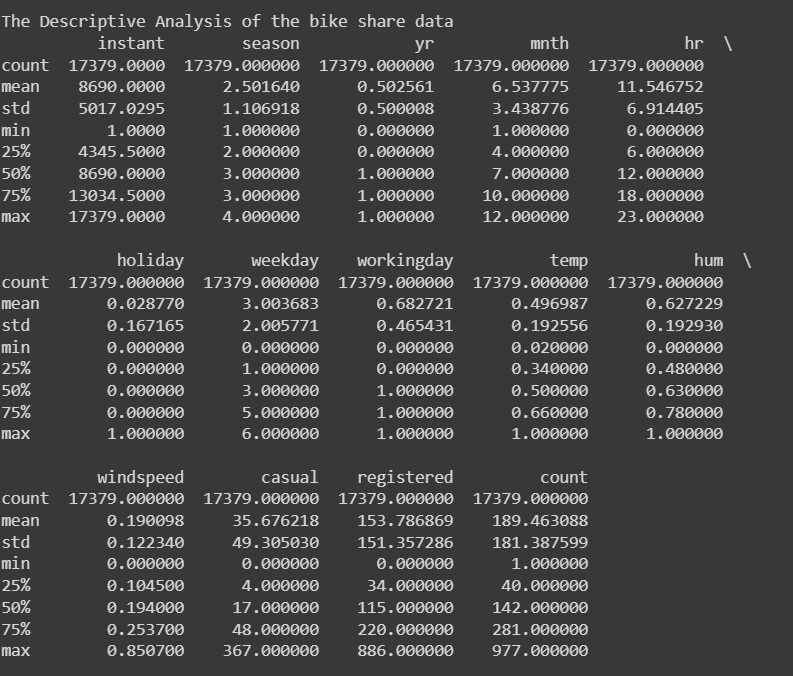
**Project**

# Introduction

This dataset encompasses bike share information gathered by a company operating in Washington, D.C., spanning the years 2011 and 2012. Each entry is identified by a unique 'instant' ID number and includes various hourly recorded attributes such as date ('dteday'), season, year ('yr'), month ('mnth'), hour ('hr'), and binary indicators for holidays and weekdays. Meteorological metrics like temperature ('temp'), humidity ('hum'), and windspeed are normalized to a scale between 0 and 1. The dataset also provides counts of casual and registered riders, as well as a total count representing the overall number of riders during each recorded hour. Instances with zero bike usage were excluded from the dataset. This extensive set of variables enables a detailed examination of bike share patterns and their relationships with temporal and weather-related factors.

The below is the Descriptive Analysis of the bike share data



# The below is the insights from the descriptive analysis of the data

**Season:**

The data covers four seasons (1 to 4), with an average season value of approximately 2.5.

The standard deviation is around 1.11, indicating a moderate amount of variation.

Seasons are evenly distributed, as the mean is close to the midpoint (2.5).

**Casual, Registered, and Total Count:**

Casual, registered, and total bike counts (casual, registered, and count) have right-skewed distributions.

The mean count is 189.46, with a standard deviation of approximately 181.39.

The minimum count is 1, and the maximum count is 977.

**Holiday:**

The holiday variable is binary (0 or 1), indicating whether it is a holiday or not.

The dataset has a low average holiday occurrence (approximately 2.88%).

**Workingday:**

A binary variable (0 or 1) indicating whether it is a working day or not. The average suggests that around 68.27% of the instances are working days.

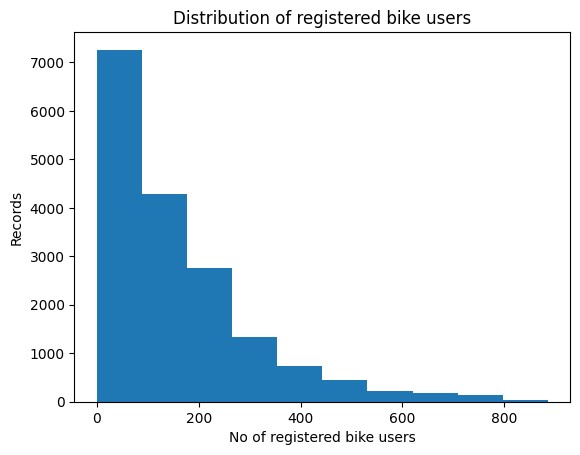
**Temperature (temp):** Temperature values range from 0.02 to 1, with an average of 0.5. The temperature distribution seems well-spread.

