CCT College Dublin

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Machine Learning |
| **Assessment Title:** | CA1 Project |
| **Lecturer Name:** | Dr. Muhammad Iqbal |
| **Student Full Name:** | Izaias de Oliveira Gomes Junior  Salvador Navarro |
| **Student Number:** | 20 |
|  |  |
|  |  |
|  |  |
|  |  |
| **Assessment Due Date:** | 26th Nov 2023 |
| **Date of Submission:** | 26th Nov 2023 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

## Introduction

Have you ever wondered how some cities are moving from cars/buses for a more environment transport such as Bikes? Nowadays it is more common to see people using a bike to explore the city, move from one side to another and now there are companies around the world renting bikes to promote a healthy and green environment! In this project we were given 3 topics to study (Covid-19, Transport and Crime) Transport have a big impact for both members of the team, we have had previously worked in logistics (transport areas) in our countries and we decided to choose it because transport is such an important topic to speak about. In the effort of many companies, goverments around the world to reduce the high comsumption of oil, many of these groups have installed areas where people can rent a bike 24/7 making this more accessible to the community, reducing the traffic and making the roads less crowder.

Maybe we might find that people rather use the bike between 5-7pm when most of them finish work and to avoid traffic they better use a bike, it is well know around the world that peak traffic in most of the cities is either they start or finish their work.

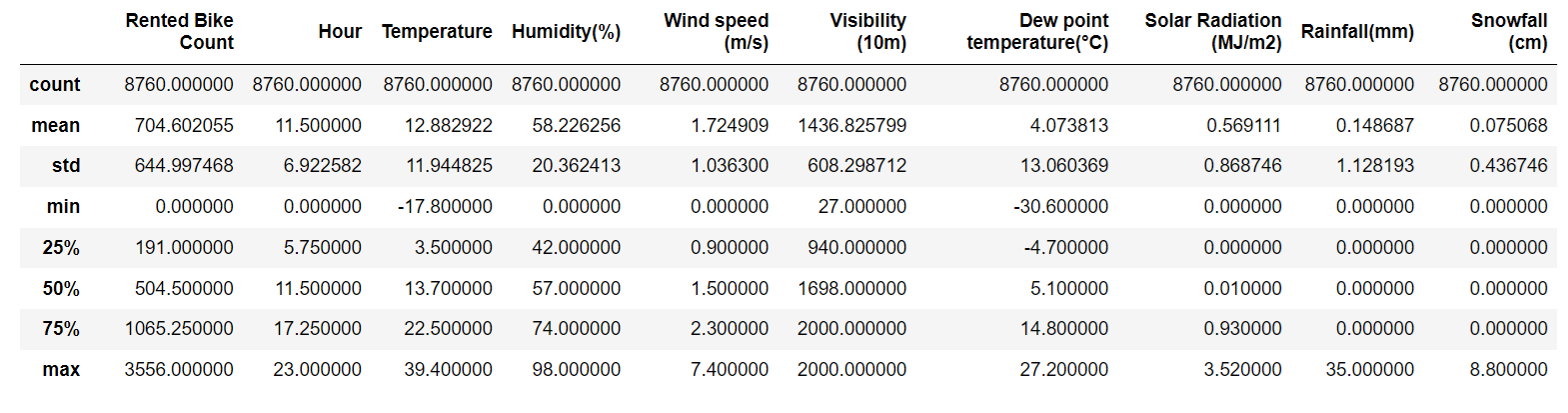
“Traffic congestion is getting worse in Seoul as the city and central governments fail to come up with effective measures.  
  
It has become a serious problem in Seoul, as well as Gyeonggi Province and Incheon."

“Even if the commute is shorter via public transportation, many commuters choose to go to work by car to avoid standing in cramped spaces on buses or subways.” Said Lee Suh-yoon

https://www.koreatimes.co.kr/www/nation/2018/05/281\_249206.html

We decided to choose this data set because we wanted to explore where exactly in the day people tends to use the bike more, if these decisions are affected by external facts such as the temperature, snow, rain etc and how this can impulse people to start using the bikes.

