

Capstone Project
Bike Sharing Demand
Prediction

Team - Dream

Team Members

Indugopal Maity

Md. Tousif Ali

Let`s analyse "Bike Sharing":

1. What is Bike sharing System
2. How does Bike sharing work
3. What are we getting from Bike Sharing
4. Where can we find Bike sharing
5. Data Overview
6. EDA
7. Feature v/s Bike Count
8. Regression Plot
9. Correlation Analysis
10. Modelling Overview
11. Feature Importance
12. Conclusion.



What is Bike Sharing??

Facilities which let people borrow bikes from a 'dock' or a bike rack and return it back at another 'dock' belonging to the same system

In 1965, Amsterdam first implemented the Concept with the name "White Fietsen"

Currently, there are around 2000 active schemes in 85 countries.

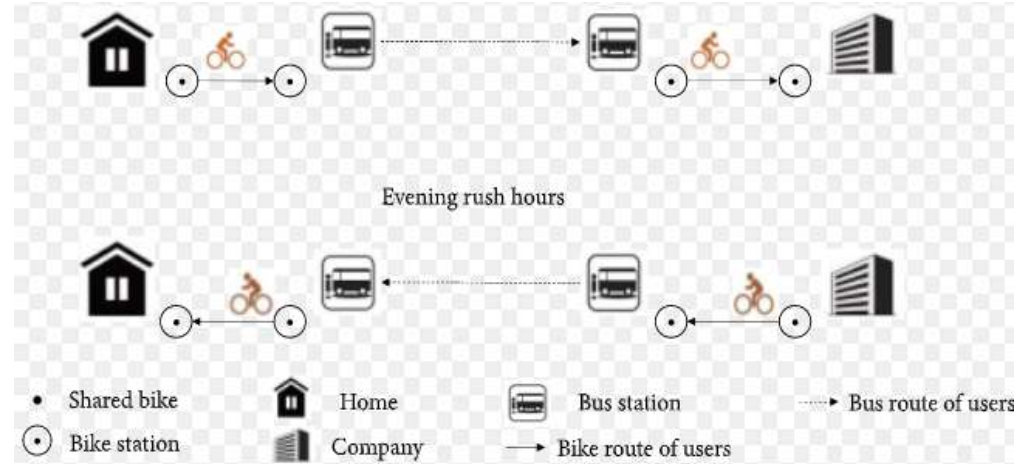


How does Bike Sharing work??

Bike sharing is a service that let Companies/ service provider to rent out their bikes to travellers/users for a certain interval time.

Benefits: Travellers can rent a bike

1. 24 Hours/ 7 days
2. Self Service Management
3. Registration using cards.
4. Web/ Smart Phone Integration



Sign In

Sign up online for a B-card or pay daily at the kiosk.



Select

Choose a bike from any B-station.



Ride

Grab your bike and go.



Return

Park your B-cycle at any B-station.

HEALTH AND ENVIRONMENTAL

BENEFITS OF BIKE SHARING



WESTCOASTTRIALLAWYERS.COM



Lowers your blood pressure.



Cuts down on greenhouse gas emissions and global climate change.

Increases your energy.

Decreases air pollutants.

Increases your flexibility and muscle strength.

Reduces noise pollution and congestion.

Reduces your stress levels.

Lessens your ecological footprint.

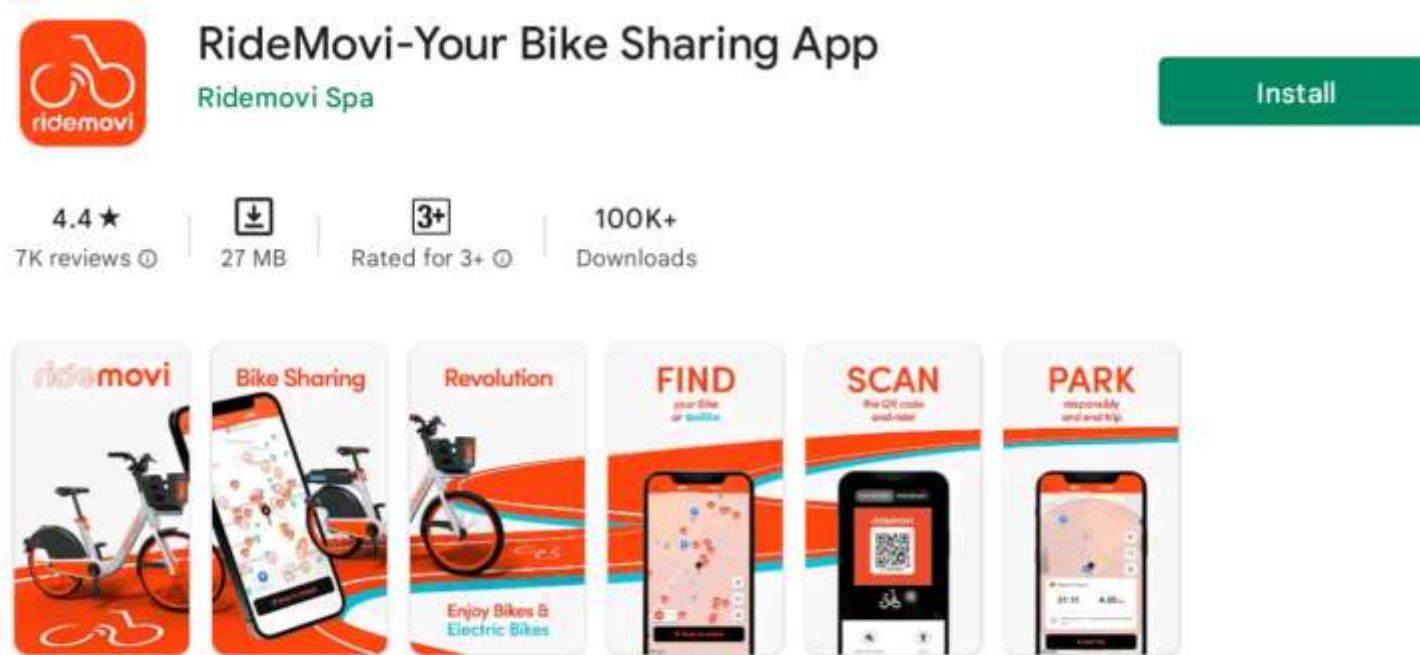
What are we getting??

Digitized Bike-Sharing Benefits



Where can we find Bike Sharing ??

