TABLEAU PROJECT

on

Bike Sharing Analysis

GROUP 11

Mayur Mahanta

Moheth Muralidharan

Vignesh Sivakumar

Zhibek Kassymkanova

Signature of Mayur: *Mayur* mahanta.m@northeastern.edu

Signature of Moheth: *Moheth* muralidharan.mo@northeastern.edu

Signature of Vignesh: *Vignesh* sivakumar.vig@northeastern.edu

Signature of Zhibek: *Zhibek* Kassymkanova.z@northeastern.edu

Submission Date: 6th December 2022

Introduction:

Bike-sharing systems are a new version of traditional bike rentals in which the entire process of membership, rental, and return has been automated. The user can easily rent a bike from one location and return it to another using these services. There are currently over 500 bike-sharing schemes worldwide, with over 500 thousand bicycles. Because of their importance in traffic, environmental, and health issues, these systems are currently generating a lot of interest.

Aside from the exciting real-world applications of bike sharing systems, the data features created by these systems make them appealing for research. In contrast to other modes of transportation such as bus or subway, the duration of journey, departure and arrival position are all explicitly recorded in these systems. This feature transforms the bike sharing system into a virtual sensor network capable of sensing motion in the city. As a result, it is envisaged that the most significant occurrences in the city will be discovered by monitoring this data.

Overview:

The goal of our project is to analyze the bike sharing that took place between 2011 and 2012 in Capital Bikeshare system, Washington D.C., USA. Bike rental process is highly correlated to the environmental and seasonal settings. Weather conditions, precipitation, weekday, season, hour of the day, and so on can impact rental behavior. We aggregated the data daily and hourly, then extracted and incorporated the necessary weather and seasonal information.

This dataset was kindly contributed by Capital Bikeshare and obtained through Kaggle. We represented our analysis using numerous plots and charts via interactive dashboards that allow viewers to choose from a range of parameters. We will be able to visualize the areas where they should concentrate their efforts to maximize profitability.

Data Description

The dataset consists of 17389 instances of the hourly and daily count of rental bikes between the years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information.

No.	Variable	Description
1	instant	Record index
2	dteday	Date
3	season	Season (1: winter, 2: spring, 3: summer, 4: fall)
4	yr	Year (0: 2011, 1:2012)
5	mnth	Month (1 to 12)
6	hr	Hour (0 to 23)
7	holiday	Whether day is holiday (1) or not (0)
8	weekday	Day of the week (1: Monday, 2: Tuesday, 3: Wednesday, 4: Thursday, 5: Friday, 6: Saturday, 7: Sunday)
9	workingday	if day is neither weekend nor holiday is 1, otherwise is 0
10	weathersit	- 1: Clear, Few clouds, Partly cloudy, Partly cloudy - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
11	temp	Normalized temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-8, t_max=+39 (only in hourly scale)
12	atemp	Normalized feeling temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-16, t_max=+50 (only in hourly scale)
13	hum	Normalized humidity. The values are divided to 100 (max)
14	windspeed	Normalized wind speed. The values are divided to 67 (max)
15	casual	Count of casual users
16	registered	Count of registered users
17	cnt	Count of total rental bikes including both casual and registered

Design Methodology

There are two datasets available on UCI: day.csv and hour.csv. The only difference is that hour.csv has ‘hr’ variable, so we decided to proceed with this dataset for further visualization. Categorical variables ‘season’, ‘yr’, ‘mnth’, ‘holiday’, ‘weekday’, ‘workingday’ and ‘weathersit’ we encoded to numerical, and we encoded them back to string values, as it is better works for understanding the visualization.

For ‘season’ we converted ‘1’ to winter, ‘2’ to spring, ‘3’ to summer and ‘4’ to fall.

For ‘yr’ we converted ‘0’ to 2011 and ‘1’ to 2012.

For ‘mnth’ we converted ‘1’ to January, ‘2’ to February, ‘3’ to March, ‘4’ to April, ‘5’ to May, ‘6’ to June, ‘7’ to July, ‘8’ to August, ‘9’ to September, ‘10’ to October, ‘11’ to November and ‘12’ to December.