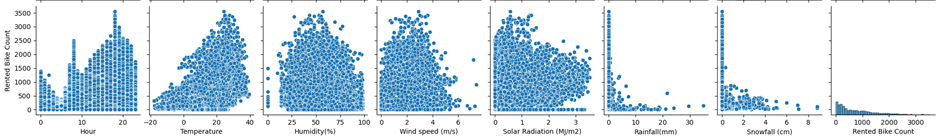
Analyzing the information.



* As first sight we can see that our mean per hour in our Renkted Bike Count is 704.6
* Our max per hour is 3556
* There are no missing values in our count.
* Temperature min in this period of time were -17.80 while max 39.40.

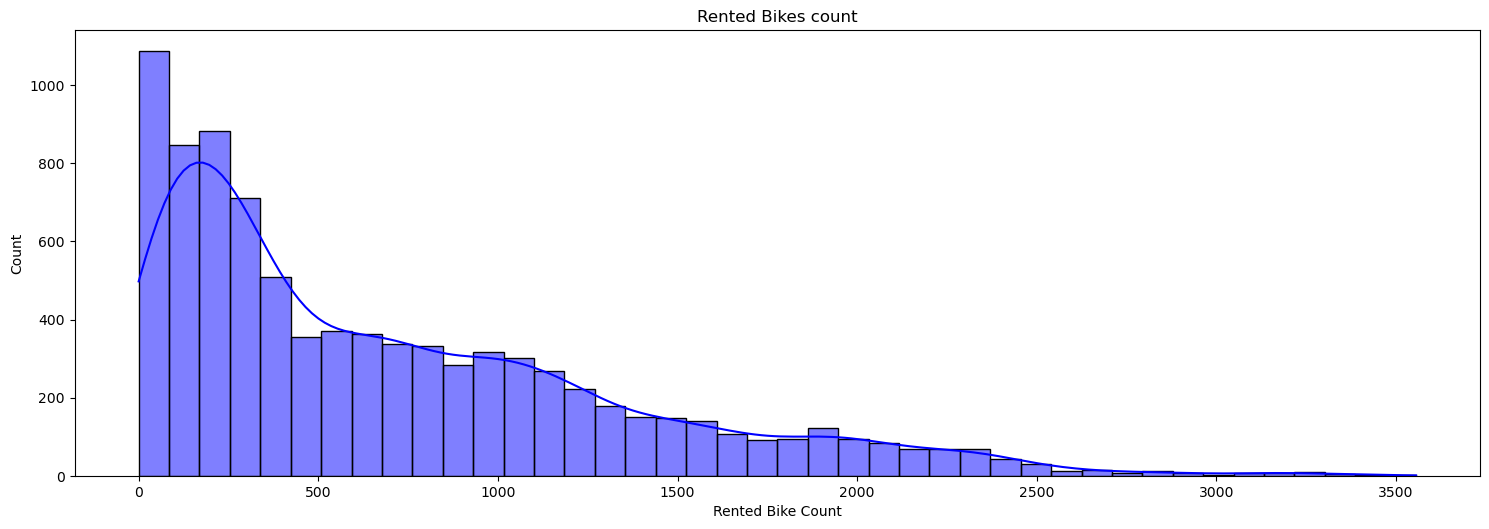
These are just a few appreciations that might be use for us in the future.

