Capstone Project

Seoul Bike Sharing Demand Prediction

By

Ravi Kumar

Problem Statement

AI

Customers that frequently require bikes for a short time might borrow them from bike rental services. Typically, people rent bikes for a few hours so they may leisurely in that area. Currently Many big cities have adopted the use of rental bikes to improve mobility comfort. It is crucial to make the rental bikes accessible and available to the general public at the appropriate time since it reduces waiting. Eventually, maintaining a steady supply of rental bikes for the city emerges as a top priority. Predicting the number of bikes needed to maintain a steady supply of rental bikes at each hour's interval is essential.

Data Description

AI

The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

- **Date : year-month-day**
- **Rented Bike count - Count of bikes rented at each hour**
- **Hour - Hour of he day**
- **Temperature-Temperature in Celsius**
- **Humidity - %**
- **Windspeed - m/s**

Data Description

- 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 Hours), Fun(Functional hours)

Insights from the Problem and Data

AI

The following elements may have an impact on the supply and demand for rental bikes:

Weather Condition:

In warmer climates with moderate sunlight, low humidity, quiet winds, and dry roads, people prefer to ride bike. People may prefer to drive a car than ride a bike if the weather is poor.

Demand for Rental Bikes:

People who don't own bike and those who migrate to cities temporarily need a bike and an easy means of transportation. Those who possess their own bike may not choose to rent one.

Insights from the Problem and Data

AI

The following elements may have an impact on the supply and demand for rental bikes:

Availability of Rental Bikes:

Bike rental prices and availability at off-peak and rush hour times, as well as on weekdays and weekends.

Stable Supply of Rental Bike:

The weather conditions, demand, and supply sides may all have an impact on the stable supply of rental bikes.