**Humidity (hum):** Humidity values range from 0 to 1, with an average of 0.63. The humidity distribution appears to be moderate.

**Windspeed:**

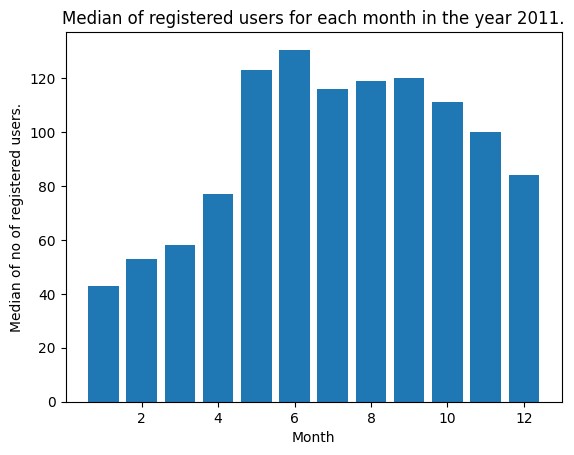
Windspeed values range from 0 to 0.85, with an average of 0.19. There is a moderate variation in windspeed.

Histogram to better understand the distribution of no of registered bike users.

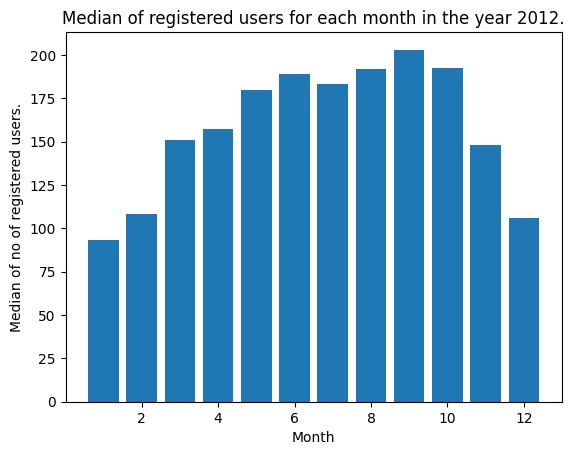


From the above histogram, we can conclude that the no of registered bike users are decreasing steadily.

A bar plot that shows the median number of registered riders (grouped by month) for each month for the year 2011.



Another bar plot that shows the median number of registered riders (grouped by month) for each month for the year 2012.



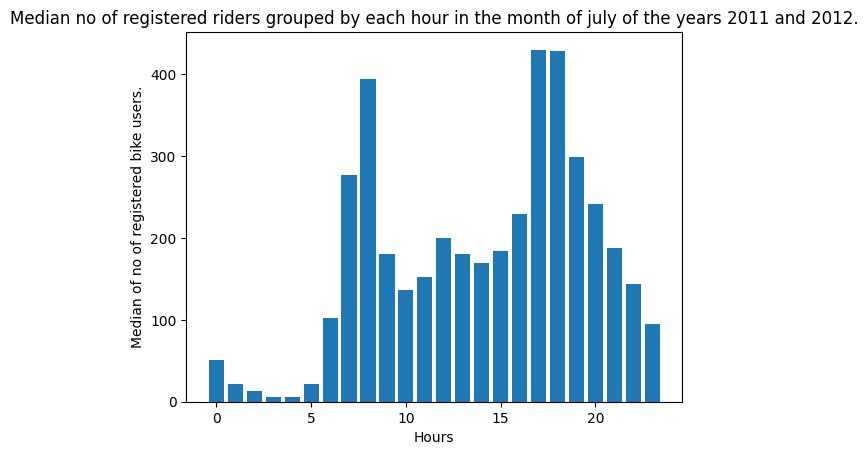
Insights from the above bar graph

1.In the year 2011, in the 6th month the median is high which is around 130. That implies there are more no of registered bike users are in the 6th month. Whereaas in the year 2012, 9th month has more no of registered bike users.

2.In both years, the no of registered bike users are less in the first month.

3.In the year 2012 has more no of registered bike users compared to the year 2011 since the highest median of 2012 is higher than the highest median of 2011.

