Capstone Project

BIKE SHARING DEMAND

By – RAHUL SINGH WALDIA


B

Bike sharing has been gaining prominence over the last few decades. More and more people are turning towards healthy and livable cities where activities like bike sharing are easily available. There are plenty of benefits from bike sharing, like the environmental benefits. It has become a greener way to travel.

.

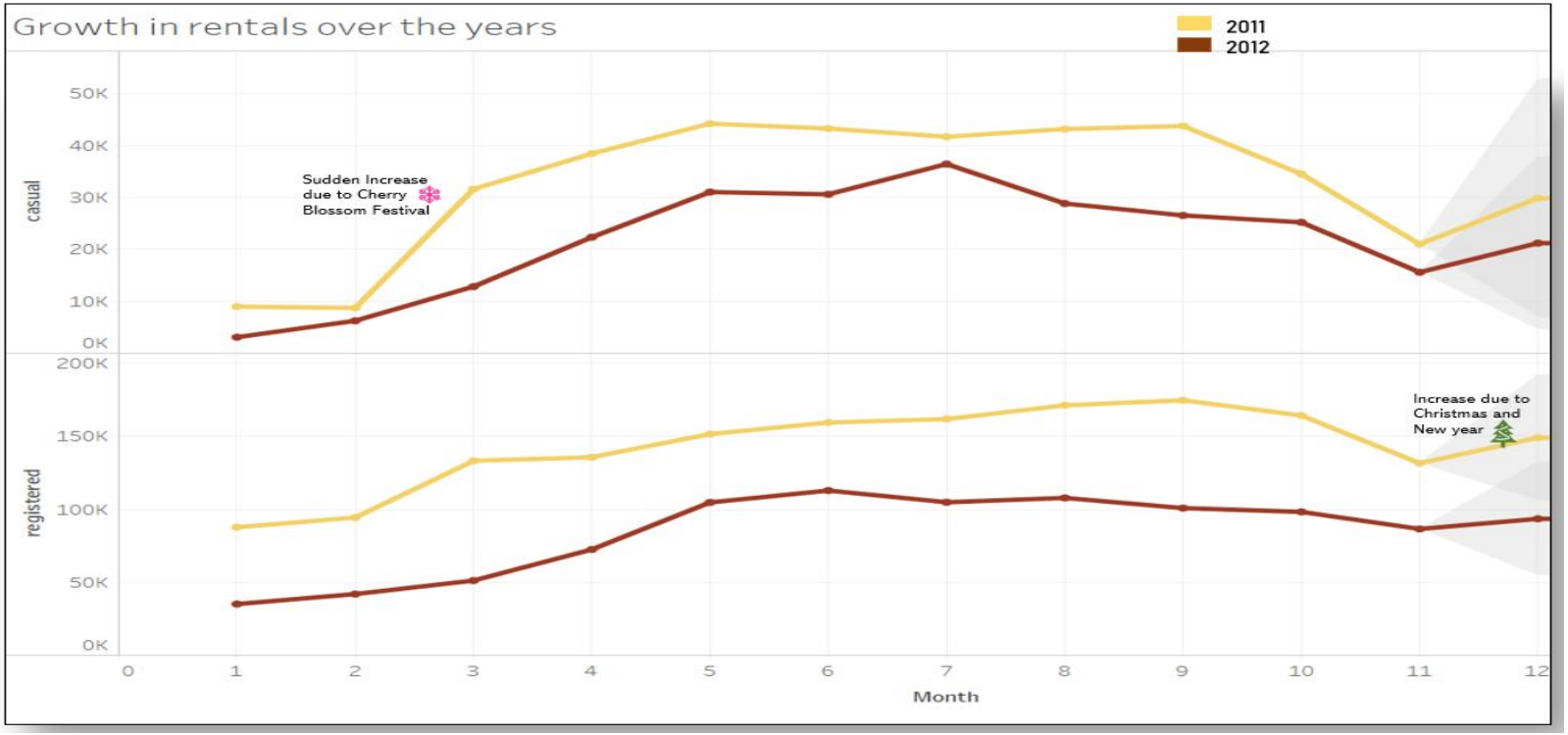
.

While there are several advantages to this system, it comes with several difficulties for the operators like the optimum demand prediction due to different factors, which leads to a demand and supply problem for the operator, where to build new stations or expand old ones. They also need to know how they can maintain an adequate inventory of bikes.