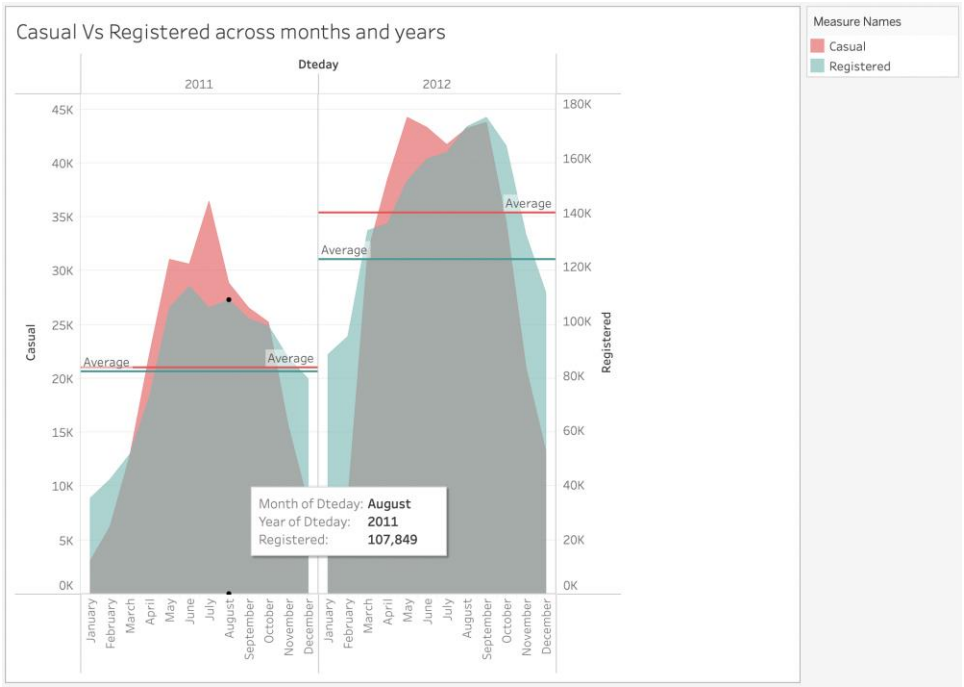
For ‘holiday’ we converted ‘1’ to Yes and ‘0’ to No.

For ‘weekday’ we converted ‘1’ to Monday, ‘2’ to Tuesday, ‘3’ to Wednesday, ‘4’ to Thursday, ‘5’ to Friday, ‘6’ to Saturday and ‘7’ to Sunday.

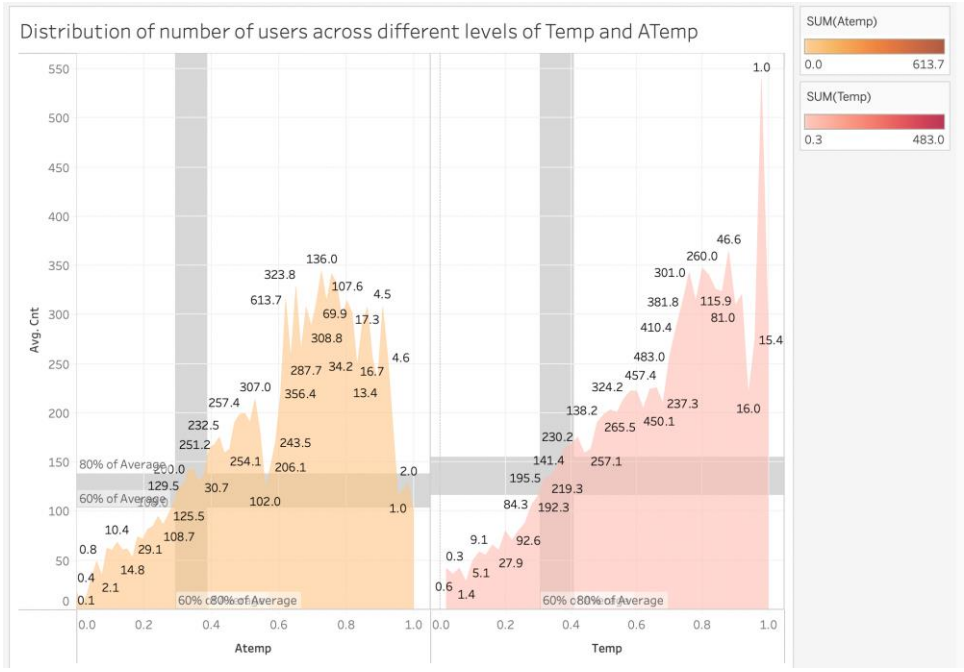
For ‘workingday’ we converted ‘1’ to Yes and ‘0’ to No.

