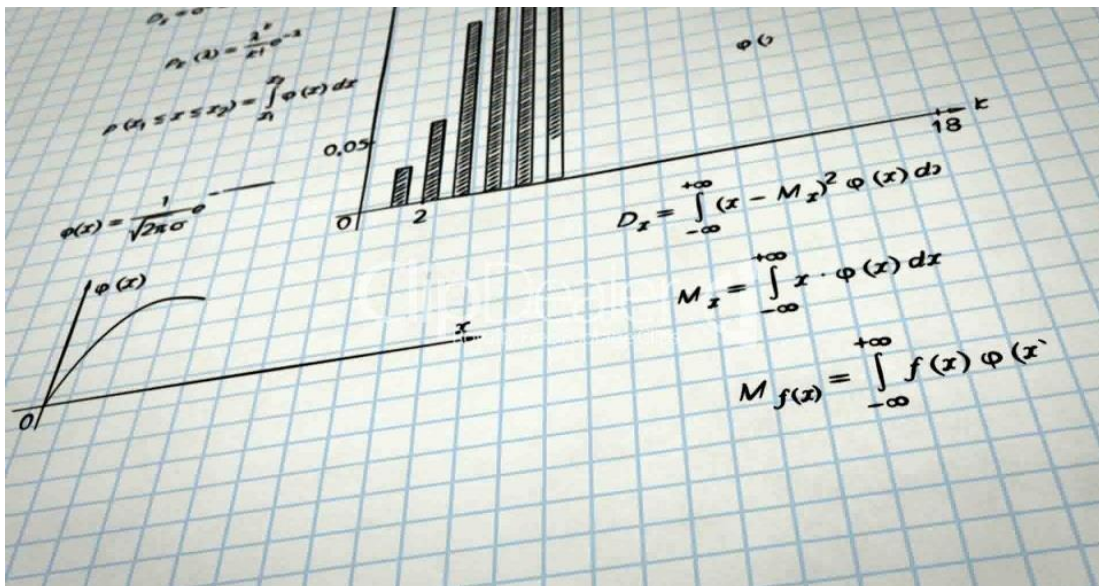**Statistical and Analysis project report on Bike Sharing Systems**

Submission Date: April 4th, 2017

Course: SC/MATH 1131: Introduction To Statistics

Submitted to: Professor Xin Gao

Department of Mathematics and Statistics

Project Report

Report prepared by:

Nafis Ahmed Awsaf

PROJECT REPORT

1. Introduction:

This report summarizes the statistical and analysis results on bike sharing systems. It shows the usage of bikes based on different temperatures over the 2 years 2011 and 2012. The purpose of this report is to predict the bike rentals based on the temperature and test the difference between bike rental from casual and registered users. A lot of variables come into play of the number of bike rentals such as time of the day, season, weekends, working days etc.

A data set of bike rentals over a 731 days period through different seasons and temperatures are given. The data set used for this project is from the source: Laboratory of Artificial Intelligence and Decision Support (LIAAD), University of Porto.

MY GOAL: PREDICT THE BIKE RENTAL BASED ON TEMPERATURE AND OTHER VARIABLES (humidity, wind speed, weather conditions, holidays, etc)

POINTS INCLUDED IN THIS REPORT:

- Testing the relationship of bike rentals with temperatures across different seasons.
- Testing the relationship of bike rentals with different weather conditions.
- Testing if different levels of humidity or wind speed affect number of bike rentals.
- Analyzing if holidays are a good predictor for bikes rented.
- Linear Regression Model of total bikes and temperature.
- Difference between Bike Registered Users and Bike Casual.

There are four seasons across one year. The temperatures vary across each season. The mean, median and standard deviations of each season are given below in (normalized temperature, converted in C).

There are four different weather conditions over the year: clear, misty and cloudy, light snow and thunderstorm and heavy snow.

