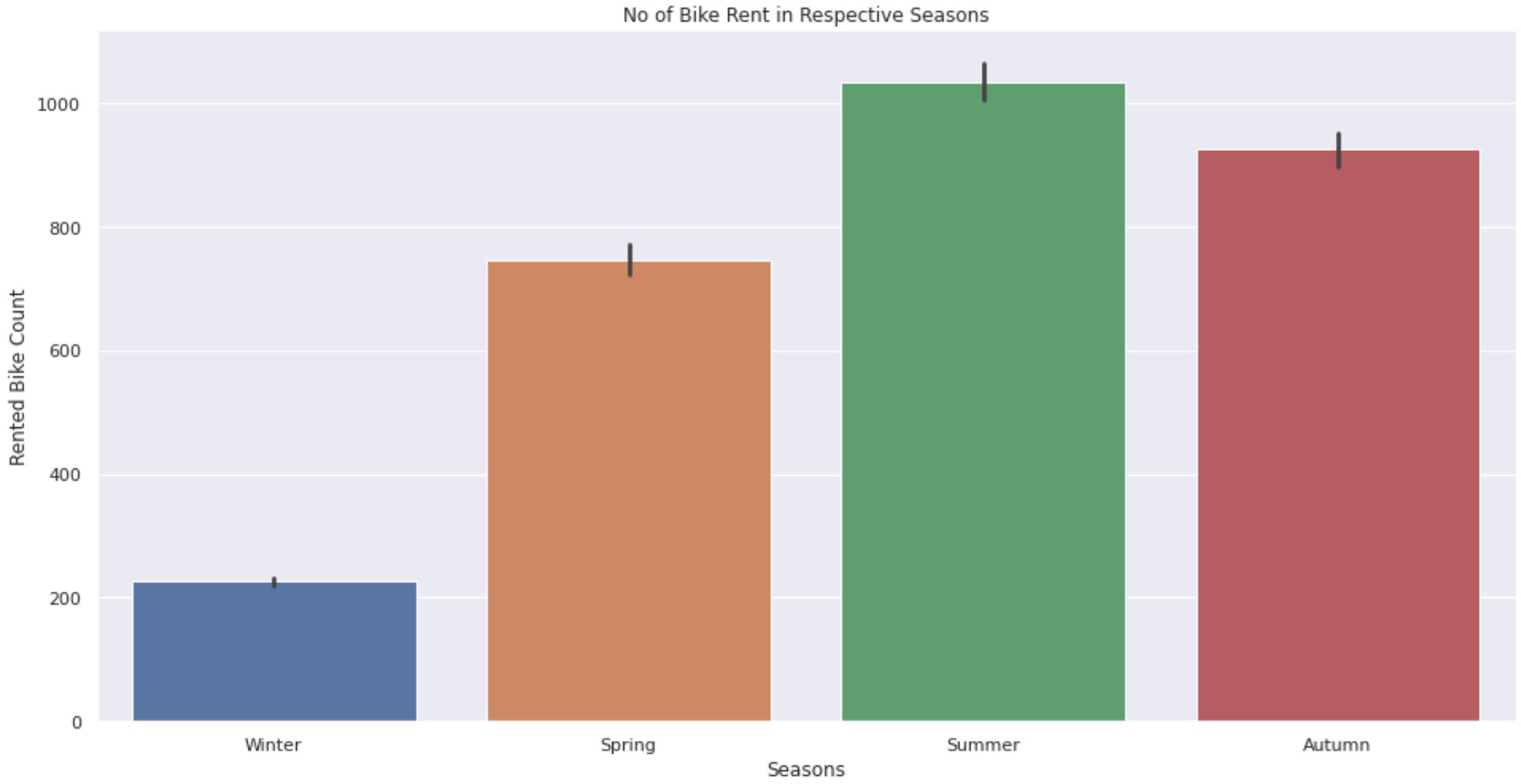
Analysis of rented bike

(HOUR)



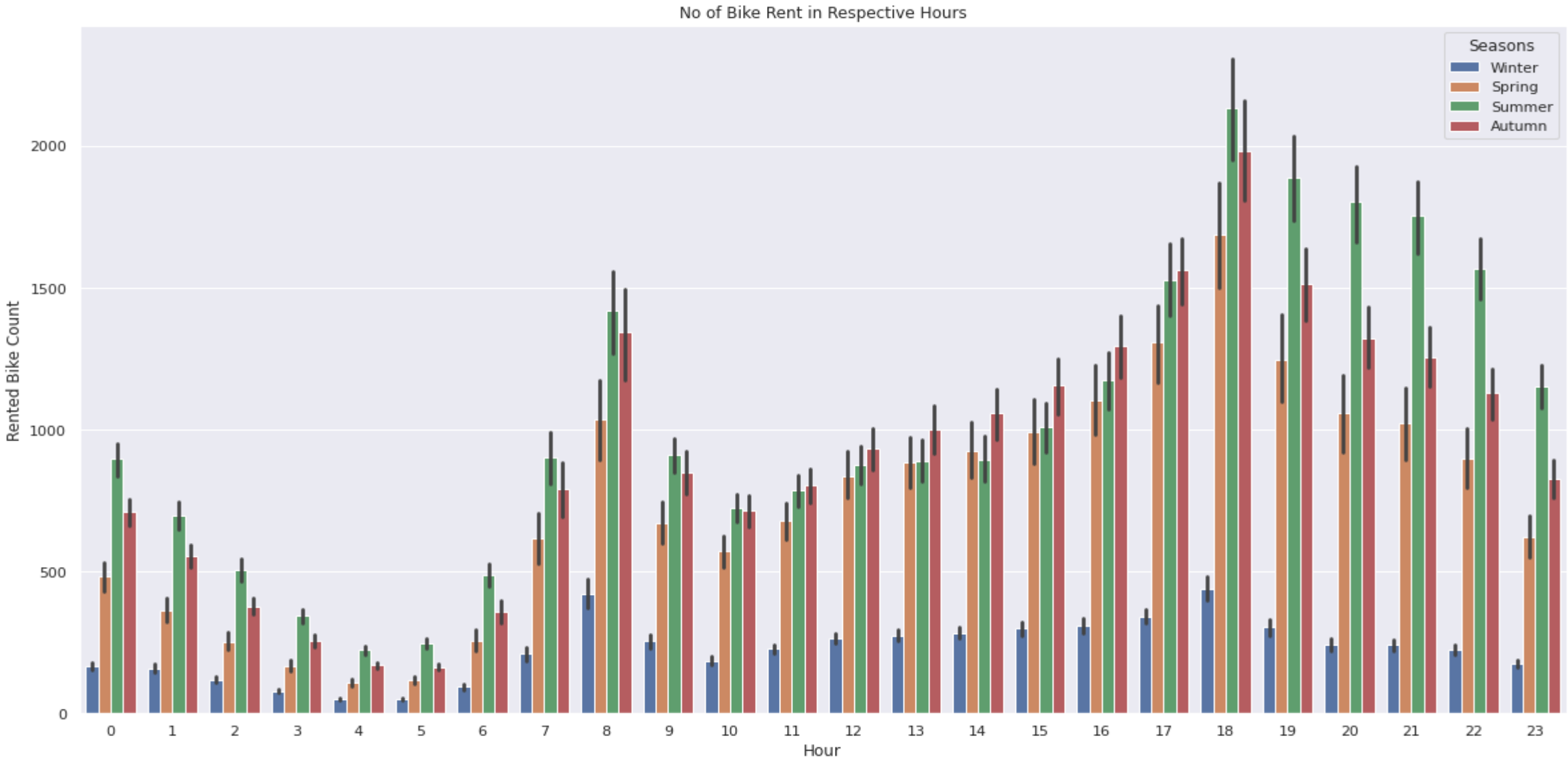
Analysis of rented bike

(SEASON)