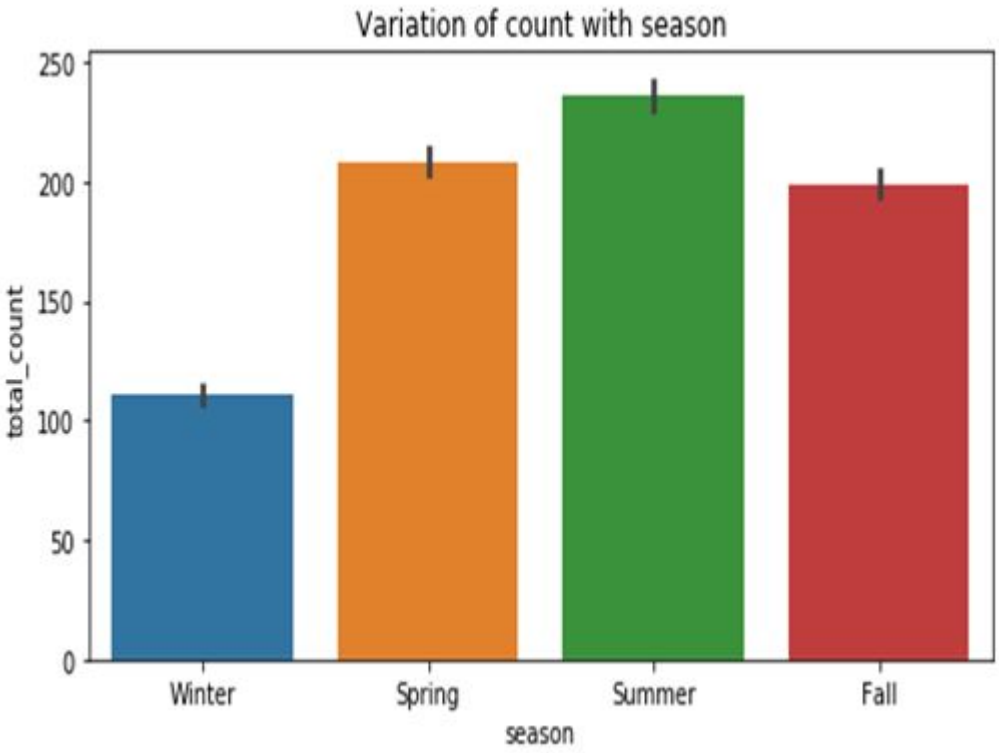
Feature Name	Description	Range	Feature Type
instant	Record Index		Numerical - Discrete
dteday	Date	all dates in 2011 & 2012	Numerical - Discrete
season	Season/Climate	(1: Spring, 2: Summer, 3: Fall, 4: Winter)	Categorical-Nominal
Yr	Year	(0: 2011, 1: 2012)	Categorical-Nominal
Hr	Hour	(0,23)	Numerical-Discrete
holiday	Whether the day is a holiday or not	(0: No Holiday, 1: Holiday)	Categorical-Nominal
weekday	Day of the week	(0,6)	Categorical-Nominal
working day	Whether the day is a working day or not	(0: Non-working day, 1: Working Day)	Categorical-Nominal
weathersit	Weather situation	(1,4)	Categorical-Nominal
temp	Normalized temperature in Celsius		Numerical-Continuous
atemp	Normalized feeling-like temperature in Celsius.		Numerical-Continuous
hum	Normalized humidity		Numerical-Continuous
windspeed	Normalized windspeed		Numerical-Continuous
casual	Count of non-registered users		Numerical-Continuous
registered	Count of registered users		Numerical-Continuous
cnt	Count of total rental bikes including both casual and registered		Numerical-Continuous

