1. **Business overview**

Autolib’ is atechnical, operational and commercial success. This electric car-sharing service started in Paris and operates in several French cities. The service was launched for ‘beta testing’in Paris on October 2nd, 2011 and at that time, there were 66 cars and 33 car rental stations. In 2015, Autolib had 1060 rental stations and 5,700 charging stations and the number was believed to increase over time.

1. **Problem Statement**

**2.1.** **Introduction of the dataset to be used**

In this study, the Autolib dataset will be used. It contains various columns such as date which contains the date of the row aggregation, dayOfWeek which is the identifier of weekdays from Monday to Sunday, day\_type which is either weekday or weekend, and Blue cars taken, Utilib taken, Utilib\_14 and they will be investigated during the study.

**2.2.** **Objectives**

**2.3.** **Main objective**

The main objective is to determine the commonly used mode of transportation in Paris.

**2.4.** **Specific objectives**

a) To improve the quality of autolib dataset by cleaning it.

b) To analyze and visualize the different types of vehicles used in transportation.

c) To determine which day of the week on which Blue Cars are mostly used

**2.5.** **Null hypothesis**

Blue Cars are not the most preferred mode of transportation compared to other modes in Paris.

**2.6.** **Specific hypothesis**

Blue Cars are the most preferred mode of transportation compared to other modes in Paris.

**2.7.** **Importance of the hypothesis**

By concluding the hypothesis we will be able to know the most preferred mode of transportation and through that, we will be able to do the prediction and hence better planning.

**3.** **Data Description**

Autolib dataset has 16085 rows and 13 columns. Out of 13 columns, two columns has object as datatype while the datatype of the rest columns is integer. The columns includes the following:

Postal code

date

n\_daily\_data\_points

DayOfWeek

Day\_type

BlueCars\_taken\_sum

BlueCars\_returned\_sum

Utilib\_taken\_sum

Utilib\_returned\_sum

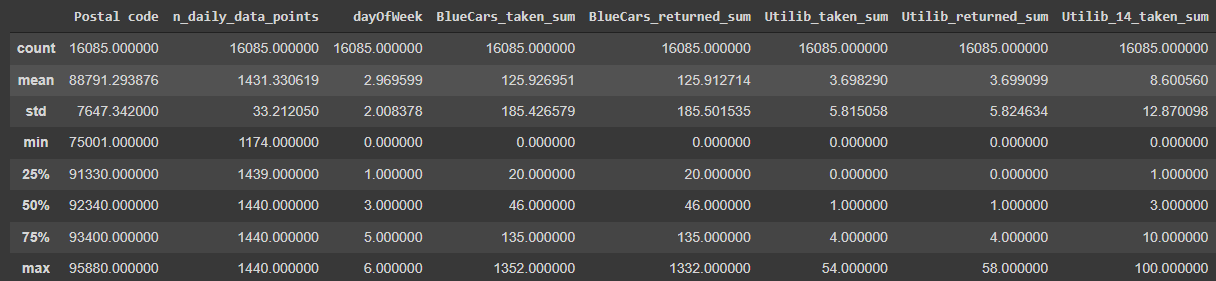
Utilib\_14\_taken\_sum

Utilib\_14\_returned\_sum

Slots\_freed\_sum

Slots\_taken\_sum

The count, mean standard deviation, minimum, maximum and percentile of various columns are as shown in the figure below:

****

The Autolib data set has been sourced from a secondary source that is on the internet:<http://bit.ly/DSCoreAutolibDataset>

**4.** **Sampling the data**

A stratified random sampling method will be used to sample the data. Out of the Autolib dataset, BlueCars\_taken\_sum will be used as strata. BlueCars\_taken\_sum has a mean of 125.926951 while the mean of the whole dataset is 90409.078209.

**5.** **Hypothesis testing procedure**

***Specify the Null Hypothesis***

Null hypothesis Ho, µ = 125

***Specifying the Alternative Hypothesis***

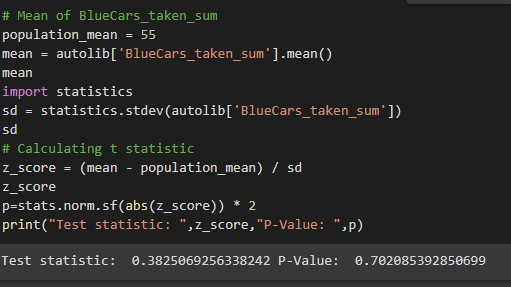
Alternate hypothesis H1, µ ≠ 125

***Setting the significant level***

α=0.05

***Calculating the Test Statistic and corresponding P-Value***

Test statistic and p-value will be calculated using the following code:

******

**6.** **Hypothesis Testing Results**

z-test will be used since the population is greater than 30.

**7.** **Discussion of Test sensitivity**

Z-test was used to test the hypothesis and the following was the outcome:

Test statistic: 0.3298092245654572 P-Value: 0.7415441168319368

Since P-value is greater than the significant level, the null hypothesis will not be rejected. There is not sufficient evidence that BlueCars is the most used mode of transportation.

**8.** **Summary and conclusion**

By removing the outliers, I was able to improve the quality of the data. I did a bivariate analysis to determine the correlation of various modes of transportation such as Bluecars and utilib. I performed a z-test to support the hypothesis which I derived.