For ‘weathersit’ we converted ‘1’ to Clear, ‘2’ to ‘Mist’, ‘3’ to ‘Light Snow / Rain’ and ‘4’ to ‘Heavy Rain’.

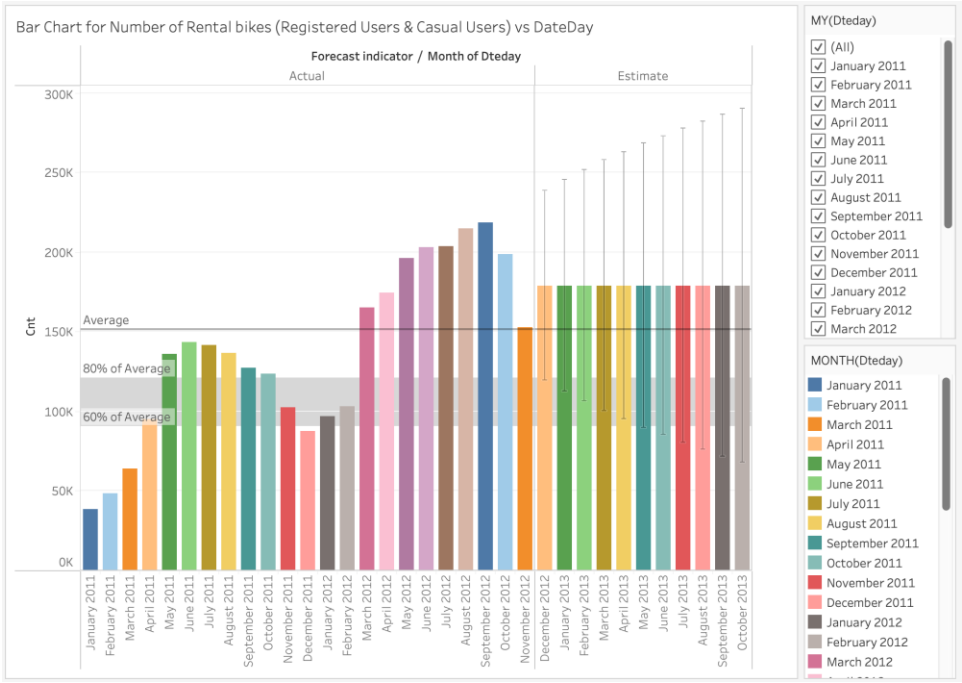
Visualizations:



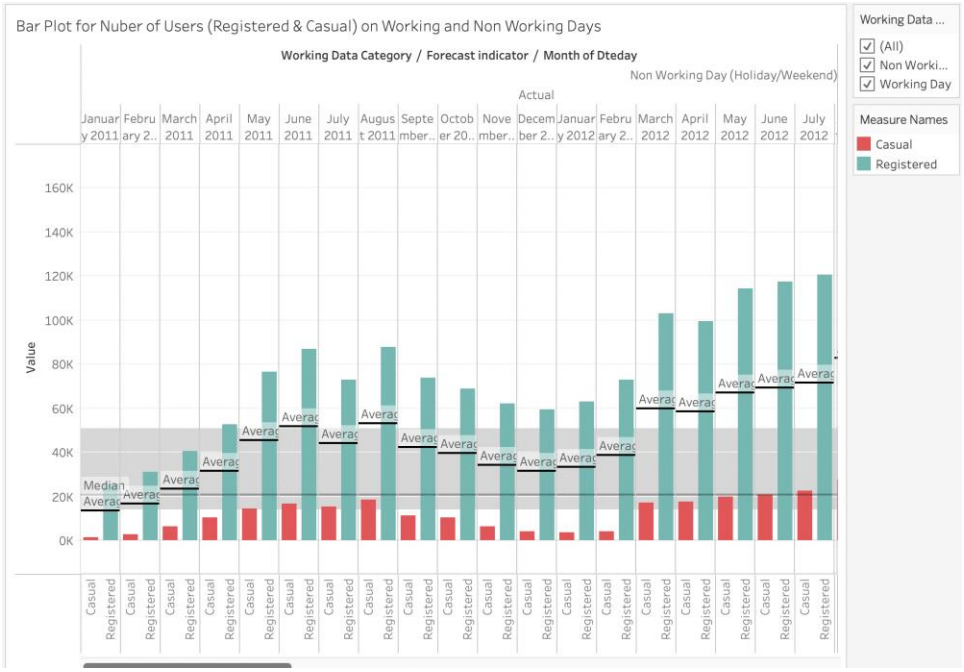
Area plot of Casual Vs Registered Across months and years



