### **BUSINESS UNDERSTANDING**

**Business Overview**

This is an electric car-sharing company in Paris that would like to understand their electric car usage over time.

**Business Objective**

The business objectives are:

* Identifying the most popular hour of the day for picking up a shared electric car (Bluecar) in the city of Paris over the month of April 2018.
* What station is the most popular?
  + Overall?
  + At the most popular picking hour?
* What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?
  + Overall?
  + At the most popular picking hour?

**Assessing the situation**

1. Resource Inventory
   * Autolib\_DDI\_DB\_description\_MoringaSchool\_w4 [[Link](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)]
   * Autolib\_dataset .csv [[Link](http://bit.ly/autolib_dataset)] (The dataset may take a bit of some time to load [~ 10 minutes])
2. Software
   * Google Collabedit
   * Python
   * Libraries such as Pandas and NumPy
3. Assumptions:
   * The data provided is correct

**Procedure plan**

| **Procedure Plan** | **Duration** |
| --- | --- |
| **Friday** | |
| Importing the libraries | **13:04** |
| Loading the dataset | **13:05** |
| Verifying the data integrity | **13:07-14:00** |
| **Saturday** | |
| Data cleaning | **12:02-13:12** |
| Data Analysis | **16:51-19:22** |

**Business Success Criteria**

Compiling a list of the most popular hours for picking cars & returning cars, the most popular station, and the postal code popular for picking blue cars(shared electric cars).

### **2. DATA UNDERSTANDING**

**Describing the resource inventory**

1. Autolib\_DDI\_DB\_description\_MoringaSchool\_w4 [[Link](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)]
   * This dataset describes the fields in the (Autolib\_dataset .csv [[Link](http://bit.ly/autolib_dataset)]) dataset.
2. Autolib\_dataset .csv [[Link](http://bit.ly/autolib_dataset)]
   * The description of the above dataset is in this [link](https://docs.google.com/spreadsheets/d/1zAfmW_H4I-MuCt3ber7EVXxLFXbXKLG1NHebkDxVaxs/edit?usp=sharing).

**Data Description**

1. The ‘Cars’ column and the ‘Bluecar counter’ contain the same values.
2. The values in the ‘Charging Status’ column contain some typos.

**Verifying data integrity**

According to the analysis, the following columns had null values:

1. Displayed comment with a total of 4889 null values.
2. Scheduled at with a total of 4953 null values.
3. The data was collected for a duration of 9 days in the month of August.
4. The data was collected for a duration of 24 hrs per day.

### **3. DATA PREPARATION**

The following dataset was used for this project:

* Autolib\_dataset .csv [[Link](http://bit.ly/autolib_dataset)]

**Data Cleaning**

1. **Validity**
   * The ‘Cars’ column was dropped since it contains the same values as ‘Bluecar counter’
   * The ‘year’ and ‘month’ columns were dropped since we already know the data was collected in the month of April 2018.
   * The ‘minute’ column was dropped since we only need the day and hour column to answer the research question.
   * Removed leading and trailing spaces in all columns, since they may hinder the analysis process.
   * Replace ‘ ’ with ‘\_’ in columns to enable easier selection of columns during analysis.
2. **Accuracy**
   * The following dataset has no calculated fields, hence I skipped this part.
3. **Completeness**
   * The following columns contain null values:
     1. ‘Displayed\_comment’ with a total of 4889 null values.
     2. ‘Scheduled\_at’ with a total of 4953 null values.
   * The following actions were taken on the null values:
     1. Dropping the ‘Displayed\_comment’ column since it contains alot of null values, hence irrelevant.
     2. Filtering out records which had ‘future’ in the ‘Scheduled\_at’ column and then dropping that column.
     3. Filtering out records which had ‘CENTER’ in the ‘kind’ column and then dropping that column.
     4. Dropping the status column since it contains only ‘ok’ after the other processes.
   * After the other cleaning process, the dataset was found to have 26 duplicates.
   * The actions taken on duplicated is that they were removed.
4. **Uniformity**
   * The following dataset was found to be uniform hence no action will be taken onto it.

### **4. DATA ANALYSIS**

1. **Using Bluecar\_counter**

The analysis will be based on the following research questions:

1. **Identify the most popular hour of the day for picking up a shared electric car (Bluecar) in the city of Paris over the month of April 2018.**
   * This is the time when the Bluecar\_counter is minimum
   * According to the analysis, the most popular hour of the day for picking up a Bluecar is the 20th hour.
2. **What is the most popular hour for returning cars?**
   * This is the time when the Bluecar\_counter is maximum
   * According to the analysis, the most popular hour of the day for returning a Bluecar is the 12th hour.
3. **What station is the most popular?**
   * Overall?
     + This is the time when the Bluecar\_counter is mininmum
     + According to the analysis, the most popular station is: Paris/Maine/4.
   * At the most popular picking hour?
     + This is the time when the Bluecar\_counter is mininmum.
     + According to the analysis, the most popular picking hour is the 19th hour.
4. **What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?**
   * Overall?
     + This is the time when the Bluecar\_counter is mininmum
     + According to the analysis, the most postal code is: 75001.
     + It does not match with the public name of the most popular station.

**II. Using Utilib\_counter**

1. **What is the most popular hour for returning cars?**
   * This is the time when the Utilib\_counter is maximum
   * According to the analysis, the most popular hour of the day for returning a Bluecar is the 8th hour.
2. **What station is the most popular?**
   * Overall?
     + This is the time when the Utilib\_counter is mininmum
     + According to the analysis, the most popular station is: Alfortville/Charles de Gaulle/16.
   * At the most popular picking hour?
     + This is the time when the Utilib\_counter is mininmum.
     + According to the analysis, the most popular picking hour is the 23rd hour.
3. **What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?**
   * Overall?
     + This is the time when the Utilib\_counter is mininmum
     + According to the analysis, the most postal code is: 92340.
     + It does not match with the public name of the most popular station.

**NB: According to the analysis the results vary while using Utilib\_counter**

**5. RECOMMENDATION**

1. Increasing Bluecars in stations in the 20th hour, in the city of paris, since they tend to be more on demand during this time.
2. Increasing Bluecars in all stations in the 12th hour since they tend to be more on demand during this time.
3. Find out reasons why the station Paris/Maine/4 is popular for picking up bluecars and applying the same methodology to other stations.

**6. EVALUATION**

The business success criteria has been reached since I was able to compile a list of the most popular hours for picking cars & returning cars, the most popular station, and the postal code popular for picking blue cars(shared electric cars).

### **7. Appendix**

Bluecars - shared electric cars

**Github link:**