



Capstone project-2

Bike Sharing Demand Prediction



Team Details:

Team Name: *Data Avengers*

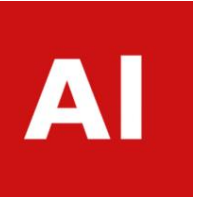
Team Members Details:

- Vivek Kumar Soni
- Ashish Pandey

Content

A red square containing the white letters 'AI' in a bold, sans-serif font.

- Individual Information
- Business Understanding
- Feature Analysis
- Data Summary
- Exploratory Data Analysis
- Data Preprocessing
- Implimenting Algorithms
- Challenges
- Conclusion



Business Understanding

Bike rentals have become 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 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 location .

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

Data summary

AI

	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0	Winter	No Holiday	Yes

- This dataset contains 8760 lines and 14 columns.
- Three categorical features 'seasons' , holiday' and functioning
- One Datetime features 'Date'.
- We have some numerical type variables such as temperature , humidity and visibility , dew point temp ,solar radiation ,rainfall ,snowfall with environment conditions at particular hour of the day.

Feature Summary

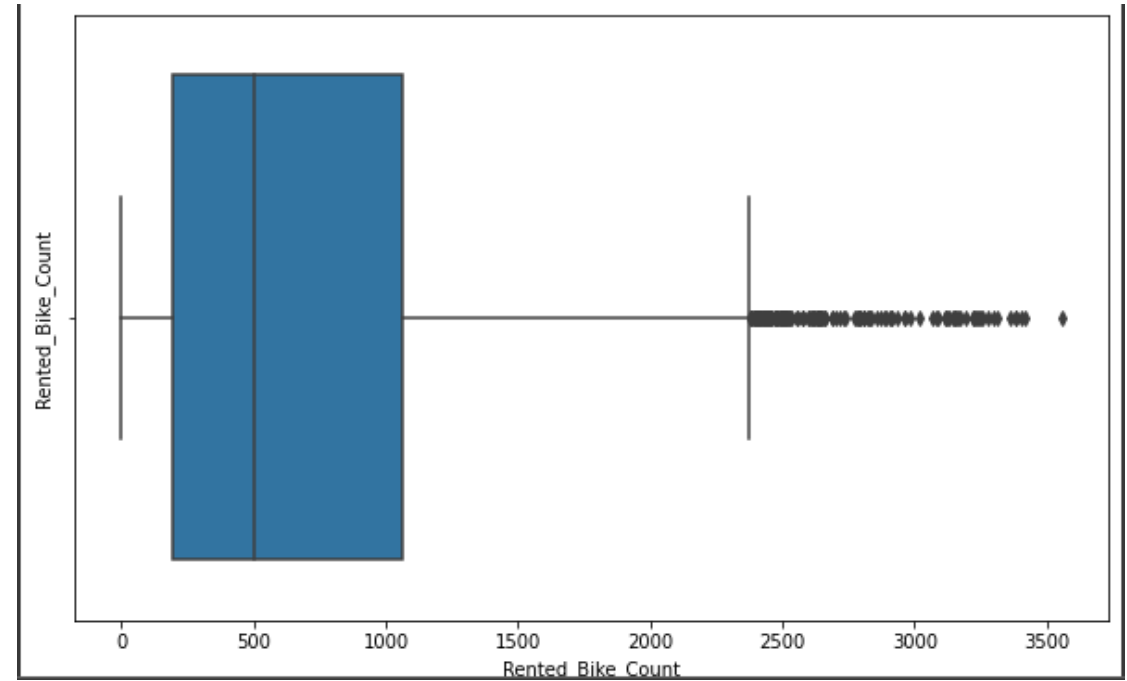
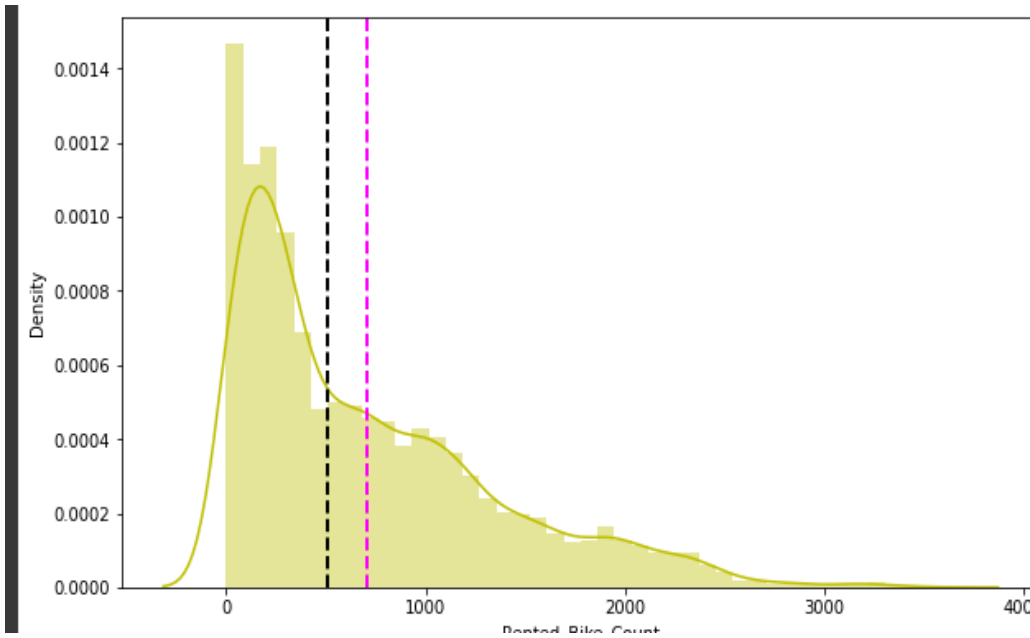
A red square logo with the white letters 'AI' inside.

