**A REPORT ON A PYTHON PROJECT TO UNDERSTAND ELECTRIC CAR USAGE**

1. **PROBLEM STATEMENT**

Was tasked as a Data Scientist to investigate the claim that the difference in the mean of the blue cars taken in postal code 75015 and 75017 is not 0. This was done with the goal of understanding electric car usage in the Autolib electric car-sharing service.

THe null hypothesis is that the difference in the mean of the blue cars taken in postal code 75015 and 75017 is 0 while the alternate hypothesis is the difference in the mean of the blue cars taken in postal code 75015 and 75017 is not 0

H0- the difference in the mean of the blue cars taken in postal code 75015 and 75017 is 0

H1-the difference in the mean of the blue cars taken in postal code 75015 and 75017 is not 0

1. **DATA DESCRIPTION**

Bluecars taken can be identified as the random variable. The data describes number of blue cars taken over different postal codes in France during the year 2018.The source of the dataset can be found <http://bit.ly/DSCoreAutolibDataset> while the description of the columns is <http://bit.ly/DSCoreAutolibDatasetGlossary>

Data was loaded cleaned, exploratory data analysis was done and sampling of 30% of the dataset was done to better test the hypothesis

1. **HYPOTHESIS TESTING PROCEDURE**

Testing will be undertaken using a sample of the total population.

The test statistic used is the z test. WIth a sample size of 5,000, there are more than 30 items in the sample. This means that the appropriate test to use is the z test. Other conditions which led to the use of the z test were;

1. **Data points should be** [**independent**](https://www.statisticshowto.com/probability-and-statistics/dependent-events-independent/#or) from each other. In other words, one data point isn’t related or doesn’t affect another data point.
2. Your **data should be normally distributed**. However, for large sample sizes (over 30) this doesn’t always matter.
3. Your data should be [**randomly selected**](https://www.statisticshowto.com/random-selection-and-assignment/) from a population, where each item has an equal chance of being selected.

Level of significance used was 0.05

1. **HYPOTHESIS TESTING RESULTS**

The results of my hypothesis test is thatthe difference in the mean of the blue cars taken in postal code 75015 and 75017 is not 0. Thus we rejected the null hypothesis.

The p value that i got was 4.1340093175145906e-20 which is less than the significance level

1. **DISCUSSION OF TEST SENSITIVITY**

Sensitivity testing was not conducted because this is not a machine learning model.

1. **SUMMARY AND CONCLUSION**

The main goal was to understand electric car usage in the Autolib electric car-sharing service.

Necessary steps to perform the analysis were undertaken.

Our conclusion from the claim that the difference in the mean of the blue cars taken in postal code 75015 and 75017 is not 0 was that the claim is true.