Correlation in Rented Bikes

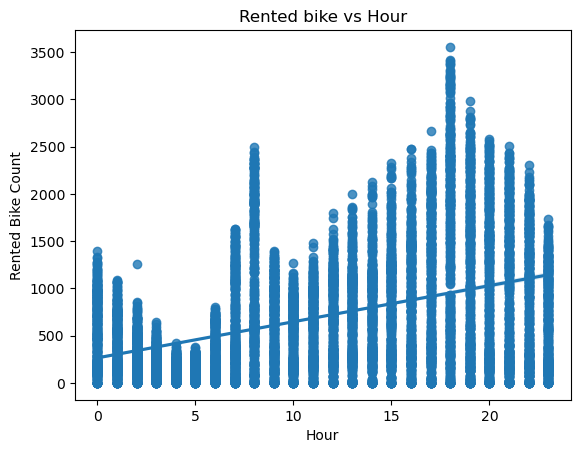


* We can see a good correlation between hour , temperature
* In the Hour it goes up in the evening reaching the peak around 18 hour
* As the temperature is warmed the more bikes are rented.

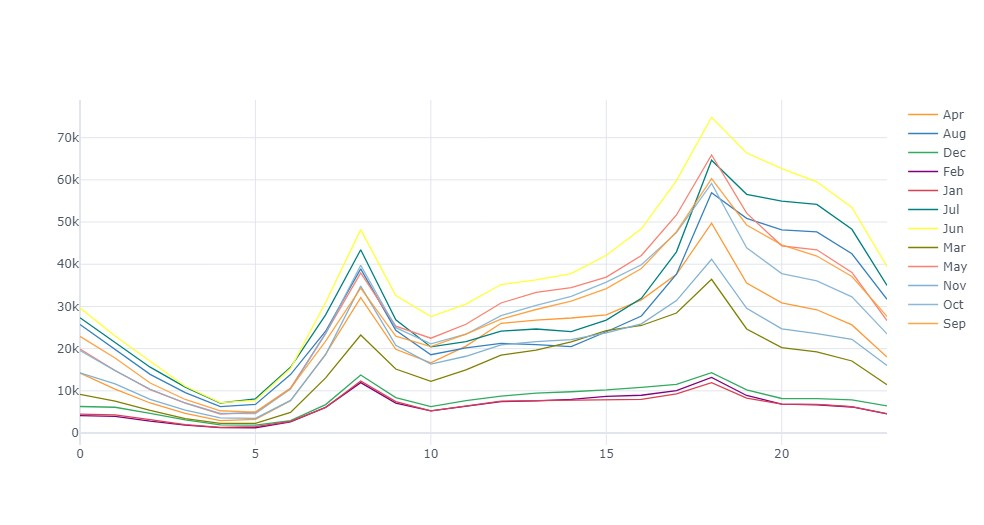
Histogram

In the following chart we can see that most of our data are skewed positive to the left, this is because most of our data are concentered in the evening

