**1.Business Understanding**

**Business Objectives**

**1.1 Business Overview**

The role assumed in this project is of a data scientist asked to process stations data for an electric car-sharing service company in order to understand electric car usage over time in the City of Paris.

**1.2 Business Objective**

This research seeks to aid the electric car company to:

* Determine the most popular stations
* Determine the most popular hour of the day for picking up electric cars in Paris
* Determine the most popular hour of returning the electric cars in Paris

**1.3 Business success criteria**

To draw insights from a 9 day dataset across various stations in the city of Paris to assist in understanding the usage of electric cars over time and maximize profits.

**Assessing the situation**

Resource Inventory

a.) Datasets

* [<http://bit.ly/autolib_dataset>]
* [[Link]](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)

b.) Software(Github, JIRA, Google Colaboratory, Python)

**Assumptions**

1. The data provided is correct and there are no missing values

**Constraints**

1. There are no constraints

**Data Mining Goals**

Our data mining goals for the project are as follows:

1. ) To identify the most popular hour for returning and picking the cars
2. )To identify the most popular station

**2. Data Understanding**

**Data Understanding Overview**

We have one dataset [<http://bit.ly/autolib_dataset>] , The dataset contains postal and physical addresses,stations, bluecar counter, picking and returning time.

**Initial Data Collection**

Data was extracted from opendataparis.com, where the Autolib availability information was available in real-time. The sample contains data from April 1 to April 9, 2018 over a period of nine days.

**Describing and Exploring Data**

The dataset contains a total of 25 columns and 5000 rows. The below table is a full description of the dataset.

|  |  |  |
| --- | --- | --- |
| Column Name | Type | Comments |
| Address | String | address of the station |
| Cars | Number | Number of cars available at the station - redundant with Bluecar counter, always the same  value |
| Bluecar counter | Number | Number of Bluecars available at the station |
| Utilib Counter | Number | Number of Utilibs available at the station |
| Utilib 1.4 Counter | Number | Number of Utilib 1.4 available at the station |
| Charge Slots | Number | Number of Charging slots available at the station |
| Charging Status | String | Whether the station is operational for recharging. |
| City | String | City |
| ID | String | ID of the station |
| Kind | String | Station, space,parking centre |
| Geo Point | String | GPS coordinates of the station |
| Postal Code | Number | Postal code of the station |
| Public Name | String | Name of the station |
| Scheduled At | String | Planned opening date |
| Rental Status | String | Whether the station is available for renting vehicles. |
| Slots | Number | Number of parking slots available at the station? |
| Status | String | “Ok”, “closed”,”scheduled” |
| Station Type | String | Was that actually a selling  point for Autolib subscriptions? |

**Verifying Data Quality**

The dataset provided is complete and does not contain any errors. All of the columns included are necessary for our analysis.

**3. Data Preparation**

Steps taken during data exploration are as follows

1. **Selecting Data**

The following dataset was used for analysis in this project based on the relevance of our goals and data quality

* [<http://bit.ly/autolib_dataset>]

We used data frames to load data from a file, examine basic statistics of the data, change and replace some values and finally output the results.

1. **Cleaning the data**

Data cleaning procedures performed during the cleaning process include:

* Dropping irrelevant columns like Displayed comment, Scheduled at, Geo point and subscription status
* I checked for variables with leading and trailing spaces and there was none in the dataset
* I checked for outliers using boxplot and found some in the Utilib and Utilib 1.4 columns and dropped them
* I changed the minute, day, hour columns to datetime for relevance.
* Finally, I checked for accuracy and the data provided is accurate.

**4. Analysis**

During our analysis, the following questions were answered.

1. What is the most popular hour of day for picking a car in Paris?

From the analysis the most popular hour of day for picking a blue car is at 9PM

1. What is the most popular hour of day for returning a blue car in Paris?

The most popular hour of day for returning a blue car is 2PM

1. What is the most popular station? The most popular station overall is Paris/Grenelle 31/Parking

**5.Recommendations**

From the analysis, the company should invest in having more blue cars available at the Paris/Grenelle 31/Parking station at the 9PM to match the demand and maximize on returns.

**GITHUB REPOSITORY LINK:** [**https://github.com/vngeno/BLUE-CARS-PARIS**](https://github.com/vngeno/BLUE-CARS-PARIS)

**KANBAN BOARD: https://vanessan.atlassian.net/jira/software/projects/MOR/boards/1**