2. ANALYSIS

The relationship of Bike Rentals with temperatures across different seasons:

Mean = 12.2

Temperatures Spring in C
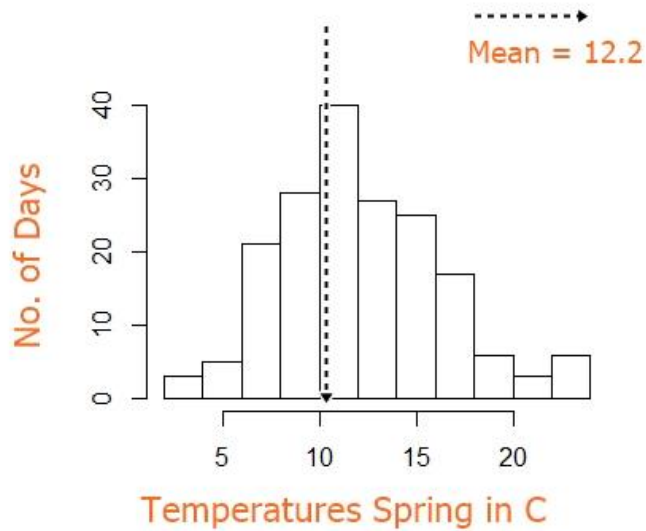
SEASON 1	TEMP SPRING (in C)
MEAN	12.2
MEDIAN	11.7
STANDARD DEVIATION	4.2

As we can see from the histogram, the mean temperature of spring is 12.2C. The mean number of bikes rented in spring season was 2605. However if we observe the summer season, the mean temperature is very high of 22.3C.

Mean = 22.3

Temperatures Summer in C
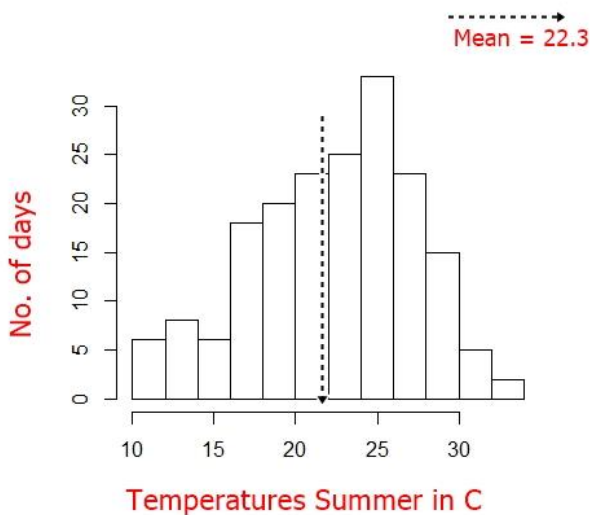
SEASON 2	TEMP SUMMER (in C)
MEAN	22.3
MEDIAN	23.0
STANDARD DEVIATION	5.0

The mean number of bikes rented in summer season was 4992. Because of the very good temperature in summer, the number of bike rentals was much higher than the spring season.

Mean = 29.0

No. of days

Temperatures Fall in C
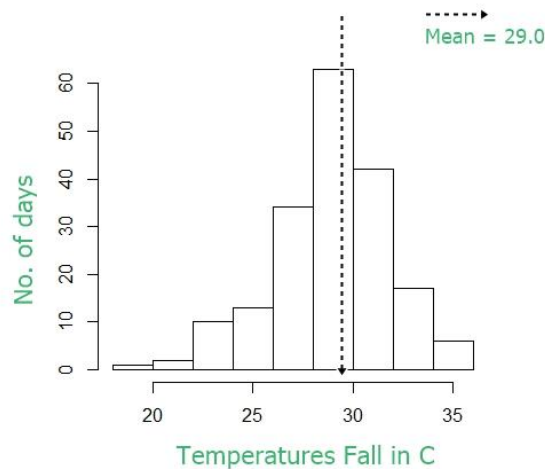
SEASON 3	TEMP FALL (in C)
MEAN	29.0
MEDIAN	29.3
STANDARD DEVIATION	2.9

The mean number of bikes rented in the Fall season was 5644. The mean temperature was 29C. This is the best season as it has the highest number of bike rentals and very good temperature.