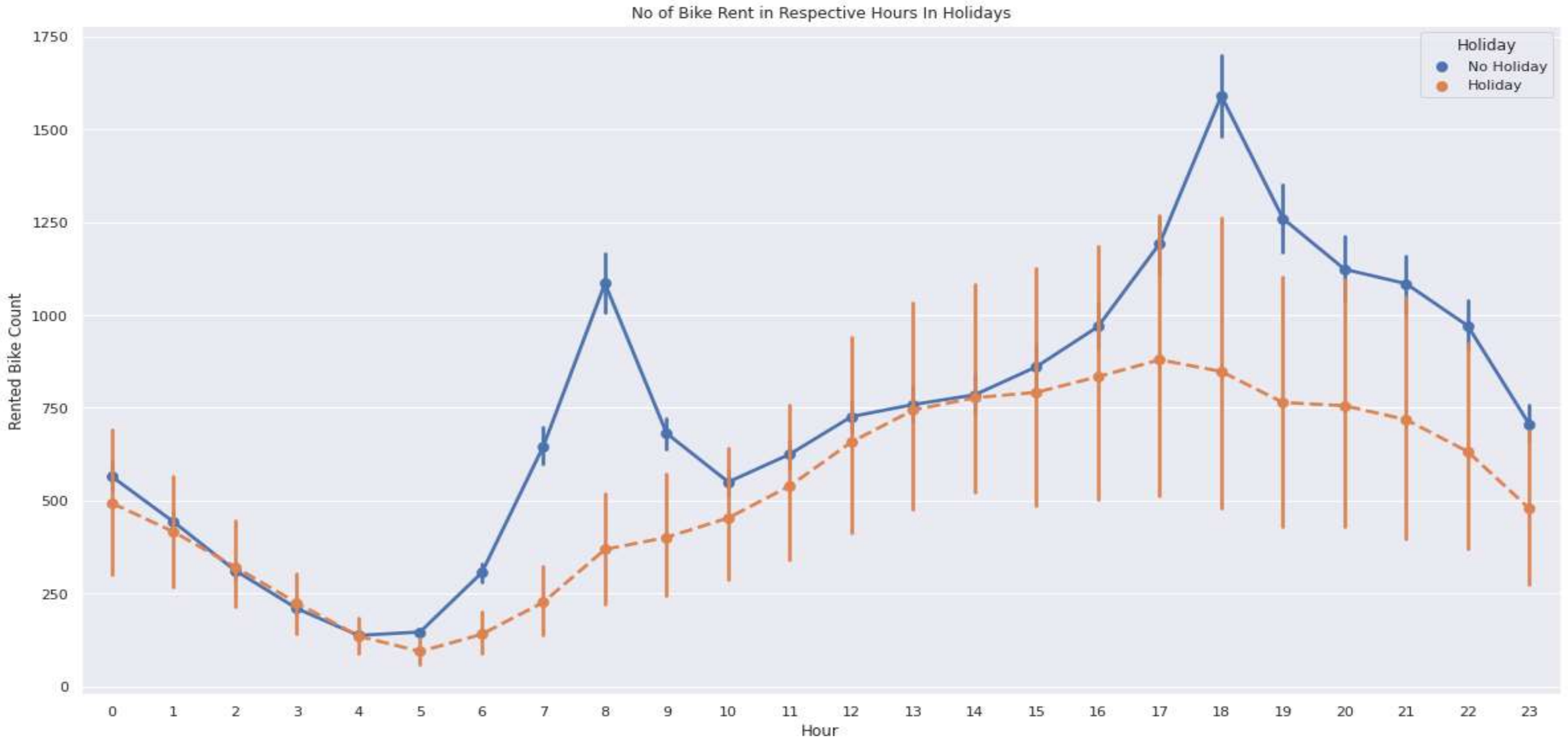
Analysis of rented bike

(Hour and season)

