**HYPOTHESIS TESTING REPORT**

**Problem Statement**

Autolib is an electric car sharing company in Paris,France.As Data scientists of the company we are going to investigate if there is a significant difference in the means of the blue cars taken on weekdays in two different postal codes..

The null and alternative hypothesis are:  
H0 : There is no significant difference in the mean of blue cars taken in two different postal codes.

H1 : There is a significant difference between the mean of blue cars taken in two different postal codes.

This hypothesis was chosen since we wanted to know if there is a difference in the number of cars taken in different postal codes.This would help the company to determine whether to supply more cars to certain postal codes.

**Data Description**

We are provided with two datasets.

1. Column explanation-This dataset gives us an explanation of the variables in the Autolib Daily Events Postal code Dataset.
2. Autolib Daily Events Postal code Dataset-This dataset shows the postal code of a station,the date the data was collected,the number of daily data points that were available for aggregation that day,the day of the week the data was collected and whether it was on a weekday or weekend,the number of blue cars taken,the number of blue cars returned,the number of utilib taken,the number of utilib returned,the number of utilib 14 taken ,the number of utilib 14 returned ,the number of recharging slots released and the number of recharging slots taken.

Column explanation-This dataset gives us an explanation of the variables in the Autolib Daily Events Postal code Dataset.

**Hypothesis Testing Procedure**

* We performed Stratified sampling on the population so that we could get a sample with the same proportion as the population.Postal codes served as the stratas.We then choose postal codes 75015 and 95880 randomly.
* We formulated the following hypotheses;

H0 : There is no significant difference in the mean of blue cars taken

in postal codes 75015 and 95880.

H1 : There is a significant difference between the mean of blue cars

Taken in postal codes 75015 and 95880.

* Since the sample size was less than 30 i.e. 11 for both postal codes we chose to use two-tailed T-test.
* We choose a level of significance(∝=0.05).
* Before running the results we first observed if the variables to be investigated are dependent or independent.We found them to be independent.We then checked for the normality test by drawing the probability plot.Since the plot did not show the results clearly we decide to test statistically using the shapiro wilk test and found both variables are normally distributed.
* We computed the T Statistics and the P value.

**Hypothesis Testing Results**

After calculating the T Statistics and the P value using the google colaboratory we found the following results;

T Statistics=19.2895

P value=2.1516

We also calculate the point estimate and the confidence interval.The point estimate was the difference between the sample means which was 846.7272.

We then calculate the 95% confidence interval by using the point estimate and the Margin of Error.The confidence interval was [717.3065,976.1481].We were 95% confident that the point estimate lied between this range.

**Conclusion**

Since the P value was greater than the level of significance(∝=0.05).We failed to reject the null hypothesis since we did not have enough evidence to reject the null hypothesis that there is a significant difference in mean of the blue cars taken.