Dataset Overview

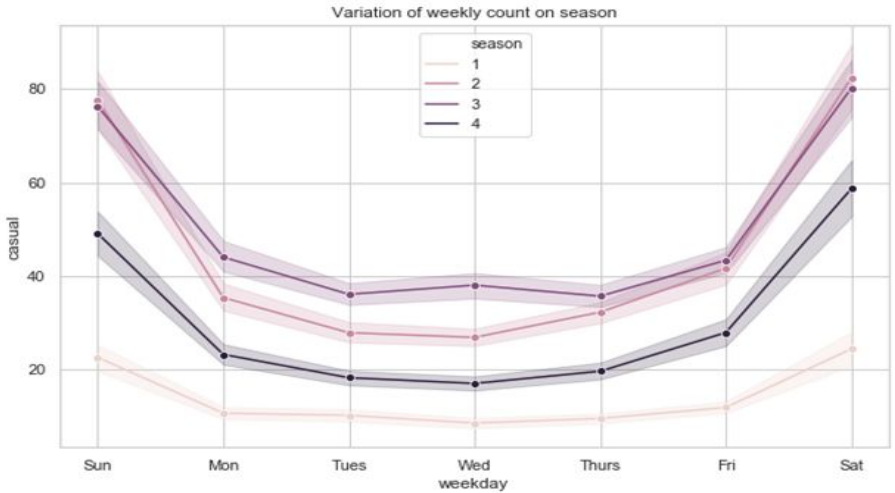
This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information. The dataset contains 17379 rows (every hour of each day for 2011 and 2012) and 17 columns (the features which are under consideration). We also included the trip data for year 2011 for analysing the station level data.

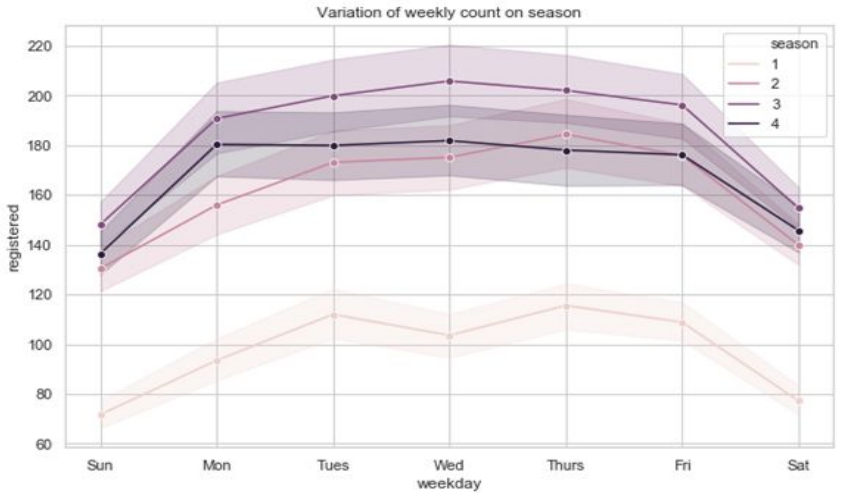


We started our EDA with a year on year comparison for both casual and registered members who rented out bikes. A clear rise can be noticed with the successive years(Increase from **12,43,103** in **2011** to **20,49,576** in **2012**).The casual members have shown an increase in numbers from February to March (Cherry Blossom Festival), indicating that they might be tourists visiting for the festival. There was also an increase in the numbers from November to December for both membership types, during the Christmas/New Year season.