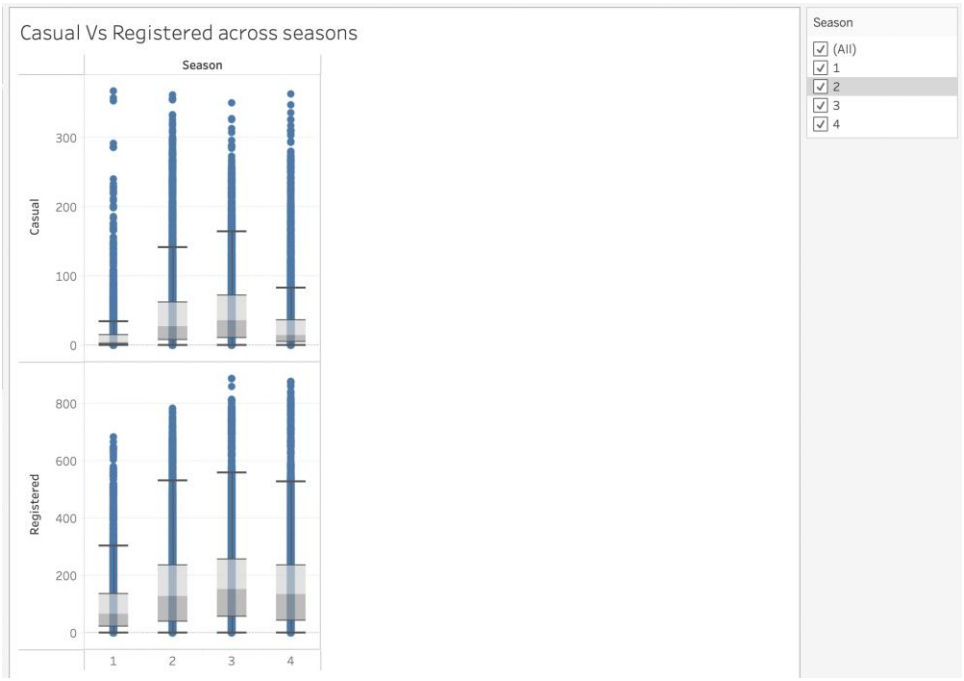
Area plot of the Distribution of number of users across different levels of Temp and ATemp



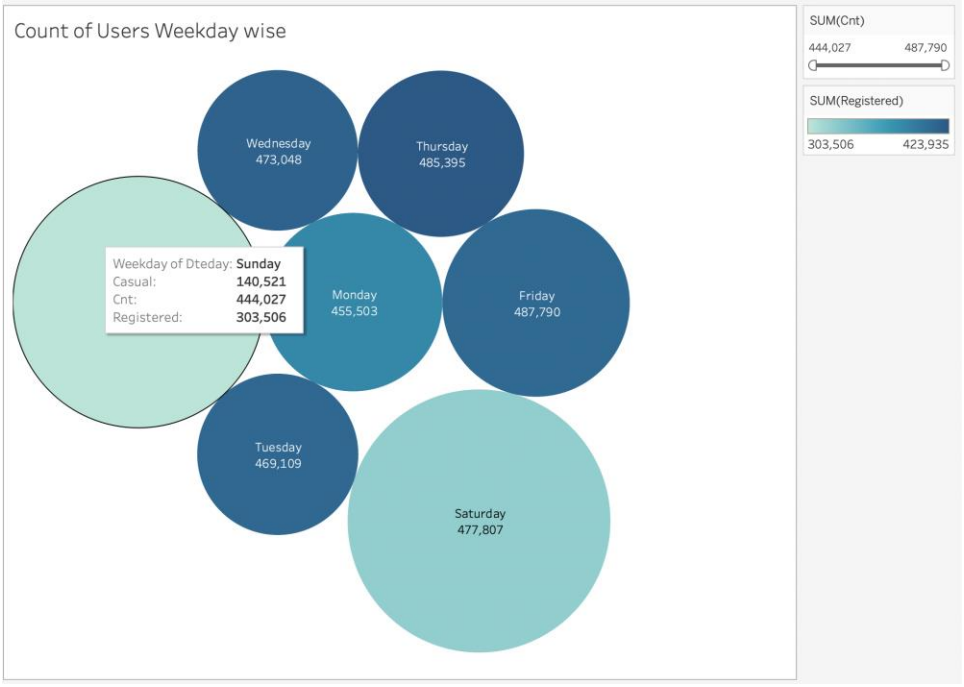
Bar Plot for the Number of Rental Bikes Vs DateDay