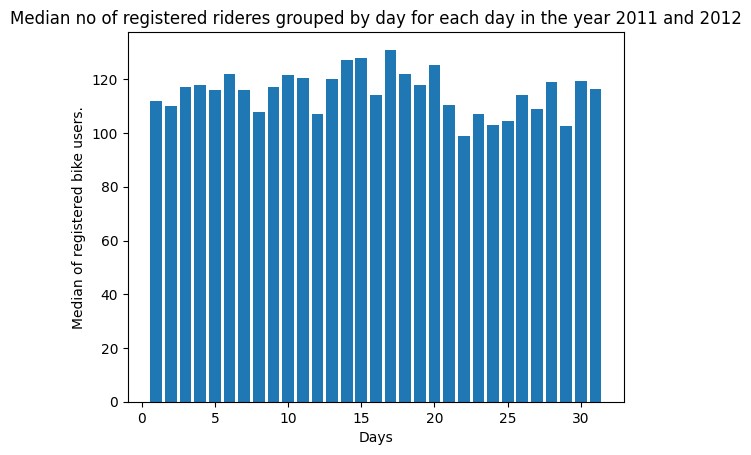
A bar plot showing the median number of registered riders (grouped by hour) for each hour for the month of July (include both years).



1. In the month of july of both year 2011 and 2012, the 16th &17th hour has more no of registered bike users.

2.At the 3rd and 4th hour there are least no of registered bike users.

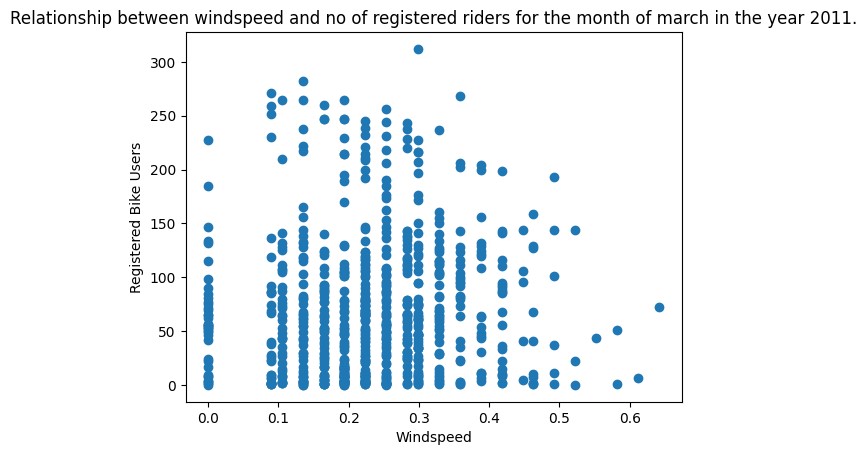
A bar plot showing the median number of registered riders (grouped by day) for each day (include both years)



1.On the 17th day, no of registered bike users are more. And it was on day 21 where the no of registered bike users are least.