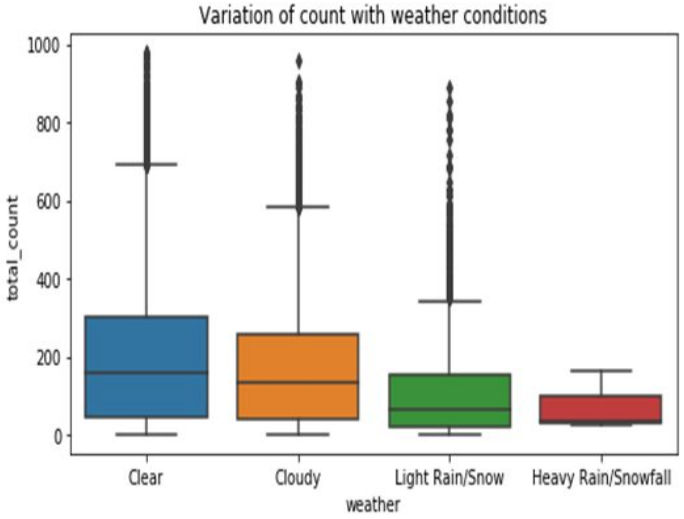


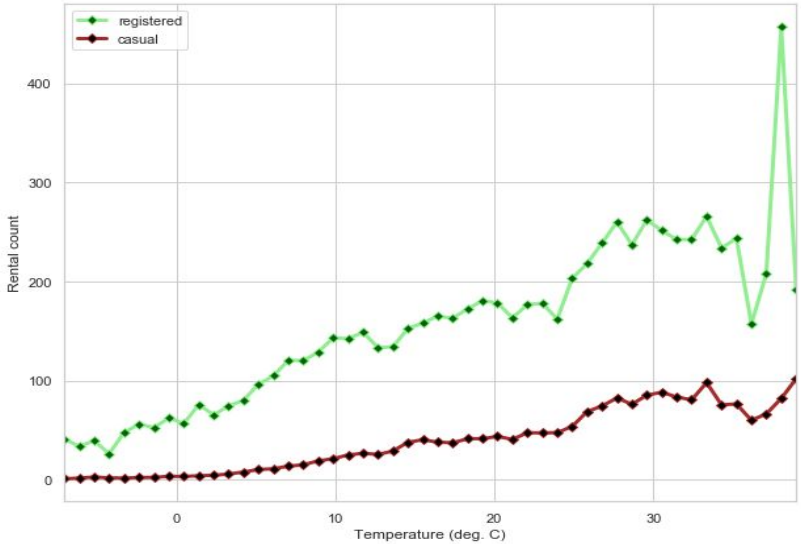
While comparing the bike rental count season-wise we observe that summer has the highest rentals followed by spring and then fall, which gives us reason to believe that the bike riders prefer warm to pleasant climates than colder climates.



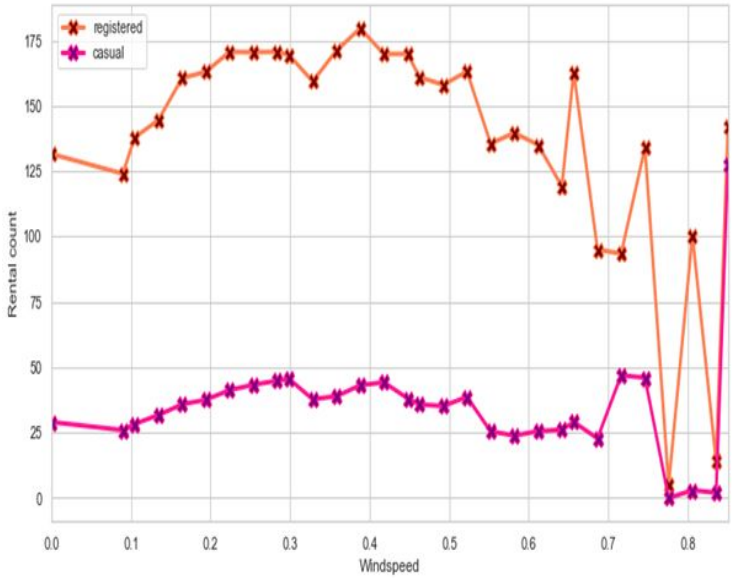


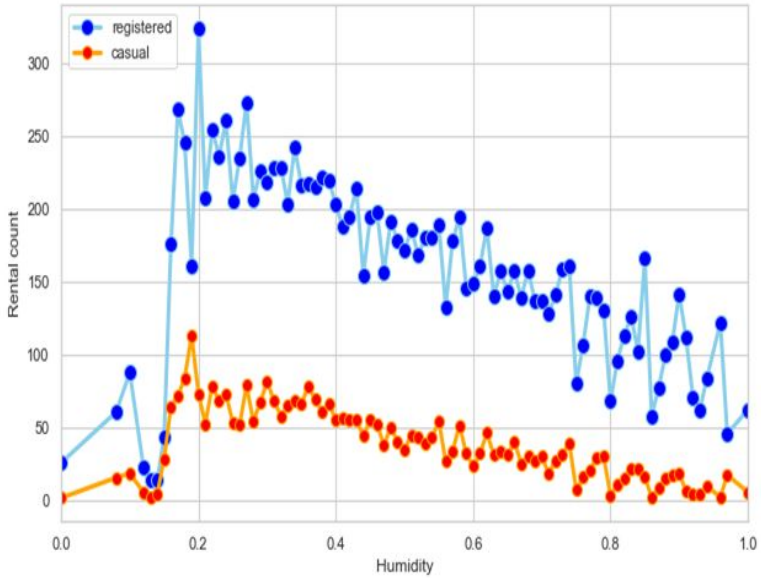
There was no significant relationship between the weekdays and the ridership. But it can be inferred from the below graphs that there is a significant difference in the behavior of casual and registered members while contributing to the ridership. The casual riders increase during the weekend and have a drop during the week. The opposite can be seen from the registered customers, where the peak is during the middle of the working week and drops during the weekends. From these results we can infer that the registered customers might be the employees who are commuting to work or college students taking a ride to the college, or any similar categories of people who follow a regular routine over the week. The casual users are weekend riders or tourists who are visiting during a short duration and do not require a regular membership.