- Date : Year-Month-Date
- 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
- Functional Day- NoFunc(Non Functional Hrs),Fun(Functional Hrs)

Insights From Our Dataset

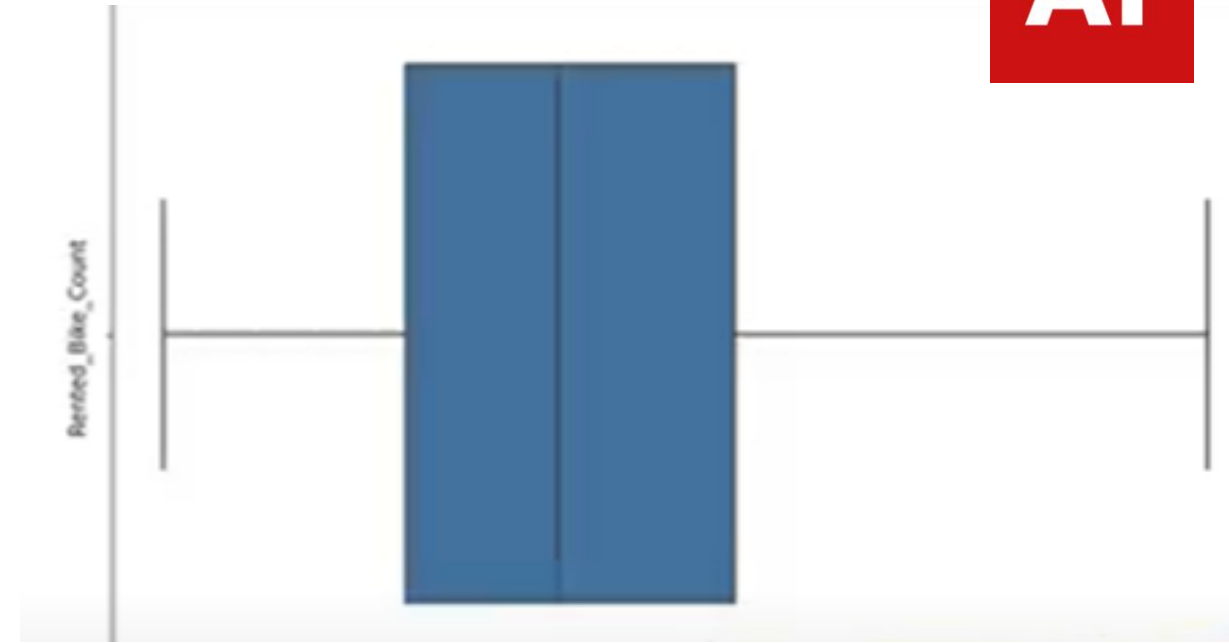
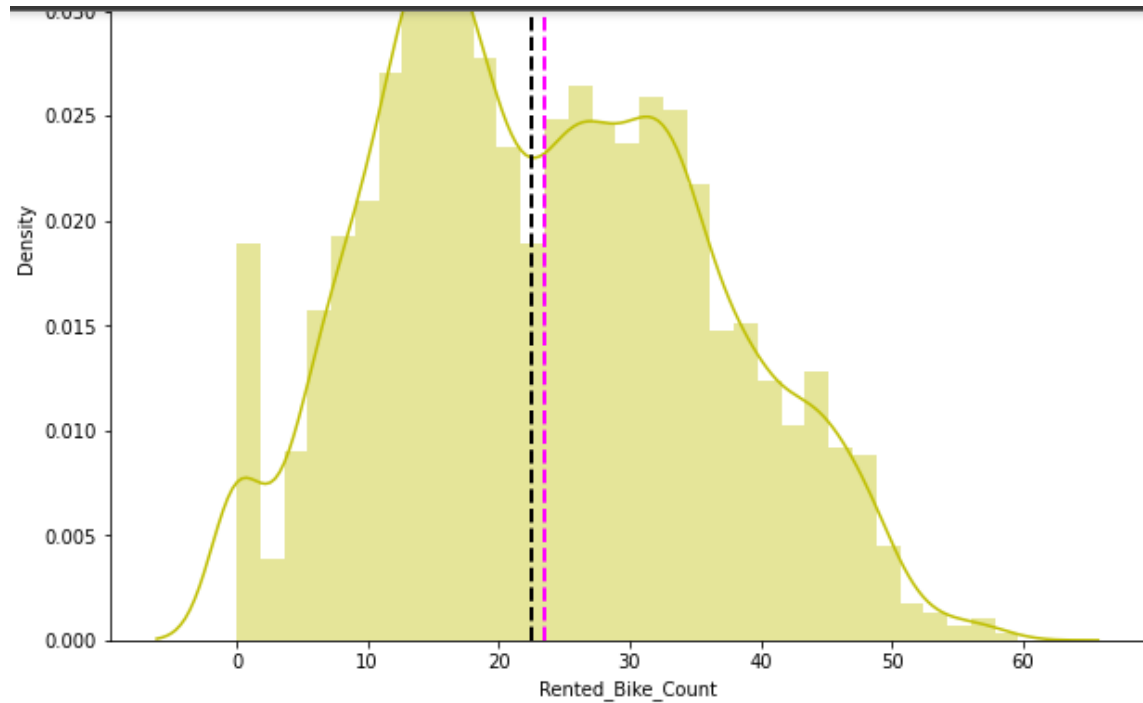
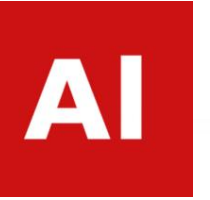


- There are no missing values present.
- There are no missing values present.
- There are no null values.
- And finally we have 'rented count' variable which we need to predict for new observation.
- The dataset shows hourly rental data for one year 1December 2017 to 31 November(2018)(365days).we consider this as a single year data.
- So we convert the “ date” column into 3 different column.
- We change the name of some feature for our convinence, they are 'Rented_bike_count' , 'Hour' , 'Temperature' , 'Humidity' , 'Wind_speed' , 'Dew_point_temperature' , 'Solar_Radiation' , 'Rainfall' , 'Snowfall'