Plot Rented Bike vs Hour

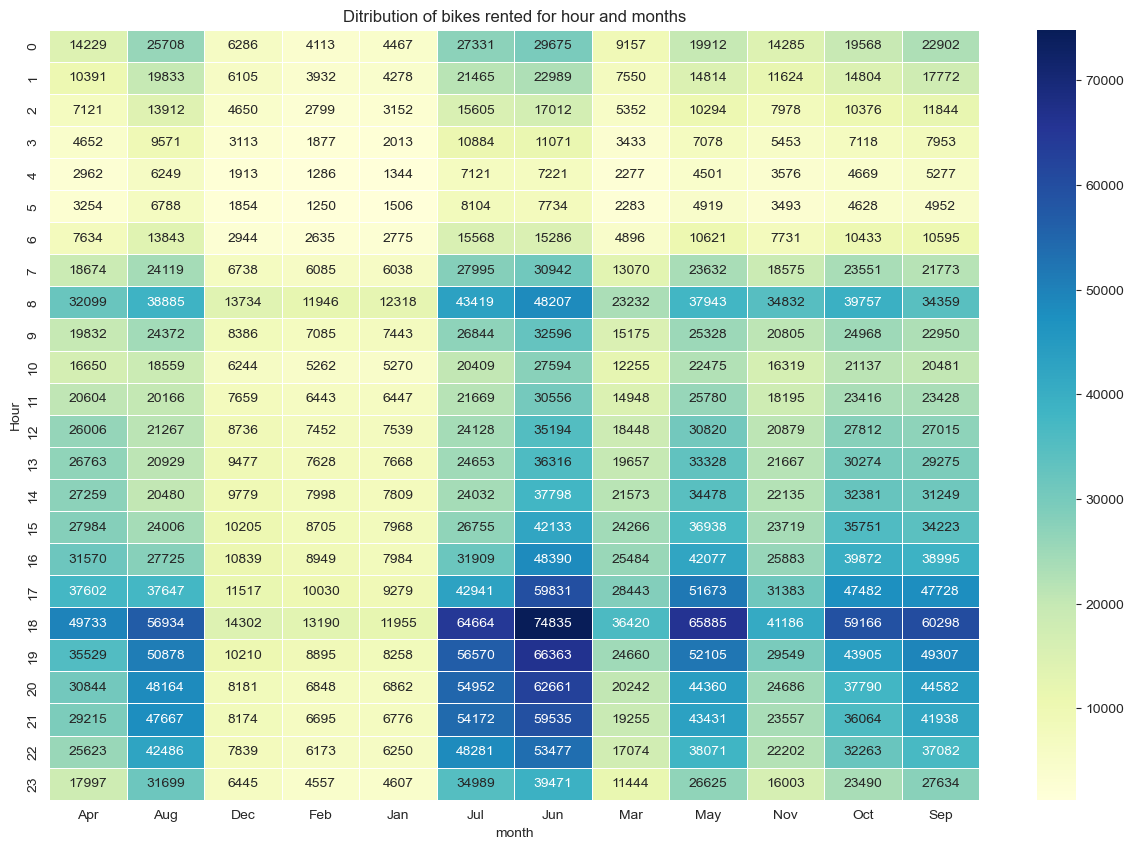
Here in the next graphic we can see that these data we have rented 3500 bikes in the hour 18:00. This is considered by the chart as an outlier in the 24 hours of the day this is where we have the most rented bike per hour.

Interactive Graphic

We graphic the hours by months we can see that the yellow line in the month of June, hour 18 was our most renkted bike per hour.

Please explore with our dynamic graphic in our jupyter notebook.

Correlation between Hour / Month

We can clearly see that in the month of June, hour 18:00 we have the most amount of rented bikes with 74,835 in the month.

|  |  |  |
| --- | --- | --- |
| Month | Hour | Total |
| Dec | 18:00 | 14,302 |
| Feb | 18:00 | 13,190 |
| Jan | 08:00 | 12,318 |
| Mar | 18:00 | 36,420 |
| Apr | 18:00 | 49,733 |
| May | 18:00 | 65,885 |
| Jun | 18:00 | 74,835 |
| Jul | 18:00 | 64,664 |
| Aug | 18:00 | 56,934 |
| Sep | 18:00 | 60,298 |
| Oct | 18:00 | 59,166 |
| Nov | 18:00 | 41,186 |

* Here we can see that in 11 months except in Jan, the hour where we rent most of the bikes is at 18:00 with our highest rent of 74,835 in the month of Jun.
* Jan peak hour was 08:00 with the total count of 12,318.

