## **Electric car share Data analysis in Paris**

### **Business Understanding.**

#### **Problem definition**

From the dataset and the research question for this project it requires the data scientist to process stations data to understand electric car usage over time by solving for the following questions:

#### **Business Objectives**

The research aims at discovering the usage of electric car per station in city of Paris,the following questions will assist in coming up with a guide from analysis to help make the right decision at the end and be able to give a recommendation:

* What is the most popular hour of returning cars
* What station is the most popular overall and at the most popular picking hour
* What postal code is the most popular for picking up blue cars? And does the most popular station belong to that postal code? Either overall and at the most popular hour
* Do the result change if you consider Utilib and Utilib 1.4 instead of blue cars

#### **Assess current situation**

For a project to be successful and to go through all the phases without missing out the process as a data scientist we have to consider the risk management, identify our resources,time,computer and software availability and other constraints that may impede the process of analysing the data and giving the accurate/appropriate recommendation which is crucial for determining the decision making.

Data is obtained from Paris electric car sharing

Constraints involved in this project time to wrangle through the dataset

Computer resources and software python programming Language and its packages especially pandas ,Numpy ,sklearn (LabelEncoder)and Matplotlib

#### 

#### 

#### **Data Mining Goals**

#### **Project Plan**

|  |  |
| --- | --- |
| Data sourcing | Download the data to be used is provided from the research |
| Data Understanding | Data collection,data description and data verification |
| Data Preparation | Select important attributes of the data,  Clean the data  Construct,integrate and format the data |
| Analysis | Answer main objective and subsequent research questions. |
| Recommendation | Best recommendation from the insight delivered from the dataset |
| Conclusion |  |

### **Data understanding.**

#### **Data collection**

The data is provided from the research and it look as follows:

df\_auto = pd.read\_csv('Autolib.csv')



#### **Data description**

There are 5000 entries and 24 columns

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Address | Address of the station |
| Cars | No. of Cars Available at Station |
| Bluecar Counter | No. of blue cars available at station |
| Utilib Counter | No. of Utilibs Available at the Station |
| Utilib 1.4 counter | No. of Utilibs 1.4 Available at the Station |
| Charge Slots | No. of charging Slots available at Station |
| Charging Status | Status of the charging cars nonexistent,operational,broken or future |
| City | City |
| Displayed comment | Comments |
| ID | Id of the station |
| Kind | Either Center,Space and Parking center |
| Geo Point | GPS coordinates |
| Postal Code | Postal Code of the Station |
| Public name | Name of the Station |
| Rental status | Whether the station is Available for renting vehicles |
| Scheduled at | Planned Opening date |
| Slots | No.of Parking slots available at Station |
| Station type | Station type either station,full\_station,sub\_centre |
| Status | Status of the station ok,closed or scheduled |
| Subscription Status | Can either be non-existent,Operational,broken and future |
| Year | Year |
| Month | Month |
| Day | Day |
| Hour | Hour |
| Minute | minute |

#### **Data Verification**

After keenly going through the dataset there were no missing values that would cause any impact toward drawing the insight for this project,however there was a column that was irrelevant for my study and it will be dropped in the next phase.

### **Data Preparation**

#### **Data Selection**

The dataset to be used in this project will only miss one column of the original dataset. The other columns are relevant for drawing the necessary insight for the project.

#### **Data Cleaning**

The dataset selected is accurate,complete,consistent and it's also uniform as these are the factors crucial for a dataset that can give correct insight and will eliminate errors in decision making.

The dataset remains with 23 columns and 5000 entries

### **Data Analysis.**

This phase is conducted after obtaining clean data that is carefully selected from the above processes which in turn will help to draw insight that is crucial during decision making.

The research question to be answered that led to this project are:

Determining the count of the most popular city by the hour; i discovered that hour 21 was the most common in the most common city which is paris

The Most popular hour of returning vehicles.

### **Recommendations**

### **Conclusion.**

[Github Link](https://github.com/ArthackA/moringaIP4) for the project

[Jira link](https://arthacka.atlassian.net/secure/RapidBoard.jspa?projectKey=EL&rapidView=7&atlOrigin=eyJpIjoiOTM0MWM0MTRjOTRjNDEzNWJiZTdjY2FkNGNjNjdhZGEiLCJwIjoiaiJ9)