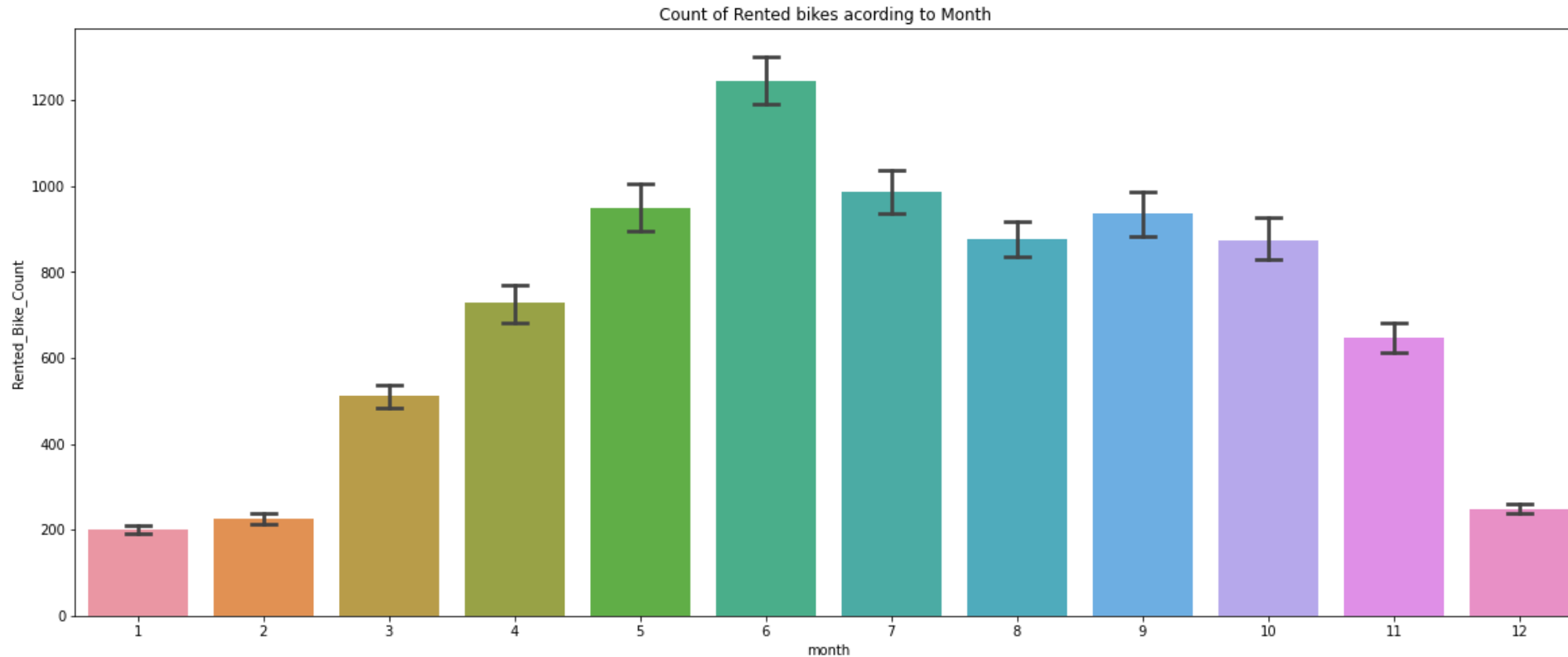
- The above graph shows that rented bikes count has moderate right
- The above boxplot shows that we have detect outliers in rented bikes
- Since the assumption of linear regression is that 'the distribution variable has to be normal' , so we should perform square root normal.

Analysis of rented bike column



- After applying square root to the skewed rented bike count.
- After applying square root to the rented bike count column .