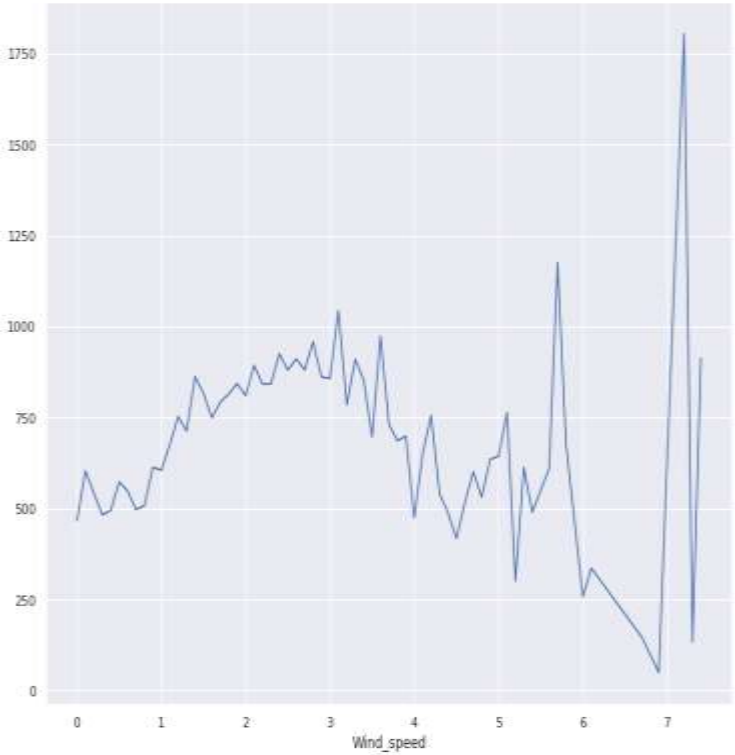


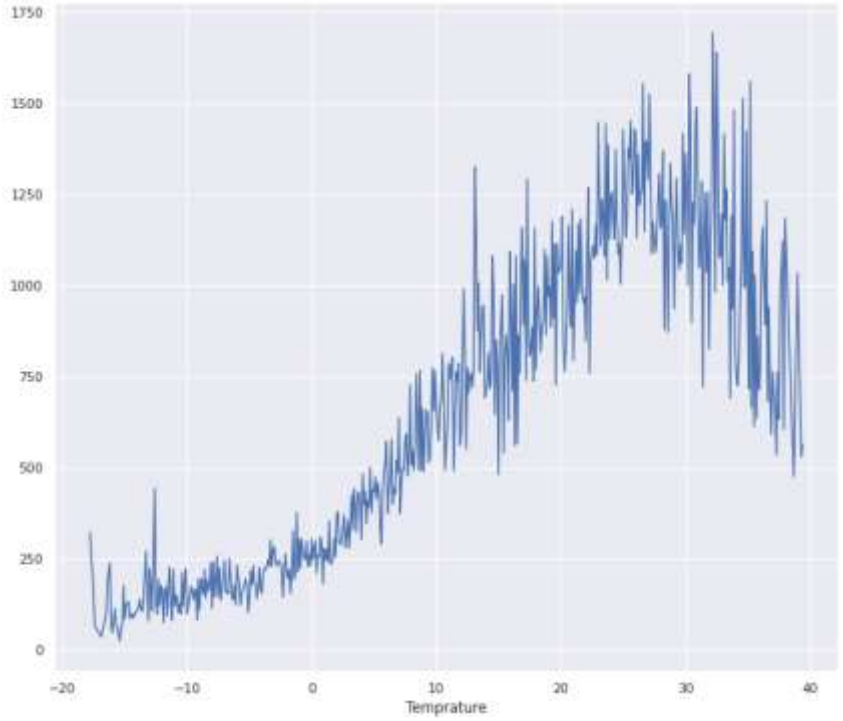
Analysis of rented bike

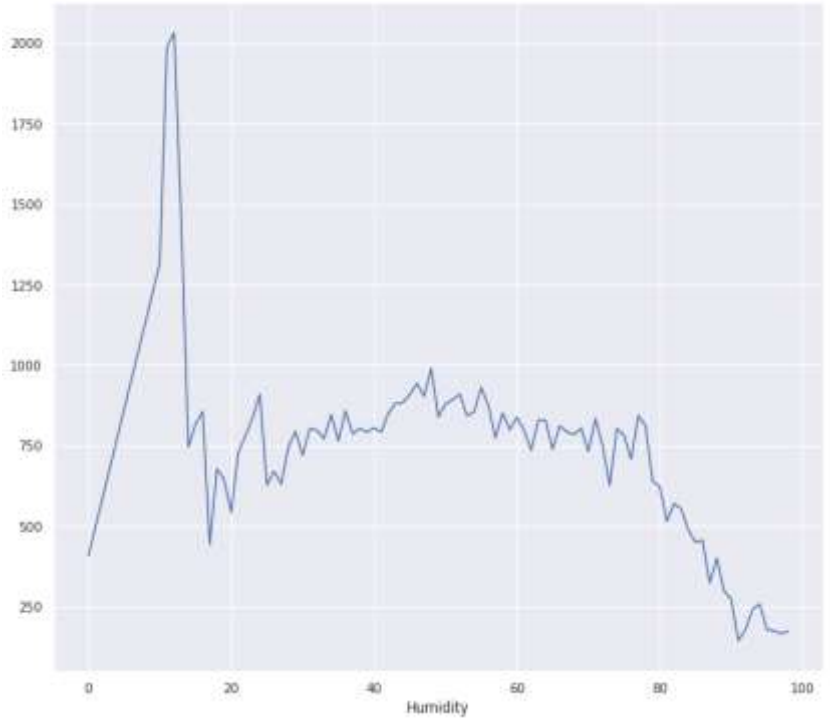
(Hour and holiday)



