Capstone Project

Bike Sharing Demand Prediction

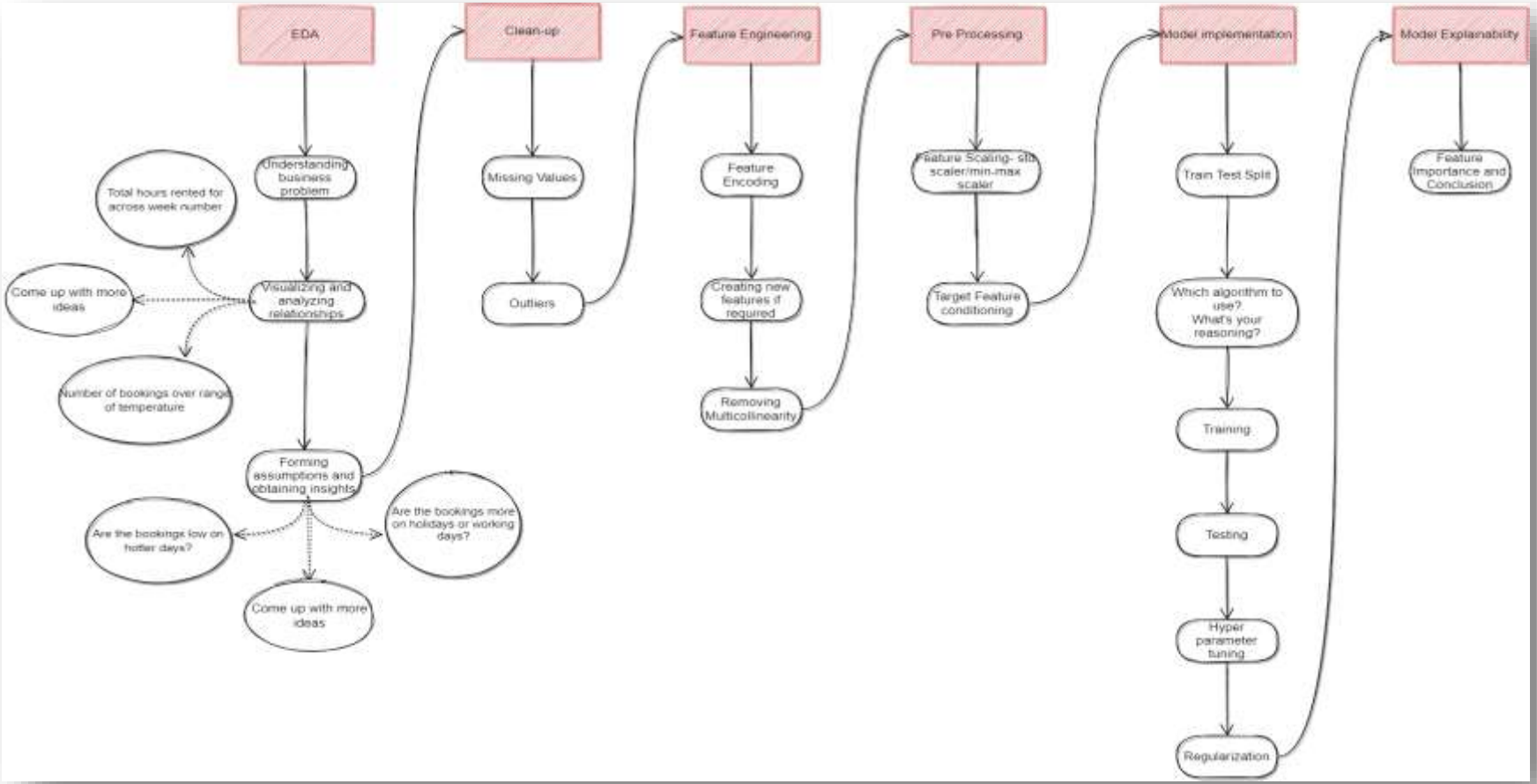
By:

Shubham Chougule
Akashada Phunde
Bhojraj Jadhav
Shruti Jain
Suraj Singh

Point of Discussion

1. Introduction.
2. Problem statement.
3. Dataset.
4. Feature engineering.
5. Data preprocessing and exploration.
6. Data Transformation and visualization.
7. Model Building :- Linear Regression, Ridge Regression , Decision Tree, Random Forest.
8. Conclusion.