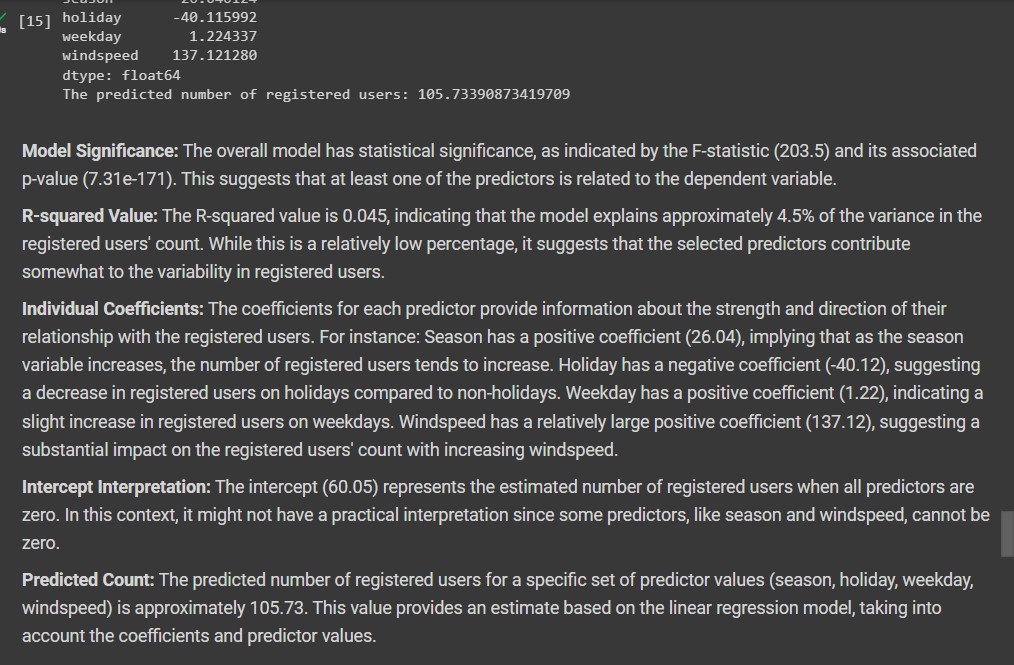
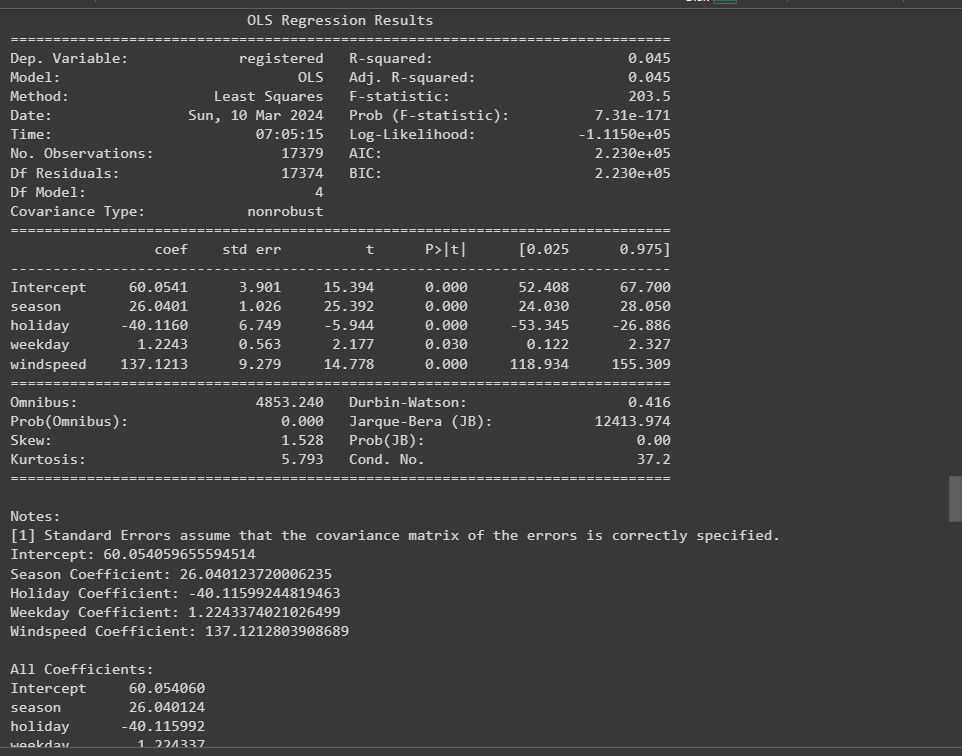
2.Almost every day has almost similar no of registered bike users.

A scatter plot to show the relationship between windspeed and the number of registered riders only for the month of March in the year 2011.



1. As windspeed increases, the no of registered bike users decreases. 2. The more no of registered bike users are at the windspeed 0.3

The linear regression model to predict the no of registered bike users.



**Conclusion:**

In conclusion, the descriptive analysis of the bike share dataset for Washington, D.C., spanning 2011 and 2012, reveals valuable insights into the factors influencing bike usage patterns. The examination of variables such as season, temperature, humidity, windspeed, and holiday indicators provides a comprehensive understanding of their impact on registered bike users. The temporal analysis, focusing on monthly and hourly trends, highlights variations in user behavior, with notable peaks during specific months and hours. Additionally, the observed relationships between windspeed and registered riders in March 2011 indicate a potential influence of weather conditions on bike usage. The constructed linear regression model serves as a predictive tool, offering a valuable framework for estimating the number of registered bike users based on the identified influential factors. Overall, this analysis contributes to a nuanced comprehension of bike share dynamics, enabling informed decision-making for bike-sharing service providers and urban planners.