AI

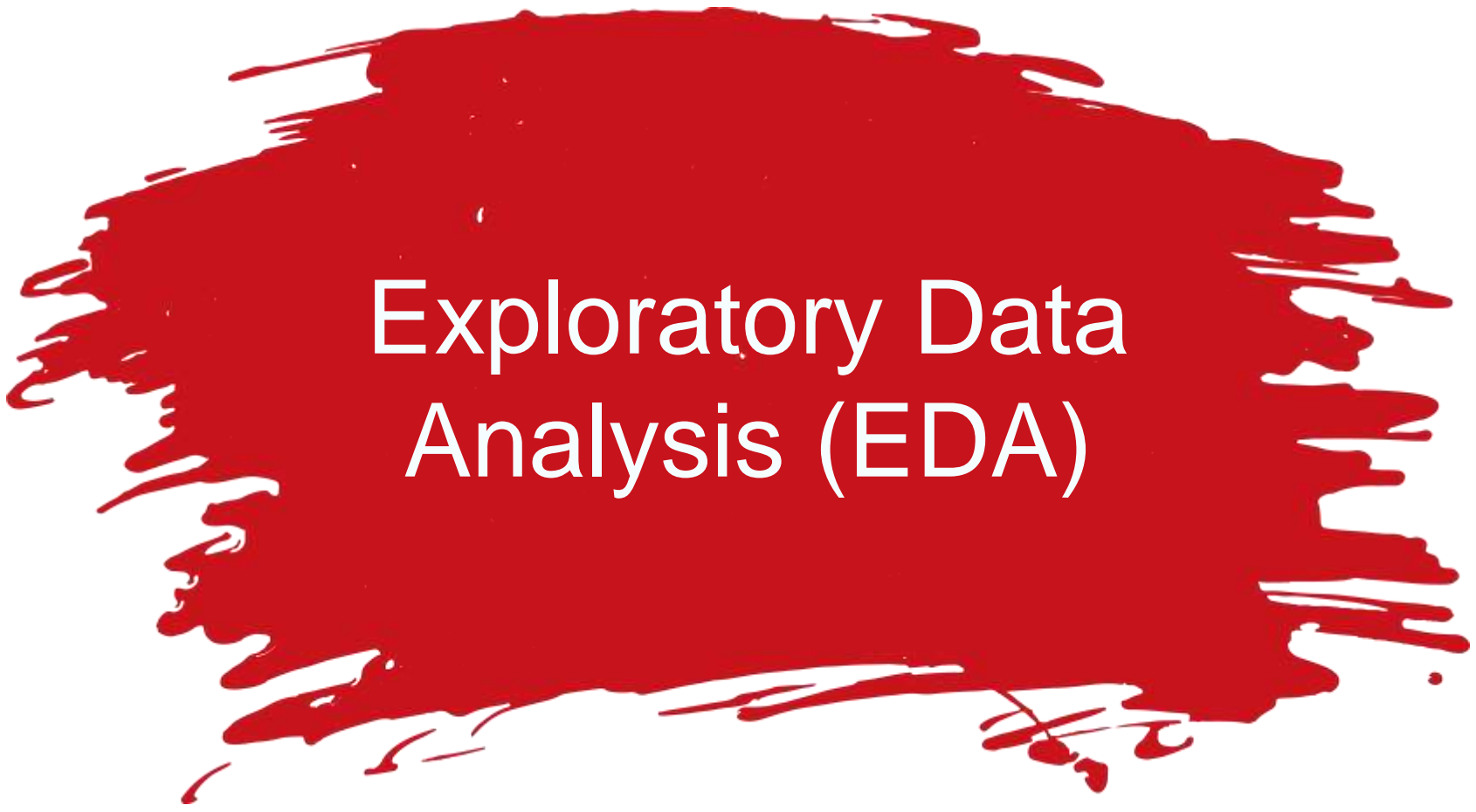


Can be found in “Application” and in “Website”

Data Overview:

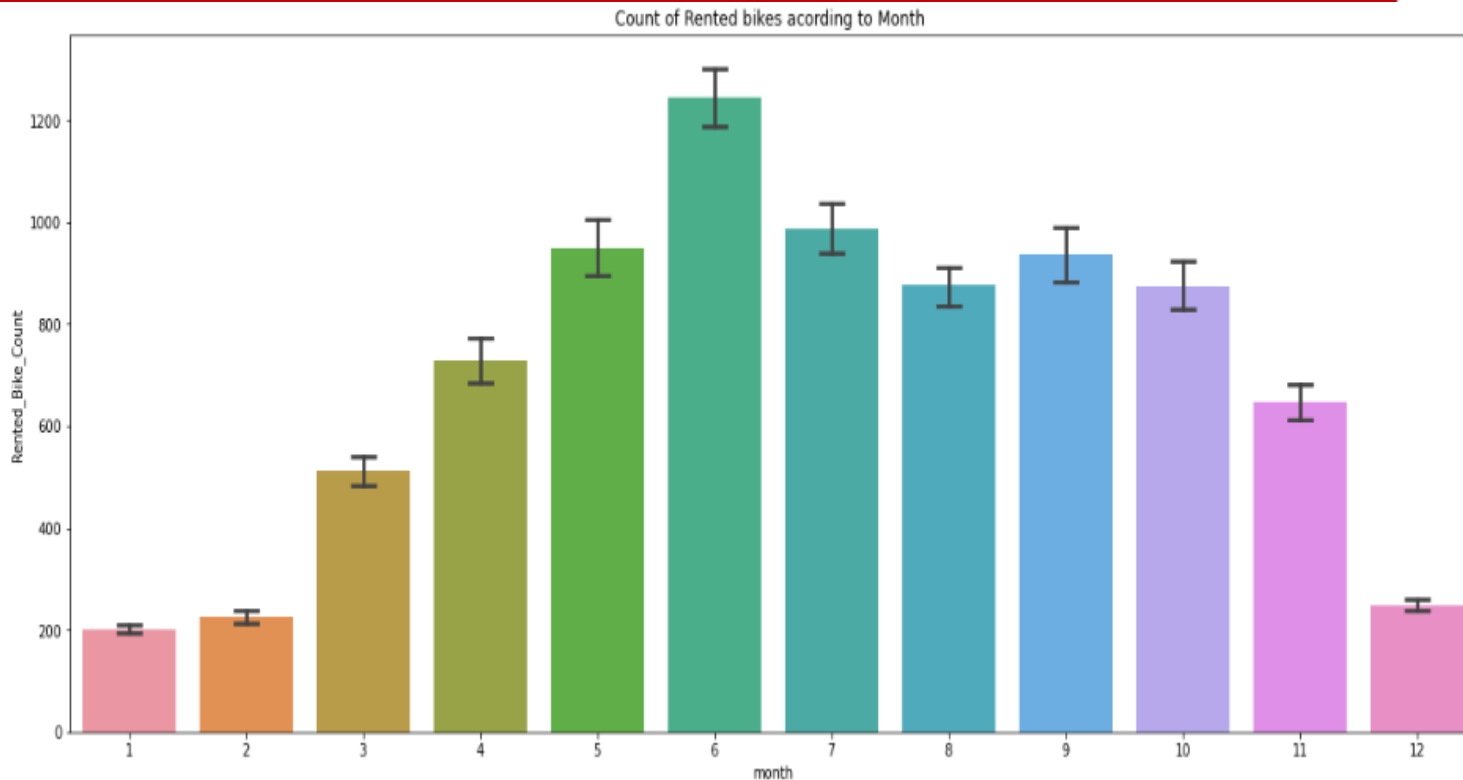
- Data set obtain from **“SeoulBikeData.csv”**
- Provided Feature Set
 - Rented Bike Count
 - Weather Conditions: Temperature(`C), Humidity(%), Wind Speed(m/s), Visibility(10m), Dew Point Temperature(`C), Solar Radiation(MJ/M2), Rainfall(mm) Snowfall(cm)
 - Day – Working Day or Not with Season
 - Time of The Day and Date