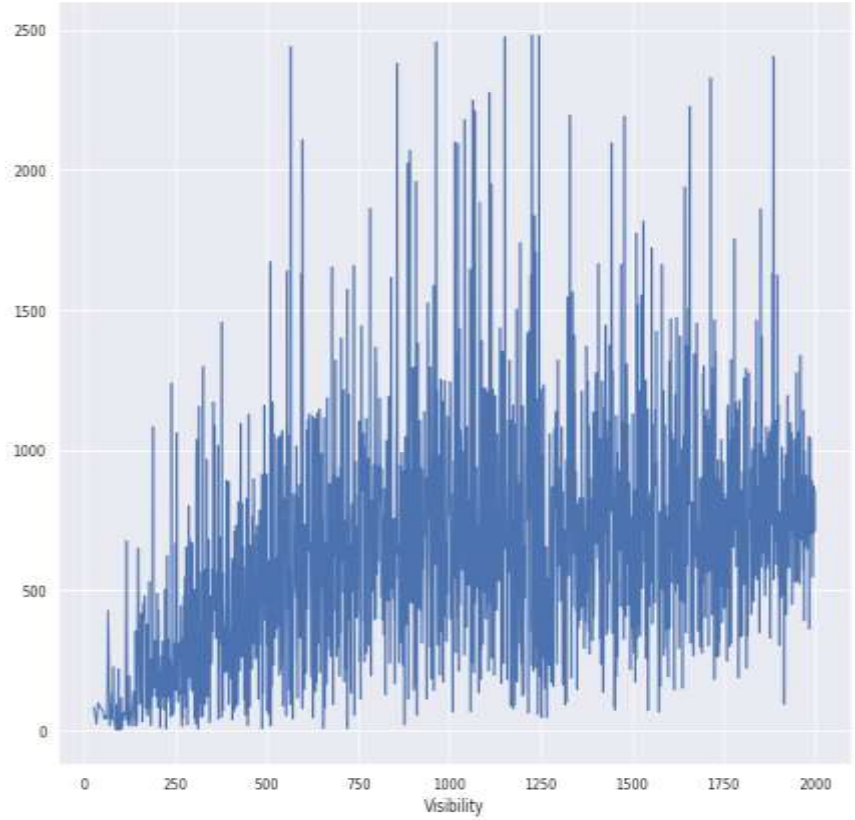
Numerical vs Rented bike count

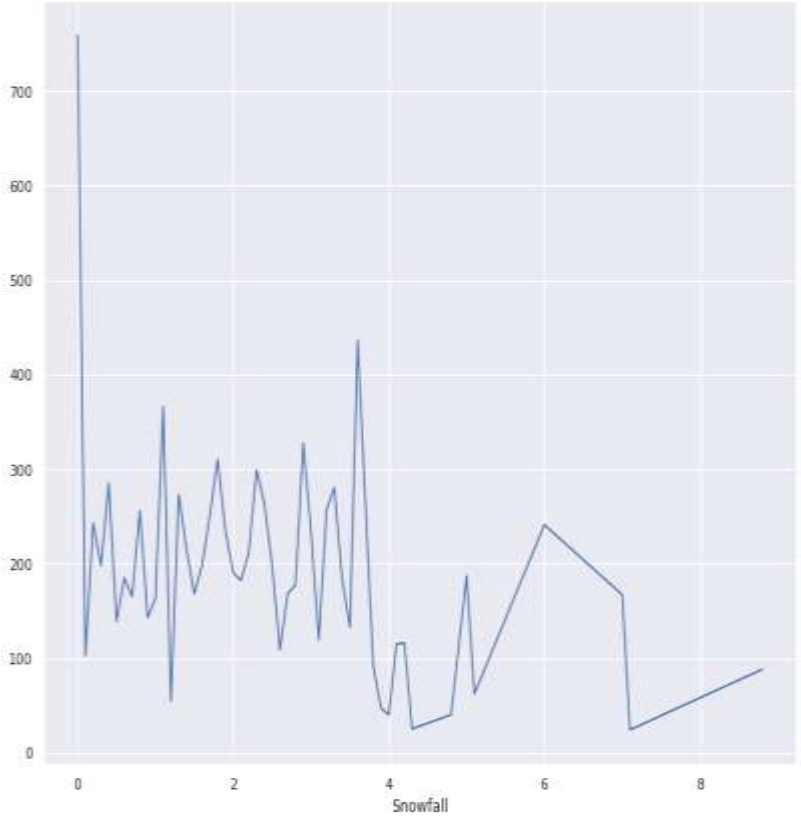


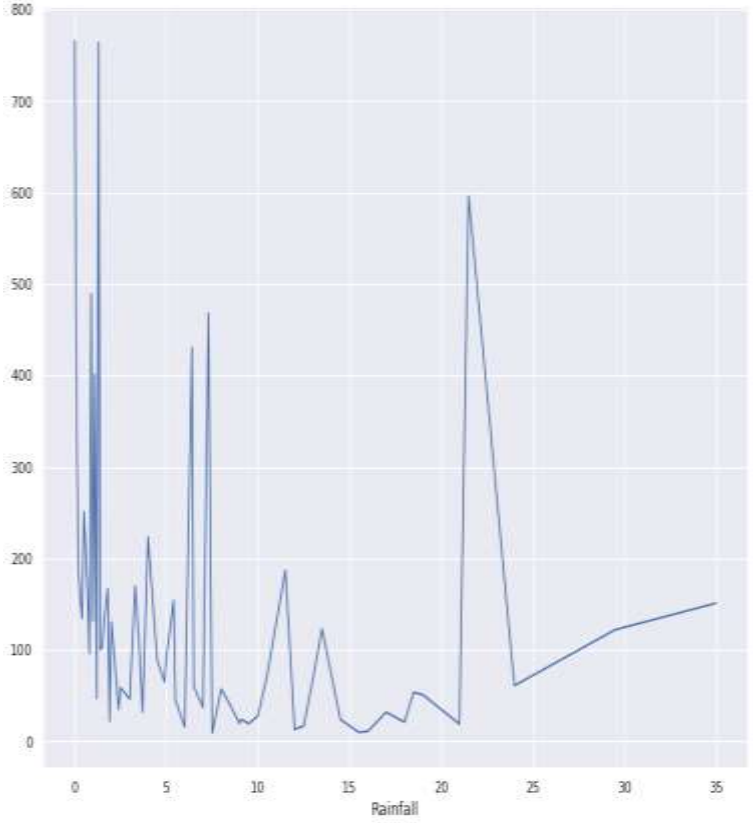




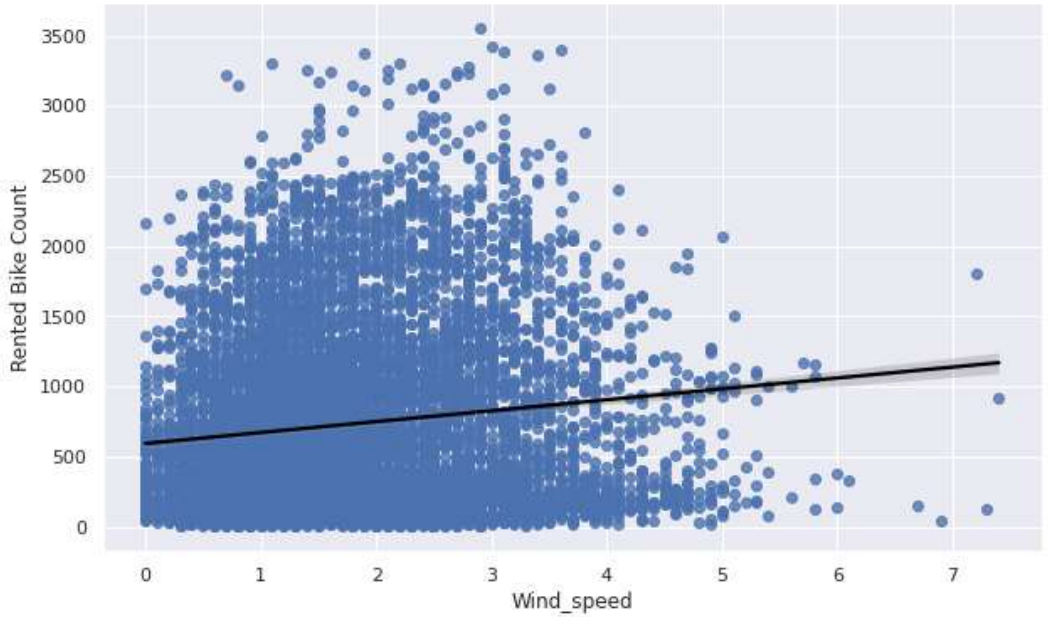
Numerical vs Rented bike count

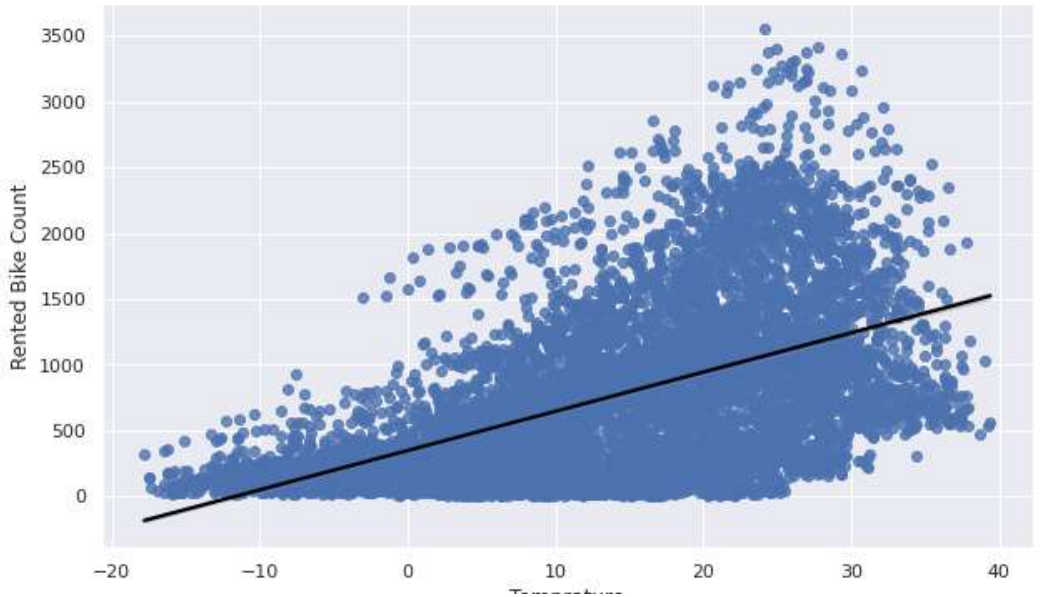


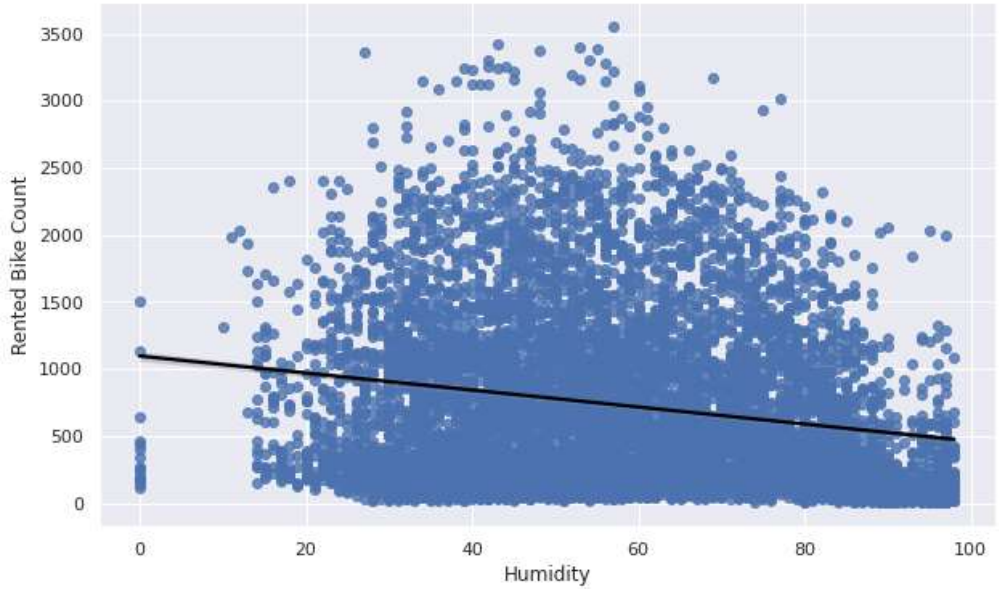


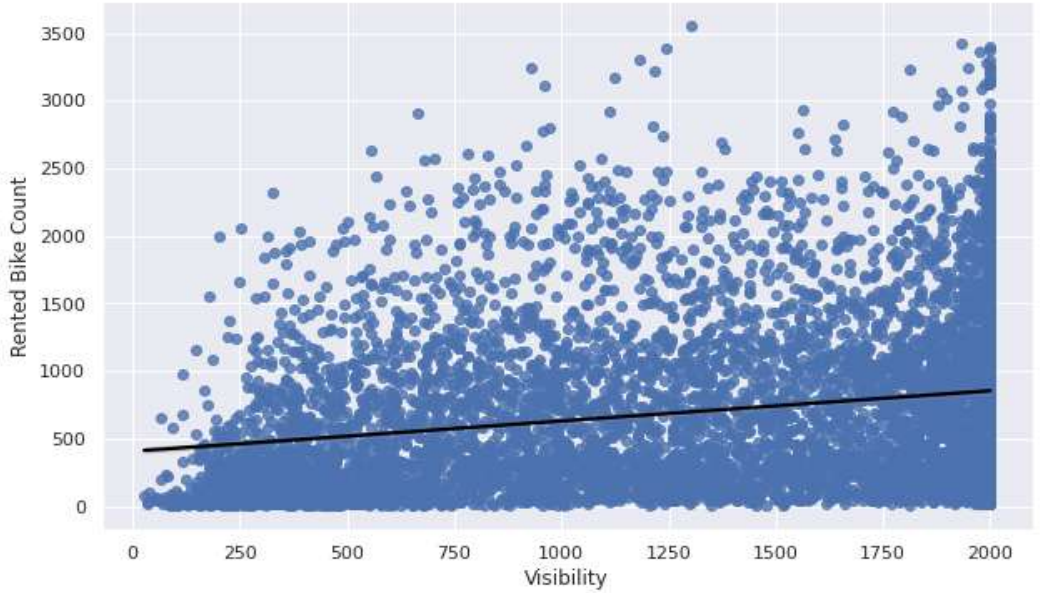


Regression Plot for numerical variable