Roadmap for the project



Introduction

- ✓ Currently Demand of rental bikes are increasing throughout the world and specially in fast growing cities .
- ✓ It helps individuals to Move easily without waiting for any public transport and it helps reduce dependency of public transport and reduces pollution.

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.

Data Collection And Understanding

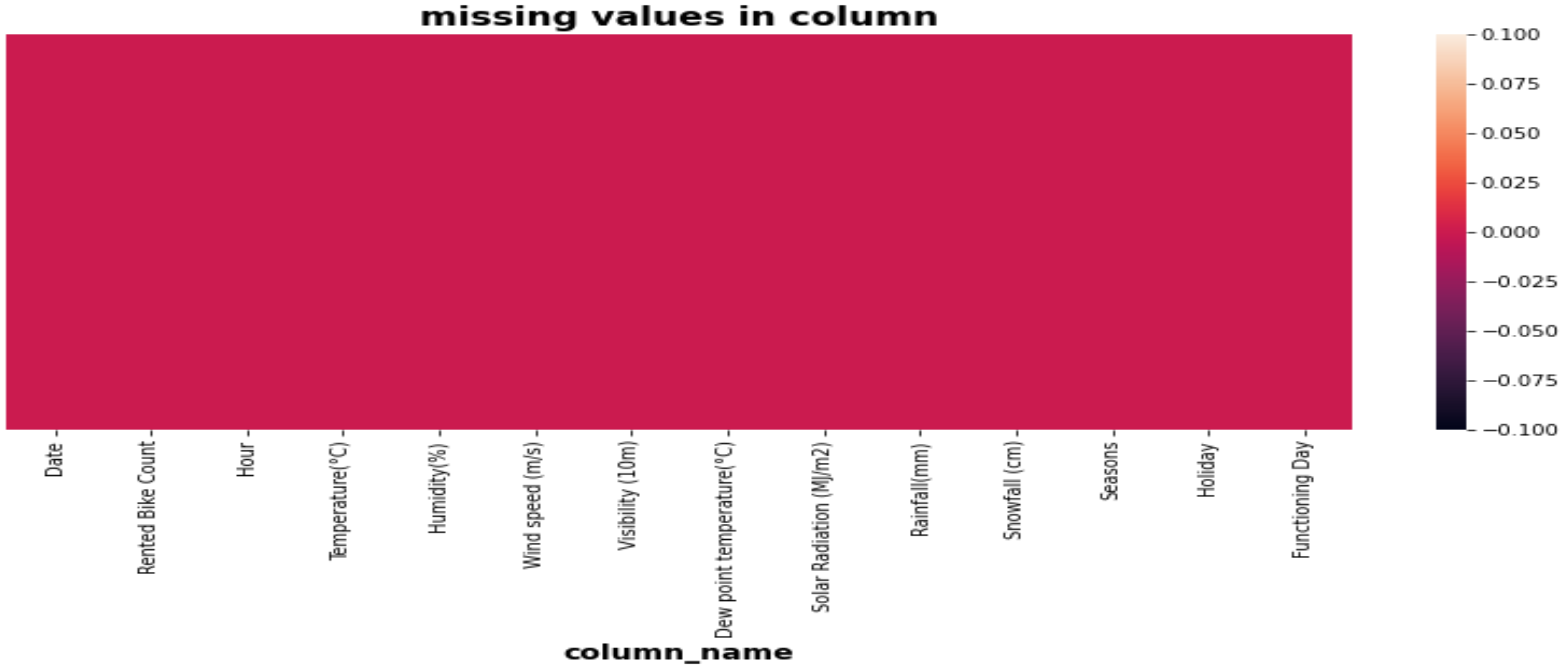
Columns Description:

1) **Date**-It shows the dates on which bike were book on rent.
2) **Rented bike count** - It shows the count of number of bikes were booked on the basis of per hour.
3) **Hour** - It shows the format in 24 hours .
4) **Temperature(°C)** -Temperature in Celsius .
5) **Humidity(%)** -Relative humidity.
6) **Wind speed (m/s)** -Wind speed in meter per second.

