## **Business Understanding**

### **Business Overview**

To work as a data scientist for an electric car sharing company. Tasked to process the electric cars usage data over time .

### **Business Objective**

The main objective of this report is to identify the most popular hour of the day for picking shared car in Paris for the month of April 2018

### **Business Success Criteria**

To identify the most popular hour of the day for picking shared blue cars.

### **Assessing the Situation**

1. **Resource Inventory**
   1. Datasets:
      1. Autolib datasets <http://bit.ly/autolib_dataset>
      2. Autolib descriptor [Autolib\_DDI\_DB\_description\_MoringaSchool\_w4.docx](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)
   2. Software( Github, Google Collaboratory, JIRA)
2. **Assumptions**
   1. The data provided is correct and up to date
3. **Constraints**
   1. There are no constraints

### **Data Mining Goals**

Our data mining goals for this project are as follows:

* To get a subset if the dataset where conditions match criteria
* To get a grouping of the sum of cars which match criteria .

**Data Mining Success Criteria**

Our success criteria will be measured by the following criteria;

* We get the hour when the number of cars picked up from stations is the highest

## **Data Understanding**

### **Data Understanding Overview**

For this project, we are using the availed dataset by the company. These datasets are

* Autolib datasets - This dataset gives the number of cars and hours
* Autolib descriptor - This dataset show the schema of the autolib datasets

### **Verifying Data Quality**

Autolib datasets had missing values but according to the autolib descriptor the missing values were expected

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded Autolib datasets from the CSV and loaded it into google collabs. The Autolib descriptors dataset was loaded into google docs

#### **Cleaning Data**

* Column names were formatted to change to lowercase, remove trailing and leading blank spaces and replacing ‘\_counter’ with ‘\_nos’
* Checked for null values and only found expected ones
* Checked for duplicates and found none
* Dropped columns deemed unecessary

#### **Deriving New Attributes**

Created new column called timestamp containing Year, month, day and time

## **Analysis**

During our analysis, we were able to identify the most popular hour of the day for picking up an electric car was when the highest number of cars were removed from the stations compared to the previous hour

Hour 22 i.e 10pm

Cars -71

## **Recommendations**

From our analysis, we would recommend that the company to prioritise increasing the number of cars available for sharing at 10 pm to meet the surge in demand

1. **Evaluations**

The results from data mining show 10pm is when demand is the highest and as such the company should increase the volume of cars available at thea hour to maximise on demand.

# 

[Collab notebook](https://colab.research.google.com/drive/1WMV3BZxCwbjfTtKY4Fn4IVamgHHu8ap4?usp=sharing)

<https://kevo.atlassian.net/jira/software/projects/IW4/boards/1>