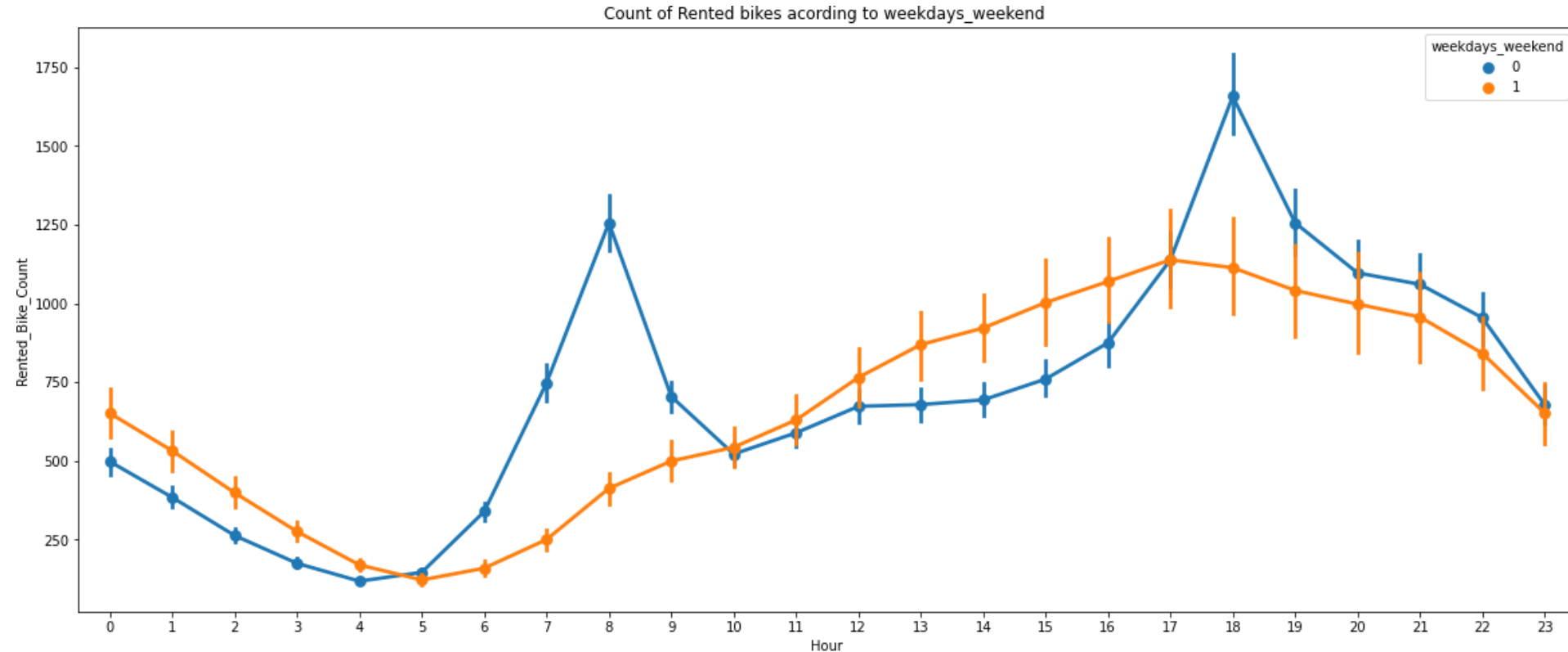
ANALYSIS OF MONTH VARIABLE



From the above bar plot we can clearly say that from the month that demand of the bike is high as compare to other month are comes inside the summer season.

ANALYSIS OF WEEKDAYS WEEKEND VARIABLE

AI



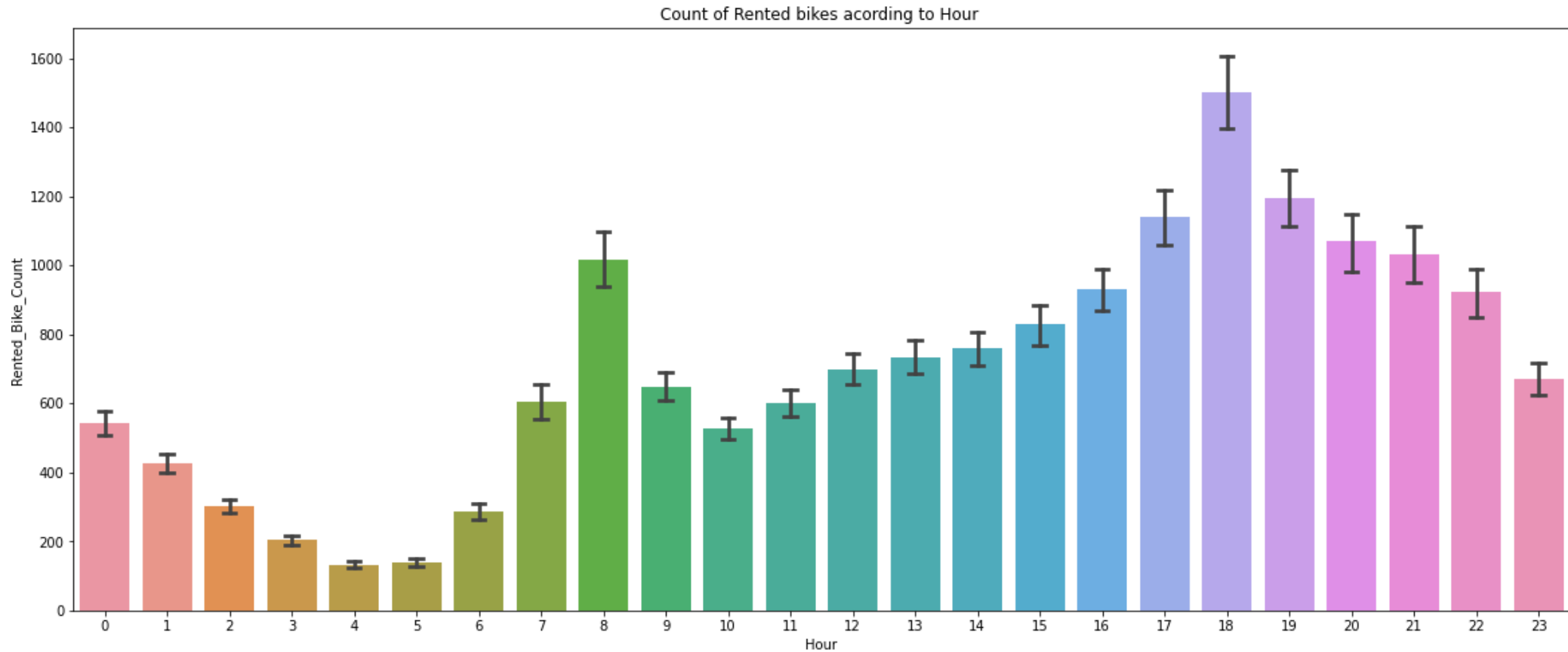
From the above point plot and bar plot we can say that in the weekdays which represent in blue colour show that the demand of the bike higher because of the office.

Peak Time are 7 am and 5 pm to 7 pm.

The orange colour represent the weekend days, and it show that the demand of rented bikes are very low specially in the morning hour but when the evening start from 4 pm to 8 pm the slightly increase.

