### **HYPOTHESIS TESTING REPORT FOR AUTOLIB**

### **Business Overview**

The City of Paris was among the first major European city, which successfully deployed a public electric car-sharing program. The Autolib program had more than 3,000 cars operating on the streets of Paris and within the whole region. There were around 860 Autolib stations where users can subscribe, pick up or drop off the cars. As well, there were 4,400 parking spaces and charging points reserved exclusively for ”Bluecars”. Electric car-sharing services greatly contribute to the preservation of the environment and facilitate mobility in the bustling city of Paris.

**Problem statement**

A statistician in the Autolib company believes that the Bluecars picked on weekdays are more than Blue cars picked on weekends. As a Data Scientist for the Autolib electric car-sharing service company,I have been tasked to investigate a claim about the blue cars from the provided Autolib dataset.

**Null Hypothesis;** The number of Blue cars taken on weekdays are more than the number of Blue cars taken on weekends

**Alternative Hypothesis**The number of Blue cars taken on weekdays are not more than the number of Blue cars taken on weekends.

The reason why this hypothesis is interesting is because we assume that weekdays are usually the busiest and thus we expect the blue cars taken on weekdays to be more than the cars taken on weekends.

This Hypothesis is important because it will inform us on when the cars need to be made available to the public in plenty.This hypothesis test will be helpful to the company as it will help them make informed decisions based on facts.

**Data Description**

The dataset provided was an open source Autolib dataset which contained information on the electric car usage in Paris during a period of six months between 1st January 2018 to 19th June 2018.

These datasets are

1. **Autolib Datase**[<http://bit.ly/DSCoreAutolibDataset>]- the dataset is a daily aggregation, by date and postal code, of the number of events on the Autolib network (car-sharing and recharging).It entailed the total number of electric cars, that is, bluecar, Utilib and Utilib 1.4 taken and returned.
2. Autolib Dataset Glossary [[Link]](http://bit.ly/DSCoreAutolibDatasetGlossary) which gives us the description of our columns

Below are the descriptions.

**Data Preparation**

These are the steps followed in preparing the data

1. Loading the data
2. Checking the data
3. Cleaning the data
4. Exploratory Data Analysis
5. Hypothesis Testing

**Hypothesis Testing Procedure**

The dataset had a total of 16085 records consisting of the following fields ; postal code, number of blue cars taken and returned to the respective addresses, the dates that the electric cars were taken or returned as well as the days of the week that electric cars were taken and returned. All this information was useful in helping us understand the electric car usage in Paris.

The first step was to define our hypothesis population by locating the Blue cars picked on weekends and weekdays.I went ahead and picked a random sample from the population I defined. I selected 100 samples from the Blue Cars picked on weekend and 100 samples from the Blue cars picked on weekdays which translated to 200 records. I used a simple random sampling method because each individual in the population has the same probability of being selected. It ensures that there is no bias and also makes sure that the proportion of weekdays and weekends data picked is equal in our grouped sample.

After getting my sample,I imported the from statsmodels.stats.weightstats import z test as ztest and performed two sample z test which helped in attaining the p\_value. I used z test because my sample was large and was greater than 30.

Having set the significance level at 5% (0.05), I compared the p-value to the significance level with the aim of deciding whether to reject or fail to reject the null hypothesis.

The above analysis was done using python. The full analysis and hypothesis testing can be found in the following notebook [[Link]](https://colab.research.google.com/drive/1f6B2it4FSM0mH-8P8Kb6qU36ED6fHN-j?usp=sharing)

**Hypothesis Testing Results**

After deciding to use the z-score as my test statistic, I calculated the z-score which was -1.4437992453388084 and the p value was 0.14879545435592487.

We therefore fail to reject the null hypothesis since the p\_value was greater than the significance level and clearly state that the We have enough evidence to say that more Blue cars are taken on weekdays than on weekends.

**Discussion of Test Sensitivity**

This is only relevant when the null hypothesis is false. The statistical power of a hypothesis test is the probability of correctly rejecting a null hypothesis or the likeliness of accepting the alternative hypothesis if it is true. So, the higher the statistical power for a given test, the lower the probability of making a Type II (false negative) error.

For our case, we failed to reject the null hypothesis which means that we don't have to perform the test sensitivity.

**Summary and Conclusions**

I performed data cleaning and exploratory data analysis on our dataset before choosing a sample using the simple random sampling technique which ensures there is no bias in our sample size.I then performed hypothesis testing as the implementation to our research question.

From the hypothesis testing, we can conclude that The autolib company should avail more cars during the week as we have a huge market base that needs to be tapped because We have enough evidence to say that more Blue cars are taken on weekdays than on weekends.