Data Collection And Understanding

7. **Visibility (10m)** -Visibility measure 10 meter per unit.
8. **Dew point temperature(°C)** -Dew point temperature measure.
9. **Solar Radiation (MJ/m2)** -Solar radiation measure.
10. **Rainfall(mm)** -Rainfall in mm.
11. **Snowfall (cm)** -Snowfall in cm.
12. **Seasons** -It has shown 1=spring,2=summer,3=fall,4=winter.
13. **Holidays** -Whether a holiday or not.
14. **Functioning Day** -whether a functioning day or not.

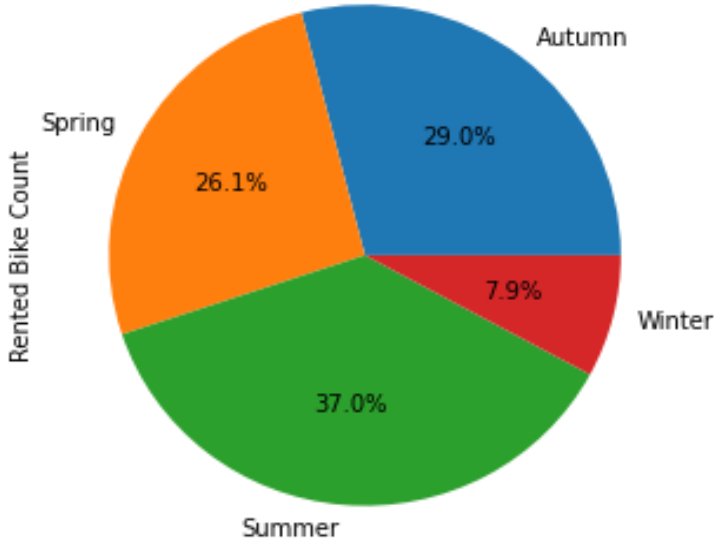
Feature Engineering and Data Visualization



- The entire colour of heatmap is same ,Hence we conclude that there is no missing values present in our dataset.

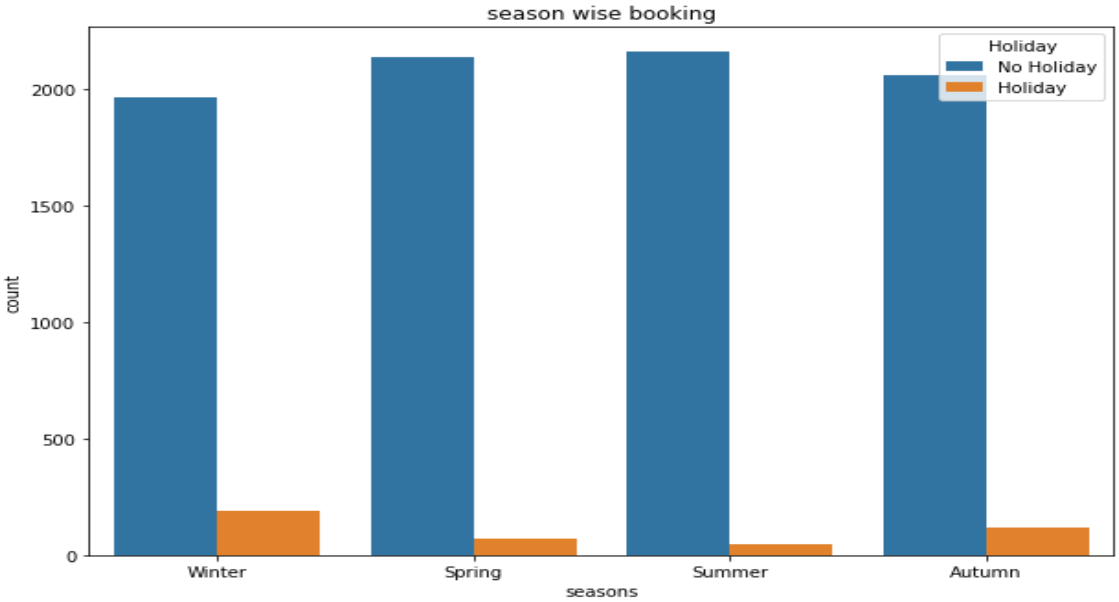
Season wise rented bike counts:-

- ✓ This shows the total count of rented bikes throughout all seasons in a pie chart.
- ✓ Most bikes have been rented in summer Where as during winter least number of bikes have been rented.
- ✓ Around 1790002 in autumn 1611909 in spring 2283243 in summer 487169 in winter.