ANALYSIS OF HOUR VARIABLE

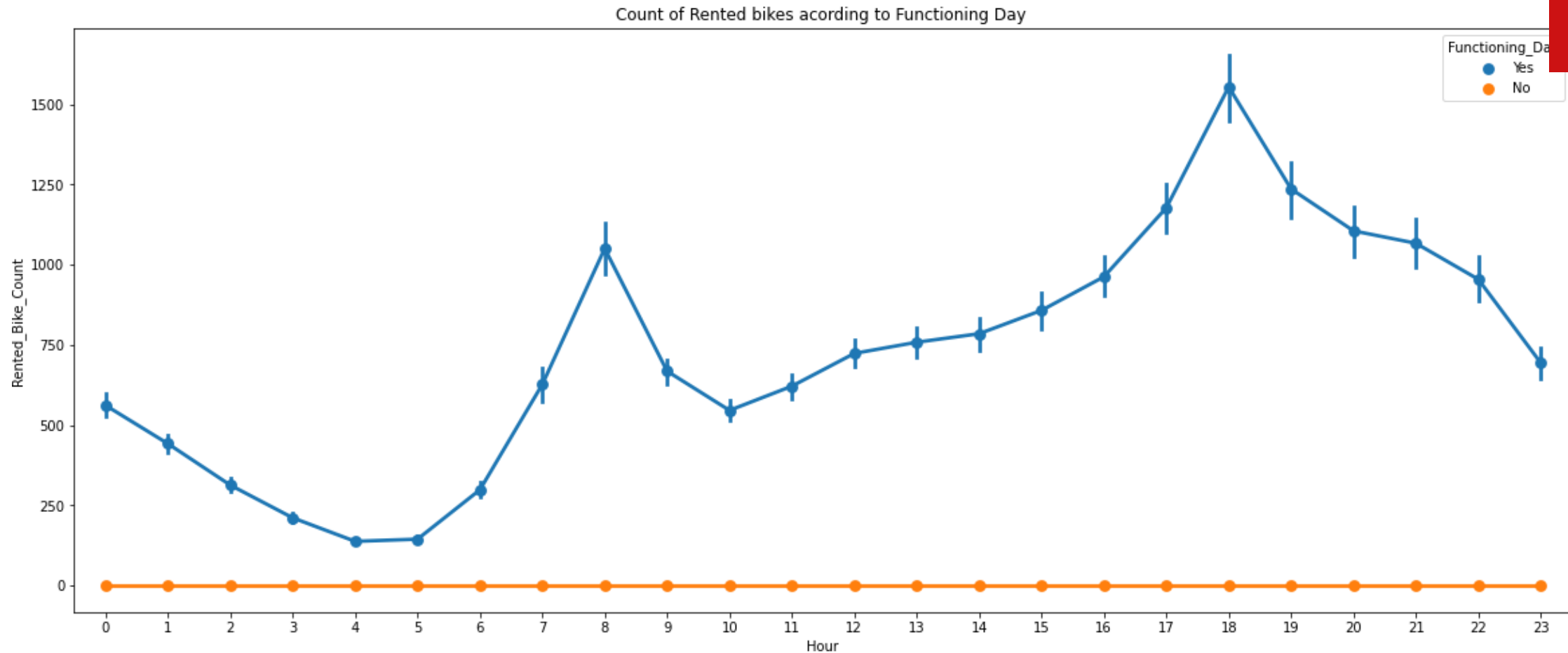
AI



- In the above plot which shows the use of rented bike according the hour and the data are from all over the year.
- Generally people use rented bikes during their working hour from 7am to 9pm and 5pm to 7 pm.