Season Hourly Distribution of counts

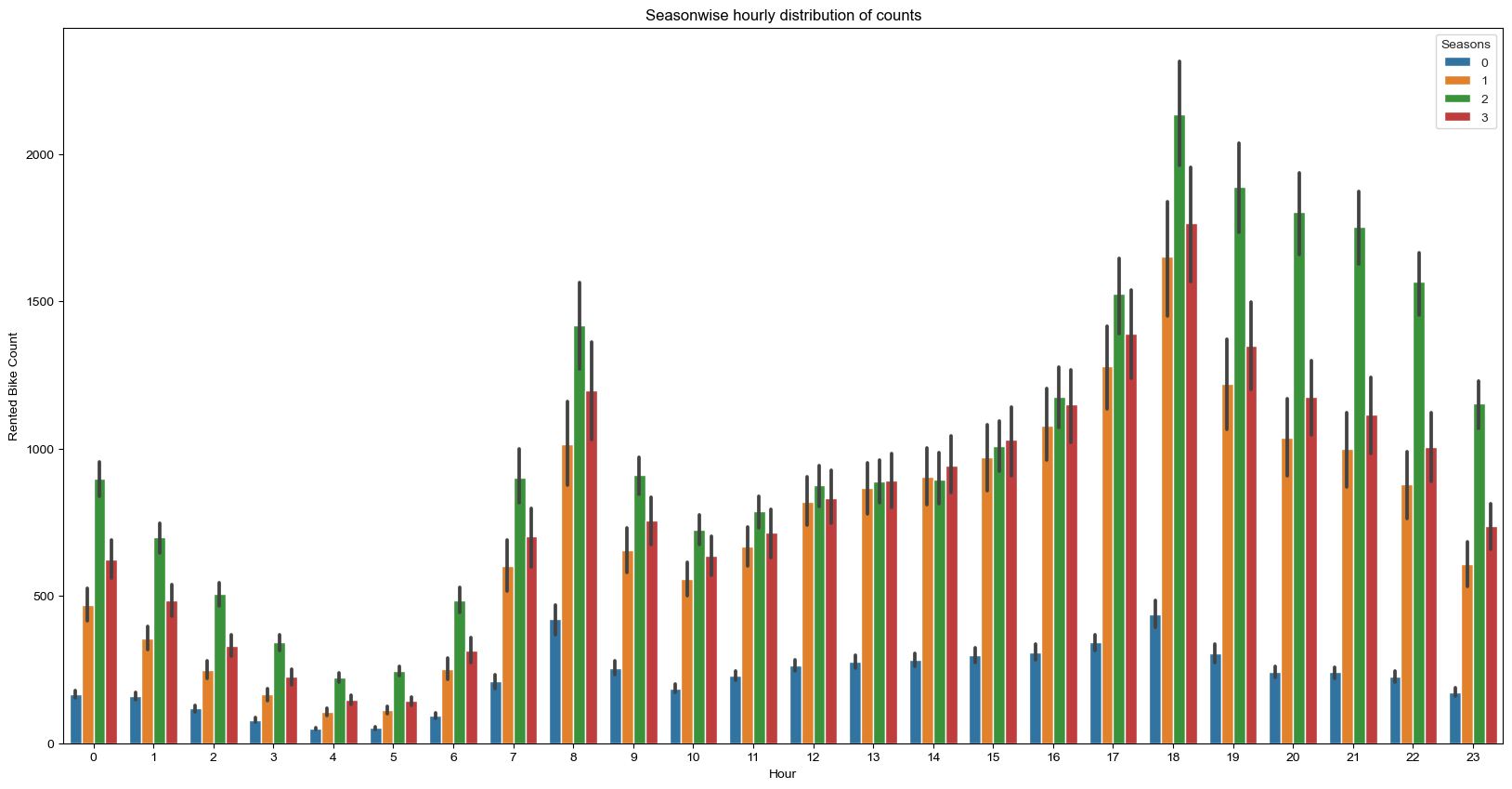
In the next plot we can see the Hourly distribution by seasons and we can again confirm that in summer the 18:00 is the most import, but between 13:00-15:00 people use them more in spring. And winter is by far the lowest season people tends to use the Bikes due the weather conditions

