**1.0 PROBLEM STATEMENT**

The dataset that i’ll be working on is an Autolib electric car sharing service company. The electric car sharing service company operates in different areas of Paris and has recorded their operations of their three different electric cars: BlueCars, Utilib and the Utilib 14.

I would like to know whether there is a significant difference that BlueCars that were taken on Monday were less than that of Friday. Despite the difference in means, it is really interesting to find out that Monday has an equal number of BlueCars being taken because of the surge of numbers of people going to work on Monday hence the need for a car sharing service.

I created a null hypothesis that there is no difference between the two days and the alternative hypothesis that there is actually a difference in the two days.

H₀ μ(Monday) = μ(Friday)

H₁ μ(Monday) **≠** μ(Friday)

**2.0 DATA DESCRIPTION**

The dataset contains 16,085 entries defined by 13 columns which are:

* Postal Code which is the postal code of the different areas in Paris
* The Date which records the date where the action was taken.
* N\_daily\_data\_points which is the number of daily data points that were available for aggregation that day.
* Day of the week which is labelled as: 0 for Monday, 1 for Tuesday, 2 for Wednesday, 3 for Thursday, 4 for Friday, 5 for Saturday and 6 for Sunday.
* The day type which is either weekday or weekend
* BlueCars\_taken\_sum which shows the number of blue cars taken that date in that area.
* BlueCars\_returned\_sum shows the number the number of bluecars returned that date in that area.
* Utilib\_taken\_sum shows the number of Utilib cars taken that date in that area.
* Utilib\_returned\_sum shows the number of Utilib cars returned that date in that areA
* Utilib\_14\_taken\_sum shows the number of Utilib 14 cars taken that date in that area.
* Utilib\_14\_returned\_sum shows the number of Utilib 14 cars returned that date in that area.
* Slots\_freed\_sum shows the number of recharging slots released that date in that area.
* Slots\_taken\_sum shows the number of recharging slots taken that date in that area.

The mean of the BlueCars taken is 116. The mean of the BlueCars returned is 115.

The mean of Utilib Cars taken is at 3. The mean off Utilib Cars taken is 3.

The mean of Utilib 14 cars taken is at 8. The mean of Utilib 14 cars returned is at 8.

The number of Utilib and Utilib cars picked is rather low than the BlueCars. Hence the BlueCars seem to be making more revenue than the others making the reason for making it the variable of interest.

The data was sourced from the Moringa School Learning Management System hence no information on the data collection procedures.

**3.0 HYPOTHESIS TESTING PROCEDURE**

I’ll be investigating the weekdays hence I’ll create a new dataset with weekdays alone.The number of rows in the dataset reduced from 16,085 to 11,544 rows therefore it became our new population.

I created a new dataframe with Mondays alone which had 2313 rows. I decided to use random sampling to be represented by 231 entries. Since the sample number was fairly large I decided to use the z-test in testing my hypothesis at a 95% confidence level putting the alpha at 0.05. There are a number of assumptions in place that our data follows a normal distribution and that the samples taken are completely independent of each other.

The reason I chose this because I decided to find out if indeed the number of BlueCars taken on Monday is not equal to the ones picked on Friday. This is because there is a surge in the number of workers going to work on Monday hence making it a fairly equal to the number of BlueCars taken on Monday deserving an equal number of resource allocation in marketing and the number of cars reserved for that day.

I’ll be testing the difference in means of BlueCars picked on Monday and Friday. Hence my variable of interest is the mean of Monday.

The point estimate for my Monday population is 94 and the point estimate for the Monday sample is 92. The variation between the point estimates is 2.5 which was rather small compared to other sample sizes.

I also got the interval estimate of the population mean which was at 89.5 to 99.9 at a 95% confidence level. This satisfies with our population mean and sample means.

**4.0 HYPOTHESIS TESTING RESULTS**

The test statistic is at 153.0225 which was used in calculating the p-value which I found was 0.0.

There was sufficient evidence at 95% significance level to reject the null hypothesis. Therefore the cars taken on Monday are not equal to the ones taken on Friday.

Since the test was a two tailed test, the alpha level used was divided into two for both ends of the test and I used 0.025 as my alpha level.

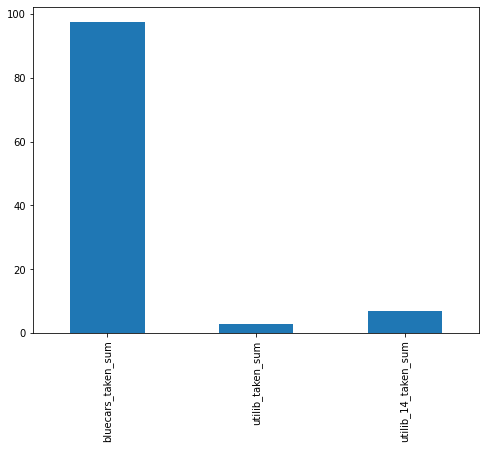
The point estimate for the Monday population is at 94.8 and the sample point estimate is at 92.3. The interval for the Monday population is 89.5 to 99.9, which was done at a confidence level of 95%.

**5.0 DISCUSSION OF SENSITIVITY**

The possibility of rejecting the null-hypothesis while its true is at 5% which is fairly low . The power of the test is at 95%.

**6.0 SUMMARY AND CONCLUSIONS**

After the analysis it was concluded that BlueCars experienced the highest number of traffic.



It was also apparent that the hypothesis that Monday could have the same traffic as Fridays did was a false one and in the hypothesis.

H₀: μ(Monday) = μ(Friday)

H₁: μ(Monday) **≠** μ(Friday)

I have sufficient evidence to support the alternative hypothesis that indeed the mean number of BlueCars taken on Monday and Friday are not equal. The power of the test is at 95%.