	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



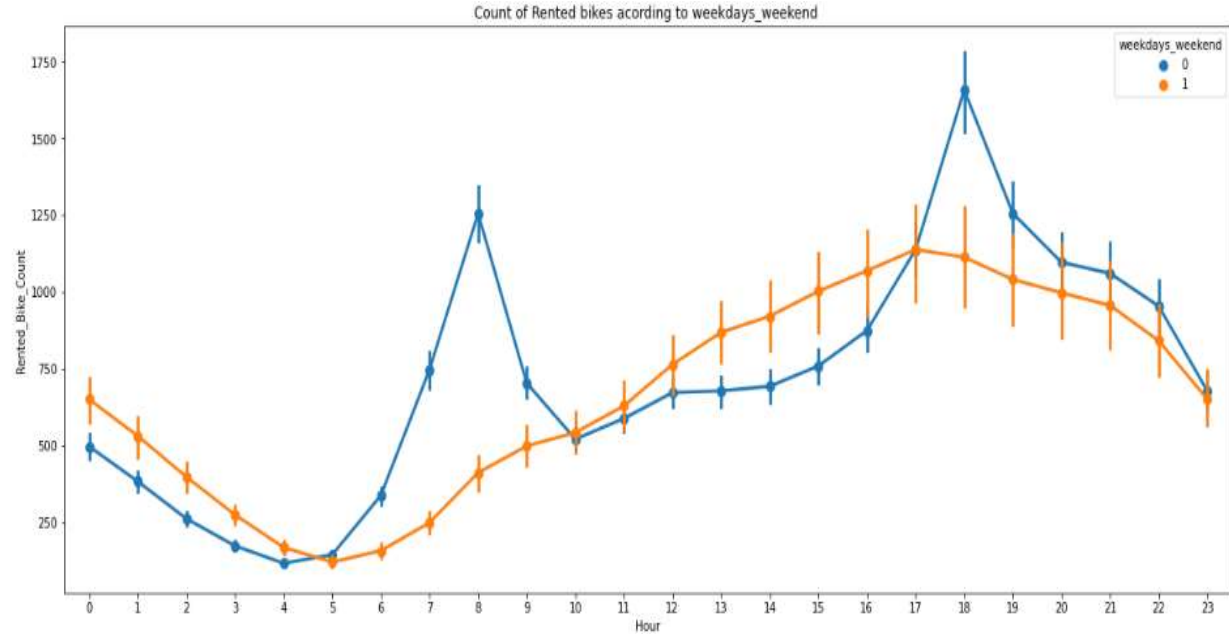
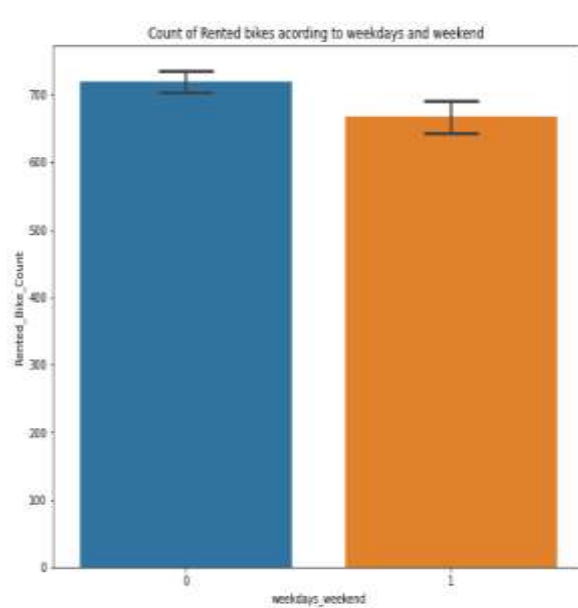
Exploratory Data Analysis (EDA)

Relationship: Rented Bike to Different Month



We can see from the data set most number of rented bike demands in between the month Of **“5-10”**, which belongs to **“summer season”**.

Relationship: Rented Bike to Weekdays/Weekend

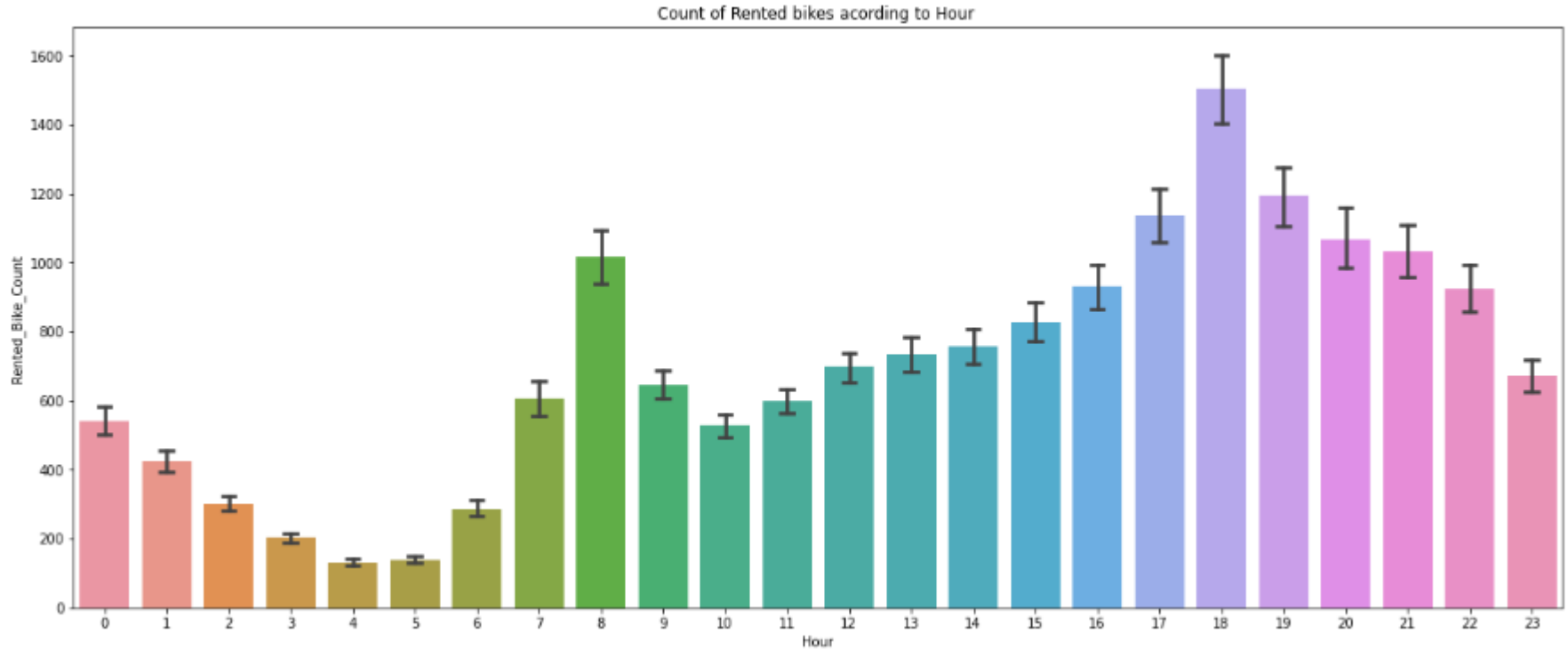


In the above two plots, “weekdays” is shown in “Blue” and “Weekend” is shown in “Orange”

Peak time for “Weekdays” is “7-9AM” and “5-7 PM”, key reason may be “Office Hour”

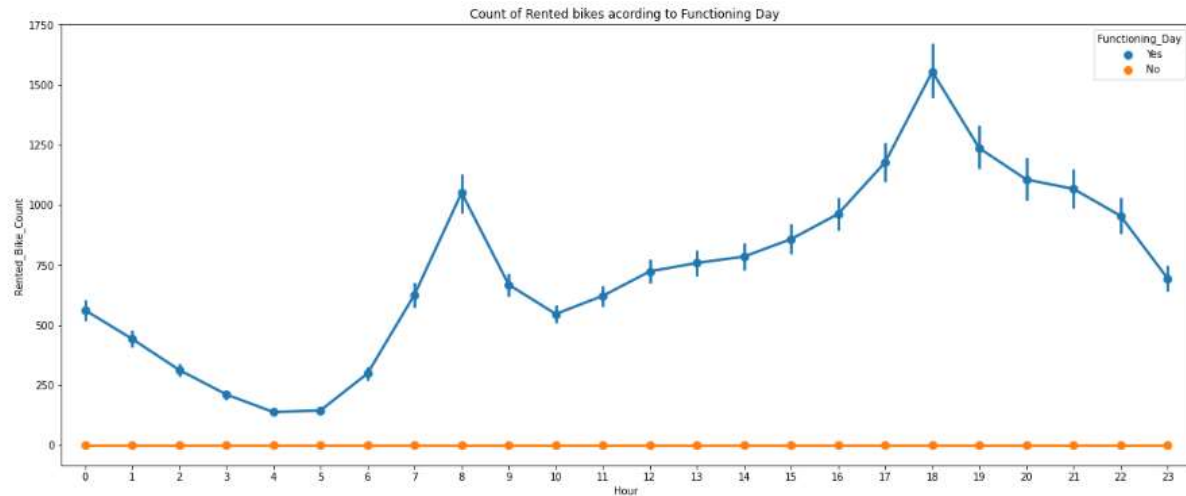
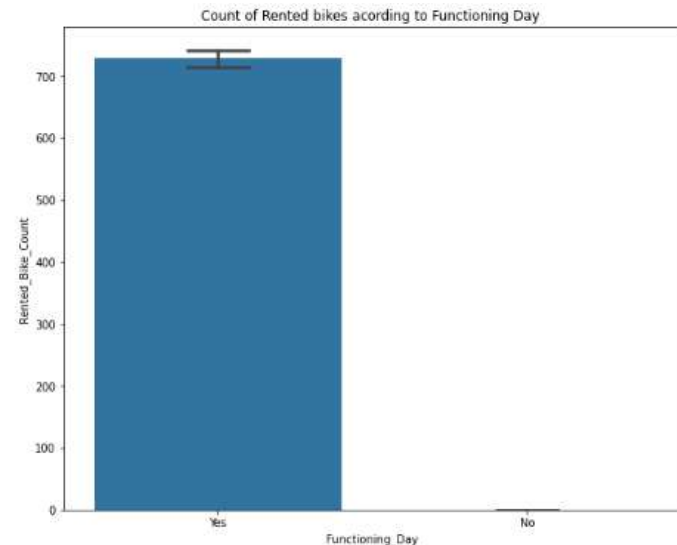
In “Weekend” demand is quite low slightly increase in between “4-8 PM”