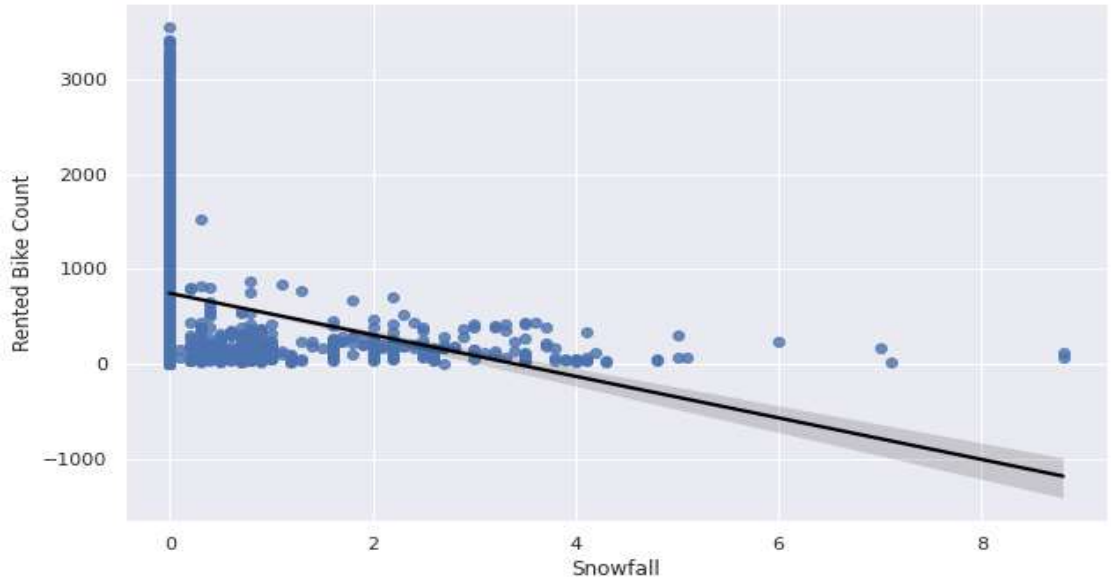


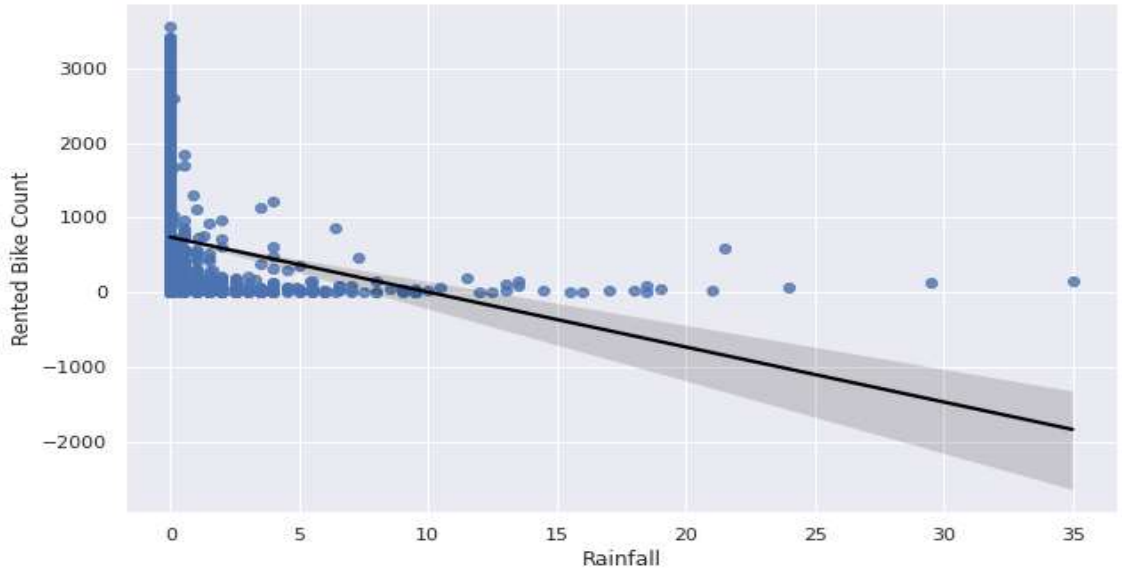


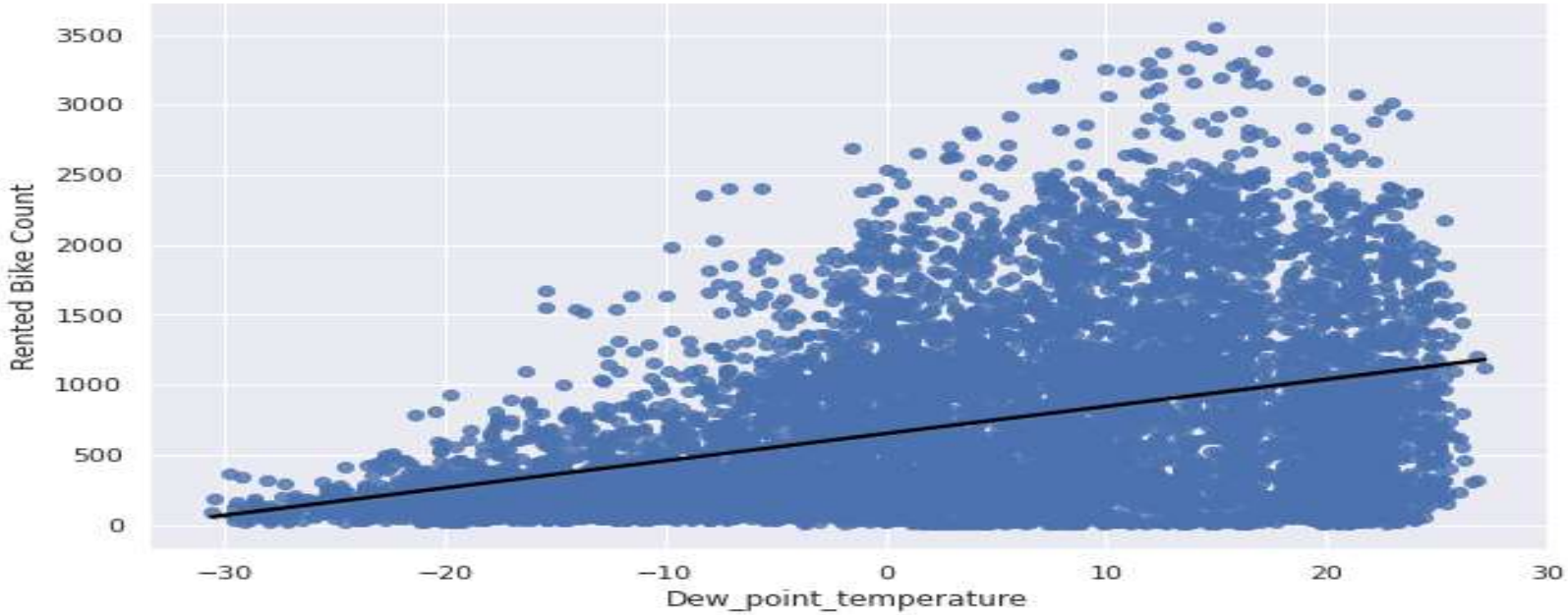




Regression Plot for numerical variable





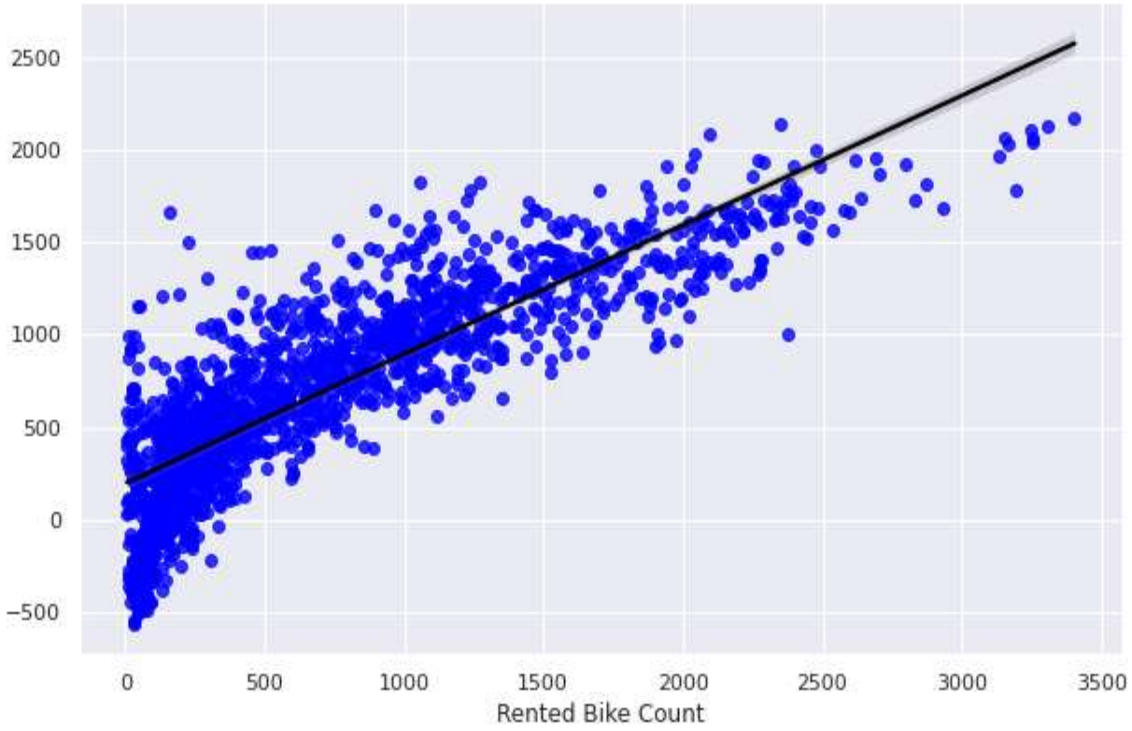


Correlation Heatmap

Model building

- LINEAR REGRESSION
- LASSO REGRESSION
- RIDGE REGRESSION
- DECISION TREES REGRESSOR
- RANDOM FOREST REGRESSOR
- GRADIENT BOOSTED REGRESSOR