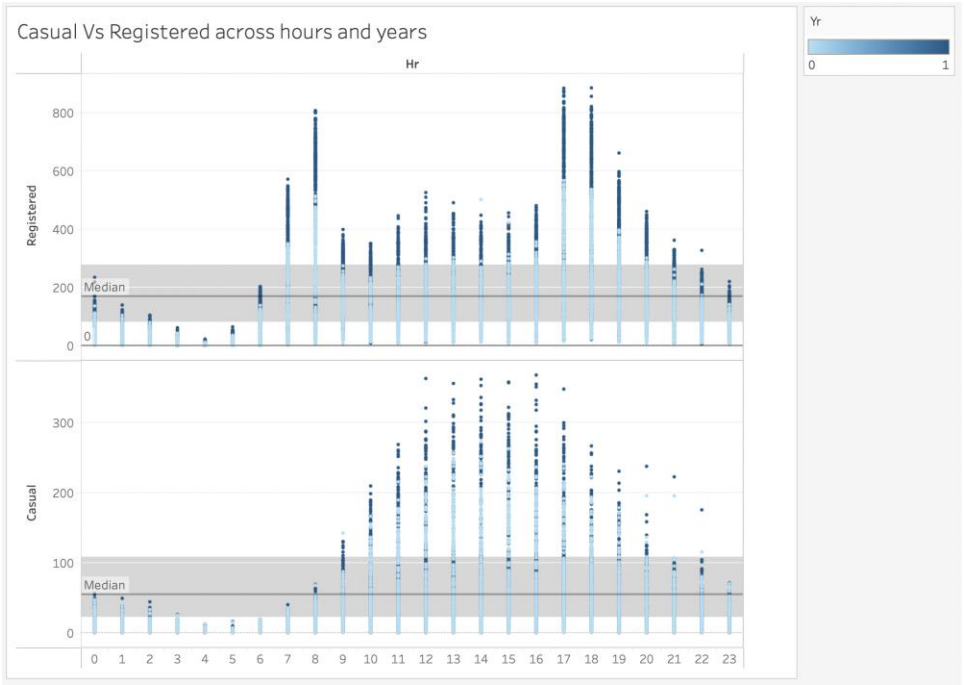
Bar Plot for the number of Users on Working and Non-Working Days



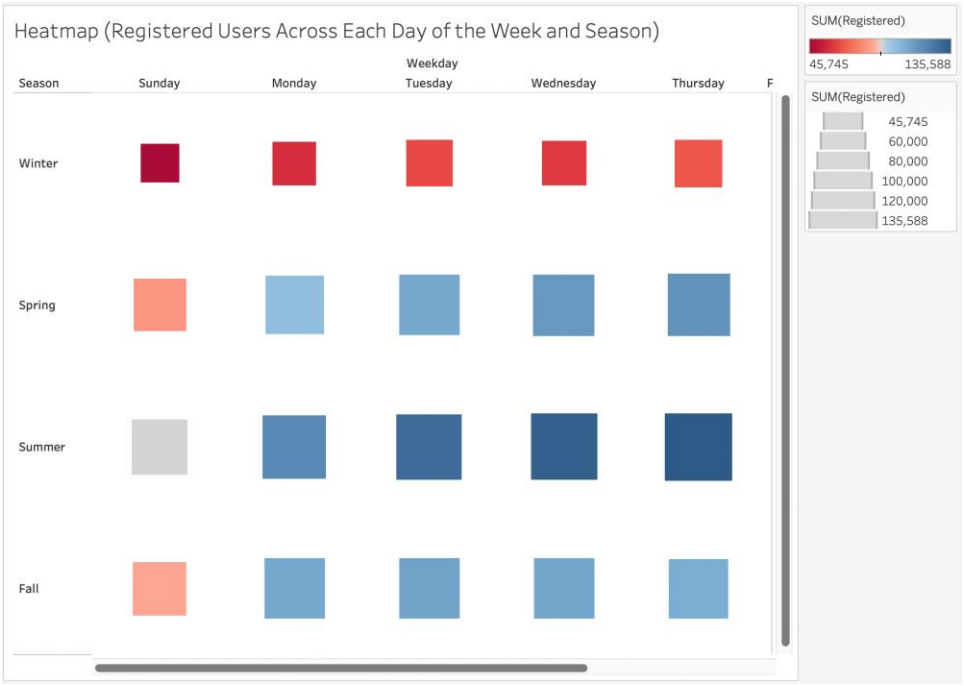
Boxplot of Casual Vs Registered- across Seasons



