

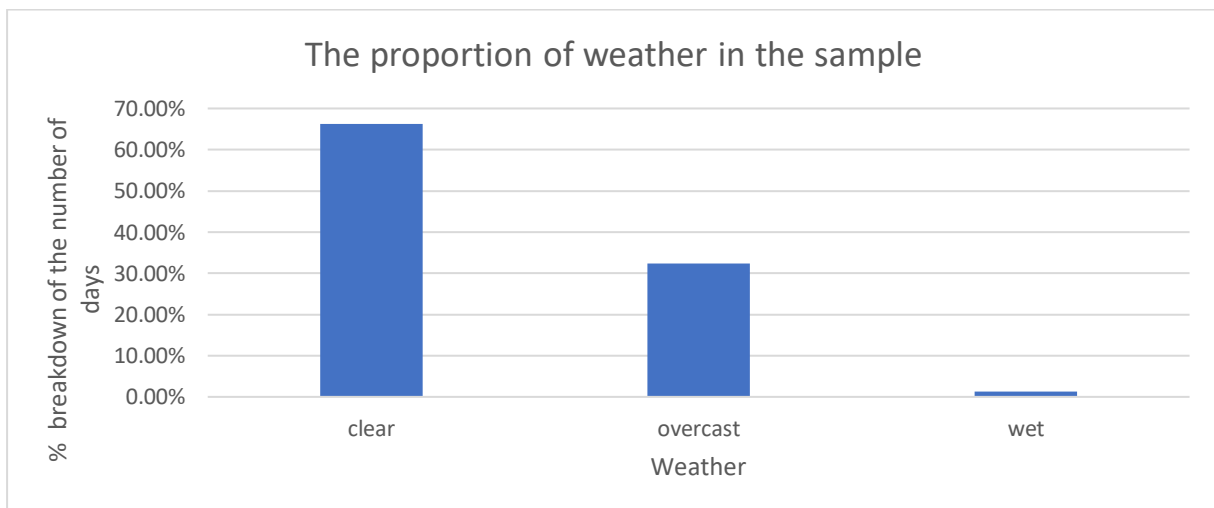
MA1670 Final Excel Assignment

When I write this report, I am going to use the data about the bikes in the bike sharing system in Washington DC between 2011 and 2012. As I write the report, I will talk about the number of days of the weather and the variability between some of the variables. My birthday is on 22nd December. The subset of the data that is going to be used starts from the date of my birthday in 2011 to the data of my birthday in 2012.

Firstly, there is data for three different weathers in my sample. From the two pivot tables on the 'Question 1' sheet of Excel, there is only data for the clear weather, overcast weather and the wet weather. On the pivot table that talks about the number of days of the weather, there is data such as 243 days of clear weather and 5 days of wet weather. These facts show that there are no signs of stormy weather in the sample which also proves my point.

Weather	Number of days
clear	243
overcast	119
wet	5
Grand Total	367

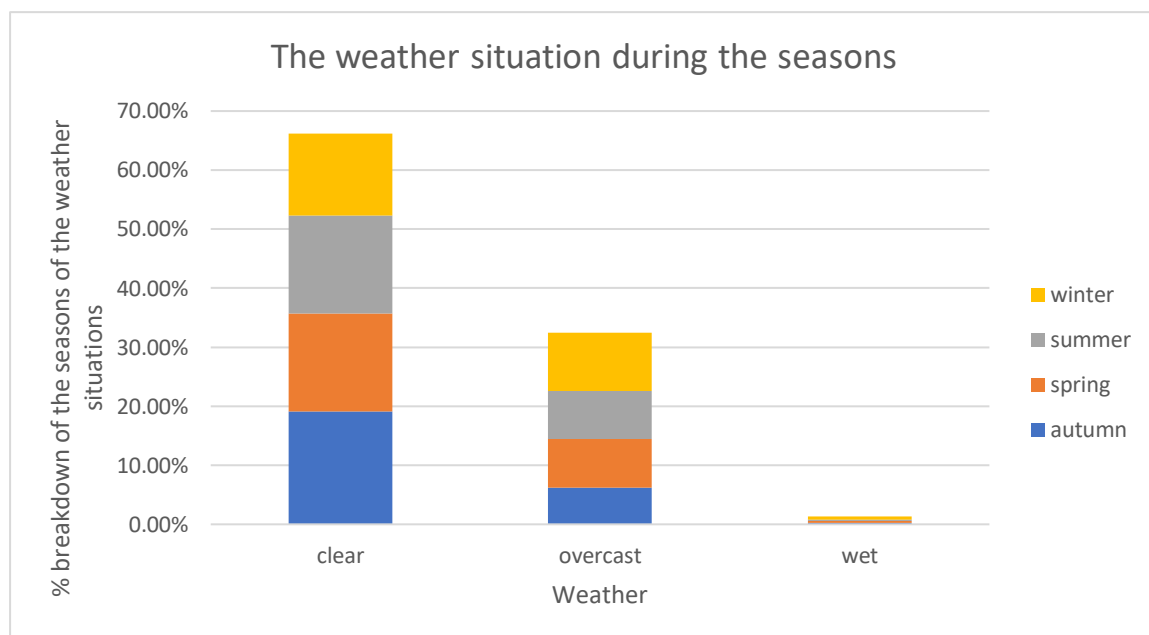
I have concluded that the weather is mostly clear in my sample. On the bar chart on the 'Question 1' sheet, the bar is at its highest at value closer to 70% when the weather is clear. It implies that the weather is mainly clear in the sample. The length of the bar at clear proves that the weather is clear most of the time.