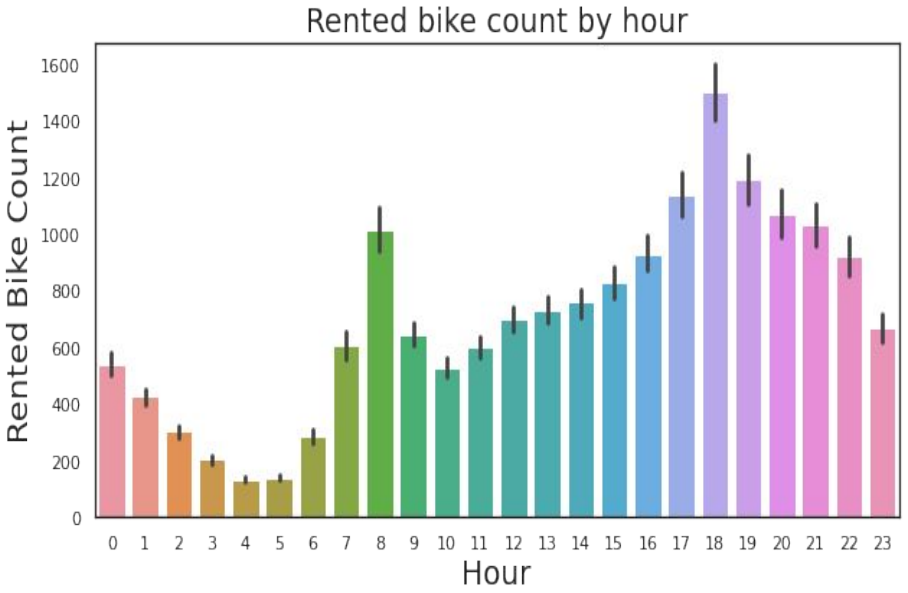
Split the Dataset into three Categories

AI

- Date and Time Variable
- Numerical Variables
- Categorical Variables

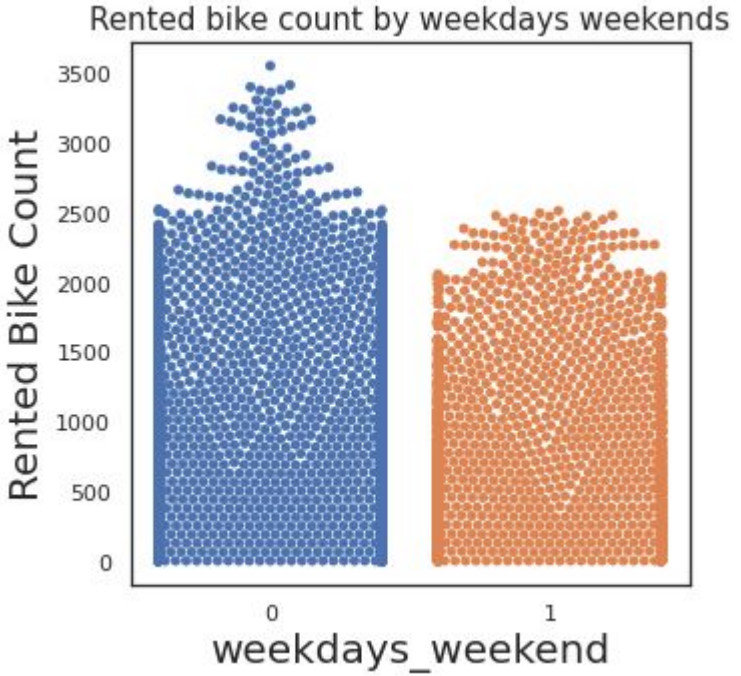
Date and Time Variable

- **Barplot is being used to count hourly rented bikes.**
- **From the plot up above, it is clear, Between 8:00 a.m. and 9:00 p.m., people prefer to ride bikes, which increases demand for bike rentals.**



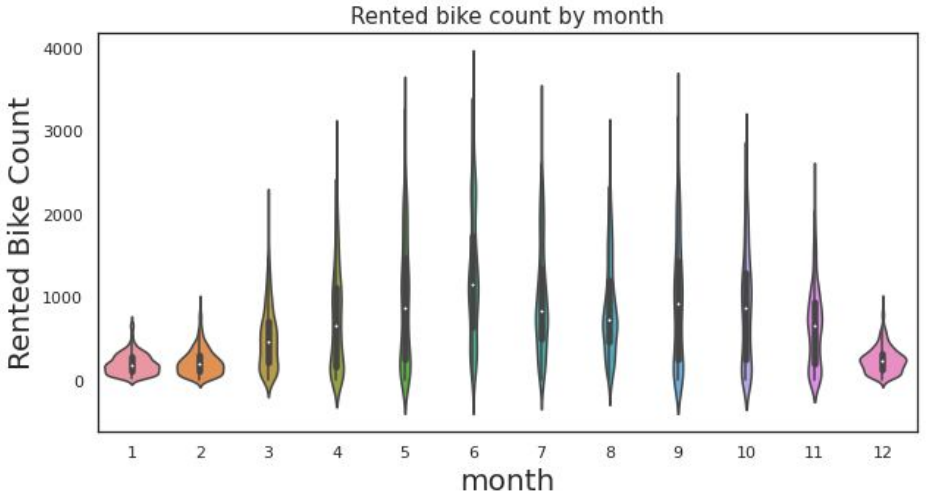
Date and Time Variable

- **Swarmplot is being used to plot weekdays_weekend rented bikes.**
- **From the plot above it is clear, weekdays rented more bikes than weekends.**



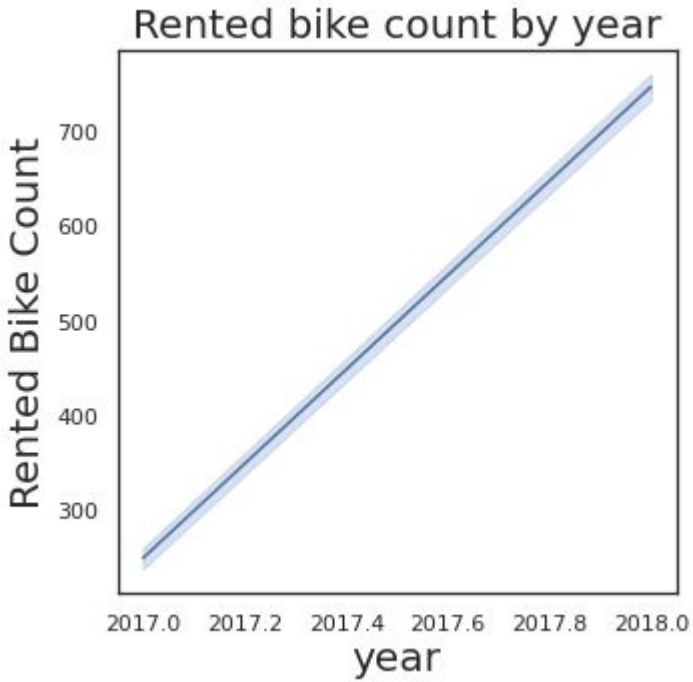
Date and Time Variable

- Violinplot is being used to plot monthly rented bikes.
- From the plot above it is clear, In the sixth month, people rented more bikes and took a ride