Relationship: Rented Bike to Hours



Peak time is “7-9AM” and “5-7 PM”, key reason may be “Office Hour”

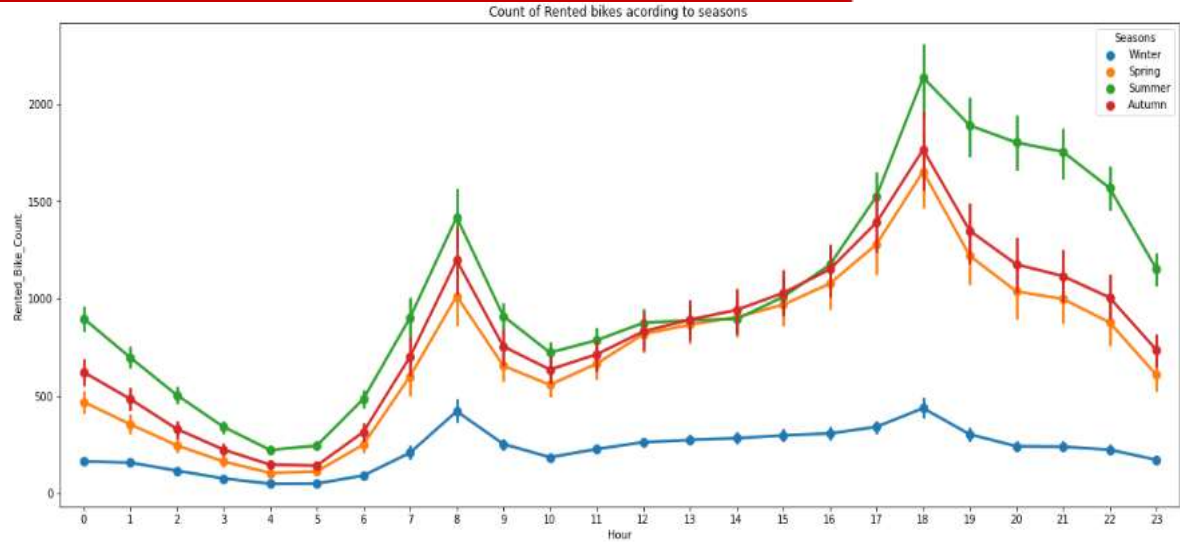
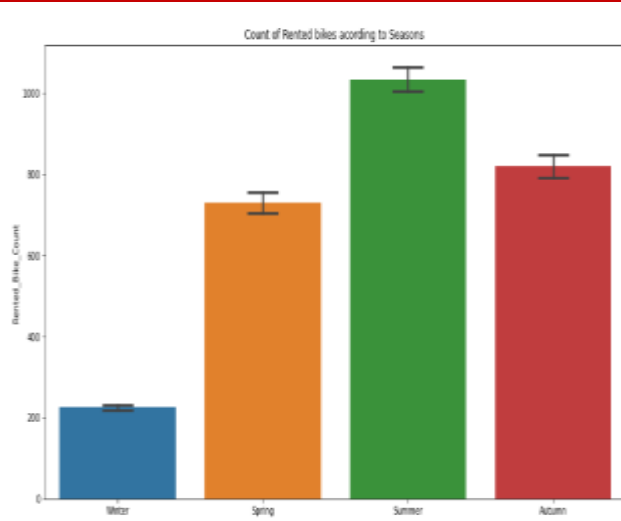
Relationship: Rented Bike to Functioning Day



In the above two plots, “Functioning Day” is shown in “Blue”
and “Non-Functioning Day” is shown in “Orange”

From the above bar plot and point plot, People **don't use** rented bike on “Non-Function Day”

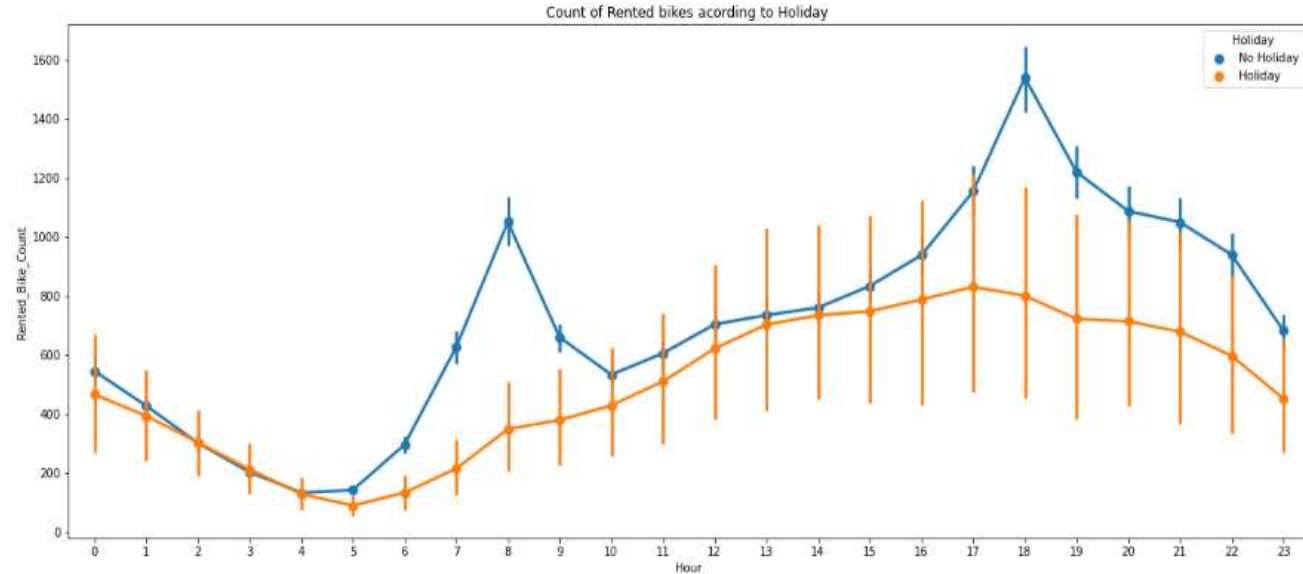
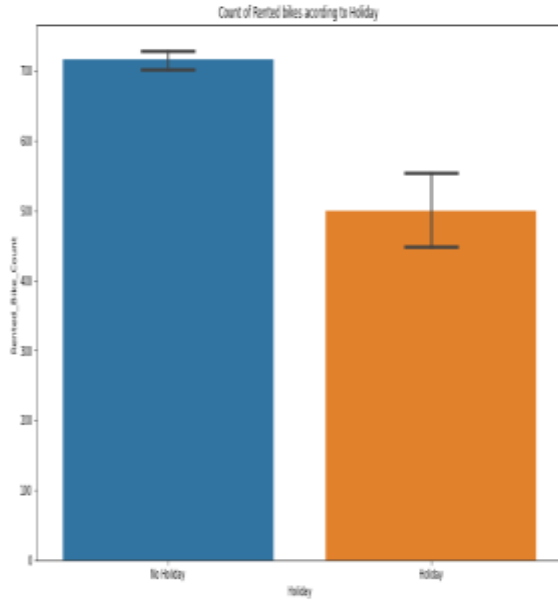
Relationship: Rented Bike to Seasons



In the above two plots, “Winter” is in “Blue”, “Spring” is in “Orange”
“Summer” is in “Green” and “Autumn” is in “Maroon”.