Winter:0 = Blue Line

Spring:1 = Orange Line

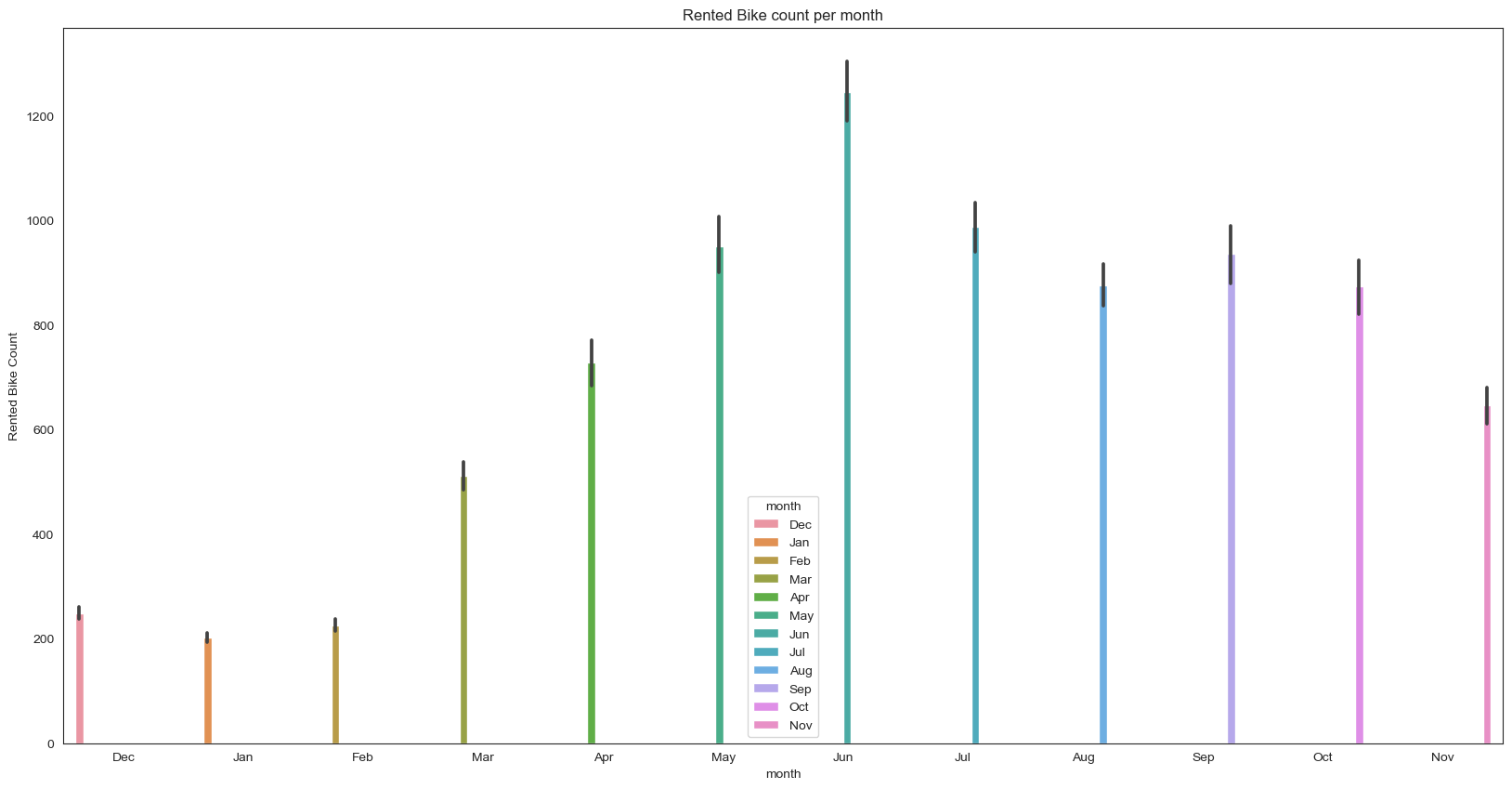
Summer:2 = Green Line

Autumn:3 = Red Line



Rented Bike Count per Month

Clearly we can see that in the summer is where most of the people tends to use the bikes.



Outliers

I know that variable is skewed and it has outliers but these outliers occur in the hour 18:00 when is our peak hour. Saying that they wont be remove.