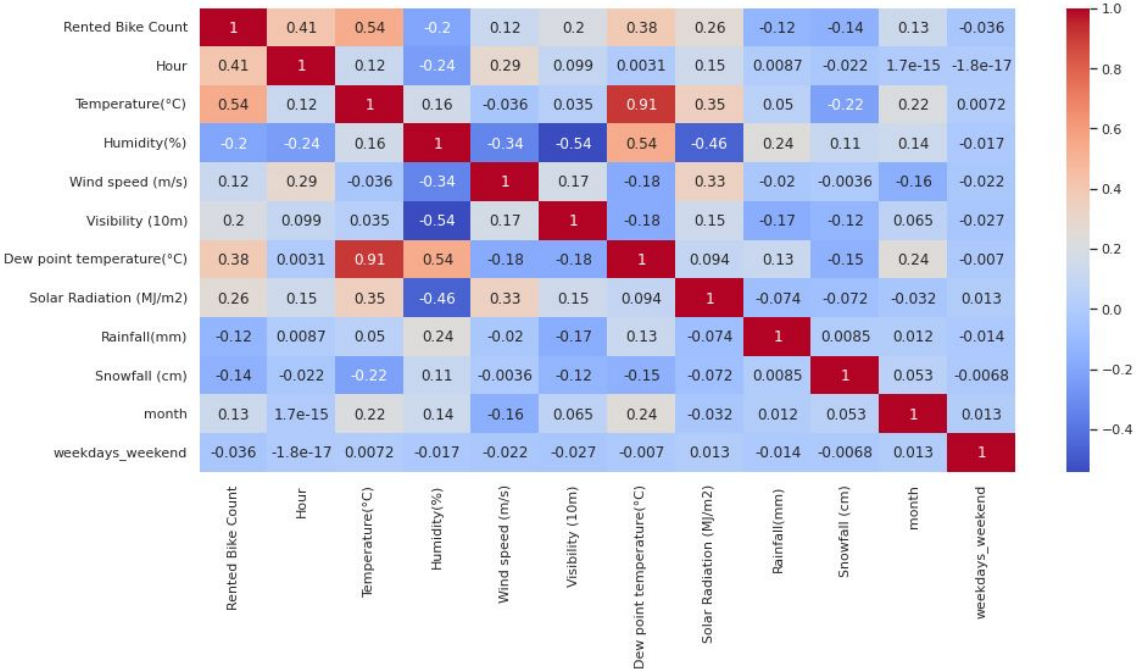
Date and Time Variable

- Lineplot is being used to plot month rented bikes.
- From the plot above it is clear, The number of bike rentals increased from 2017 to 2018 has the business grows.



Date Variable

Correlation Analysis:



Date Variable

AI

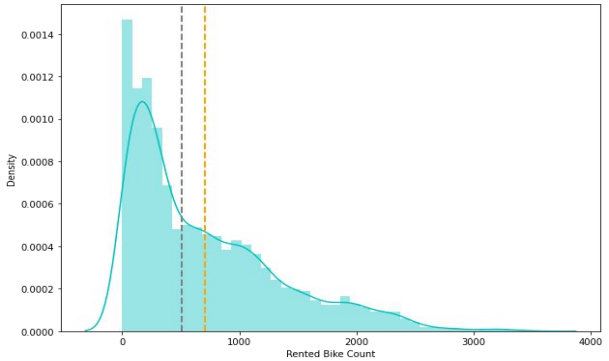
Correlation: You can only determine how much two variables are linearly reliant on one another by looking at their correlations.

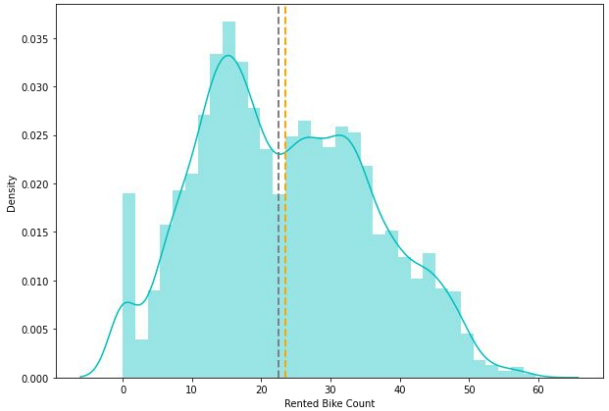
- **The variables are said to be positively connected if the value is positive.**
- **The variables are said to be negatively linked if the value is negative.**
- **It is said that there is no correlation between the variables if the value is zero or very near to it.**
- **Dew point temperature(C) and temperature is highly positive correlation, as temperature increases dew point temperature also increases. Hence, we remove this column, our analysis' results are unaffected.**

Analysis of Dependent Variable

Checking the skewness of the rental bike count distribution

- **We use distplot to check distribution for the output variables in the dataset and if it is skewed, we should do transformation to make it normal distributed.**
- **We can determine the direction of the skewness. The tail of a distribution curve has a longer right side when there is a positive skew and if it is left side, it is negatively skewed.**
- **We reduce the skewness, by square root.**