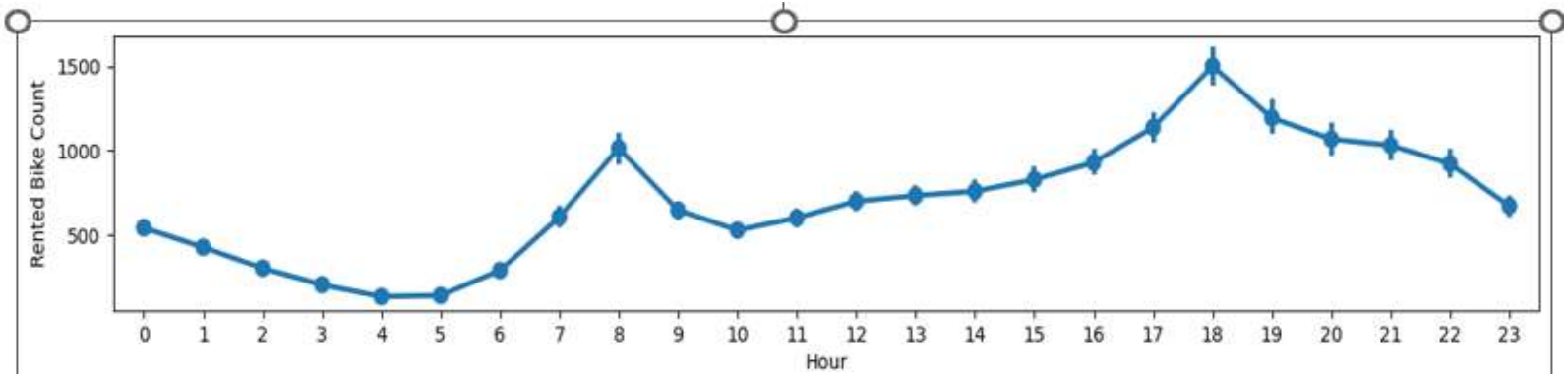


Holiday in Seasons:-

- Season wise differentiation of rented bike on Holiday day and Non holiday.
- It is shown that number of holidays is maximum in winter season as compare to other seasons.



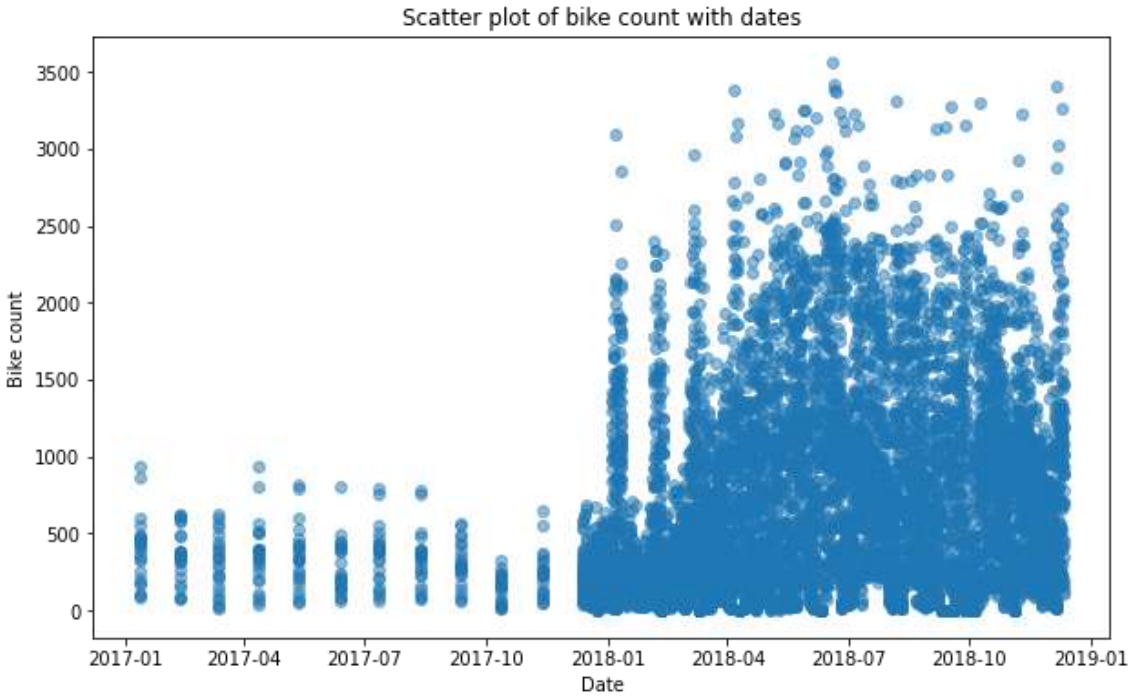
Bike Demand as per hour:



- The demand of rented bikes is more in morning 8:00 AM and evening at 6:00PM.
- The demand of rented bikes is decline at evening and lowest at morning 4:00Am .

Bike count with Date:

- Scatter plot of bike counts on different dates.
- Using these graph we can say that in 2017 the demand of bikes is low as compare to 2018 because in 2018 demand rises very sharply.

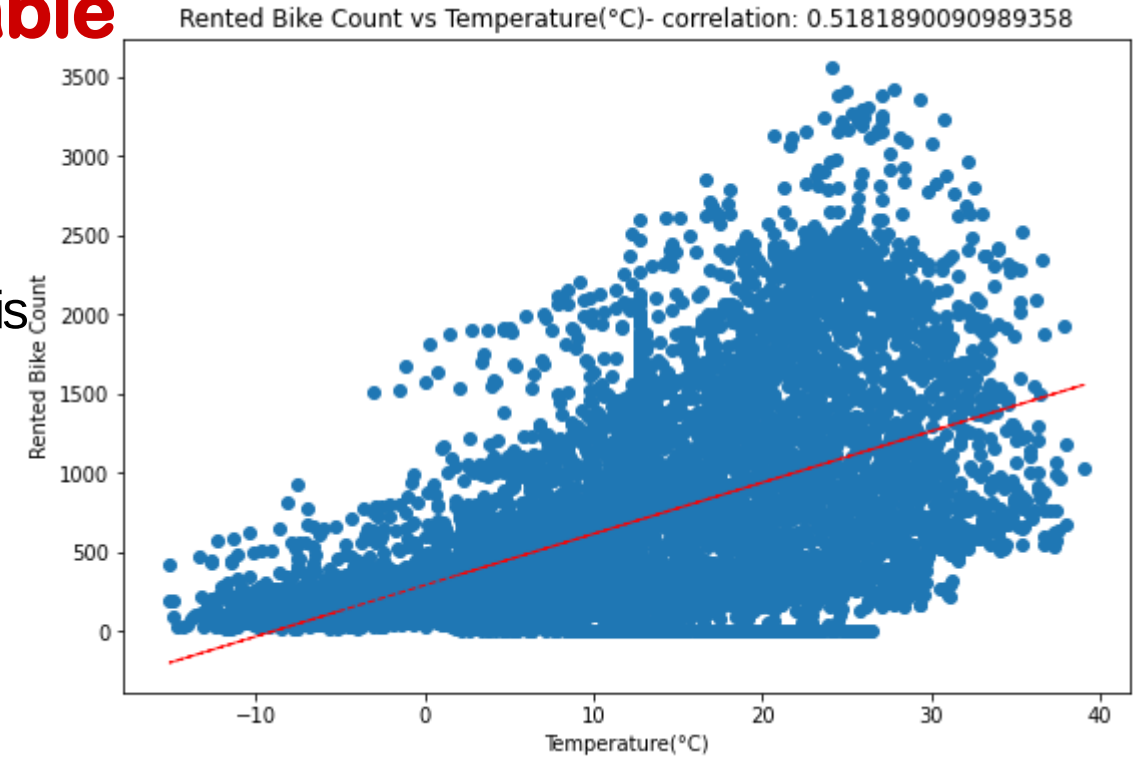


Multicollinearity:

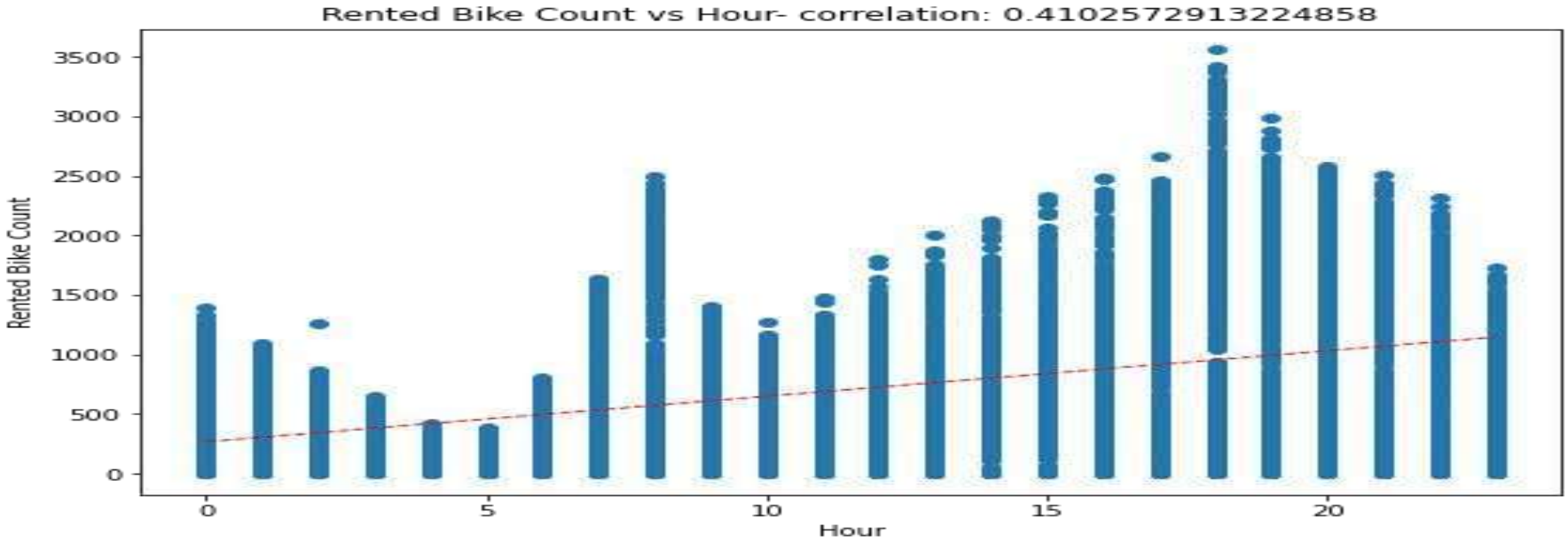
- Variance inflation factor has been used to detect multicollinearity and summer feature is highly collinear so, it is dropped.

Correlation Plot Between Dependent and Non Dependent Variable

- We can say that when temperature is low there is less demand of bike because people prefer to Stay in home.
- When the temp is high people go out and take bike on rent.