Secondly, there is a variation between the weathersit and the seasons. By looking at the table about the number of days of each season for each weather situation on 'Question 2' sheet on the Excel data, the outputs of the percentage of days of season for each weather situation is at its highest when the weather is clear. This implies that clear weather has the highest percentage number of days and wet weather has the lowest percentage number of days for each season, which explains how there is a variation between the two variables.

The weather situation during the seasons	Season				
Weather	autumn	spring	summer	winter	Grand Total
clear	19.07%	16.62%	16.62%	13.90%	66.21%
overcast	6.27%	8.17%	8.17%	9.81%	32.43%
wet	0.27%	0.27%	0.27%	0.54%	1.36%
Grand Total	25.61%	25.07%	25.07%	24.25%	100.00%

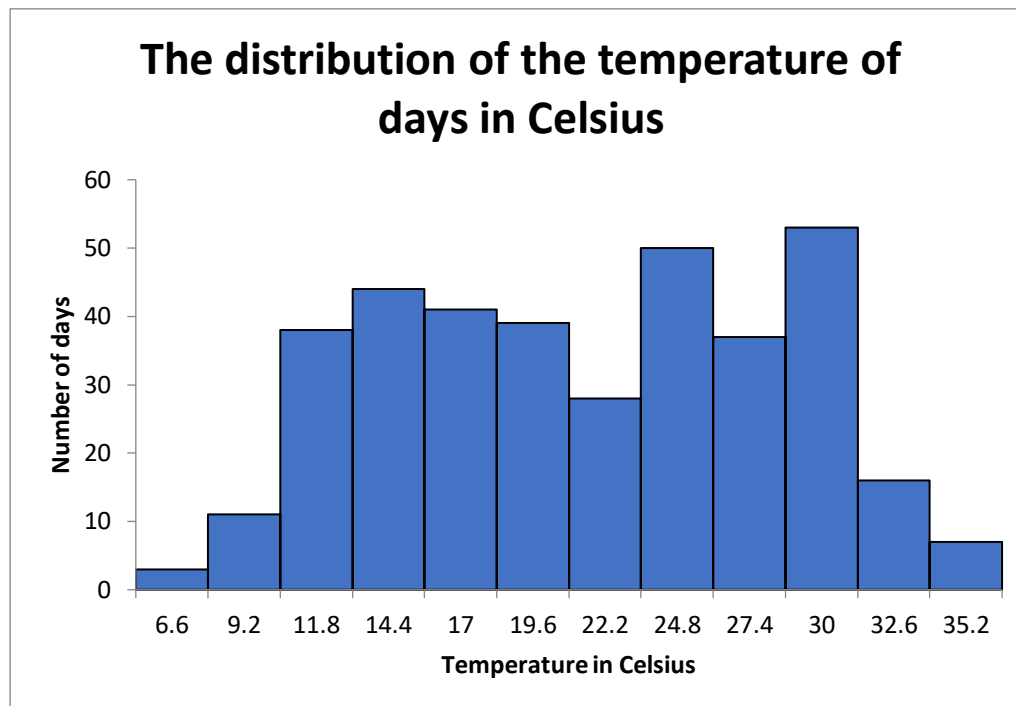
Moreover, I think that the weather situation can vary in the different seasons in my sample. In the 'Question 2' sheet, the clustered bar chart shows the percentage number of days of each season for each weather. This shows that the likelihood of a certain weather taking place can be different for each season.



Thirdly, the temperature for the location in my sample was quite warm. From the descriptive statistics table on 'Question 3' sheet in excel, the output for the mean temperature is 20.23 and the output for median temperature is 20.56, which indicates the weather has been reasonable during this period and it explains how warm it can be at a certain temperature.