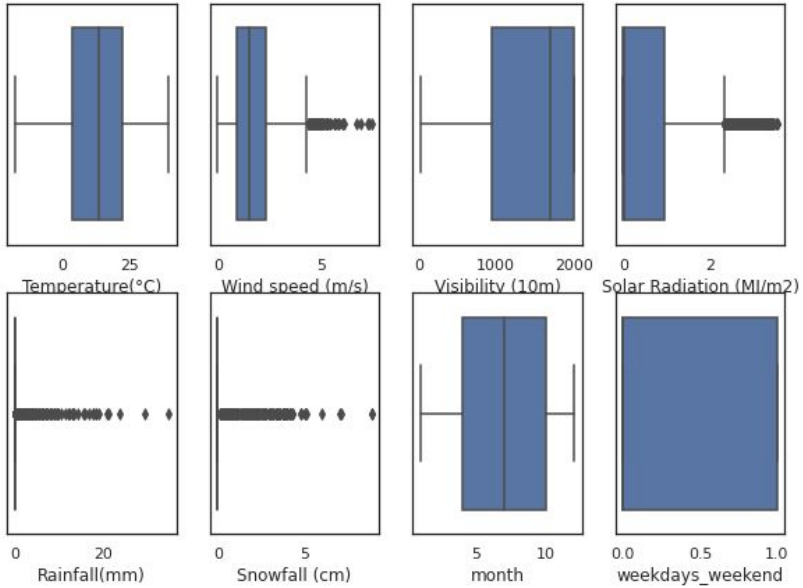




Numerical Variables

Handling Outliers:

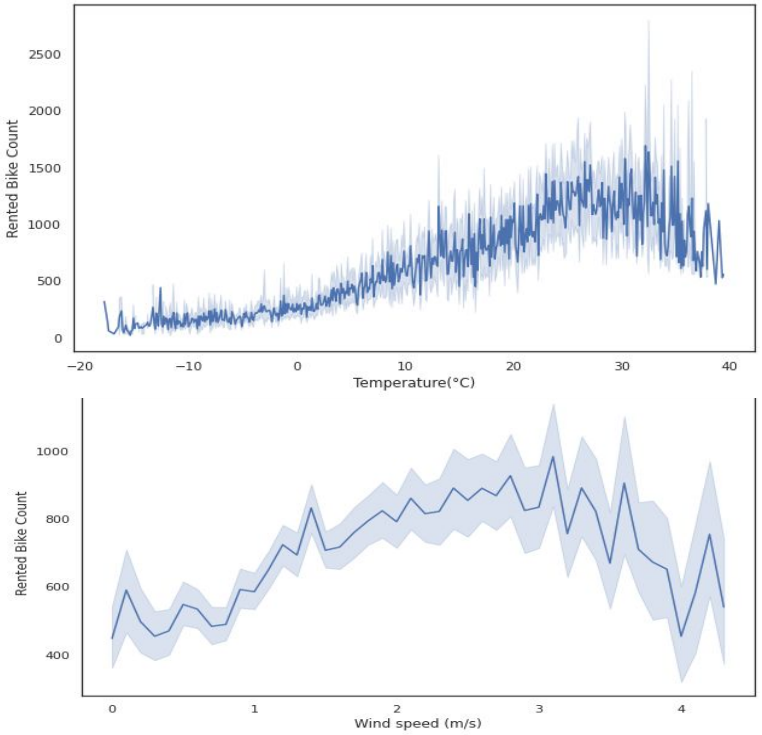
- For identifying outliers, scatter plots and box plots are the most used visualisation techniques.Here we use boxplot.
- Here we can see that the columns that contain outliers are Rainfall, Snowfall, Windspeed and Solar Radiation
- We removed the outliers through IQR



Numerical Variable

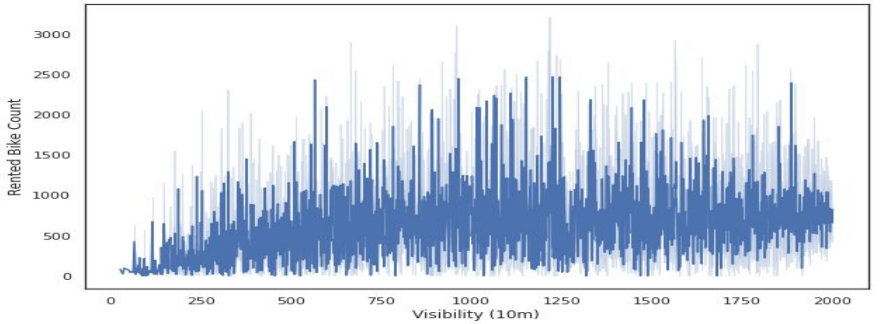
From the plot between the numerical variable and the dependent variable shown:

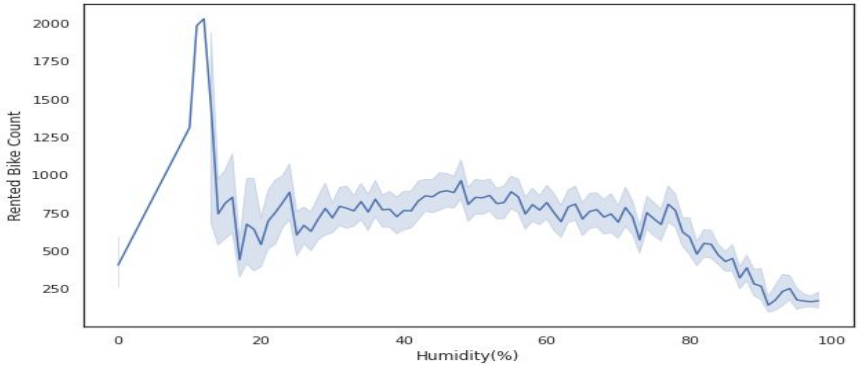
- **People enjoy riding bike in hot weather between 23° and 30°C**
- **As the wind speed is increases between 2 and 3 (m/s), people prefer to hire and ride bike**



Numerical Variable

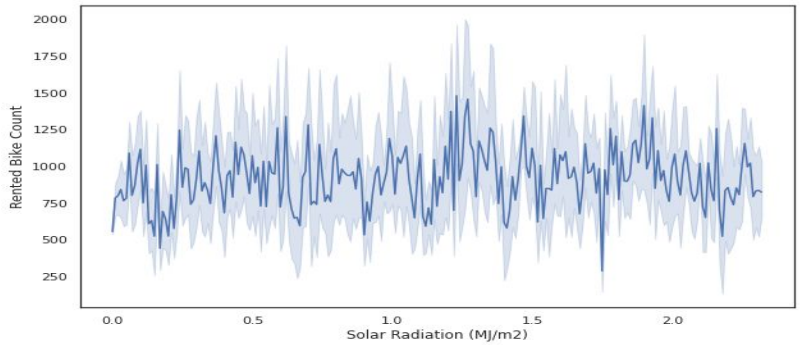
- People favour biking in areas with great visibility.
- People prefer to ride bikes where humidity below 20%

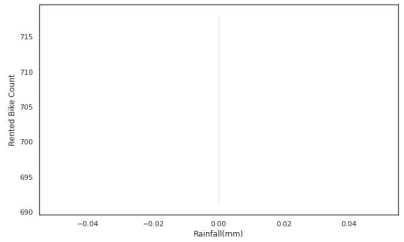


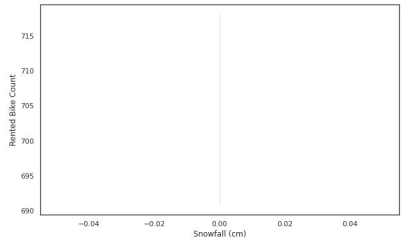


Numerical Variable

- When there is solar radiation, there are a lot of leased bikes—the rental counter is approximately 1000
- People prefer to hire and ride bikes when there is no rainfall and snowfall