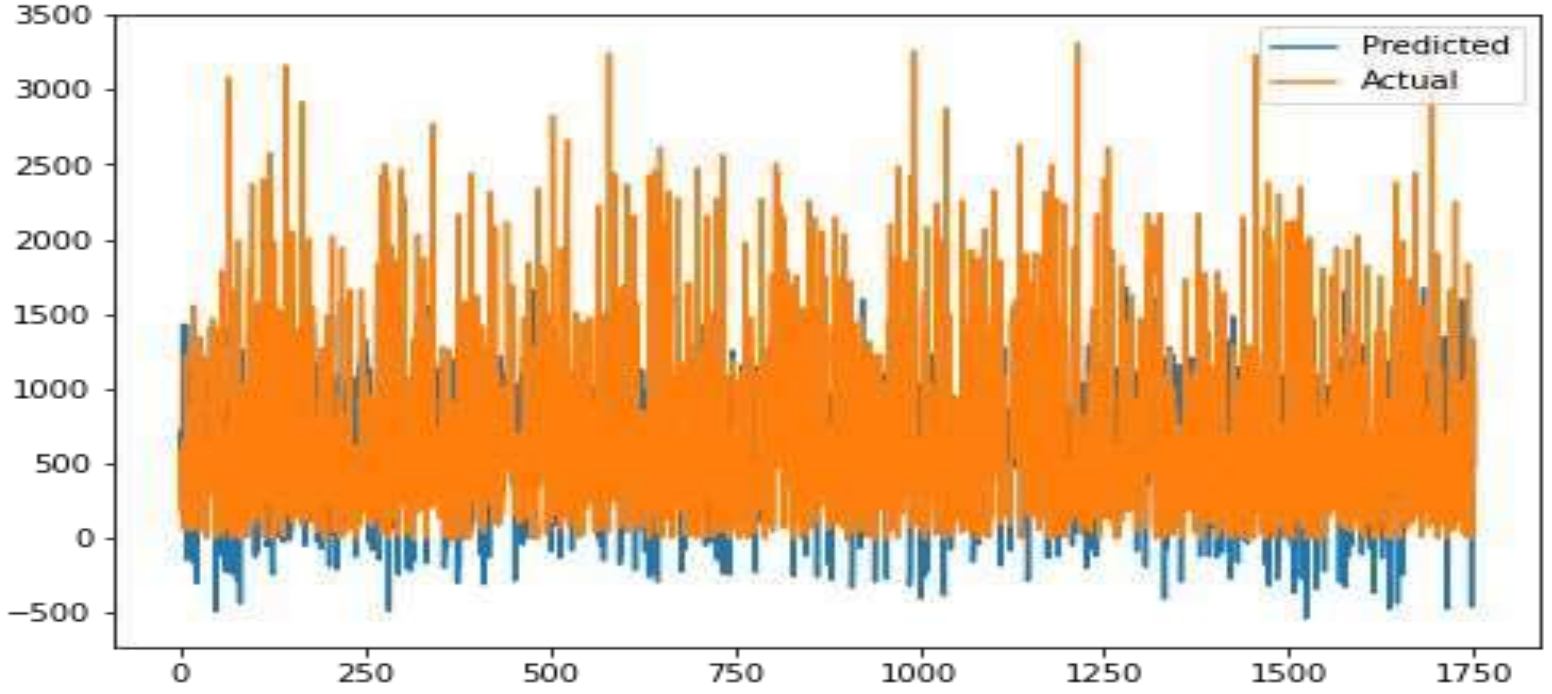
Correlation between Hour and Rented bike count.



- This shows the rented bike count per hour.

Model Building :- Linear Regression Model

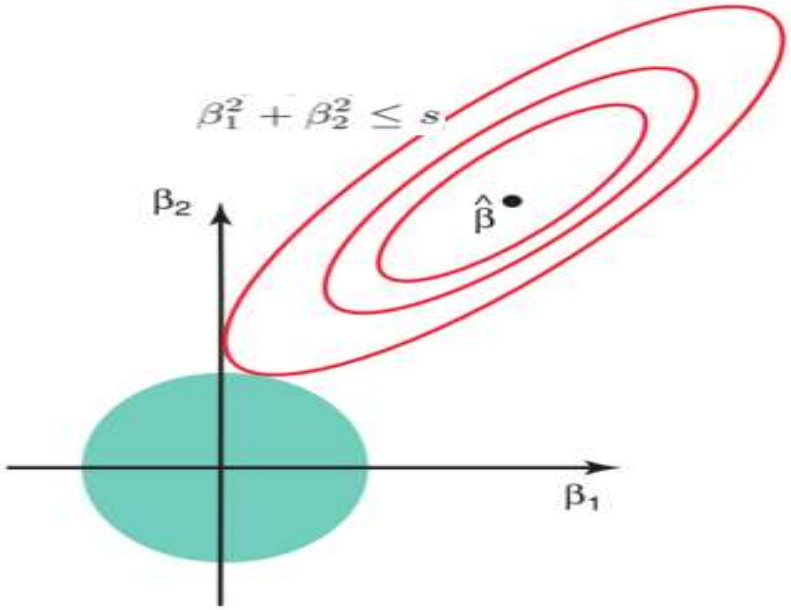
- After selecting rented bikes feature as the independent feature and the rest other columns as dependent feature, we split the data into train set and test set , later we transformed the data using minmax scalar and we fitted LINEAR REGRESSION MODEL to the dataset.
- The following results were obtained,
- MSE : 213299.99590594516
- RMSE : 461.8441251179289
- R2 : 0.48805490941404095
- Adjusted R2 : 0.484522223337542