Mean = 17.3

No. of days

Temperatures Winter in C
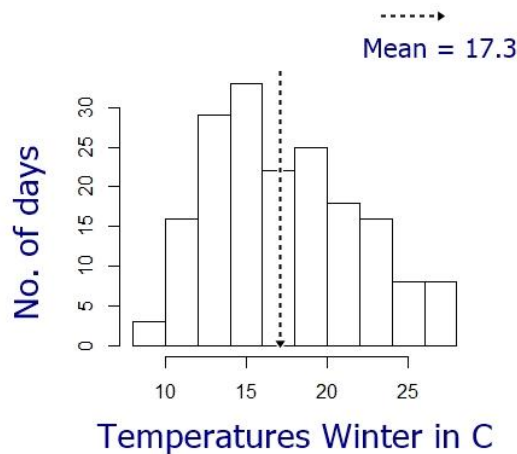
SEASON 4	TEMP WINTER (in C)
MEAN	17.3
MEDIAN	16.7
STANDARD DEVIATION	4.4

By analyzing the four graphs and tables, we can detect that the temperature has been quite high throughout the whole year. By comparing the seasons, highest mean temperature was 29C in fall. The lowest mean temperature was 12.2 C in spring. Temperature has been in the range $(12.2 < x < 29)$ C. Each season has different temperatures. Winter (cold weather) and summer (high weather), spring are very good variables that shows there is a relation between the total number of bike rentals and temperature. We can easily see this from the data set.

On April 4th 2011, instant =94, season=2 (summer), converted temperature = 23.5C the total number of bikes rented was 3115. However on 1st January 2011, instant =1, season =1 (spring), converted temperature = 14.1 C, the total number of bikes tented was only 985.

 We can see that temperature is in fact related to number of bike rentals. Therefore, good weather conditions indicates a higher number of bike rentals compared from the other seasons. From the histogram we can analyze that season =3, fall has had the best temperature with a mean of 29C. If we go through the data set of season 3, they have the highest number of bikes

rented in the fall season ranging from 4000 to a max of 6000. Although in some days of season 3, there have been low bike rentals of 1000 due to weather conditions such as snow and rain.

Mean number of Bikes Rented under different Weather Conditions:

Weather 1 = Clear

Weather 2 = Mist and Cloudy

Weather 3 = Light Snow and Thunderstorm

Weather 4 = Heavy Snow

The mean number of total bike rentals under weather 1 was 4877.

The mean number of total bike rentals under weather 2 was 4036.

The mean number of total bike rentals under weather 3 was 1803.

There were no rental bikes under weather 4.

From here we can analyze that the number of bikes rented occurs mostly in good weather conditions such as clear, misty and cloudy. However on poor weather conditions such as snow and thunderstorm, the amount of bikes rented is very low with a mean value of only 1803.

The pie chart below summarizes the percentage of bike rentals under different weather conditions.
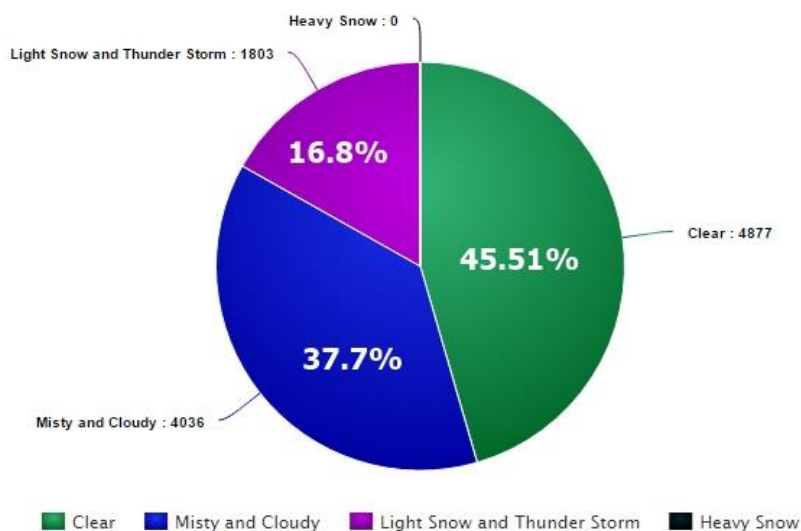
From the following pie chart, 45.51% of total bikes counts were rented under clear weather conditions. 37.7% of total bike counts were rented under misty and cloudy. There were no Bikes rented under snow and lowest of 16.8% under light snow conditions.