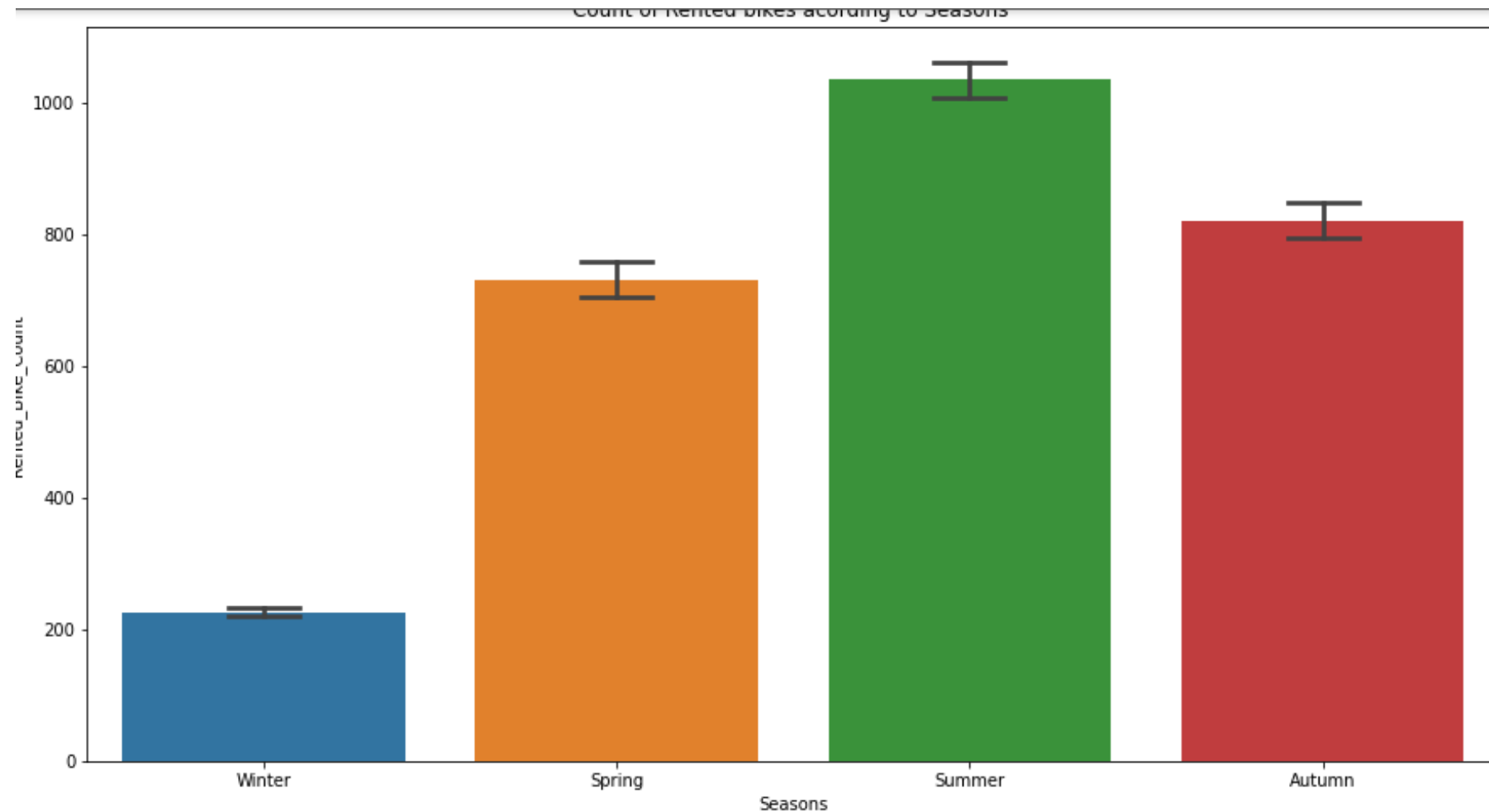
ANALYSIS OF FUNCTIONING DAY VARIABLE

AI



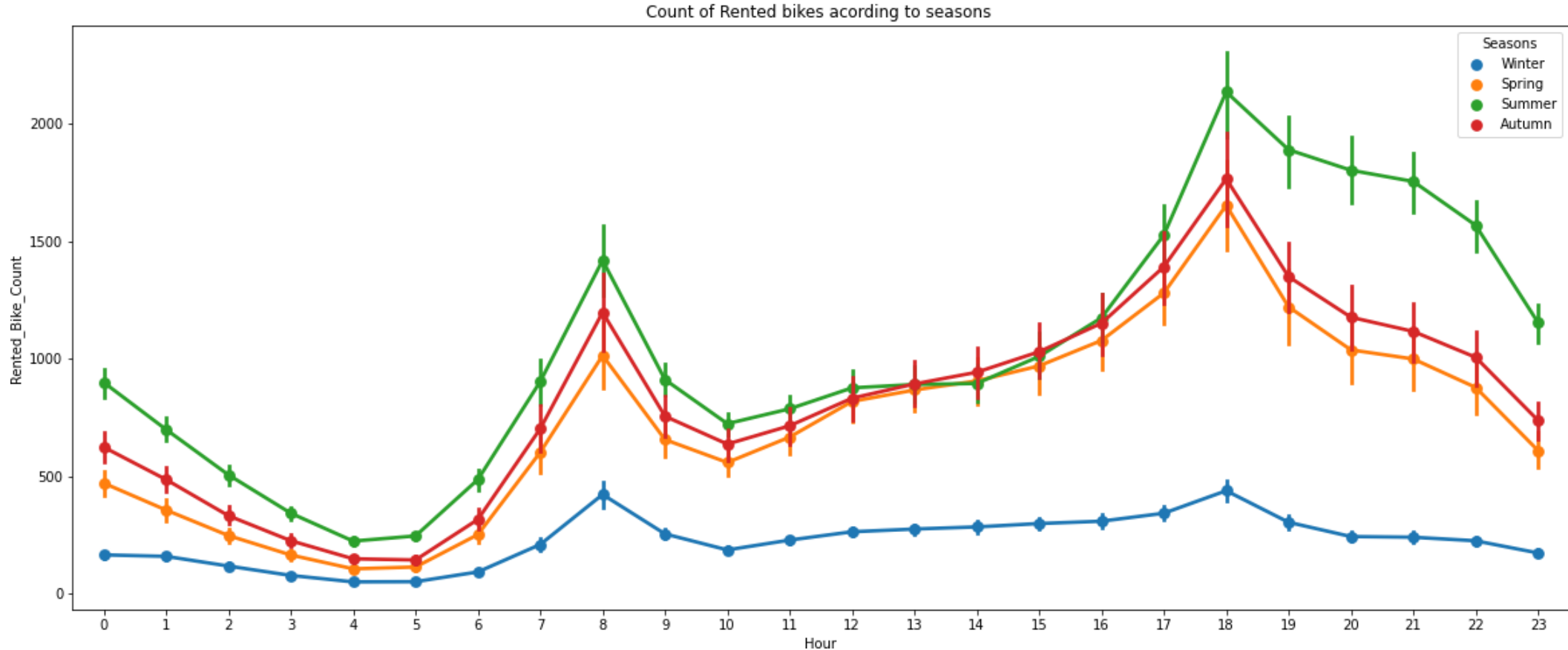
In the above point plot which shows the use of rented bike in functioning days or not , and it clearly shows that,
Peoples don't use rented bikes in no functioning day.

ANALYSIS OF SEASON VARIABLE



- This above bar plot shows the distribution of rented bike count
- And we can clearly see that the peoples love to ride bike in summer and antumn season
- But in winter season peoples don't take any rented bike