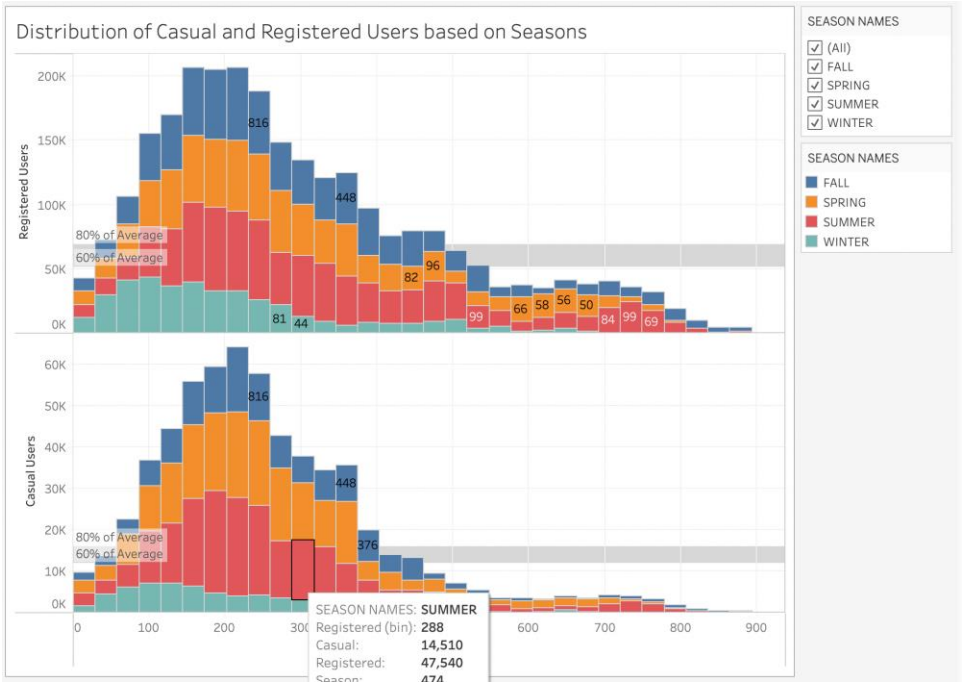
Bubble plot of Count of Users in the days of a week