<u>Linear Regression Model</u>

Prediction graph of the number of bike rentals associated with bike temperature.
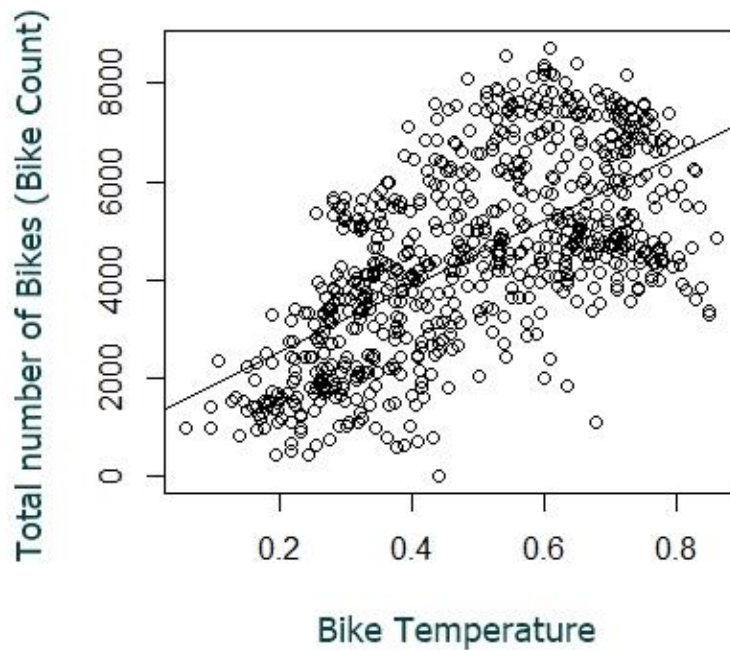
From the graph we can analyze that high number of bike rentals is associated with good temperature. As we can see in the plot, as soon as the temperature increases the amount of bike rentals increase.

Therefore based on temperature we can assume a rough estimate of the number of bikes that will be rented on that certain period. The plot confirms the strong role of temperature.
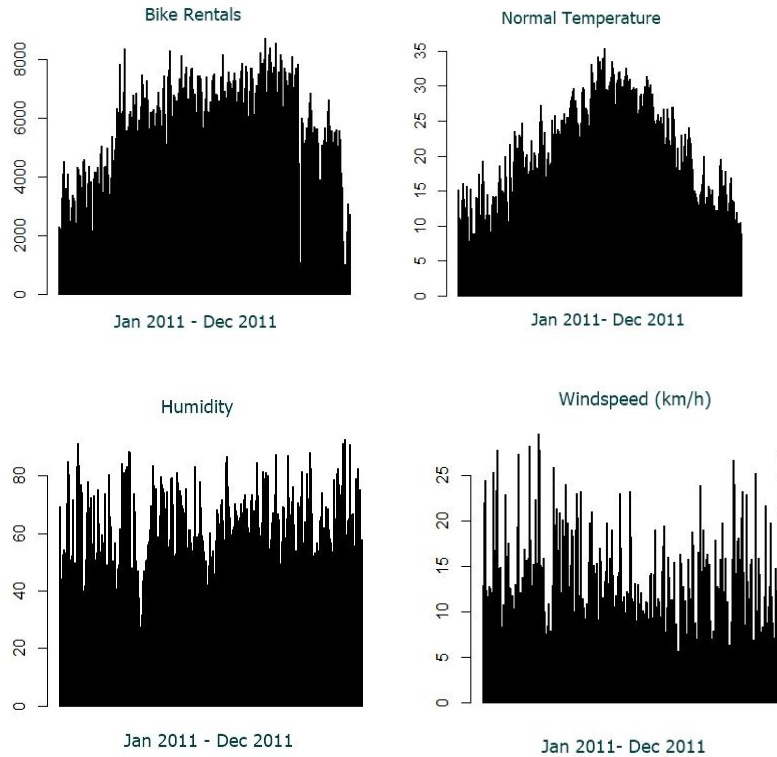
Coefficients: (Intercept)

