**1.BUSINESS UNDERSTANDING.**

* **Business Overview**

Dalberg is a global working group that aims for inclusivity and sustainable growth around the world in terms of providing an innovative mix of advisory, research, investments, analytic and design services.

The group believes in the power of complex global problems requiring new solutions.

* **Business Objective**

The objective of this report is to draw insights from stations data of an electric car sharing service company and to understand the usage of electric cars over a period of one month.

**Assessing the Situation.**

**Requirements, Assumptions and Constraints.**

* **Resources**
* Personal (Technical support, Data mining experts)
* Project Datasets used

**a.** <http://bit.ly/autolib_dataset>

**b.**  <https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing>

* Computing resources
* Software ( Google Colaboratory, Google Drive, Github, Jira)
* **Assumptions**
* Data sampled will be an accurate representation of the whole Dataset.
* Data given represents a uniform trend in the electric car usage in the area.

**Data Mining Goals**

Our data mining goal is to determine the most popular hour of the day for picking up the electric Blue Car for the month of April 2018.

Potential questions for research include:

* What is the most popular hour for returning cars?
* Which station is the most popular overall and at the most popular picking hour?
* What postal code is the most popular for picking up?
* Do results change if we consider Utilib and Utilib 1.4 instead of Blue car?

**Project Plan**

Cross Industry Standard Process for Data Mining will be used as a guidance for conducting this research. Here is an overview of the plan.

| **Phase** | **Time** | **Resources** | **Risks** |
| --- | --- | --- | --- |
| Business Understanding | Less than an hour | Project Datasets  Data Analysts/Scientist |  |
| Data Understanding | 1 hours | Project Datasets  Data Analysts/Scientist |  |
| Data Preparation | 8 hours | Project Datasets  Data Analysts/Scientist |  |
| Data Modelling | Less than an hour | Project Datasets  Data Analysts/Scientist |  |
| Evaluation | 1 hour | Project Datasets  Data Analysts/Scientist |  |

**2.DATA UNDERSTANDING**

**Data understanding Overview**

The existing dataset consists of sample data collected from the station’s data about the usage of electric cars for the month of April in 2018. We will not need any additional dataset apart from what we have to meet our objective. The datasets used were:

1. <http://bit.ly/autolib_dataset>
2. <https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing>

**3. DATA PREPARATION**

**Collecting Initial Data**

The source of the data collected was the company’s database.

**Describing and exploring Data**

There are two datasets available. One of the datasets describes the columns and values of the data sets. The other dataset contains all of the details of the cars, locations and users information including date and time.

A further description of the provided datasets is as follows:

**Auto lib Data set -** This dataset contains the details of the information of the company’s data. It contains no null values, it has 19 columns and 2807 entries. All the values are represented in the data set and the data set takes some time to load, approximately 10 minutes.

**Autolib DDI DB description -** This dataset contains the description of the auto lib data set. It has details about the fields in the auto lib dataset.

A data exploration report can be written for further investigation.

**Verifying Data Quality**

The dataset does not contain missing values but it has some errors. For example there are some unnecessary columns we had to drop and rename some fields.

**Selecting Data**

The following datasets were used for analysis in this project based on the relevance of our goals and data quality.

* Auto lib data set
* Auto lib DDI DB Description

Used data frames to load data from a file, examine basic statistics of the data, change some values and finally get the output which is the results.

**Cleaning Data**

Data cleaning procedures performed during the analysis included:

* Dropped all irrelevant columns for analysis so we could remain with the beneficial columns. Also added relevant columns.
* Dropped duplicate values and information from the dataset.
* Checked for null values in the data and resolved it.
* Dropped all rows where the values of blue car were less <0 and >7

**Constructing New Data**

* New data was created upon combining the date and time columns and renaming them then dropping some irrelevant columns.
* New data was also created by creating a new data frame where the city was Paris. Then we had to reset indexes and assign new dates and new time to get the most popular hour.
* A new column was created to get successive blue cars.

**4. DATA ANALYSIS AND MODELLING**

**During analysis, the following question was answered.**

1. Which was the most popular day for picking up the blue car? Sunday
2. Which was the most popular hour? 1840 hrs and 2100hrs
3. Most popular station? Porte de Montrouge
4. Most popular postal code? 75015

The above analysis was done using the python programming language. This analysis can be found in the following python notebook.

<https://colab.research.google.com/drive/1SrAmkrFpakeJc-0H1K10MdZ6mRKeQgab#scrollTo=PxDYpGkA3reJ>

**5. RECOMMENDATION**

Following my analysis, the following recommendations are provided:

* The days with the highest frequency , like on Sunday at around between 6 and 9pm should be considered first. There should be a sufficient number of cars at the time.

**6. EVALUATION**

From our business overview we have been able to determine the frequency of cars getting picked up, returned or if nothing happened. Our approach was to be able to figure out most used cars, at what hour and the frequency of usage. For purposes of further investigation we have also provided further questions that can provide more insight while mining data.