**Problem Statement**

We have been tasked to understand electric car usage by testing an alternative hypothesis that is generally held as true and draw insights and make a decision whether to reject or fail the null hypothesis.

In order to do and achieve this, we had a null hypothesis and the claim to be tested. Our null hypothesis was the average number of blue cars taken is no different compared to that of the blue cars returned at postal code 75001 whereas our alternative hypothesis is the average number of blue cars taken is different that of the blue cars returned.

**Data Description**

The data used for this project involved a dataset that is a daily aggregation, by date and postal code, of the number of events on the Autolib network which is a car-sharing and recharging service company. We focused on blue cars specifically. The sampling method used was stratified random sampling. Most of the columns in the dataset have been explained below

Postal code - postal code of the area (in Paris)

Date - date of the row aggregation

n\_daily\_data\_points - number of daily data points that were available for aggregation, that day

dayOfWeek - identifier of weekday (0: Monday -> 6: Sunday)

day\_type - weekday or weekend

BlueCars\_taken\_sum - Number of bluecars taken that date in that area

BlueCars\_returned\_sum - Number of bluecars returned that date in that area

Utilib\_taken\_sum - Number of Utilib taken that date in that area

Utilib\_returned\_sum - Number of Utilib returned that date in that area

Utilib\_14\_taken\_sum - Number of Utilib 1.4 taken that date in that area

Utilib\_14\_returned\_sum - Number of Utilib 1.4 returned that date in that area

Slots\_freed\_sum - Number of recharging slots released that date in that area

Slots\_taken\_sum -Number of recharging slots taken that date in that area

**Hypothesis Testing Procedure**

We came up with a null hypothesis and an alternative hypothesis. Which would be the basis of a couple of our tests. The null hypothesis and the alternative hypothesis were as:

* Null hypothesis(Ho): The average number of blue cars taken is no different compared to that of the blue cars returned at postal code 75001
* Alternative hypothesis(Ha): The average number of blue cars taken is different that of the blue cars returned.

Carried out a z test because the sample in question fell above 30 which is supposed to be the threshold for which we are to carry out a t-test. We will use a significance level of 5% and a corresponding value of 95%.

**Hypothesis Testing Results**

The p value obtained from carrying out the relevant tests and computations for our sample was 0.36 and the z-score obtained was -0.357.

Since the p-value we got was higher than our significance level,we fail to reject the null hypothesis since there is no significance.

**Discussion of Test Sensitivity**

An increased sample size would lead to a corresponding decrease in our p-value

**Summary and Conclusions**

The project saw us start working on the dataset from reading the data, then to cleaning the data we had by checking for the consistency, completeness, uniformity and validity then we carried out exploratory data analysis by looking at univariate and bivariate analysis. We then carried out hypothesis testing to check whether we would reject the null hypothesis or whether we’d fail to reject the same.

We carried out a couple of tests dictated by our sample size and from those we deduced that our p value was higher than the significance level hence the decision to uphold the null hypothesis