In “Summer” the “Rented Bike” is maximum and in “Winter” the “Rented Bike” is minimum
“Winter” has less “Rented Bike” due to “Snowfall”

Relationship: Rented Bike to Holiday



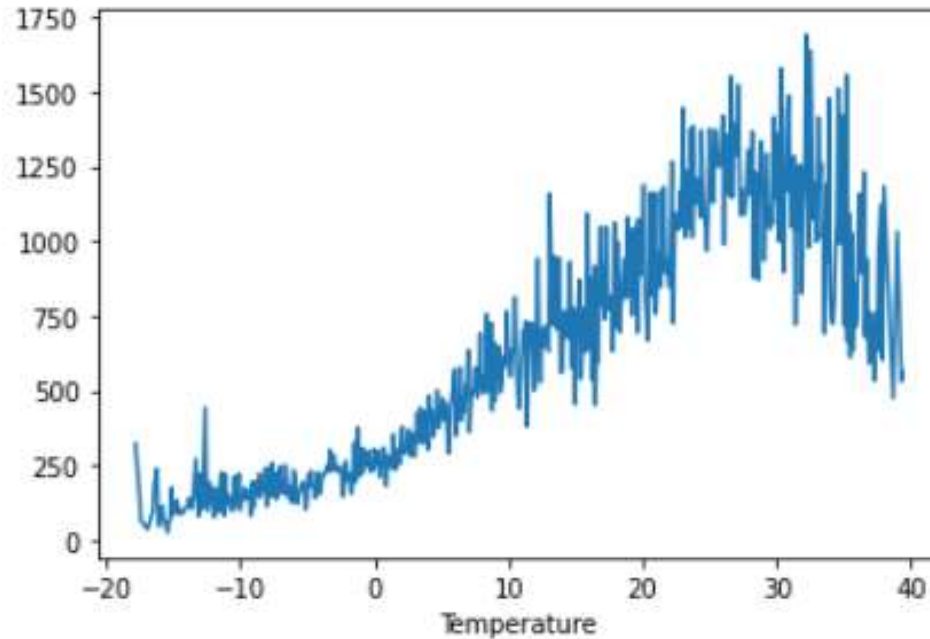
In the above two plots, “No Holiday” is shown in “Blue” and “Holiday” is shown in “Orange”

From the above bar plot and point plot, People use rented bike on “Holiday” in between “2-8 PM”



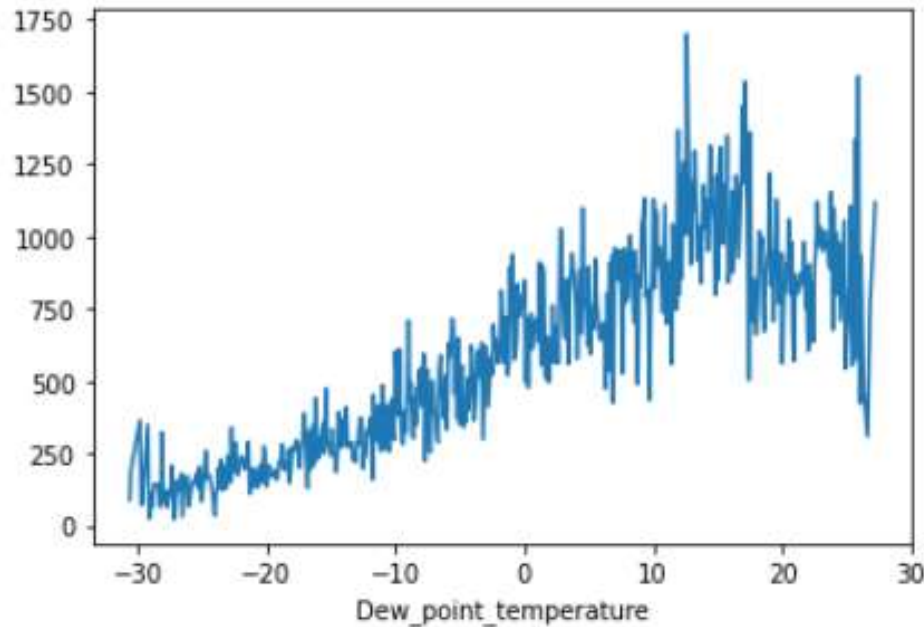
Numerical(Features)
v/s Rented Bike Counts

Relationship: Rented Bike to Temperature



People like to ride Bike when it is pretty hot around, like **“25°C to 30°C”** in average

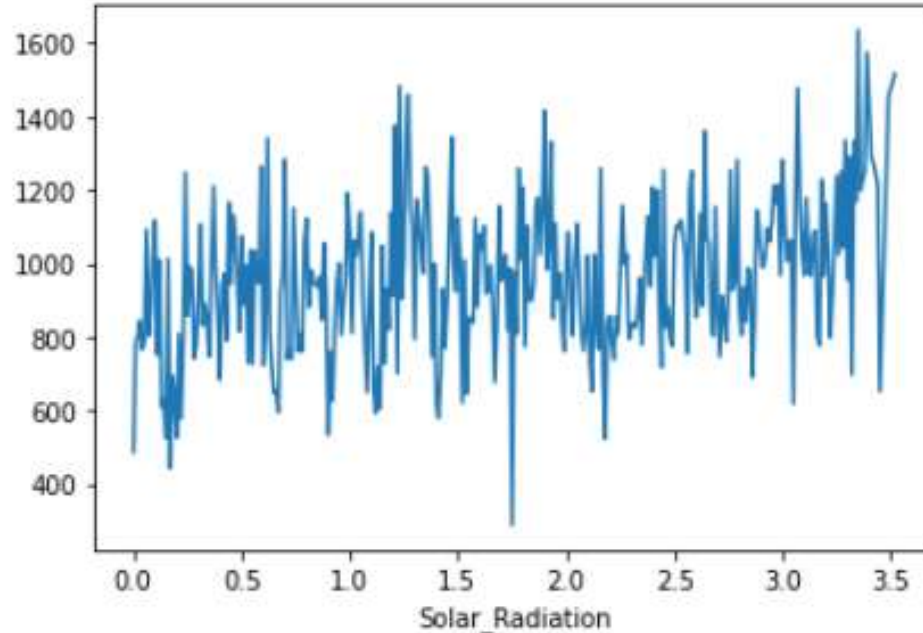
Relationship: Rented Bike to Dew Point Temperature



“Dew Point Temperature” works similar to “Temperature” with “Bike Sharing”
People like to ride Bike when it is pretty hot around, like “25°C to 30°C” in average

