**1)** **Problem statement**

We used autolib dataset which contained information about the operation of cars within the city of Paris.It had information about the specific dates when blue cars were picked and returned to specific postal addresses.

The random variable being investigated (claim) was whether the average number of blue cars picked was different from the number of blue cars returned or the same within the dataset time frame.

Null hypothesis:The average number of blue cars taken is the same as the average number of cars returned.

Alternative hypothesis:The average number of blue cars taken is different from the average number of blue cars returned.

The CLAIM was set to alternative hypothesis.The importance and the end goal of hypothesis testing was to come up with a clear conclusion on blue cars operation at autolib.

**2) Data Description**

Wewere provided with one open dataset url and a detailed description of the dataset url.The dataset contained information about cars.We have three agencies Blue car,Utilib and Utilib 1.4 where the cars are taken and returned.

The dataset column names which are our variables are hours ,postal code and dates to help perform data manipulation as required.

The dataset timeframe was january 2018 to July 2018.

The days were classified as either week day or weekend .

We choose to work on weekends.

Since we were provided with a dataset url , it was an open dataset where data had already been collected .

Without the limitation I would have collected my own primary data.

**3)Hypothesis Testing Procedure**

The population was equivalent to the dataset provided which had 16,085 rows and 13 columns.

From the population size,I came up with a sample size using the stratified sampling method.

In stratified sampling data is grouped into stratas then randomly picked without bias.

To determine the hypothesis I used dates and counts of the cars that were picked and the ones that were returned on a daily basis.And to simplify the numbers , I worked on their mean(average) .Hence,using the average of cars picked and returned on a daily basis, I was able to investigate the claim.

My sample size was 296 rows and 4 columns from stratified sampling. My n was 54.34 which is greater than 30.Therefore,I used z-score to determine p-value.

My alpha level was 0.05.

I did normality tests to check if the data was normally distributed.

**4) Hypothesis Testing Results**

I failed to reject the null hypothesis since there was no sufficient data to prove the CLAIM; the average number of blue cars taken is not different from the average number of blue cars returned.

I obtained a z-score of -1.5945

With this data the null hypothesis was not rejected.

The point estimation was 76.9236.

The p- value was 0.05541

The z-critical value:1.9510 with a confidence interval: (27.879,70.1272)

**5)Discussion of Test Sensitivity**

Test sensitivity is the true positive rate ,the recall.

In our dataset it was 80%.

**6)Summary and Conclusions**

The steps that were followed were data exploration ,analysis, implementation and hypothesis testing.

The z-score value obtained was -1.5945

Hypothesis test was not rejected because there was no enough evidence to reject it .