1. **Business Understanding**

**Business overview**

Autolib was formed to complement the bike-sharing service and was the first electric car-sharing program in the France region. The electric cars are referred to as ‘Blue Cars’, and have stations spread across the region where users can subscribe, pick or drop off cars. This project requires us to use the stations’ data to understand the electric car usage over time.

Below is a link to the GitHub repository that contains the data preparation and analysis:<https://github.com/Caleb83-Moringa/Independent-Project---Week-4>

**Business objective**

Main aim is to draw insights from the provided data set to identify an hour of the day when the most blue cars were picked in the month of April, 2018 in Paris.

**Business success criteria**

To gain an understanding on the electric car-sharing usage in Paris over time.

**Assessing the situation**

I have signed up and done the Jira Assessment with my personal email: muindecaleb83@gmail.com because the account I created with my school email a few months ago free trial is expired**.**

Below is a link to the JIRA CSV FILE EXPORT download in my drive:<https://drive.google.com/file/d/18Gxv31E4Iwb6GCbz5LU9h43RAsk2K0SH/view?usp=sharing>

Resources

1. Personnel: Technical and data science support

2. Datasets

<http://bit.ly/autolib_dataset>

[Link](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)

3. Computing resources

4. Software-Python , Collaboratory , GitHub , JIRA , Google docs

**Assumptions**

The available data is sufficient and will accurately represent the Autolib entire dataset.

Data will answer the research question and give a usage trend of the electric cars both in the short and long run.

**Risks and contingencies**

There is possibility of bias due to missing values or human errors

**Cost/Benefit Analysis**

The costs are minimal in this project

**Data mining goals**

Goal of the project is to determine the extent of electric car usage in the France region over time.

**Potential questions to be considered;**

● What is the most popular hour of the day to pick up a Bluecar in Paris over the month of April,2018?

**Other questions to consider**

➔ Most popular hour for returning cars?

➔ 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?

➔ Do the results change if Utilib and Utilib 1.4 are considered instead of Blue cars?

**Project Plan**

The CRISP-DM will be used as a guide in this project. Below is a plan towards how the project will be conducted;

| Phase | Time | Resources | Risks |
| --- | --- | --- | --- |
| Business understanding | 45 minutes | Autolib dataset and description |  |
| Data understanding | 1 hour | Autolib dataset and description |  |
| Data preparation | 2 hours | Autolib dataset |  |
| Data analysis | 2 hours | Autolib dataset and description |  |
| Recommendation | 30 minutes | Dataset insights |  |

1. **Data Understanding**

**Overview**

There are two data sets available for review. One is a description of the main data set to be used in answering the research question. Data set is a sample collected on the usage of shared electric cars in France.

The data sets provided are as follows,

1) <http://bit.ly/autolib_dataset>

2) [Link](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)

Collecting initial data

Data was sourced from opendataparis.com and is a sample from April 1 to April 9, 2018. The data was collected at three different times and was accessed from a database called Stations Autolib.

**Description and Exploration of the data**

Autolib dataset.csv

This provides the autolib information on stations and the usage of the shared electric cars over 9 days in April 2018. Data provided has null values attributed to a failed download during the collection period. Data set has 5,000 entries and 25 columns.The record data types are strings and integers.

The most resourceful columns in answering the research question are, the bluecarcounter , date column , and the public name column which contains the name of the station .

**Verifying data quality**

The data set does not contain a lot of missing values. The data is consistent as it does not contain any duplicates.

1. **Data Preparation**

Data set used in the analysis autolib.csv. Data was loaded to Python and a delimiter used to separate the comma separated values into columns.

**Cleaning data**

1. Removing columns that seemed irrelevant in the data set.

2. Changed the column names to a uniform naming scheme

3. Changed respective columns to appropriate data types

4. Checking for null values and found none missing.

**Integrating the data**

Merging the day ,month, year, minute, and hour columns into one column and converting them to date time data type.

1. **Analysis**

Analysis answered the following questions,

Most popular hour for picking up blue cars in Paris: 11p.m

Most popular station overall : Paris/Porte de Montrouge

Most popular postal code for picking up blue cars : 75015

Does the most popular postal code belong to the most popular station? No. Paris/Porte de Montrouge is in the code 75014

Do the results change when Utilib 1.4 and Utilib are considered? No

Below is a link to the GitHub repository that contains the data preparation and analysis:<https://github.com/Caleb83-Moringa/Independent-Project---Week-4>

1. **Recommendation**

1.Most popular station should have all charging slots working and not broken

2. Most popular postal code ,75015 , should have more stations.

3. Optimization such that at the time that most cars are being picked , it is a smooth process.

1. **Evaluation**

Using insights from the data, we have been able to gain an understanding on the usage of electric cars through an eclectic car sharing company in France. From the analysis, the most popular station should be maintained in such a way that it never has broken charging slots and its status is always operational.