# Problem Statement

Autolib is an electric car-sharing service company based in France. It maintains a fleet of electric blue cars for public use based on a paid subscription basis employing a citywide network of parking and charging stations. Our dataset given includes Postal codes and days of the week corresponding to the number of blue cars taken. We shall be using a random generated postal code to get a count of the average number of blue cars taken from that specific postal code.

For our research question, our Null Hypothesis(H0) states that the average number of blue cars taken from postal code 75015 on a weekend is equal to the number of blue cars taken from postal code 75017 on a weekend. On the other hand our Alternative Hypothesis(H1) states that the average number of blue cars taken from postal code 75015 on a weekend is not equal to the number of blues cars taken from postal code 75017 on a weekend.

The hypothesis stated above is important as it helps the researcher identify whether the postal codes randomly selected have an equal number of electric blue cars taken from them. If the number of blue cars taken is not equal from the postal codes further research can be conducted to identify the factors leading to the disparity.

# Data Description

Our dataset consists of 13 columns and their description is as follows:

* Postal code = postal code of the area (in Paris)
* Date = date of the row aggregation
* N\_daily\_data\_points = number of the daily data points that were available for aggregation that day.
* Day of week = identifier of weekday(0: Monday - > 6:Sunday)
* Day type = weekday or weekend
* Blue cars taken sum = Number of blue cars taken that date in that area
* Blue cars returned sum = Number of blue cars 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

The criteria used when conducting the hypothesis testing was first categorizing the type of days column to weekends. This would help in determining whether the average number of blue cars taken at postal code 75015 would be equal to the number of blue cars taken at postal code 75017 on a weekend which would result in accepting our null hypothesis.

The test statistic used was the two tailed z test. The latter test statistic was agreed on based on the sample size of the blue cars taken at a particular postal code during the weekend which was more than 30. The significance level to be used in the test is 5% i.e 0.05

# Hypothesis Testing Results.

On conducting the hypothesis test, we concluded that we should reject the null hypothesis. The p-value of the test statistic was < 0.05 which was our significant level.

# Summary and Conclusions.

As a result of the p-value from the test which is less than the significance level (0.05), we reject the null hypothesis H(0) : which states that the average number of blue cars taken at postal code 75015 is equal to the average number of blue cars taken at postal code 75017. Thus we accept the alternative hypothesis (H1) : the average number of blue cars taken at postal code 75015 is different to the average number of blue cars taken at postal code 75017.