- Results obtained from linear regression model

Ridge Regression:

- MSE : 213299.99590594516
- RMSE : 461.8441251179289
- R2 : 0.4880549094140409 5
- Adjusted R2 : 0.484522223337542



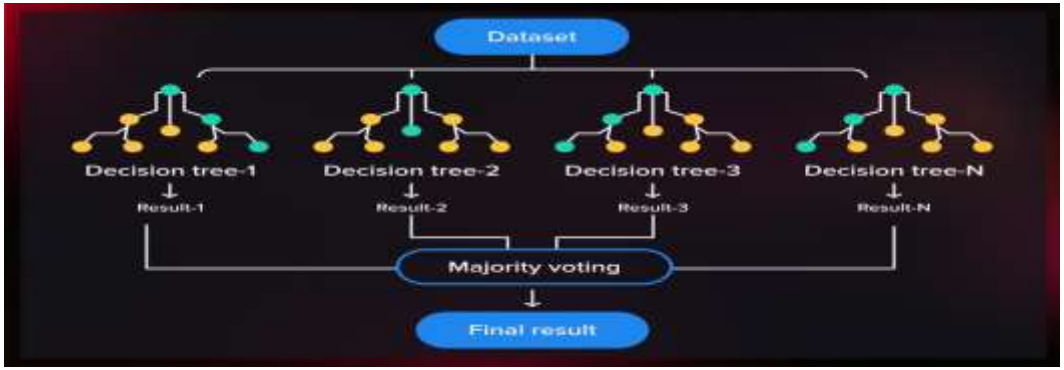
Ridge Regression

Decision Tree Model:

- After applying the decision tree model we obtained,

- The best Decision Tree R2 score is 0.7764042828411412 with max depth 10
- The best R2 test score is : 0.7720060217704434 with max depth of10

Random Forest Regression model:



- The best Random Forest R2 train score is : 0.8331665844685252 with n estimators =17, max
  - depth : 15, min samples split :4 and min samples leaf : 2

- The best Random Forest R2 test score is : 0.8115687467028467 with n estimators =17, max
  - depth : 15, min samples split :4 and min samples leaf : 2

Conclusion:

- The exploratory data analysis and modelling for Bike sharing demand prediction dataset has been successfully done and the following inferences have been made from the obtained visualizations and also from the dataset.

- **For the linear regression model the obtained results are,**

- **MSE : 213299.99590594516**
- **RMSE : 461.8441251179289**
- **R2 : 0.4880549094140409 5**
- **Adjusted R2 : 0.484522223337542**

For the decision tree model the obtained results are:

- The best Decision Tree R2 score is 0.7764042828411412 with max depth 10
- The best R2 test score is : 0.7720060217704434 with max depth =10

For the random forest model the obtained results are

- The best Random Forest R2 train score is : : 0.8331665844685252 with n estimators =17, max depth : 15, min samples split : 4 and min samples leaf : 2
- The best Random Forest R2 test score is : 0.8115687467028467 with n estimators =17, max depth : 15, min samples split : 4 and min samples leaf : 2

- The results are obtained and compared thoroughly and the best predictions are made.

- Most numbers of Bikes were rented in Summer, followed by Autumn, Spring, and Winter .
- May-July is the peak Bike renting Season, and Dec-Feb is the least preferred month for bike renting.
- Majority of the client in the bike rental sector belongs to the Working class.
- Temperature of 20-30 Degrees, evening time 4 pm- 8 pm.
- Humidity between 40%-60% are the most favorable parameters where the Bike demand is at its peak .
- Temperature, Hour of the day, Solar radiation, and Humidity are major driving factors for the Bike rent demand .
- Most preferred temperature for bike renting is 20-30 Degree Celsius .
- Humidity of 40%-60% is most favorable for bike sharing.


Thank You...