- GRADIENT BOOSTING REGRESSOR WITH GRIDSEARCHCV

Linear Regression

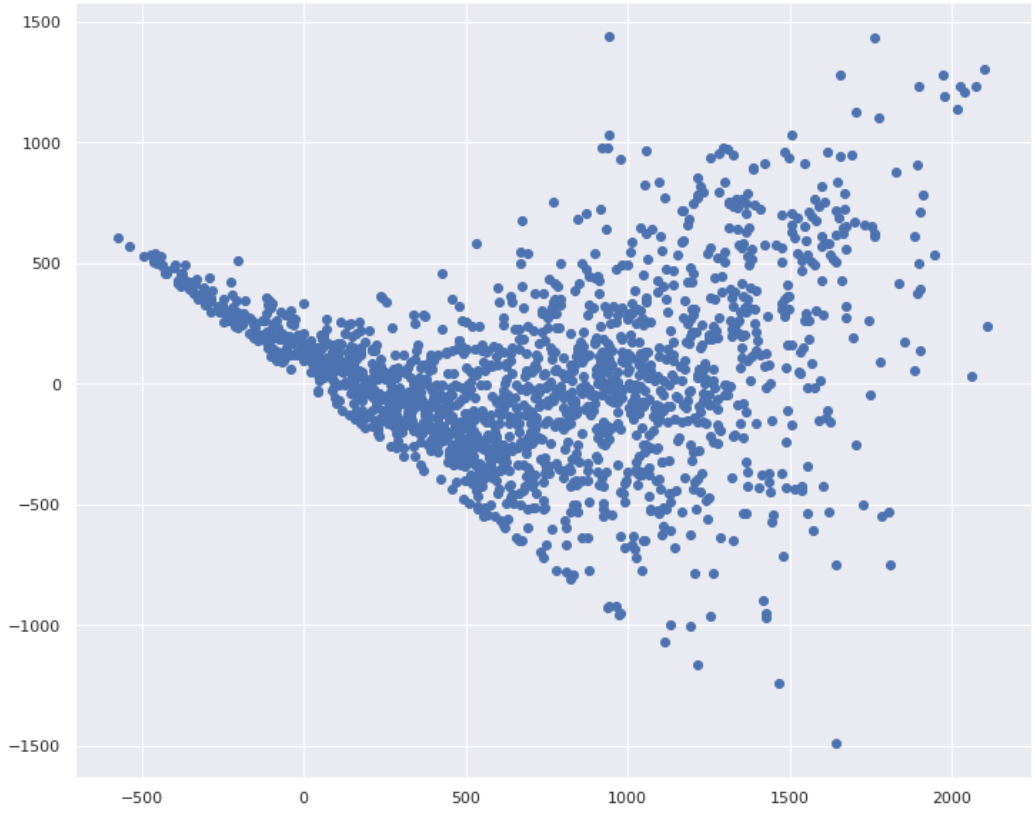


MSE : 123688.20499160145
RMSE : 123688.20499160145
MAE : 266.2024881866509
R2 : 0.713929384414175
Adjusted R2 : 0.7050387071473394

Decision Tree

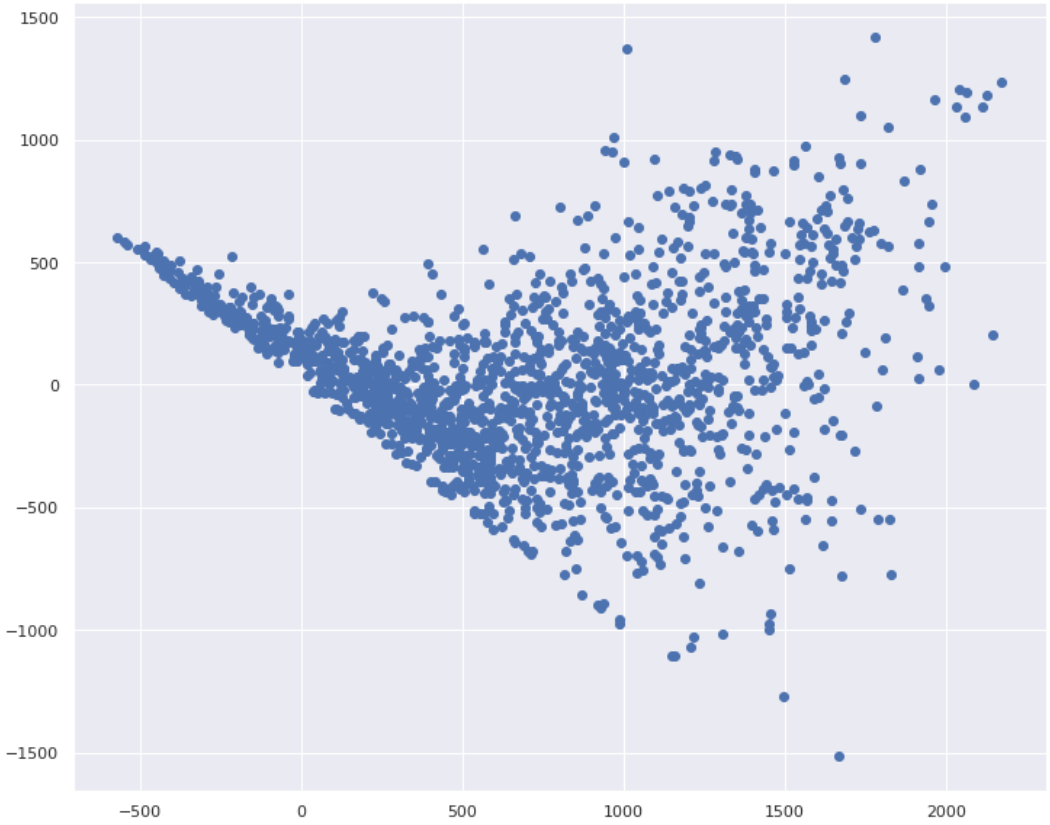
MSE : 203680.2100098523
RMSE : 451.3094393094967
MAE : 314.06006792277816
R2 : 0.5289209422666861
Adjusted R2 : 0.5142804596680273