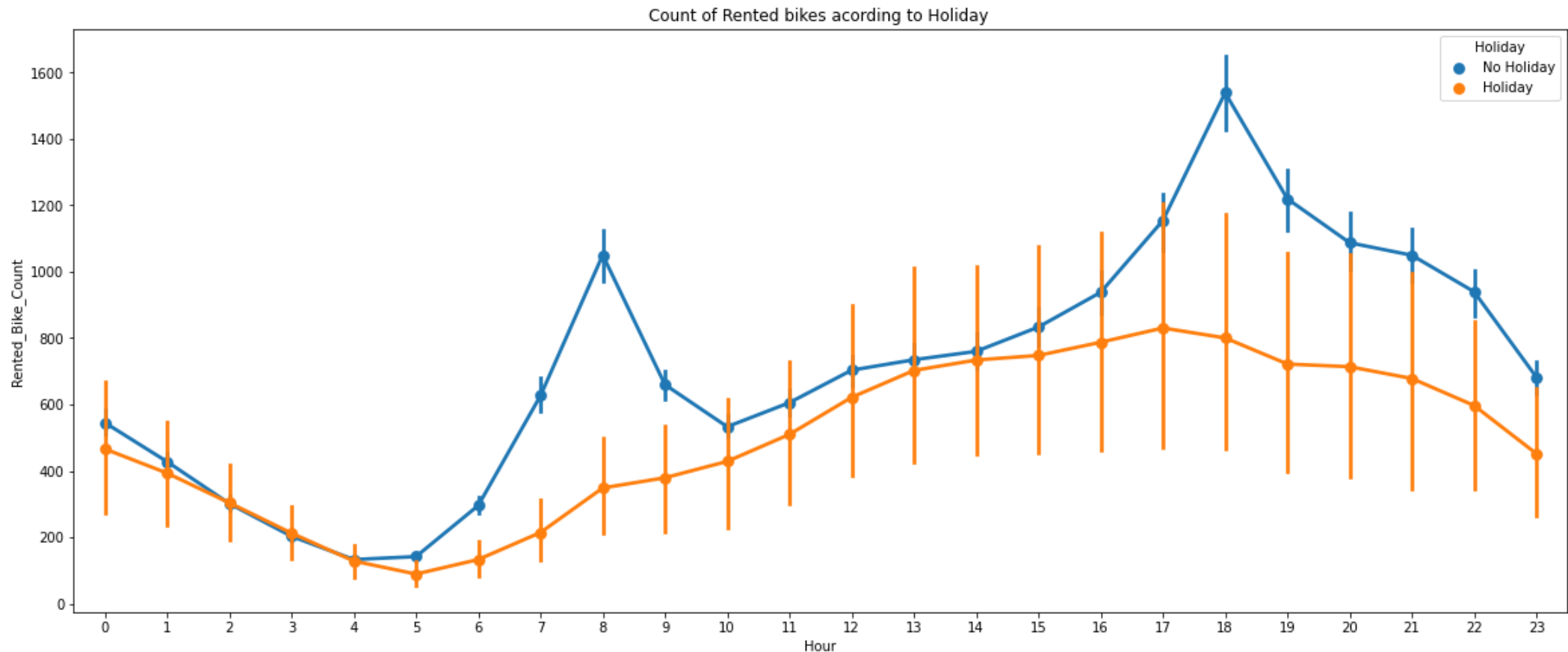
ANALYSIS OF SEASON VARIABLE



- In the above bar plot and point plot which shows the use of rented bike in four different seasons , and it clearly shows that,
- In summer season the use of rented bike is high and peak time is 7 to 9 am and 5-9 pm.
- In winter season th use of rented bike is very low because of snowfall.

ANALYSIS OF HOLIDAY VARIABLE

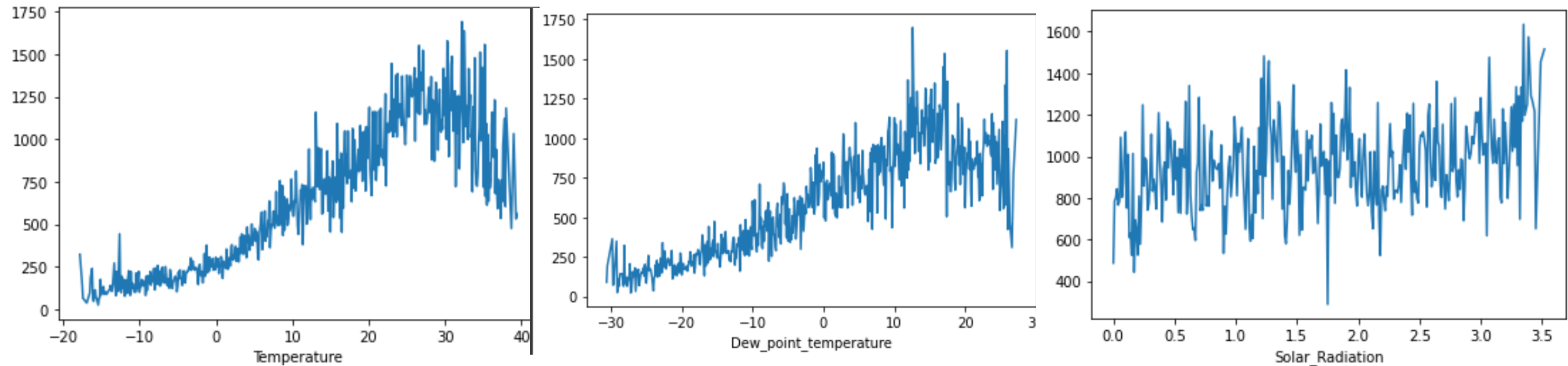
AI



In this above bar plot and point plot which shows the use of rented bike in a holiday and it clearly shows that,
Plot shows that in holiday people uses the rented bike from 2pm to 8pm.

NUMERICAL VS RENTED BIKE COUNT

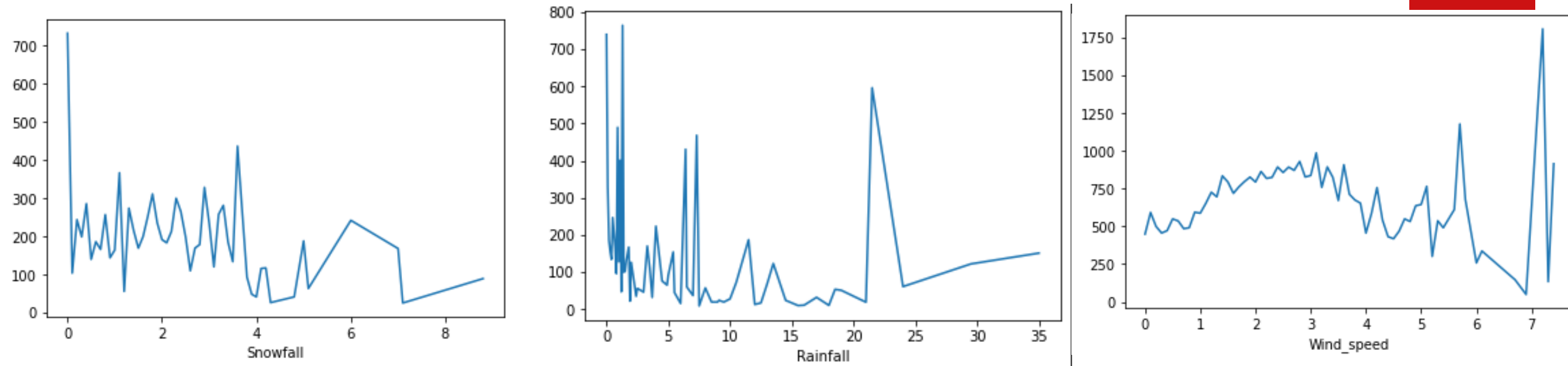
AI



- From the above plot we see that people like to ride bikes when it is around 25 degree in average.
- From the above plot of 'Dew_point_temperature' is almost same as the 'temperature' there is some similarity present we can check it in our next step.
- From the above plot we see that ,the amount of rented bikes is huge , when there is solar radiation the counter of rents is around 1000

