# BLUE CAR HYPOTHESIS TESTING REPORT

## Problem Statement

### 1.Background

Autolib is operated by the Bolloré Group enterprise, which won the contract to develop the service and to supply Paris with electric cars and stations. The program started in 2011 with an initial fleet of 250 eco-friendly electric Blue Cars. The program today has more than 3,000 cars operating on the streets of Paris and within the whole region. There are around 860 Autolib stations where users can subscribe, pick up or drop off the cars. As well, there are 4,400 parking spaces and charging points reserved exclusively for Blue cars.

### 2.Hypothesis

Analysing 2018 dataset, Data Scientists at Autolib car sharing company want to test a claim that there is a difference between the number of blue cars taken on weekdays from postal code 75011 and postal code 75015.

- **Null hypothesis**;There is no difference between the number of Blue cars taken on Weekdays from postal code 75011 and postal code 75015.
- **Alternative hypothesis**;There is a difference between the number of blue cars taken on weekdays from postal code 75011 and postal 75015.
- **Significance level** ; a= 0.025 due to it's two tailed nature.
- **P-value** ; Calculated as 0.05

### 3.Variables

This dataset contains the various postal codes that Autolib offer their services to as well as the sum of cars received and taken in the various stations on the weekdays and the weekends. This data is from the customer database in the company and is collected on a daily in the company during service provision. The random variable we will be investigating is the blue cars taken variable, with relation to the day type and postal code categorical variables.

Variables will use to test the hypothesis will be;

- Postal code
- Day type
- Blue cars taken

# Hypothesis Testing Procedure

## 1.Hypothesis procedure

We will test our hypothesis by first cleaning our data and narrowing down to the variables we need the most. Then we will conduct both univariate and bivariate analysis of our variables before we sample our data to enable us to perform our hypothesis test.

Our null and alternative hypothesis are interesting because we were interested to know if the two postal codes had differences in  traffic as always assumed due to their close proximity.

## 2.Test statistic

The test statistic we will use is the t-test statistic. We will be testing a hypothesis about a count and we will use a sample size of 1000 which is about 10% of our dataset which will correctly represent our population. We first converted our data which was skewed to the right into a normal dataset using log transformation so that we are able to work with the t-test statistic. All this will be able to satisfy the assumptions of the t-test statistic therefore, giving us a go ahead with the test.

## 3.Alpha

We chose and used an alpha level of 0.025. This is because our distribution will be two tailed.

# Hypothesis Testing Results

## 1.Test results

Our test results were that we got a p-value of 0.0248. Since our p-value was less than 0.05, we rejected the null hypothesis and favour of the alternative hypothesis that the difference between the number of blue cars taken on weekdays from area code 75011 and area 75015 is statistically significant with a p-value of 0.05 and below my alpha of 0.025.

We also had a critical value of 1.96 which was larger than our t statistic of 0.17 which promoted the rejection of our null hypothesis

Therefore, we conclude that our study supports the alternative hypothesis that there is a difference between the number of blue cars taken on weekdays from area code 75011 and area 75015.

## Discussion of Test Sensitivity

We were not able to do a test sensitivity on our hypothesis test but since we chose a commonly used alpha of 0.025 for our two tailed distribution, this means that there is only a total of 5% chance that a type 1 error will occur. The power of our test will also be increased by increasing our sample size therefore reducing the chances of a type 2 error from occurring.

## Summary and Conclusions

### 1.Summary.

| **Day** | **Activity** |
|---|---|
| Day 1 | Data cleaning and Data Validation. |
| Day 2 | Univariate and Bivariate analysis with Data Visualization. |
| Day 3 | Sampling.<br>Choosing a test statistic and fixing the data to work with the test statistic. |
| Day 4 | Hypothesis testing and conclusion |

### 2.Conclusion

The conclusion that there is a difference between the two postal codes75011 and area 75015, will be of benefit to the autolib dataset as it will be able to inform the company on ways they can increase the demand for blue cars in the two areas and how they can be able to work and understand the surrounding social, environmental and economical conditions that surround their current demand.