**A REPORT ON AN ELECTRIC CAR-SHARING SERVICE COMPANY**

This report is a rigorous study on an electric car sharing company in France. This fact finding study sets out to assess the situation so as to meet the company's business objectives while still catering to its customers while appraising how data mining techniques will govern these achievements.

I chose to employ the Cross-Industry Standard Process for Data Mining Model (CRISP-DM), an all inclusive data mining methodology suitable for this study. It will be broken down into 6 well known phases as seen below:

**Business understanding**

This requires subject matter expertise i.e. understanding the problem at hand as a business problem rather than an analytical problem, so as to reach the main objective.

The car sharing company’s objective is to make an upgrade so as to avoid churning, improve their customer retention, recuperate their operational value and accelerate their profitability.Here, I am able to clearly lay bare any hindering factors that may govern the end-result of the operation.

*Desired outputs:*

1. Set Objectives- it is necessary to detail the foremost priority of the company from a business point of view while still addressing the sub-objectives that may arise from the main objective.
2. Provide a project plan- it is believed that a dream without a plan is just a wish, therefore to set the ball rolling, I have to depict a well thought out plan that will aid in the achievement of the data mining and business goals. The plan will clearly detail the approach to be taken through out the project.
3. Business success criteria-so as to ensure a successful completion of the project, quantifiable terms that are admissible to the end-user,customers and the stake-holders are put into place .

Subsequently, I assess the company’s situation, moreso the financial situation, the personnel they are working with, the kinds of data they may have, their computing resources as well as the relevant software.

**Data understanding**

This phase contains the station’s data that will aid in the understanding of the electric car usage over a period of 9 days. The data was extracted from opendataparis.com and was available in real-time,produced by Autolib.

The sample dataset contains data from April 1st to April 9th in the year 2018. During the collection periods, some data failed to download on some occasion hence creating some gaps in the data. There are 549 missing files in the month of April (1.27%), whereas 98.73% of the data for the same month is available.

This dataset contains a total of 26 columns and 14421829 entries with 13 registered integers and 13 objects.

The fields in this dataset are as follows:

|  |  |  |
| --- | --- | --- |
| **Name of field** | **Description of field** | **Data types** |
| Address | This displays the address of the station | object |
| Cars | This displays the total number of cars available at the station | integer |
| Bluecar Counter | This displays the total number of Bluecars available at the station | integer |
| Utilib Counter | This displays the total number of Utilib cars available at the station | integer |
| Utilib 1.4 Counter | This displays the total number of Utilib 1.4 cars available at the station | integer |
| Charge slots | Number of charging slots available at a station | integer |
| Charge status | Addresses whether a station is available for recharging | object |
| City | Displays the name of the city | object |
| Displayed comment | Displays comments made | object |
| ID | Displays the ID of the station | object |
| Kind | States whether the location is a station, a parking area, a space or a center | object |
| Geo Point | Displays the gps coordinates of the station | object |
| Postal code | Displays the postal code of the system | integer |
| Public Name | Displays the name of the station | object |
| Rental Status | States whether or not a station is viable for renting out a vehicle | object |
| Scheduled at | States the planned opening date | object |
| Slots | Number of parking slots available at the station | integer |
| Status | States whether station is operating as a full station, a station or a sub center | object |
| Subscription status | States whether it is possible to subscribe to the autolib service at a particular station | object |
| Year | Specifies the date the car was picked up | integer |
| Month |
| Day |
| Hour | Specifies the time of the day the car was picked up | integer |
| Minute |

**Data preparation**

Once I was able to understand the data within the given dataset, I had to prepare the data by cleaning it up. First and foremost I created a datetime object from the year,month and day columns as well as the hour and minute columns and therefore giving rise to a new column, the timestamp column, that displayed both the date and time. I thereafter dropped columns that I considered irrelevant for the analysis step. The columns dropped are:- ‘Displayed comment’, ‘Unnamed: 0’, ‘year’, ‘month’, ‘day’, ‘Scheduled at’, ‘cars’, ‘Charge Slots’,’Charging status’ and ‘Station type’.

Subsequent to that I counted the missing values of which I found to be zero and that is solely because all the null values came from one column, i.e. the Scheduled at column.

To avoid any redundancy, I looked for any duplicated data in the given dataset then went ahead to delete the duplicates so as to be left with the original data. Afterwards, I set all the columns to lowercase and added underscores to columns that had two names for a neater look.

Once all that was done I exported the cleaned dataset for the next phase which is the analysis of the data.

**Data analysis**

Once the data was cleaned I immediately embarked on the analysis. I created three new columns that highlighted how and when the cars were accessed by the company’s subscribers throughout the nine days,so as to analyze each separately. I started with the Blue car counter,as I tried to decipher how and when it was either in use or not use.

For the Bluecar I found that the most popular hour of the day that the car was picked up was 4am whereas the most popular return time was at 2100hours(9pm). The most popular station for picking the Blue cars overall is the **Paris/UzÃ¨s/1.** However, during the most popular hour the most popular station turned out to be **Paris/Bosquet/41.**

75015 was seen to be the overall most popular postal code for picking up Blue cars, however this postal code does not belong to the most popular station i.e. **Paris/UzÃ¨s/1,** whose postal code is 75002.

The most popular code for picking up Blue cars at the most popular hour was 75016.

For the utilib car, the most popular picking hour was seen to be 5 am whereas the returning hour was 7am. The most popular postal code overall was seen to be 75015, which is similar to the most popular overall for the Blue car. In addition to that, the postal code was retained for the most popular hour. The station at the most popular hour was **Paris/Commerce/5** , therefore having the most popular code used at the most popular picking hour belonging to the most popular station.

Lastly for the utilib 1.4, the most popular picking and returning hour was maintained at 6 am. Its most popular picking postal code is 75015 in both the overall and most popular picking hours. The most popular picking station for the utilib 1.4 is **Gennevilliers/Vieux Chemin de Saint-Denis/106**. However the most popular postal code does not belong to the most popular station

**Evaluation**

As seen in the analysis above, the common denominator for all three cars is seen to be the most popular postal code i.e. 75015 which belongs to the **Paris/Commerce/5** station, which was the most popular station for picking up the Utilib car at the most popular hour.

**Recommendation**

If the car company can maximize on the use of the **Paris/Commerce/5** station, they might increase profits especially for the Utlib 1.4 that seems to be lagging behind.

**GITHUB REPOSITORY LINK**

<https://github.com/StephanieO-jpg/IP4>