Lasso Regression



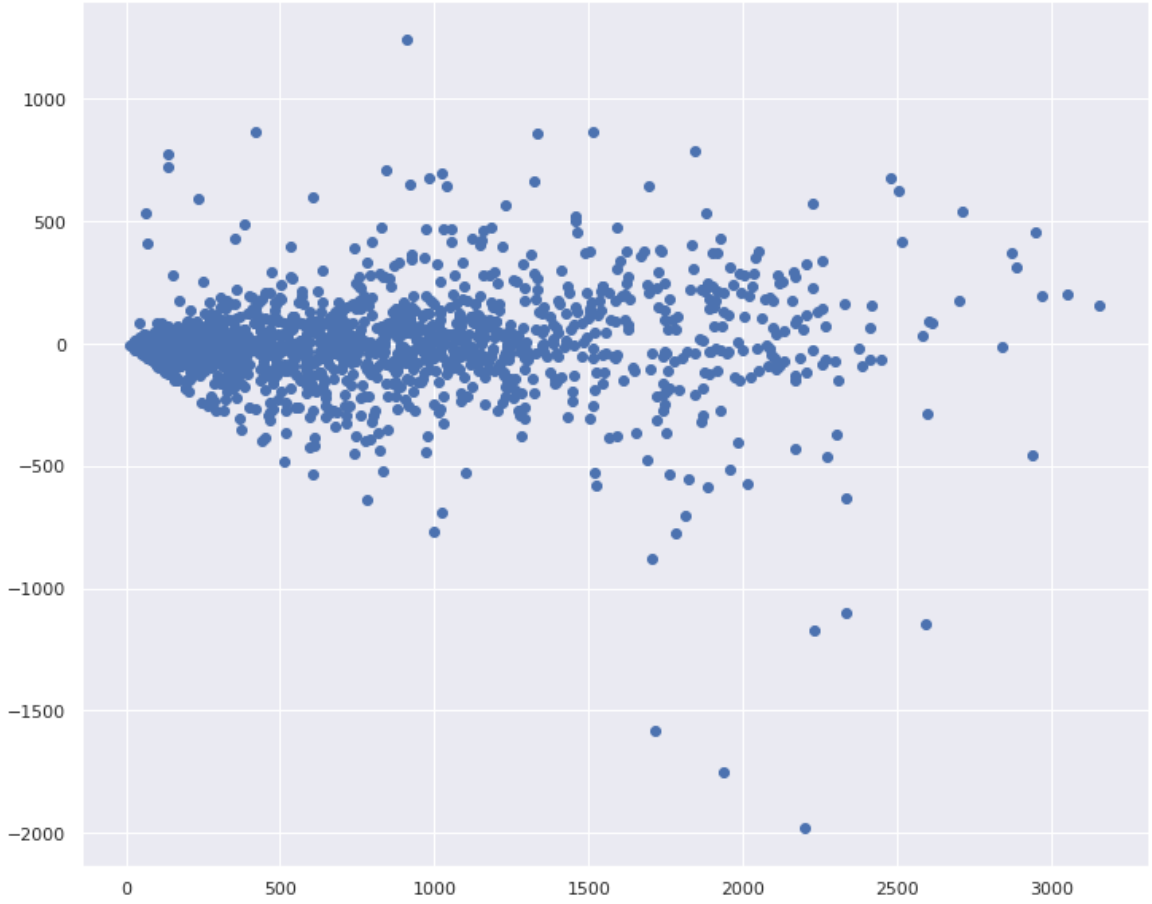
MSE : 126977.54006523405
RMSE : 356.3390801823932
MAE : 268.42480406043046
R2 : 0.7063216896509918
Adjusted R2 : 0.6971945758010227

Ridge Regression



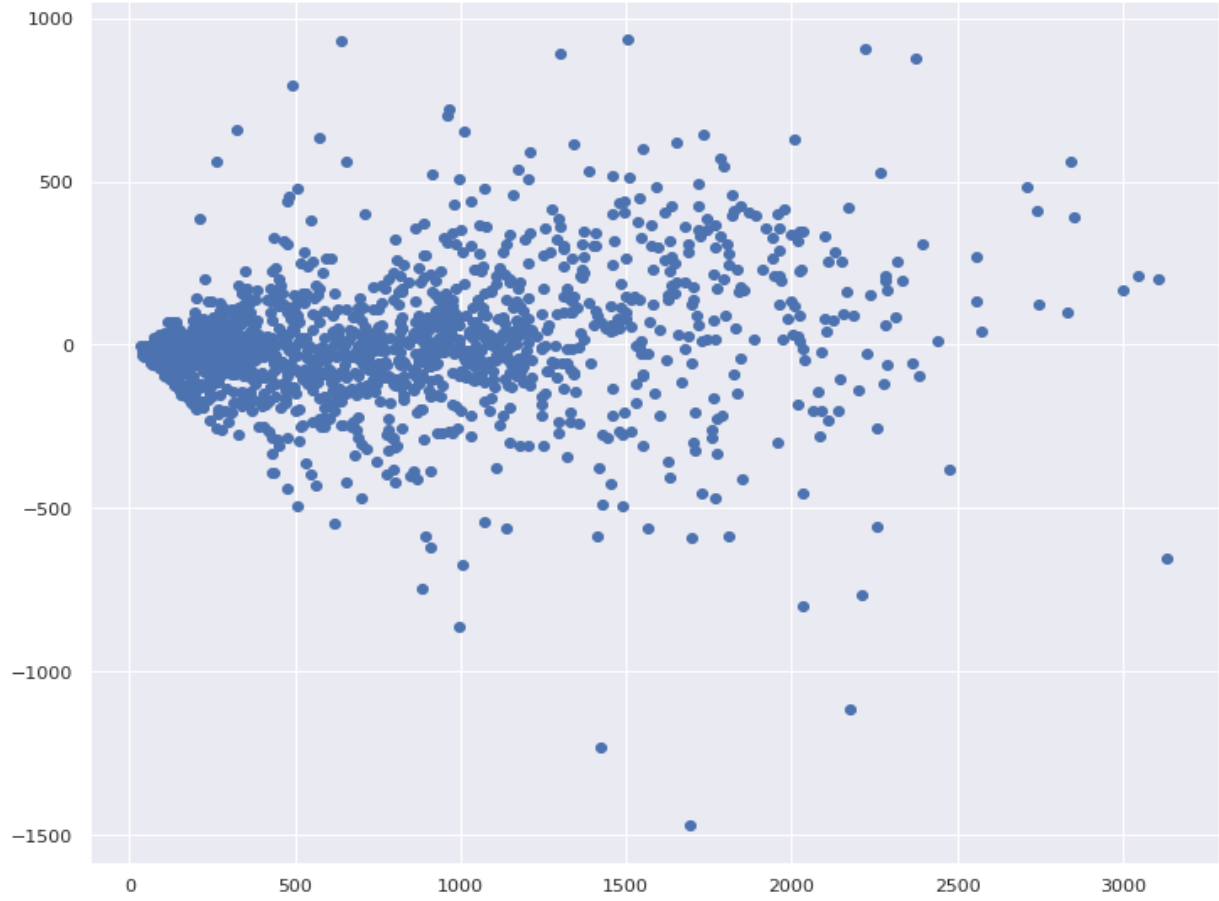
MSE : 123730.48290489083
RMSE : 351.7534405018533
MAE : 266.21508579231676
R2 : 0.7138316025061814
Adjusted R2 : 0.7049378863135033

Random Forest



MSE : 37965.17489125812
RMSE : 194.84654190223165
MAE : 115.0099291199055
R2 : 0.9121927515020274
Adjusted R2 : 0.9094638242178125

