# **1.** **Business Understanding**

## **1.1 Business Overview**

Autolib is an electric car sharing company managed by the Bolloré group enterprise. The service was introduced to complement an existing bike-sharing service, Velib. Autolib has expanded its operations through Europe with its major success being in the city of Paris, France.

Autolib wishes to expand its operations within Europe and beyond. The operations team has been tasked by the senior leadership team to develop a strategy that will guide in this process.

## **1.2 Problem Statement**

Autolib owns 3,000 cars (BlueCars, Utilib and Utilib\_14), 860 stations where users can subscribe, pick up or drop off the cars and 4,400 parking spaces.

As part of their expansion strategy, the operations team seeks to verify that the measures they have put in place to ensure that there is even distribution of vehicles, stations and parking spaces across the city of Paris are effective.

To achieve this, they have requested for the services of our Data Science team to conduct research and provide recommendations on the strategy to distribute their resources. In order to achieve this, we will responding to the following claims:

1. On a given weekday, the average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is more than the average number of vehicles being returned.
2. The number of vehicles taken on weekdays is less than the number of vehicles taken on weekends.
3. On a given day, the number of BlueCars taken is more than the total combined number of Utilib and Utilib\_14 cars.
4. The number of vehicles taken in postal code 75015 is equal to the number of vehicles taken in postal code 75017.

These claims are important in guiding the team on determining where to focus their efforts to ensure even distribution of resources in Paris.

**2.** **Dataset and data description**

**autolib\_daily\_events\_postal\_code.csv**

This dataset is an extract of the number of BlueCars, Utilib and Utilib\_14 vehicles taken and returned on every day of 2018 in the different postal codes in Paris. It also highlights the number of charging slots taken and freed for each of the days.

The fields are as highlighted below:

|  |  |
| --- | --- |
| Field Name | Description |
| Postal code | Postal code of the area (in Paris) |
| date | Date of the row aggregation |
| n\_daily\_data\_points | Number of daily data points that were available for aggregation, that day |
| dayOfWeek | Identifier of weekday (0: Monday -> 6: Sunday) |
| day\_type | Weekday or weekend |
| BlueCars\_taken\_sum | Number of BlueCars taken on that date in that area |
| BlueCars\_returned\_sum | Number of BlueCars returned on that date in that area |
| Utilib\_taken\_sum | Number of Utilib taken on that date in that area |
| Utilib\_returned\_sum | Number of Utilib returned on that date in that area |
| Utilib\_14\_taken\_sum | Number of Utilib 1.4 taken on that date in that area |
| Utilib\_14\_returned\_sum | Number of Utilib 1.4 returned on that date in that area |
| Slots\_freed\_sum | Number of recharging slots released on that date in that area |
| Slots\_taken\_sum | Number of recharging slots taken on that date in that area |

The distribution of the three cars is shown during our analysis in the python notebook.

<https://colab.research.google.com/drive/1bBougi_nO9b76GWLKK3be2oLAQFobFbc?usp=sharing>

# 

# 

# 

# 

# 

# 

# **3.** **Data preparation**

## **Loading of data**

The dataset outlined above was loaded into the Jupyter notebook using python and necessary data frames created from it.

## **Data Cleaning**

Detailed examination of the data provided resulted in identifying the following:

|  |  |  |
| --- | --- | --- |
| Finding | Detail | Data Cleaning and Preprocessing activity |
| Irrelevant Data Observation: | The values in these attributes will not be required in the analysis:  ● n\_daily\_data\_points | ● These attributes were dropped. |

## **Sampling**

**a)** **Target population**

The target population in this study constituted 16,085 records linked to postal codes in Paris for the year 2018.

**b)** **Sampling Method - Simple random sampling**

This study will focus on weekdays. In the dataset, weekdays are represented by 0 to 4 where 0 is Monday and 4 is Friday.

Simple random sampling was used to select one value between 0 and 4 and number 3 was selected.

**c)** **Sample Size**

Using the random number ‘3’ selected, all entries that were made on the 3rd day of the week were selected from the total population. This resulted in a sample size of 2,374 records.

