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

### **Business Overview**

The electrical car business is the latest trend to be embraced in mobility and movement of people and goods. Most governments and business establishments are slowly picking up vehicles driven by batteries as opposed to fossil fuels. Growing this business and trend is a high priority for most governments and businesses.

Electric vehicles are commonly used for taxi, public transport and private transportation. This research focuses on the ride sharing industry that has taken up electrical vehicles as mode for transportation.

### **Business Objective**

1. Introduce the data you will be describing and the random variable that you are investigating.
2. State very precisely the null and alternate hypothesis that you will be testing.
3. Provide some explanation for why this hypothesis is important and/or interesting.

### **Business Success Criteria**

The research would use the following datasets

### **Assessing the Situation**

**1.** **Resource Inventory**

a. Datasets:

* Data Description . [[Autolib\_DDI\_DB\_description\_MoringaSchool\_w4.docx](https://drive.google.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view)]
* Data Set [[http://bit.ly/autolib\_dataset (Links to an external site.)](http://bit.ly/autolib_dataset)]

b. Software( Github, Google Collaboratory)

**2.** **Assumptions**

a. The data provided is correct and up to date

**3.** **Constraints**

a. There are no constraints

### **Data Mining Goals**

Our data mining goals for this project are as follows:

- Calculate the number of Blue cars, Utilib utilities and Utilib\_1.4

- Order the equipment based on their availability, location and popularity

- Top products preferred by the users in Paris and the other areas

- Find out what stations are more important and popular and why.

**Data Mining Success Criteria**

Our success criteria will be measured by the following criteria;

- All the cars and units are readily available and have equipment to support their availability for operations.

## **2.** **Data Understanding**

### **Data Understanding Overview**

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

- Autolib Dataset - This dataset gives the number of electrical vehicles used in Paris.

### **Data Description**

We have one dataset available for this project. A detailed description of the datasets is provided as follows:

1. Autolib Dataset**:**

Data was extracted from opendataparis.com, where the Autolib availability information was available in real-time. The accessed database was the following:

Name: Stations Autolib: Disponibilité en temps réel

Producer : Autolib

Date :

License : Open Database License

Any further publication should include this mention and respect the terms of the Open Database License.

### **Verifying Data Quality**

Some of the data in the dataset had missing values. There was also no known data errors in the dataset.

## **3.** **Data Preparation**

These are the steps followed in preparing the data

#### **1.** **Loading Data**

Loaded the datasets from the CSV and then created an SQLite database from them.

#### **2.** **Cleaning Data**

1. **Checking for Null Values:**

* The data set did not have null values that would affect the analysis of the project.

**1.1. Dropping the columns with Null Values**

* The null values found will be deleted as they seem to not have any usable data

1. **Checking for Duplicate Values**

- No duplicate values found in the data set

1. **Change the Column names to lowercase**
2. **Some columns were changed to appropriate data types for easier interpretation.**
3. **Checking consistency in the naming of the**
4. **Eliminating redundant data:**

* No redundant data observed

1. **Replacing of white spaces in the data set**
2. **Creating one date column by combining time, day, month and year**
3. **Dropping the now redundant day, month and year columns**

**3. Analysis**

**4. Hypothesis Testing**

## **Hypothesis**

**Null hypothesis**: The Picking up of these cars is more common at specific addresses compared to the others. (μ1 >= μ2)

**Alternative hypothesis**: The Picking up of the Taxis and cars is not different on any of the other addresses. (μ1 < μ2)

**5.Data Reduction Techniques**

The sample was selected using a random distribution based on the Interval Estimation based on the P-Values of the Data Set.

## **Recommendations**

We daily to reject the hypothesis at both 95% and 99% hence staying with the null hypothesis that states that the pick up of these cars and taxis is different on the different addresses.

The company running this business should definitely focus on the clusters of its customers and target them differently.

References

**Github Repository:**

<https://github.com/ch1b4d4/IP_1_APR_22/blob/main/ip_samson_nyiro_1_april_2022.ipynb>