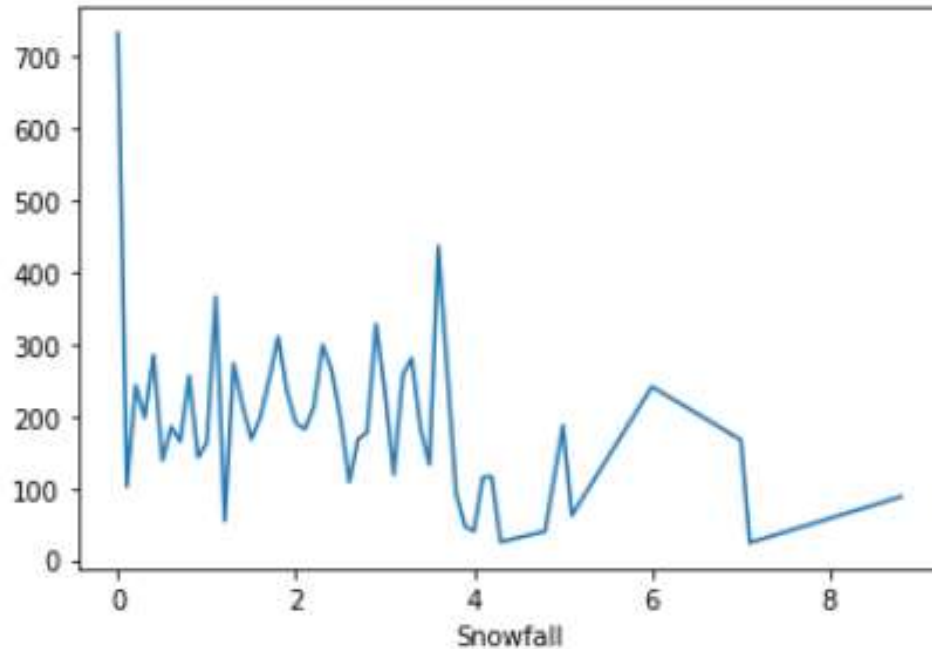
Relationship: Rented Bike to Solar Radiation

AI



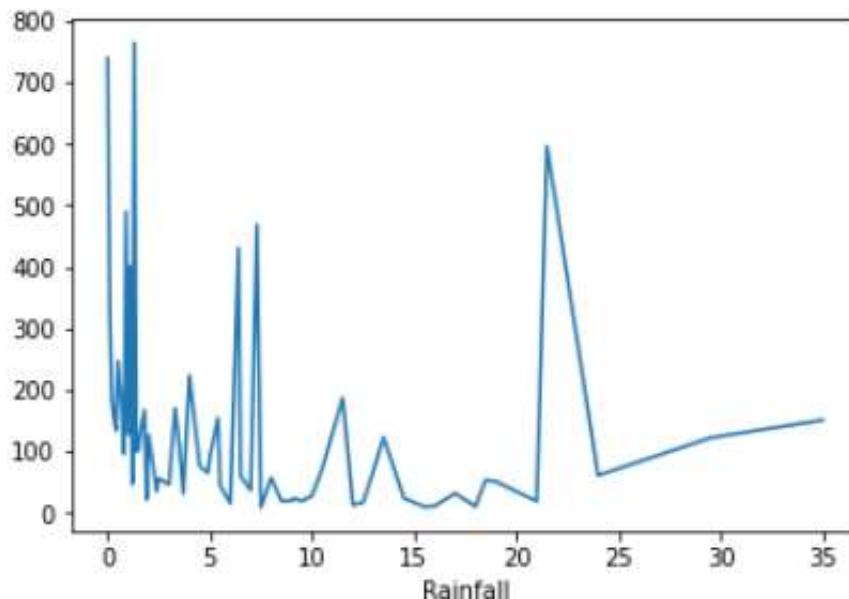
Bike rent is huge when the “Solar radiation” is in between “800-1200” in average

Relationship: Rented Bike to Snowfall



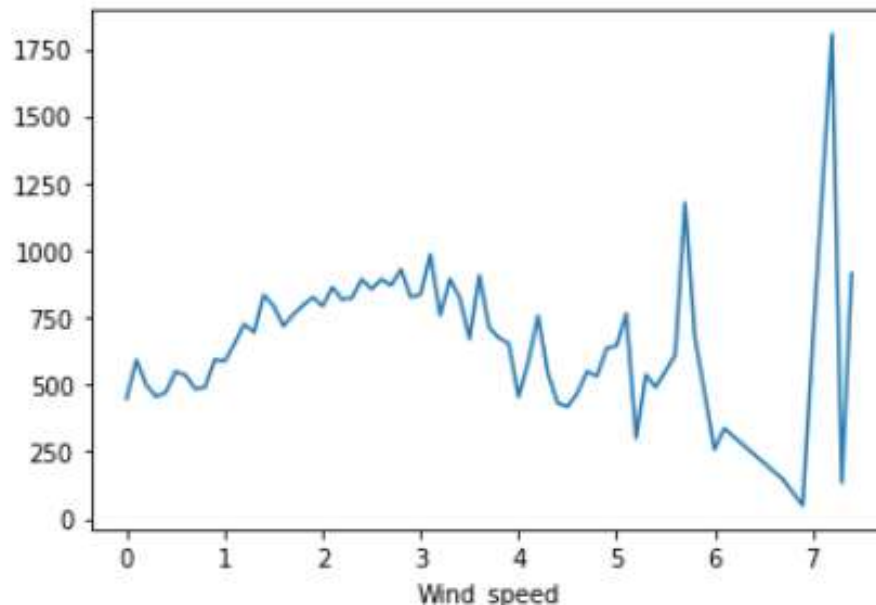
Bike rent is very low when “Snowfall” is more than “4 cm”

Relationship: Rented Bike to Rainfall



“Rainfall” does not make any difference to “Bike Rent”

Relationship: Rented Bike to Wind

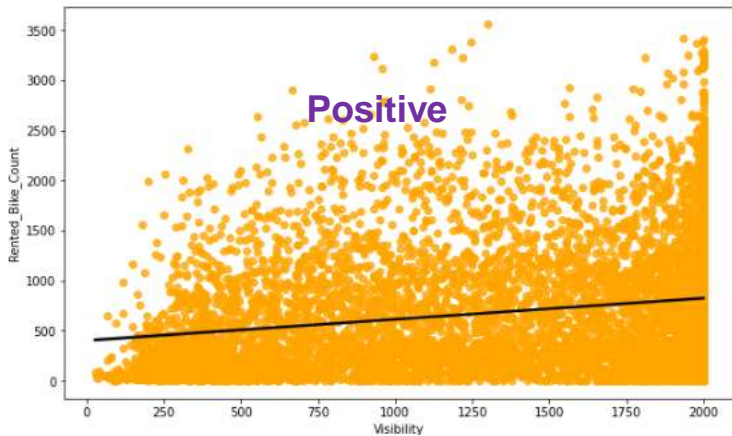
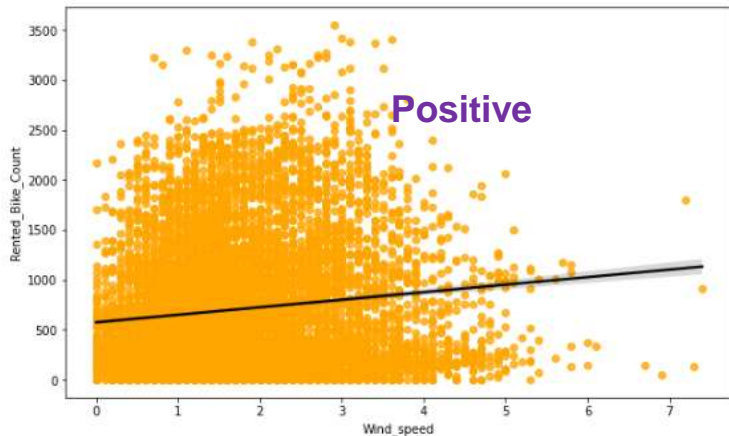
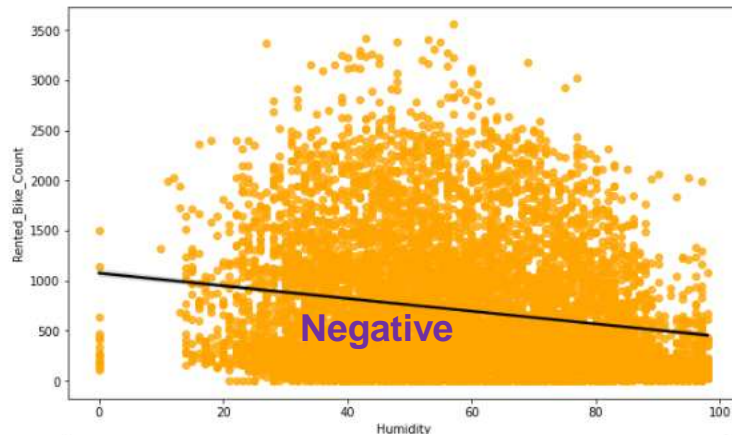
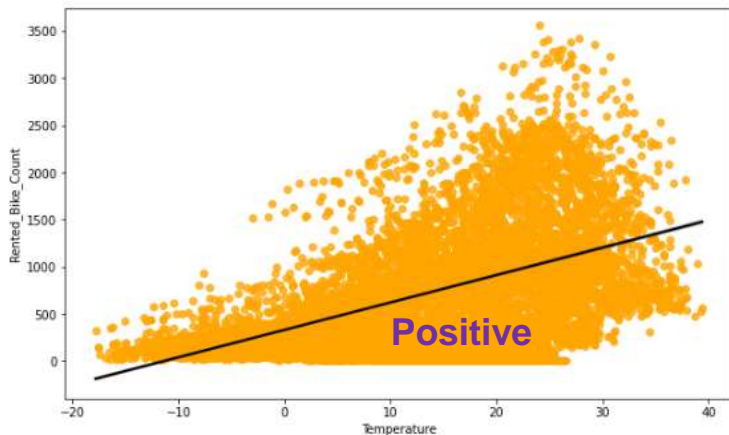