**4.** **Hypothesis Testing Procedure**

In this project, four hypotheses will be tested and will be as follows:

|  |  |  |
| --- | --- | --- |
| On a given weekday, the average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is more than the average number of vehicles being returned. | | |
| Null Hypothesis (H0) | Alternate Hypothesis (Ha) | Test statistic & Level of significance |
| The average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is less or equal to the average number of vehicles being returned.    H0: µ <= average number of vehicles being returned | The average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is less or equal to the average number of vehicles being returned.    Ha: µ > average number of vehicles being returned | The z statistic will be used in this test because the number of samples selected is more than 30. |
| The confidence level for this test will be 0.95 hence an alpha value of 0.05. |

|  |  |  |
| --- | --- | --- |
| The number of vehicles taken on weekdays is less than the number of vehicles taken on weekends. | | |
| Null Hypothesis (H0) | Alternate Hypothesis (Ha) | Test statistic & Level of significance |
| The average number of vehicles taken on weekdays is more than or equal to the average number of vehicles taken on weekends.    H0: µ => average number of vehicles taken on weekends. | The average number of vehicles taken on weekdays is less than the average number of vehicles taken on weekends.      Ha: µ < average number of vehicles taken on weekends. | The z statistic will be used in this test because the number of samples selected is more than 30. |
| The confidence level for this test will be 0.95 hence an alpha value of 0.05. |

|  |  |  |
| --- | --- | --- |
| On a given day, the number of BlueCars taken is more than the total combined number of Utilib and Utilib\_14 cars. | | |
| Null Hypothesis (H0) | Alternate Hypothesis (Ha) | Test statistic & Level of significance |
| The number of BlueCars taken is less than or equal to the total combined number of Utilib and Utilib\_14 cars.      H0: µ <= total combined number of Utilib and Utilib\_14 cars. | The number of BlueCars taken is more than the total combined number of Utilib and Utilib\_14 cars.      Ha: µ > total combined number of Utilib and Utilib\_14 cars. | The z statistic will be used in this test because the number of samples selected is more than 30. |
| The confidence level for this test will be 0.95 hence an alpha value of 0.05. |

|  |  |  |
| --- | --- | --- |
| The number of vehicles taken in postal code 75015 is equal to the number of vehicles taken in postal code 75017. | | |
| Null Hypothesis (H0) | Alternate Hypothesis (Ha) | Test statistic & Level of significance |
| The number of vehicles taken in postal code 75015 is equal to the number of vehicles taken in postal code 75017.      H0: µ = the number of vehicles taken in postal code 75017. | The number of vehicles taken in postal code 75015 does not equal the number of vehicles taken in postal code 75017.      Ha: µ ≠ the number of vehicles taken in postal code 75017. | The z statistic will be used in this test because the number of samples selected is more than 30. |
| The confidence level for this test will be 0.95 hence an alpha value of 0.05. |

# 

# 

# **5.** **Hypothesis Testing Results**

Tests that were conducted on the hypothesis revealed the following:

## **5.1 Hypothesis test 1:**

On a given weekday, the average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is more than the average number of vehicles being returned.

### **The value of the test statistics and the result of the accept/reject decision.**

The z-statistic was used to test this hypothesis and the calculated score was **-0.084.**

From the Z-Score above, the calculated p-value was **0.52**. This value was greater than the alpha value of 0.05. For this reason, the null hypothesis that claims the average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is less or equal to the average number of vehicles being returned was accepted.

## **5.2 Hypothesis test 2:**

On a given weekday, the number of BlueCars taken is more than the total combined number of Utilib and Utilib\_14 cars..

### 

### **The value of the test statistics and the result of the accept/reject decision.**

The z-statistic was used to test this hypothesis and the calculated score was **-0.62.**

From the Z-Score above, the calculated p-value was **0.73**. This value was greater than the alpha value of 0.05. For this reason, the null hypothesis that claims that on a given weekday, the average number of vehicles (BlueCars, Utilib and Utilib\_14) taken is more than the average number of vehicles being returned was accepted.

## **5.3 Hypothesis test 3:**

On a given weekday, the number of vehicles taken in postal code 75015 is equal to the number of vehicles taken in postal code 75017.

### **The value of the test statistics and the result of the accept/reject decision.** The z-statistic was used to test this hypothesis and the calculated score was **-1.05.**

From the Z-Score above, the calculated p-value was **0.85**. This value was greater than the alpha value of 0.05. For this reason, the null hypothesis that claims the number of vehicles taken in postal code 75015 is equal to the number of vehicles taken in postal code 75017 was accepted.

# **6.** **Summary and Conclusions**

The analysis conducted revealed that the internal processes within Autolib have been stabilised and the operations team can proceed and develop a plan to scale out to the larger European region.