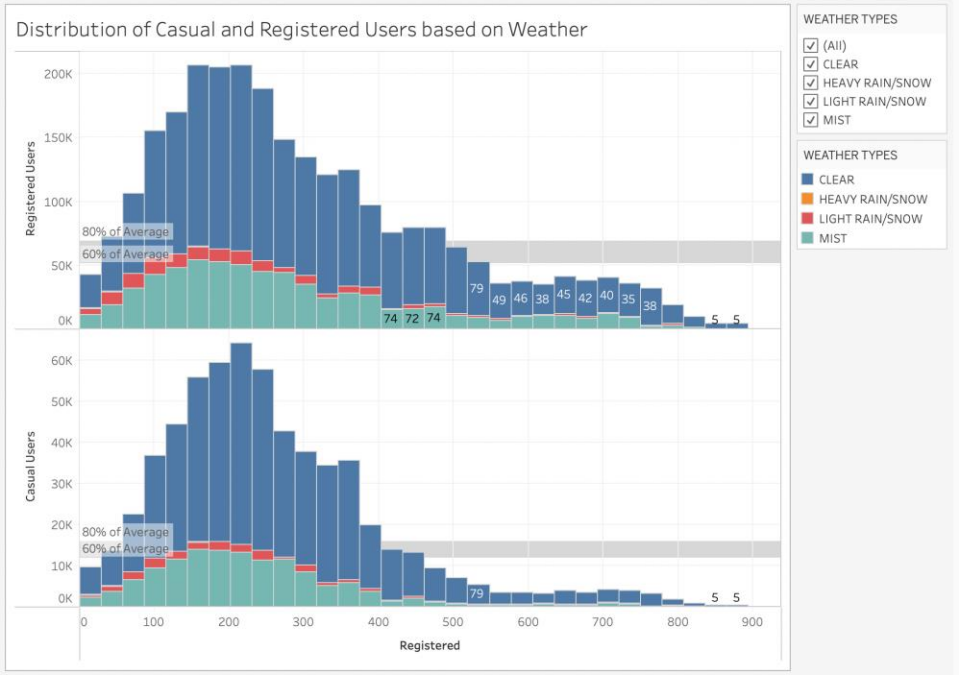
Dot plot of Casual Vs Registered Users across hours and years



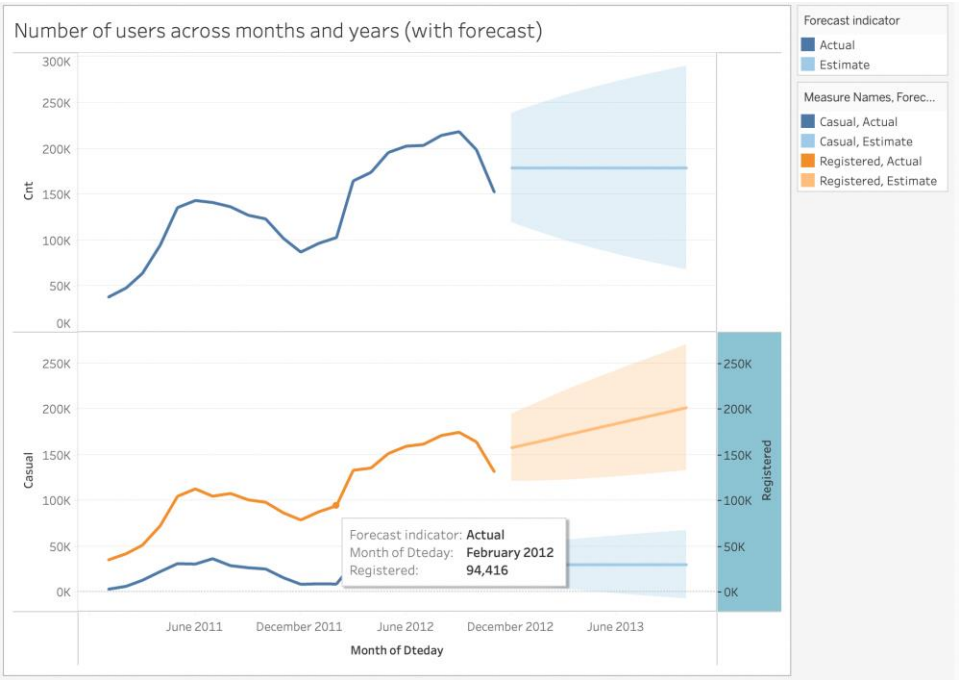
Heatmap of Registered Users across each day of the week and season