Bike$temp 1215 6641

So the linear regression model (prediction model) is: Bike$cnt=1215+6641*Bike$temp

This model can be used for future prediction of bike demand

# Testing if humidity and wind speed affects the number of bikes rented

BAR PLOTS of bike rentals, normal temperature, humidity and wind speed over the 12 month period:

By analyzing the four graphs, it can be concluded that humidity and wind speed does not affect the number of bike rentals significantly. Therefore the number of bike rentals is independent over humidity and wind speed is due to the fact that humidity levels are almost constant over the full 12 month period. We can also see that high temperature levels (good weather) are related to high number of bike rentals.

Histogram of total Bikes (casual + registered)


Mean = 4504.35
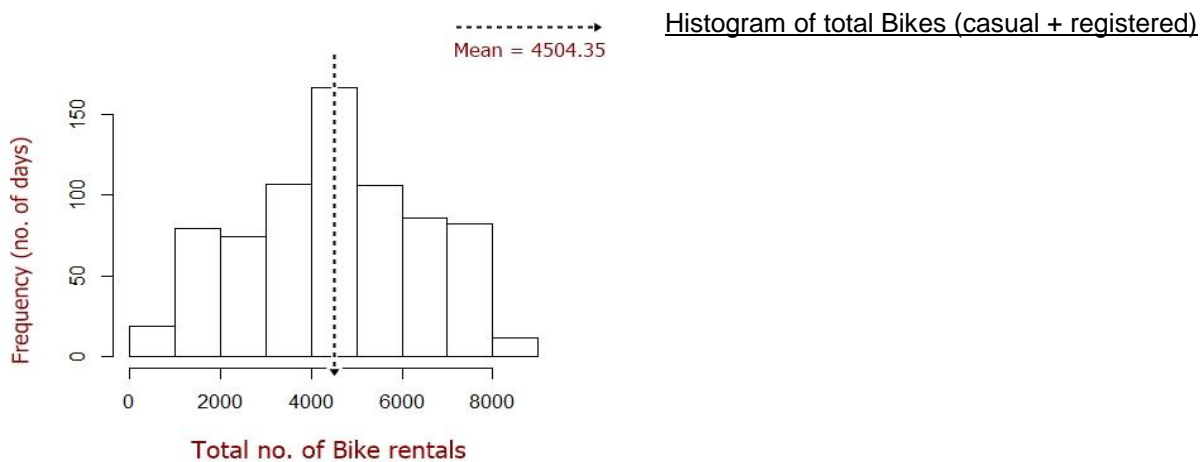
Testing if holidays, weekdays are good predictors of the number of rental bikes.
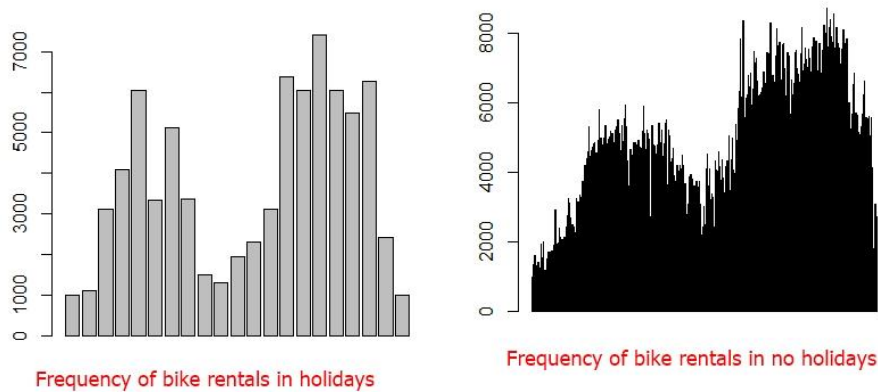

It seems holidays have no impact on the number of bikes rented. From the above 2 graphs, we can see the frequency of bike rentals is quite low when it is a holiday. But on the days which are not holidays the frequency is very high. Therefore holidays are not a good predictor for bike rentals.