Random Forest with tuning

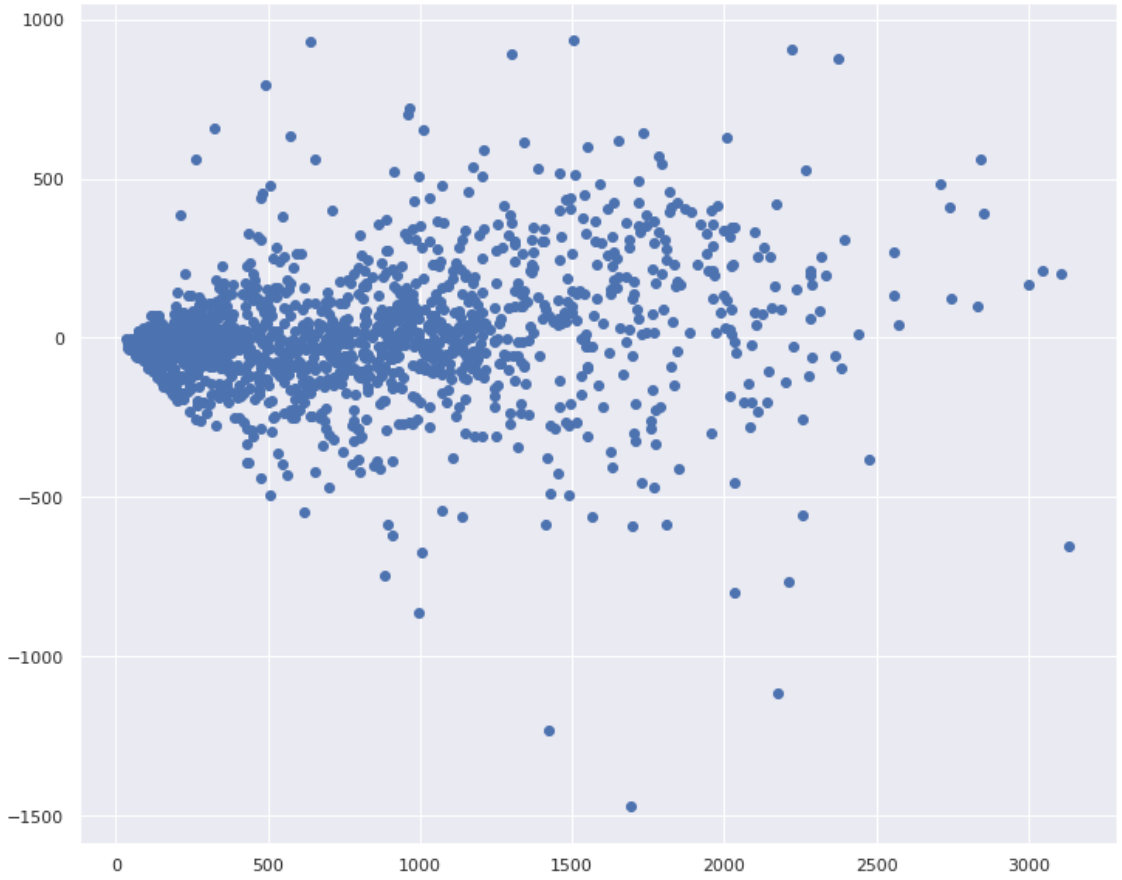


MSE : 36490.735509509745
RMSE : 191.02548392690895
MAE : 121.89040756054341
R2 : 0.9156028889650893
Adjusted R2 : 0.9129799440151927

Gradient Boosting

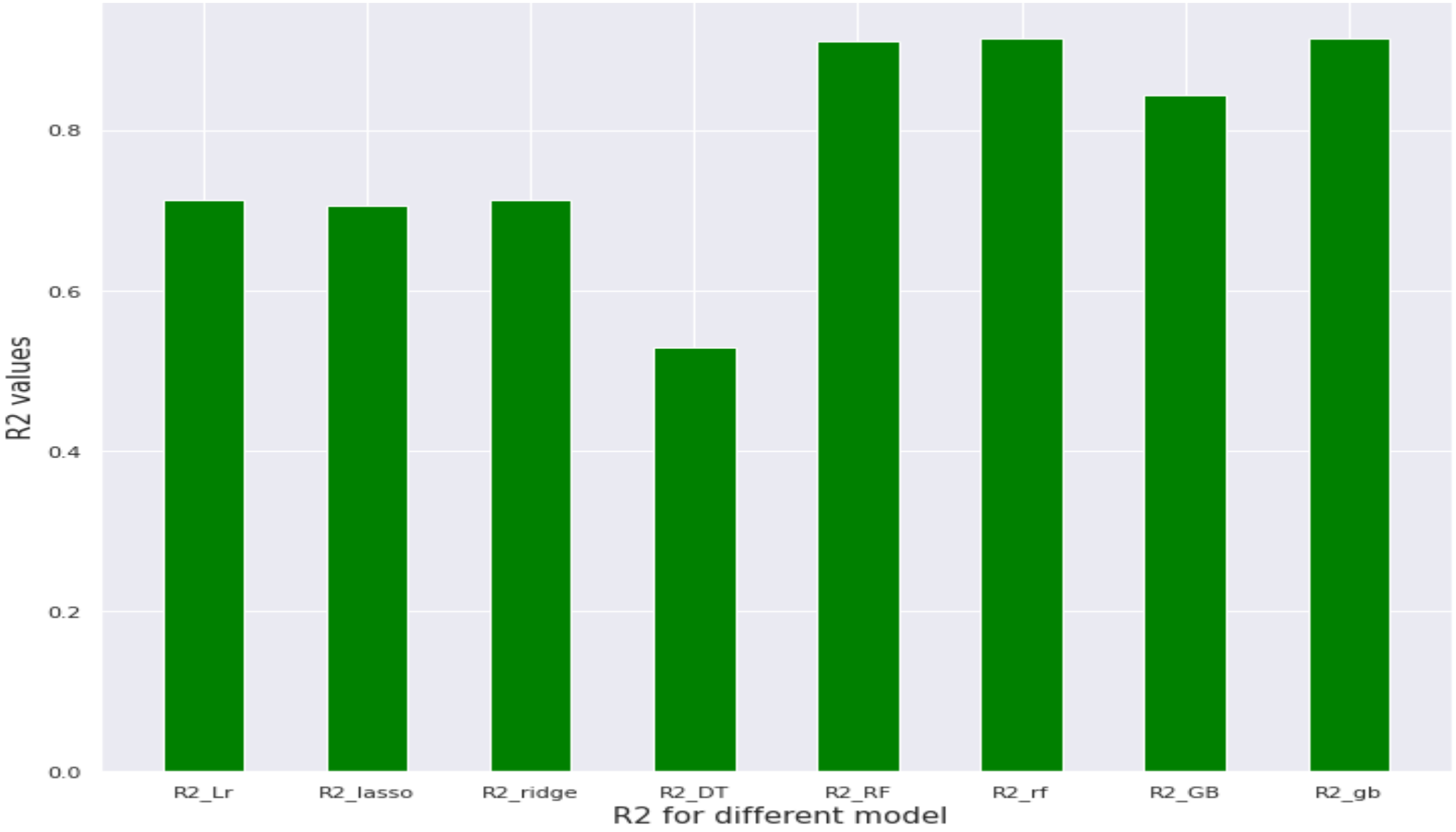
MSE : 67552.40626356455
RMSE : 259.9084574683259
MAE : 185.0604581428772
R2 : 0.843762317955589 8
Adjusted R2 : 0.8389066678737708

Gradient Boosting with tuning



MSE : 36490.735509509745
RMSE : 191.02548392690895
MAE : 121.89040756054341
R2 : 0.9156028889650893
Adjusted R2 : 0.9129799440151927

Comparison of R2 values for different model



Conclusion

➢ ‘Hour’ of the day holds the most important feature.

➢ Bike rental count is mostly correlated with the time of the day as it is peak at 10 am morning and 8 pm at evening.

➢ We observed that bike rental count is high during working days than non working day.

➢ We see that people generally prefer to bike at moderate to high temperatures, and when little windy

➢ It is observed that highest number bike rentals counts in Autumn & Summer seasons & the lowest in winter season. We observed that the highest number of bike rentals on a clear day and the lowest on a snowy or rainy day. We observed that with increasing humidity, the number of bike rental counts decreases.

Thank You