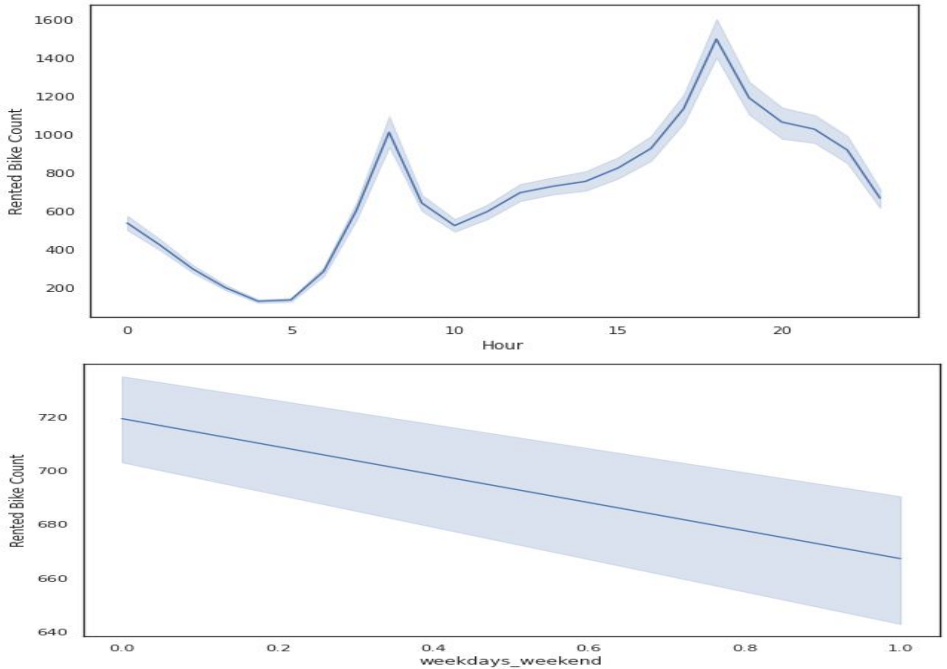






Numerical Variable

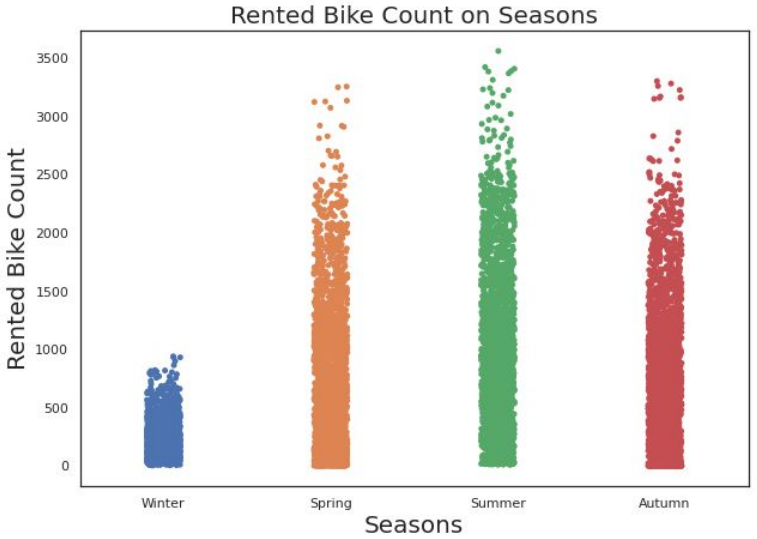
- In the sixth month, people rented more bikes and took a ride
- Weekdays are the preferred days for people to hire and ride bike than weekends



Categorical Variables

From the plot between the numerical variable and the dependent variable shown:

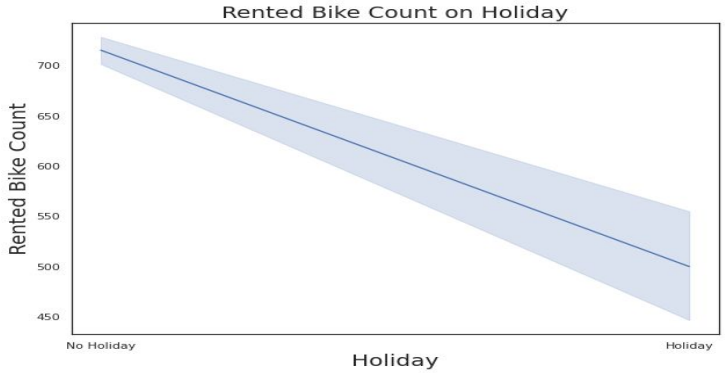
- **Summer is the season with the most demand for bi,kes followed by Autumn, Spring, and Winter.**
- **Bike rentals are highest during the summer and are lowest during the winter.**

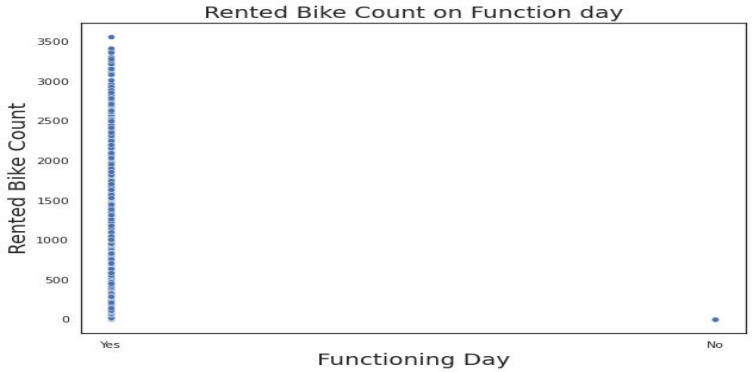


Categorical Variables

According to the figure:

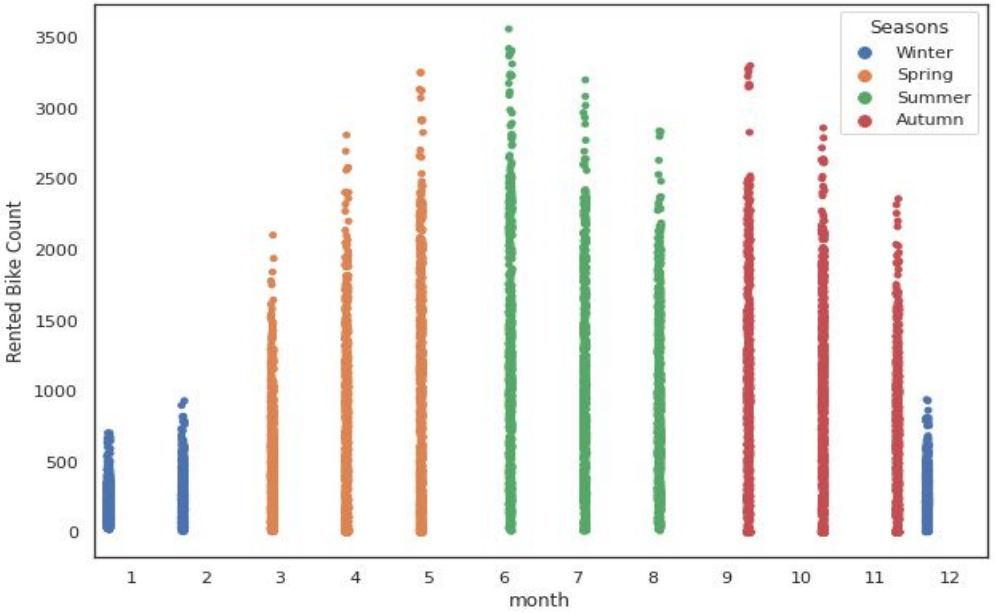
- **More people hired bikes during non-holiday times than during holiday times**
- **People leased bikes when the day was operational, but they didn't rent any on non-operational days.**





Categorical Variables

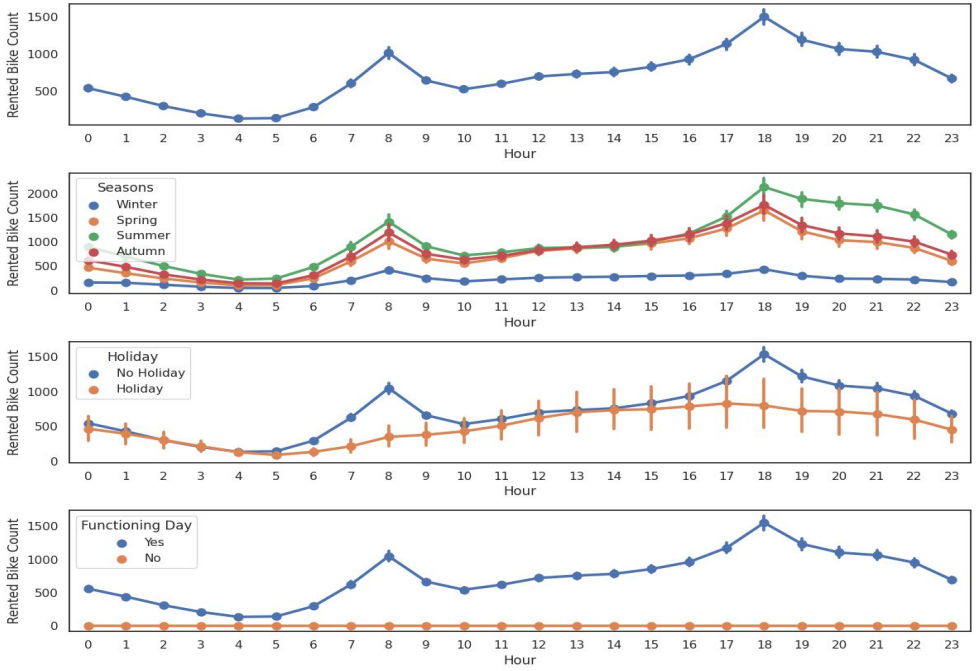
- **Summer Season is the peak season where most people hired a bike. There is a lot of demand for bikes in summer season than compare to other seasons**



Pointplot Representation

From the graph:

In the summer, when there are no holidays and the day is still operating, people choose to rent and ride more bikes at that time.Bikes are in great demand and good supply at 18 hours during the summer.