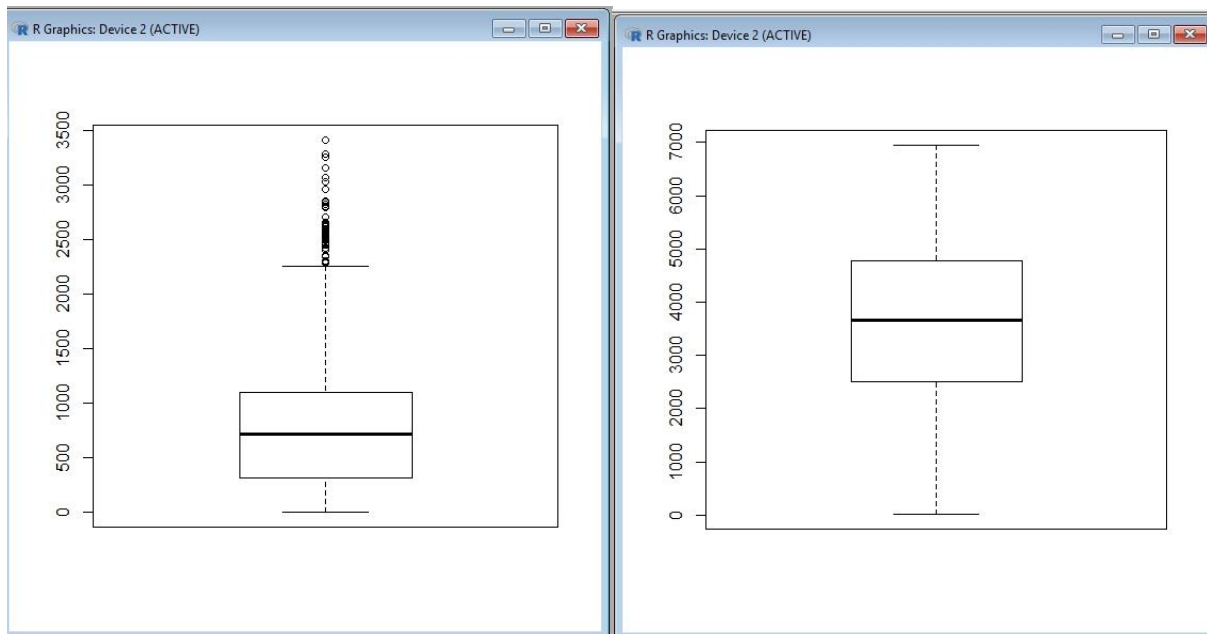## Difference of casual and registered bike users rentals:
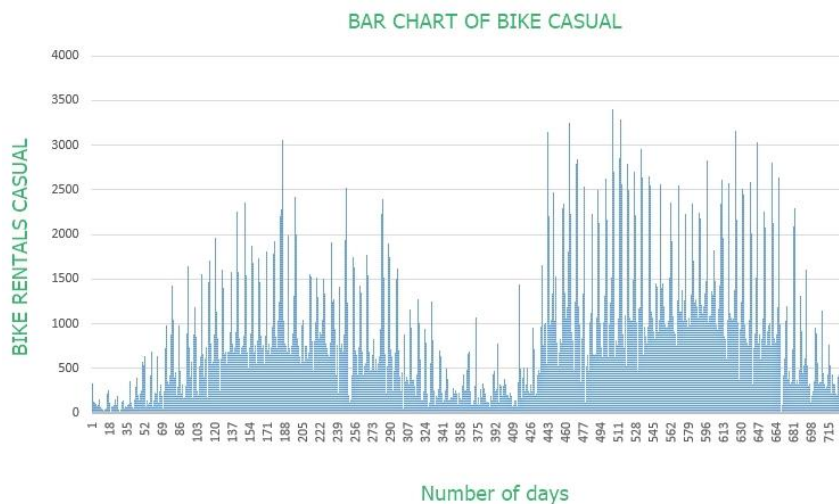
Maximum number of bike casual rentals was 3410 over the full 2 year period.

