# BUSINESS UNDERSTANDING.

### **Problem Statement**

autolib is an electric sharing company. There are BlueCars, Utilib and Utilib\_14 vehicles in 860 stations and 4,400 parking spaces. Users can pick or drop off the available 3000 cars. The Bollore group wishes to expand the european market. This brings up the need to verify the measures and conditions put in place in Paris for even distribution of all the cars, stations and parking spots are effective before moving on to other regions. research and feedback on the following will allow the group to make informed choices;

* on any given weekday, the average number of vehicles taken is more than the average number of vehicles returned
* on any weekday, the number of BlueCars taken is more than the total number of both Utilib and Utilib\_14 are taken
* the number of cars taken from postal code 75015 is equal to the number of cars taken from postal code 75017

## **Business Overview.**

The autolib system was a follow-up to the Velib’ bike sharing system which was a great success in paris. The autolib system was a huge success as car availability became a problem because many Parisians subscribed to the service. By july 2014, the system had 2,500 operational vehicles and over 150,000 subscribers.

# 

## **Business Objectives.**

As a data scientist I decided to conduct a research and analysis with the aim to determine the movement of all the cars available during this period. This should help pinpoint why the autolib system was a success and why it inadvertently failed.

### TOOLS USED IN THIS PROJECT

* Google Colaboratory notebook as the environment where I’d do my Python analysis: [[Colab]](https://colab.research.google.com/drive/1NnpW-MWEzAkfjEcdib6lidohI9fVkwSL#scrollTo=Kfi85AKrL-qa)

# DATA UNDERSTANDING.

## **Data Understanding Overview.**

We have one dataset that contains all our data and can be accessed from [source here](http://bit.ly/DSCoreAutolibDataset).

We have a couple of columns in our data and below are the column descriptions:

* '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'
* 'Total\_vehicles\_taken'
* 'Total\_vehicles\_returned'
* 'total\_Utilib\_vehicles\_taken'

## **Data Quality.**

A quick look at our dataset shows that we have no null values and no duplicate values.

# DATA VALIDITY

Validity refers to how accurately a measure measures what it was intended to measure.The validity of the data can be confirmed from a [web search](https://en.wikipedia.org/wiki/Autolib%27). I ascertained that I have a high validity in our data.

# DATA PREPARATION.

As a data scientist, I will be working majorly with data and as such need to clean our data for better and usable output. Someone rightly said garbage in garbage out. This dataset was no exception.

We first began by loading our dataset and creating dataframes that we’d use in manipulation and analysis. Next I checked for any anomalies and outliers after which I worked on the univariate and bivariate analysis to better understand the dataset. The BlueCars, Utilib and Utilib\_14 cars have a positive correlation

# HYPOTHESIS TESTING

## **Sampling**

**i) target population**

The population for this research is the entire dataset that has 16,085 entries

**II) sampling method**

For this research I am going to be using the simple random sampling technique. The question will look into weekdays (represented by 0-4 entries).

Using this [online tool](https://docs.python.org/3/library/random.html#random.random), random weekday ‘3’ was selected

**iii) sample size**

The weekday 3 selected randomly gives an output of 2,374 entries as the sample size

I came up with 3 hypothesis tests;

## **Hypothesis 1:**

### Null hypothesis(Ho):

the average number of vehicles taken is less than or equal to average vehicles returned

### Alternative hypothesis(Ha):

the average vehicles taken is greater than average vehicles returned

In order to compute this hypothesis, I had to calculate the total number of all vehicles (BlueCar, Utilib, Utilib\_14) taken and returned and find their average. Then using this online tool, I sampled a random weekday to work on. I then calculated the z score and the p value for the hypothesis at a confidence level of 0.95 and a significance level 0.05. This gave me a resulting computation for the sample of 0.5215. Since the p-value was higher than my significance level, i accept the null hypothesis

## **Hypothesis 2**

### Null hypothesis(Ho):

on any weekday, the number of BlueCars taken is more than the total number of Utilib and Utilib\_14 cars taken

### Alternative hypothesis(Ha):

on any weekday, the number of BlueCars taken is less than or equal to the total number of Utilib and Utilib\_14 cars taken.

For this hypothesis, I calculated the total number of Utilib and Utilib\_14 cars taken. This would then be compared to the total number of BlueCars taken in any given weekday. Again, the day was randomly chosen by the online tool used above. I maintained the confidence level at 0.95 and the corresponding significance level of 0.05.

I calculated the z score and the p value obtained from carrying out the relevant tests and computations for our sample was 0.7317. Since the p-value we got was higher than my significance level, i accept the null hypothesis

## **Hypothesis 3:**

### Null hypothesis(Ho):

number of vehicles taken in postal code 75015 is equal to the number of cars taken in 75017

### Alternative hypothesis(Ha):

number of vehicles taken in postal code 75015 is not equal to the number of cars taken in 75017

I calculated the z score and the p value obtained from carrying out the relevant tests and computations for my sample was 0.8534. Since the p-value we got was higher than my significance level, i accept the null hypothesis

# CONCLUSION

The people of Paris enjoyed the autolib system. This is evidenced by the rate of movement of cars to and from the stations.

The BlueCars are generally more used than the Utilib and Utilib\_14 cars

# RECOMMENDATIONS

The company can go ahead and expand their business to other regions, while maintaining or improving the system’s standards