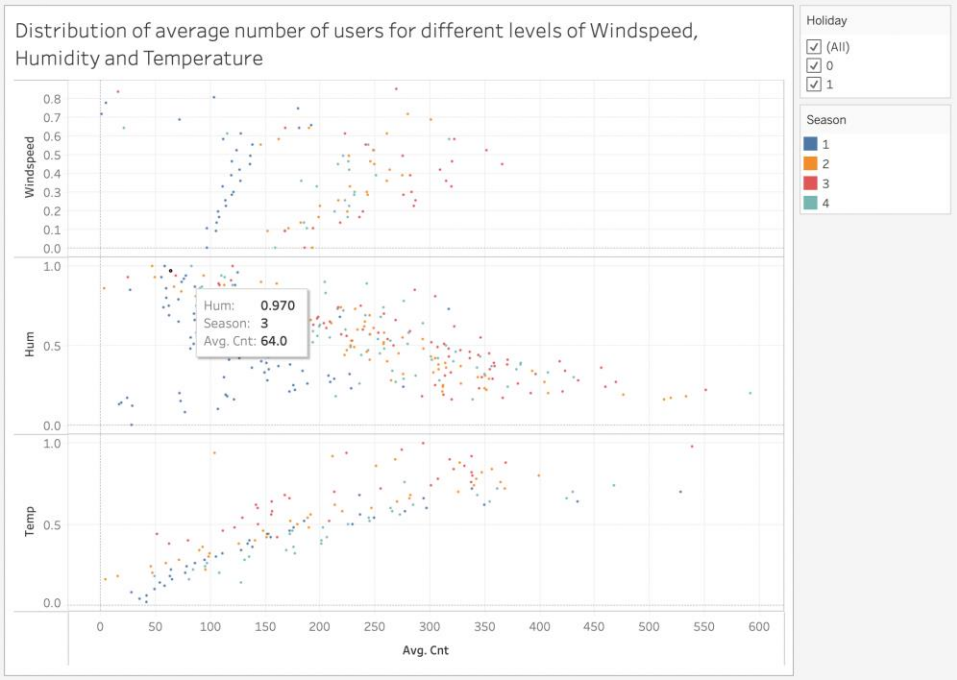
Stacked Barplot of Distribution of Casual and Registered Users based on Seasons



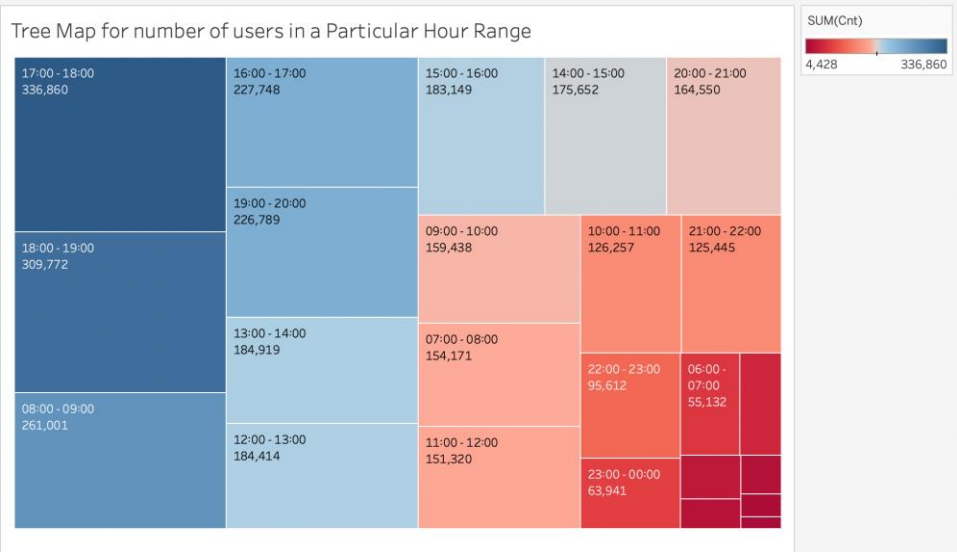
Stacked Barplot of Distribution of Casual and Registered Users based on Weather



Line plot of number of users across months and years



Distribution of average number of users for different levels of Windspeed, Humidity and Temperature



Treemap for the number of users in hour range

Conclusion:

We have inferred that external factors like seasons / weather do affect the number of bikes rented, either casual or registered. For example, we can notice that there is a drastic change between the number of renters in winter and in summer (nearly 300,000). So, with this analysis in mind, the bike lender could understand the demand and could provide supply accordingly. Bikes are the cleanest and the healthiest mode of transportation, and it would benefit both the renters (riders) and the lenders if bikes became more available.

Another insight that we made is that registered users are mostly those who use bikes as a transportation to work, with a higher number of uses right before and right after business hours.

We also observed the relationship between number of users and humidity, temperature, windspeed. There is negative relationship between number of users and humidity, meaning more users prefer riding a bike when there is no humidity. There is positive relationship between number of users and temperature, meaning more users prefer riding a bike on warmer days. And there is bell-shaped distribution of number of users across windspeed levels.number of users across windspeed levels.