BAR CHART OF BIKE REGISTERED
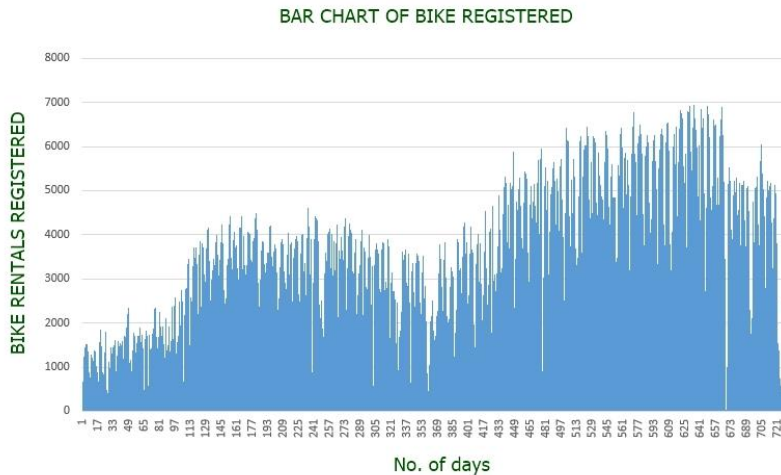
Maximum number of bike registered rentals was 6946 over the full 2 year period.

The following box plot also highlights the differences between Bike Registered users and Bike Casual users:

BOX PLOT OF **BIKE CASUAL**   BOX PLOT OF BIKE REGISTERED

The vertical axis of the bike casual shows to a number of 3500. Maximum number of bikes rented under casual users were inside the region of the rectangle in the box plot.

The vertical axis of the bike registered shows to a number of 7000. Maximum number of bikes rented under registered users are inside the rectangle.

Registered users box is more larger than casual users. Also, there are no outliers for bike registered users, but there are outliers from bike casual by comparing the above two graphs

T Test of Bike Casual vs Bike Registered

```
> t.test(Bike$casual, Bike$registered)

        Welch Two Sample t-test

data:  Bike$casual and Bike$registered
t = -44.537, df = 1002.5, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2931.719 -2684.273
sample estimates:
mean of x mean of y
 848.1765 3656.1724

> |
```

The mean of Bike Registered users is greater than the mean of Bike Casual Users.

3. Conclusion:

As a conclusion we can say, that the amount of bike rentals depends mainly on the weather and on the normal temperature. The analysis shows that there is a positive relationship between the amount of bike rentals and temperature because on average throughout the whole year temperature has been quite well. Wind speed and humidity levels had no great changes on the number of bikes rented. Holidays are not a good predictor for bikes rented, as we have tested 95% of the bikes were rented on days which were not holidays.

 Furthermore, we have seen weather conditions are a big factor in determining the number of bikes rented. From the above pie chart, we have seen majority of bikes have been rented on clear weather and on the other hand, no bikes were rented in heavy snow. This clearly shows that good weather conditions and temperature is a very good indication for business. From the linear regression graph we can see bikes increase with good temperatures.

Therefore, the mean amount of bike rentals increases and decreases with the temperature. So people mainly rent bikes on days with good temperature and clear weather conditions. This is a very important analysis to be taken into account for the bike rental stations, so that they have a very good estimates of their data. Accurate data is very good business as it gives them the good estimates of their profits and cash flow analysis.