'Bike Rent' is uniformly distributed with "Wind", People love to ride in "Mild Windy" weather

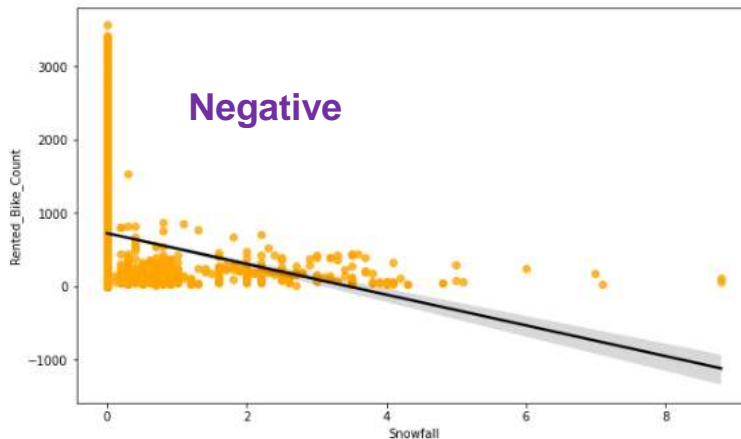
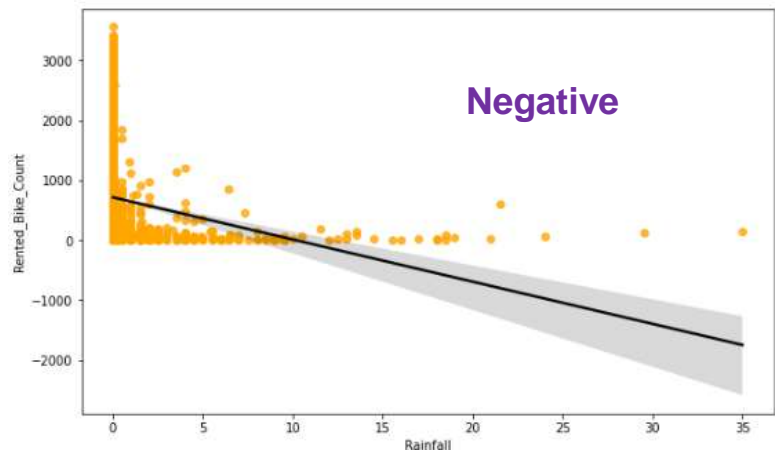
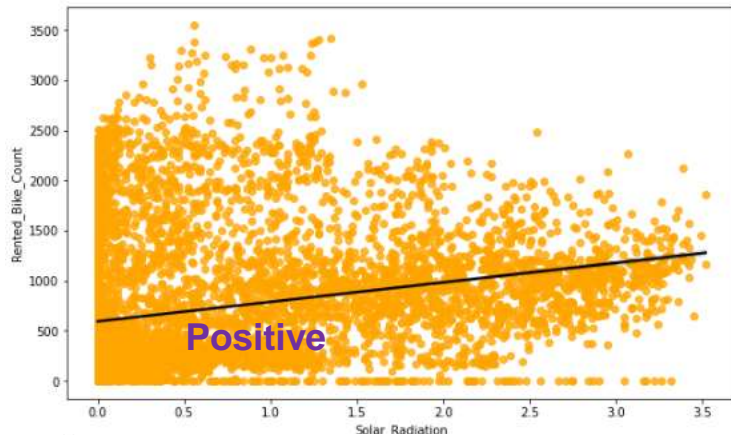
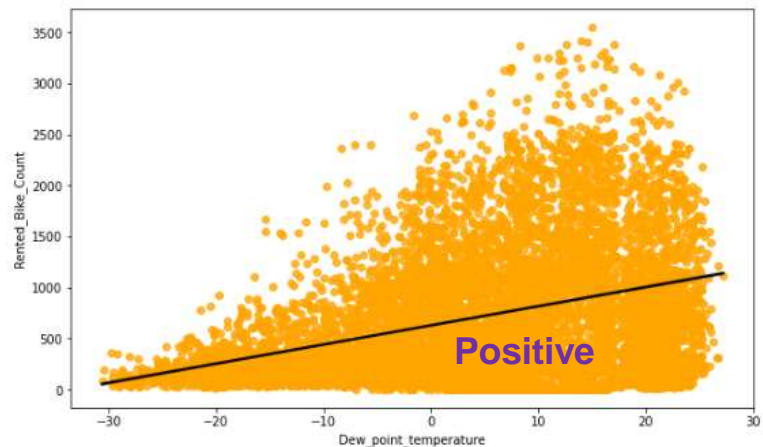


Regression Plot

Relationship: Bike to Temp, Humid, Wind & Visibility



Relationship: Bike to Dew, Solar, Rainfall & Snowfall **AI**



A large, dark, textured brushstroke background, resembling a thick application of black paint with visible bristles and some lighter areas, creating a sense of movement and depth.

Correlation Analysis

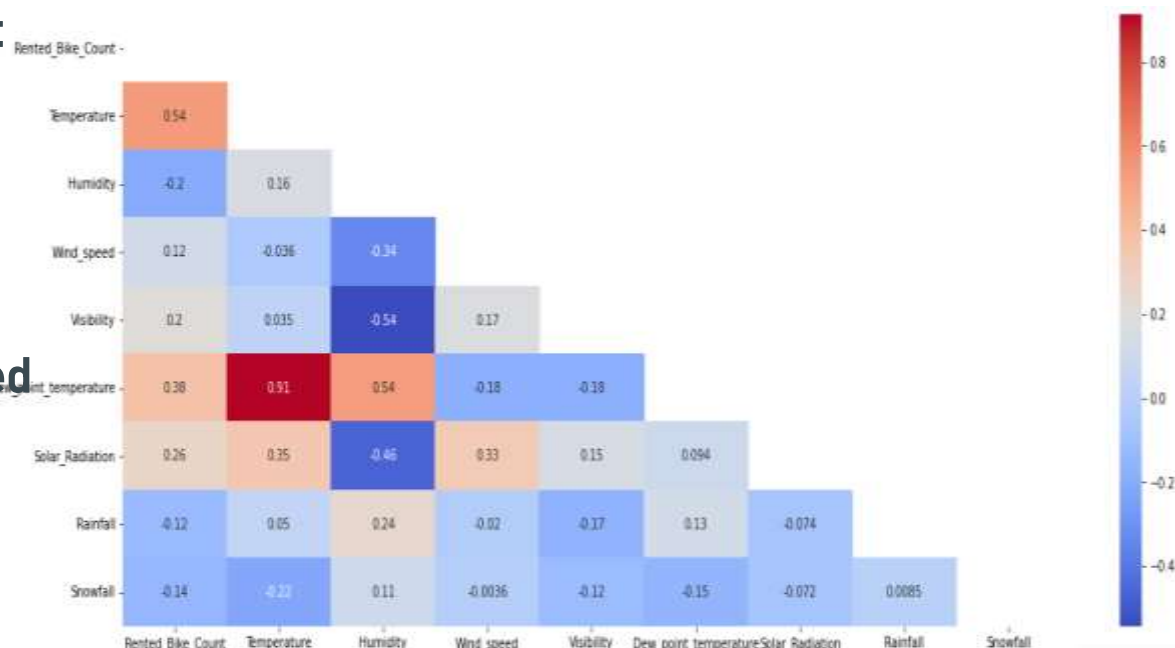
Correlation Analysis - Heatmap

- Most Positively Correlated variables to the Bike Rent are:

- Temperature
- Dew Point
- Solar Radiation

- Most Negatively Correlated variables to the Bike Rent are:

- Humidity
- Rainfall

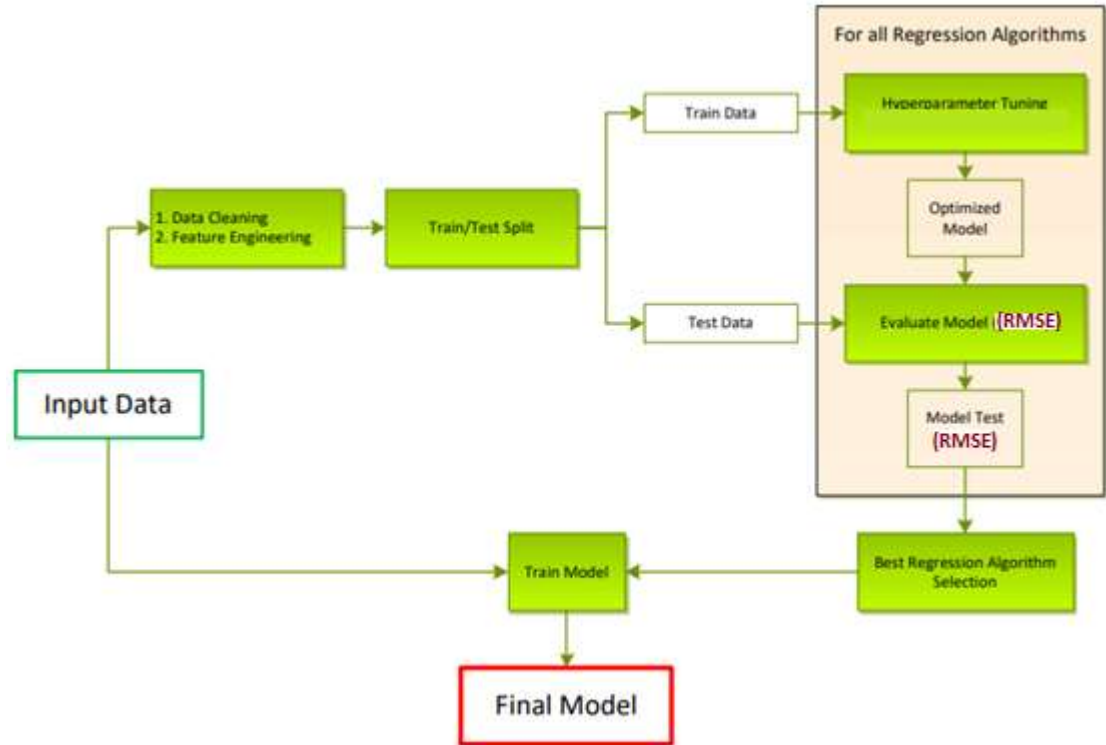




Modelling Overview

Modelling Overview

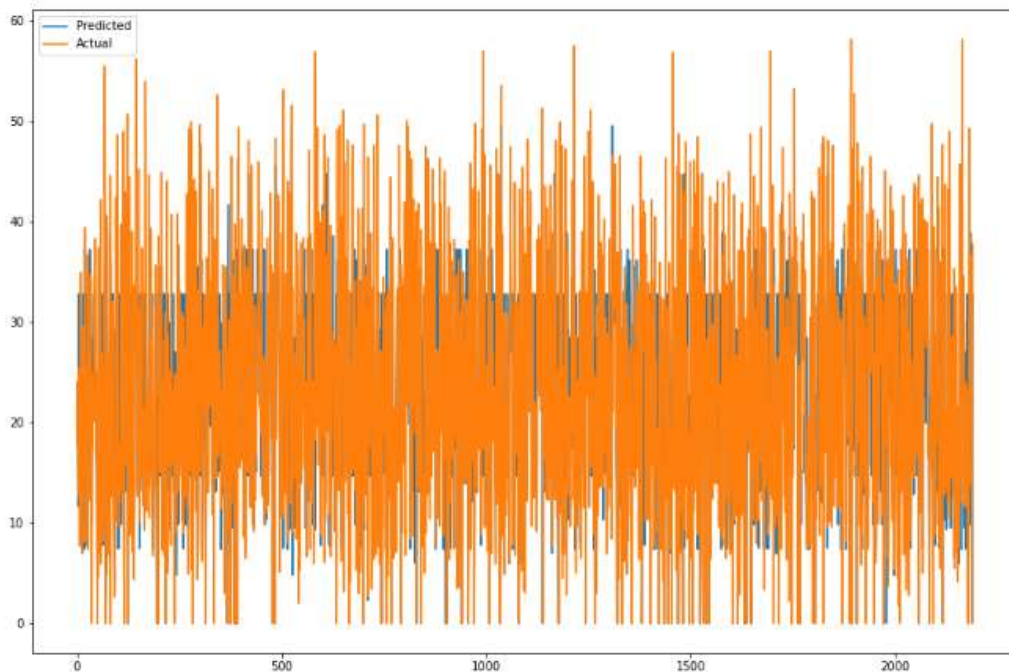
- Type: Supervised Learning using Gradient Boosting GridSearchCV
- Regression Problem: Possible Target Values $[0, \infty)$



Different Regression

- Linear
 - LASSO Regression
 - RIDGE Regression
 - ELASTIC NET Regression
- Multiple
 - DECISION Tree
- Stacking
 - RANDOM Forest

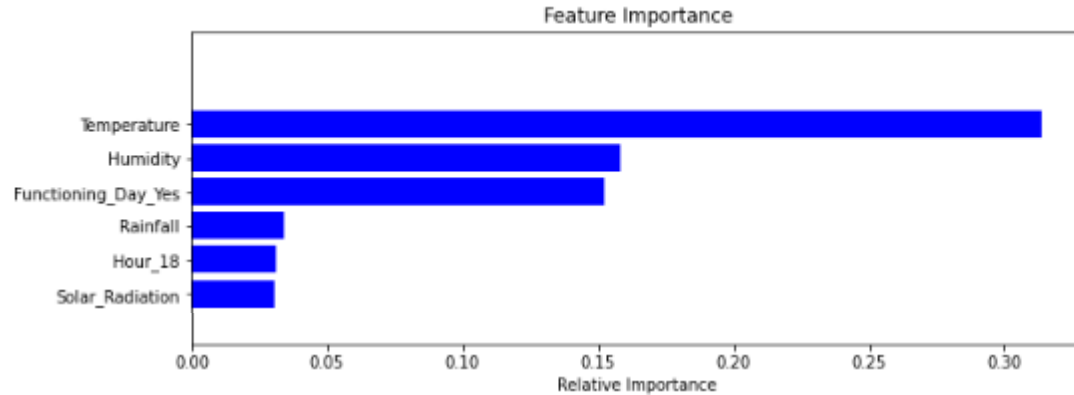
Out of all these Regression **“RANDOM Forest”** is performing the best with r2 value **“0.98”**





Feature Importance

Feature Importance



“Temperature” has the highest importance wrt others.

- Out of the five model tried RANDOM Forest performs best with “R2” value “0.98”.
- Temperature and Functioning Day hold the most importance in Prediction.
- We see two rental patterns across the days:
 - Working Day: Peak “7-9 PM” and “5-7 PM”.
 - Non Working Day: Peak “2-8 PM”.

Q & A