Additionally, the values of the lower quartile and the upper quartile are within the temperature range. On 'Question 3' sheet, the value of the Lower Quartile is 13.89 and the value of the Upper Quartile is 26.11 which is in the range between 4 and 35.2, which implies that the values are reliable. The shape of the distribution is bimodal as the number of days tend to increase or decrease at different stages. This implies that there is a variation in the data and the distribution is true. The standard deviation could be used to represent the temperature because it is used for the dispersion of the variables as seen on the histogram on 'Question 3' sheet in the excel data, which illustrates how the variation can be different. The temperature does vary in my sample as the number of days increase and decrease, which shows implications of the size of the bars changing at different temperatures and there is still

a variation. The mode could be used to describe the variability of temperature as temperatures up to 30 have the highest number which indicates a large variation.



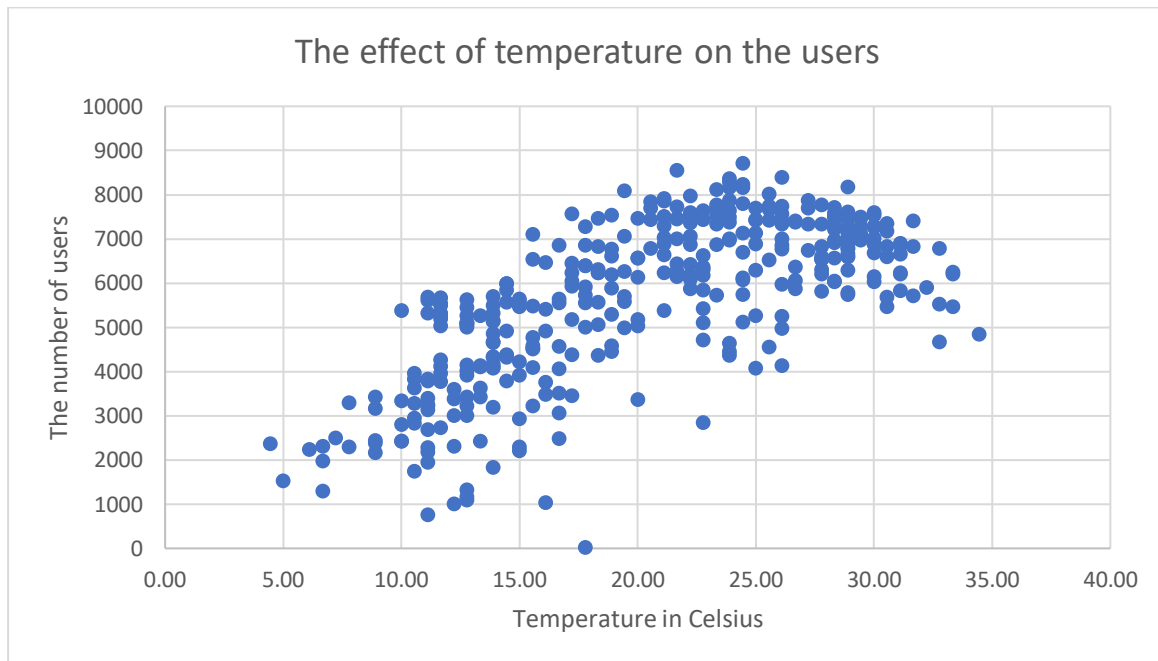
Next, there is a lot of change in the number of users per day for each season. On the table in the 'Question 4' sheet of the data, the outputs of the average for each season are different such as the average for autumn is 6824.24 and the average for spring is 3541.76, which implies that the change in the average number of users is evident. This leads to the conclusion that there is a change in the average value of users each season.

In addition to this, the number of users on average for each season are not similar. From the table on the 'Question 4' sheet of the excel data, the change in the average number of users have a big difference, which implies that the number of users for each season are different. It shows that there is not much similarity of in the number of users per day for each season.

Season	Average number of users
autumn	6824.24
spring	3541.76
summer	6209.49
winter	5791.87
Grand Total	5596.92

I believe that there is no linear relationship between the number of users and the temperature. On the X-Y plot in the 'Question 5' sheet of the Excel data, the data is scattered around the graph. This implies that it is quite difficult to work out the product moment correlation coefficient to see whether the correlation is positive or negative. This proves that there is no variability between the temperature and the number of users.

There is no evident relationship between the temperature and the number of users. I will use the data points from the X-Y plot to prove this which suggests that some of the temperatures may have the same number of users, which shows how non-linear the relationship between the variables is.



Finally, we can conclude that there is a large variation between all the variables.