Model Performance

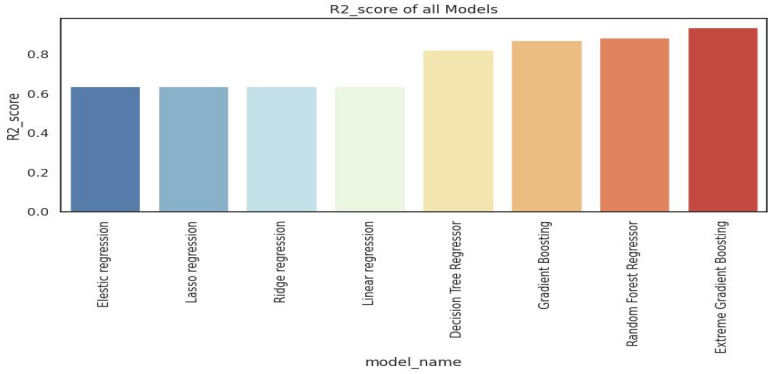
AI

- Linear regression
- Lasso regression
- Ridge regression
- Elastic regression
- Decision Tree
- Random Forest
- Gradient Boosting
- Extreme Gradient Boosting

Result of all models

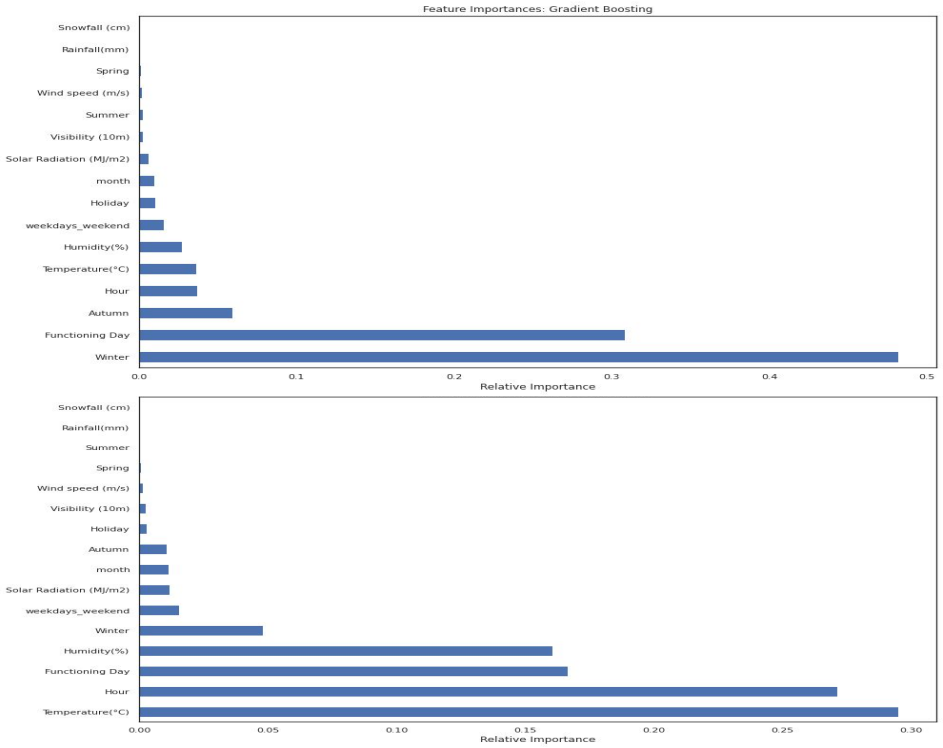
- Comparing the result of all the model, XGBoost and Random Forest gives the best R2 score.
- We can deploy this models.

model_name	MSE_score	RMSE_score	MAE_score	R2_score	Adjusted_R2_score
Linear regression	56.945815	7.546245	5.859794	0.639507	0.636852
Lasso regression	56.969414	7.547809	5.861282	0.639357	0.636702
Ridge regression	56.946640	7.546300	5.859868	0.639502	0.636847
Elastic regression	56.983494	7.548741	5.862294	0.639268	0.636612
Decision Tree Regressor	27.915300	5.283493	3.542101	0.823283	0.821982
Random Forest Regressor	18.127895	4.257687	2.957885	0.885242	0.884397
Gradient Boosting	20.139894	4.487749	3.185413	0.872505	0.871566
Extreme Gradient Boosting	10.165641	3.188360	1.997298	0.935647	0.935173



Feature Importance

- The most crucial feature taken into account when utilising a XGboost to make predictions is winter and Functioning day
- The most crucial feature taken into account when utilising a random forest to make predictions is temperature and hour



Summary and Conclusion:

AI

- In the morning and evening, many people rent bikes. The 4th and 18th hours are when the most bikes are leased when the weather is warm (20° to 28°C), the wind speed is between 2 and 3 m/s, and there is adequate visibility. More people hired bikes in the sixth months. The most bike are rented during non-holiday times.
- With a r2 score of 93%, Extreme Gradient Boosting is the model that performs the best.
- Therefore, there will be high supply and demand for bikes when there is mild sunlight, low humidity, no wind, and dry roads. People choose to hire and ride bikes.There will be a lot of demand and supply on both sides when the weather is pleasant.

Thank you