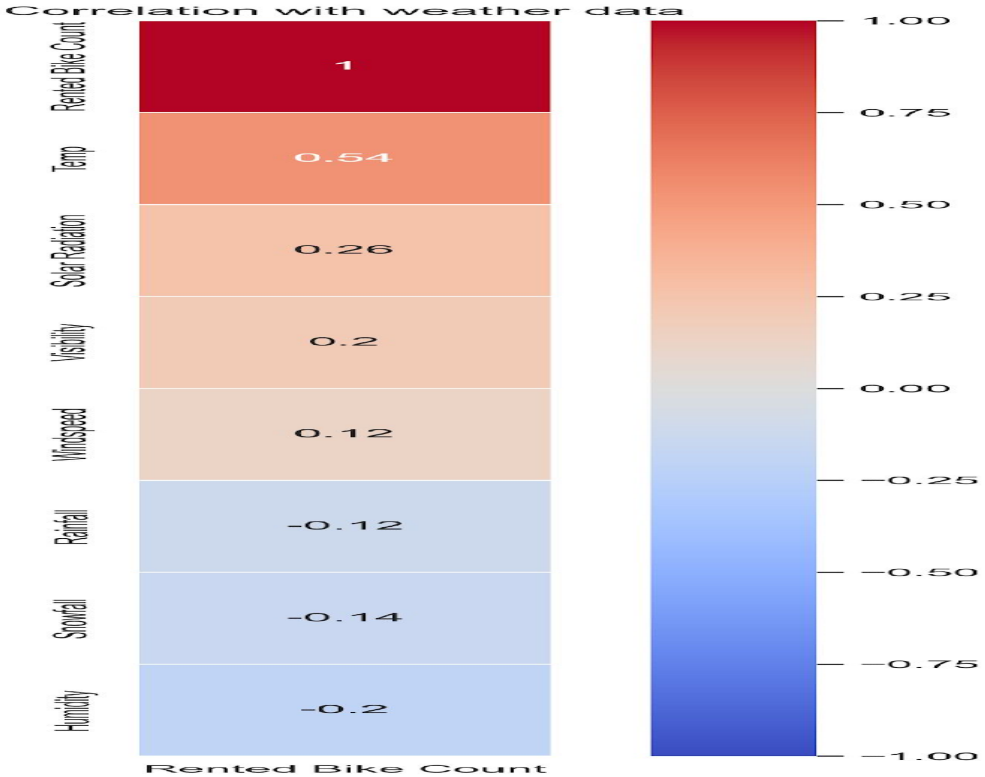


The following plots show the variations of ridership with weather conditions. As we know the ridership is high during the clear climate and decrease as the climate gets worse, this can also be inferred from the graph. An unusual trend can be noticed with the temperature, where we notice people preferring higher temperatures, than pleasant or low temperatures. A sudden drop can be observed after a certain value of temperature.





With favorable wind and humidity conditions, the number of ridership increases which is also expected. So we can say, favorable weather conditions help in increasing the number of bikes rented in a day.



The strongest correlation for bike rent is with the temperature, followed by the solar radiation and visibility while negative with humidity, snowfall and rain fall which makes totally sense. This could help to build a bike rental predictions model later.

Based on this quick exploratory analysis, we can see that people in Seoul are taking advantage of the soft mobility offered the bike rental system. We see peaks in Spring and Summer, as well as during the commuting hours in the morning and in the afternoon.

CONCLUSION

I tested a wide range of alphas for both Ridge and Lasso regressions, using cross-validation for both regression types. Ridge regression with an alpha=50 turned out to have the smallest MSE and highest R^2. Lasso regression also performed well with an alpha=7. Normalization of the data could've proven to help create a better model but wasn't part of our task.