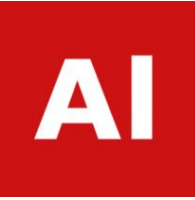
NUMERACAL VS RENTED BIKE COUNT

AI



- In snowfall plot, on the y axis the amount of rented bike is very low .when we have more than 4 cm of snow, the bike rent is much lower.
- In rainfall plot if it rains a lot the demand of rent bikes is not decrease. For example even if we have 20mm of rain there is a big peak of rented bike is count.
- In wind speed plot that the demand of rented bike is uniformaly distributed despite of wind speed but when the speed of wind was 7m/s then the demand of bike also increase that clearly means love to ride bikes when it little windy.

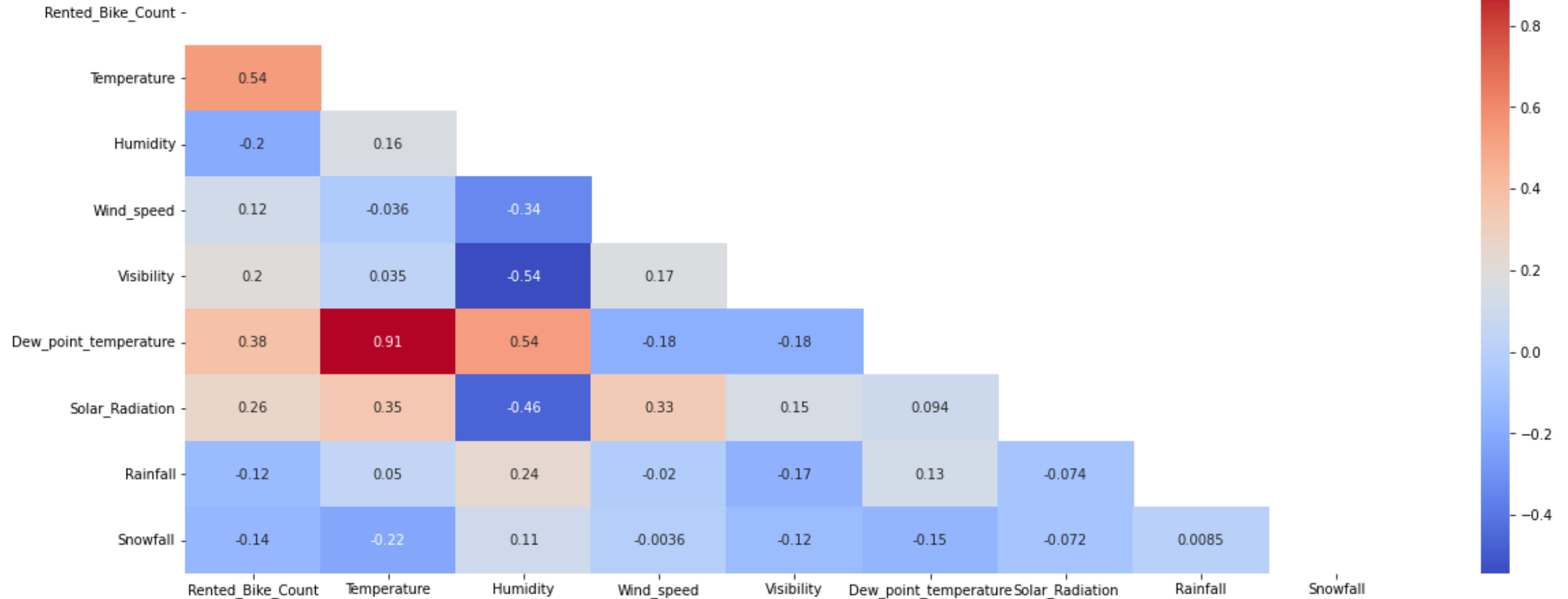
REGRESSION PLOT FOR NUMERICAL VARIABLE



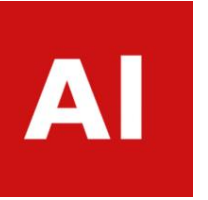
- From the above regression plot of all numerical features we see that the columns 'temperature' , 'wind_speed', 'visibility','Dew_point_temperature','solar_radiation' are positively relation to the target variable.
- Which means the rented bike count increases with increase of these features.
- 'Rainfall' , 'snowfall' , 'humidity', these feature are with the target variable which means the rented bikes decreases when these features increase.

CORRELATION MATRIX

AI



From the above correlation heatmap, We see that there is a positive correlation between columns 'Temperature' and 'Dew point temperature' i.e 0.91 so even if we drop this column then it don't affects the outcome of our analysis. And they have the same variations.. so we can drop the column 'Dew point temperature(°C)'.

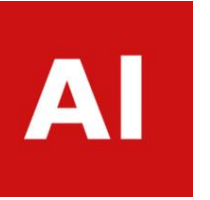


CHALLENGES

- Large dataset to handle.
- Needs to plot lot of graphs to analyse.
- Feature engineering
- Feature selection
- Optimising the model
- Carefully tuned hyperparameters as it affects

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 am 8pm at evening.
- We observed that the bike rental count is high during working days than non working days.
- We see that people generally prefer to bike at moderate to high temperature and when little windy.



THANKING YOU...