# Report on Hypothesis Testing of the Autolib cars

[Report on Hypothesis Testing of the Autolib cars](#_1bqi0y9ezblv)

[Business Understanding](#_qoloiafk3c5z)

[Data Understanding](#_ko8qfq3i2hq1)

[Data Preparation](#_usvi4kxv9sar)

[Evaluation](#_i7zrq786xm32)

[Deployment](#_4j57jfmlpcbd)

## Business Understanding

Autolib' was an electric car sharing service which was inaugurated in Paris, France, in December 2011. It closed on 31 July 2018. It was operated by the Bolloré industry and complemented the city's bike sharing system, Velib', which was set up in 2007. The Autolib' service maintained a fleet of all-electric cars which included the Blue cars, the utilib cars and utilib 14 cars for public use on a paid subscription basis, employing a citywide network of parking and charging stations.

We will work as a Data Scientist for the Autolib electric to try and answer some claim that was raised .

In an effort to do this, we need to identify some areas and periods of interest via sampling stating the reason for the choice of method, then perform hypothesis testing with regards to the claim that we will have made. Our claim which we have been asked to look into is that no. of cars taken in area postal code 75015 is greater than that of area 75017

## Data Understanding

The Dataset we will be using is an autollib dataset, which involves different sets of cars i.e Blue cars, the utilib cars and utilib 14 cars that were used in the France region. The dataset contains information from different areas with different postal codes.

The links to the dataset and descriptions that we will be using can be found below(respectively);

* <http://bit.ly/DSCoreAutolibDataset>
* [[Link]](http://bit.ly/DSCoreAutolibDatasetGlossary)

## Data Preparation

To work on this project, we will perform the following analysis with Python;

1. Find and deal with outliers, anomalies, and missing data within the dataset.

2. Plot appropriate univariate and bivariate summaries recording our observations.

3. Implement the solution by performing hypothesis testing.

## Modeling

As earlier stated, we will be using python to help us in our research.

In our model, we will carry out two major steps i.e sampling then hypothesis testing on our samples.

1. Sampling

Since we are interested in carrying out the hypothesis between two different areas i.e postal code 75015 and 75017, we will use the stratified method. This is the most efficient method since the postal codes will be our strata.

We we also only be using data from the blue cars only

1. Hypothesis Testing

We will be using the z\_test to carry out our Hypothesis testing since we are relating to a sample of greater than 30 i.e (n>30) In our case, we will be using a sample of 70.

It's also a convenient method since we know our standard deviations and mean.

A z-test is also useful when we want to determine whether two population means(in our case data from different postal codes)are different when the variances are known and the sample size is large.

In our z\_test we will be looking at blue cars taken and comparing it with the two postal codes i.e codes 75015 and 75017.

Our argument is that more blue cars were taken in Postal code 75015 compared to 75017.

Thus , our null hypothesis will be that more blue cars were taken in Postal code 75015 compared to 75017.

If we obtain different results, we will reject the null hypothesis and accept the alternative hypothesis.

## Evaluation

After performing our hypothesis testing using the z test method, we found out that we have to reject the null hypothesis since the p value was way less than our significant level.

This means that our null hypothesis which states that more blue cars were taken in Postal code 75015 compared to 75017 is false

We also carried out a power test to see how exact our analysis is. The results were that our test was 86% accurate thus making our analysis efficient.

## Deployment

After successfully performing our project which included sampling followed by a hypothesis testing, we came to a conclusion that more blue cars were taken in the area of postal code 75017 compared to that of postal code 75015

We used the Google collab to perform our analysis and posted it on github. Below are the links to both the Github repository and the google collab notebook.

* <https://github.com/frankthairu/Autolib-Research-IP-4>
* <https://colab.research.google.com/drive/11x1fahc6UimGk2O7xgApdzmDeiEAzct2#scrollTo=2P3alAP9Srt7>