**GIT-HUB LINK-** <https://github.com/RaymondKiplangat/WEEK-4-IP.git>

**DATA REPORT**

**1.BUSINESS UNDERSTANDING**

**Business Overview.**

Electric car sharing company is a company that rents electric cars to the users and makes a lot of profit in return .The money to be paid can be based on the number of hours you've rented the car or even the services you've been provided, example is ‘charging fee’ and others .

**Objectives.**

1. The objective of this project is to identify the most popular hour of the day for picking up a shared electric car (Bluecar) in the city of Paris over the month of April 2018.
2. In addition, which is the most popular hour for returning cars?,
3. station that is most popular and Its most popular picking hour?,
4. postal code the most popular for picking up Blue cars?
5. And to show if the most popular station belongs to that postal code?
6. Also experiment results with Utilib and Utility 1.4 instead of Blue cars?

The above objectives will enable the company to operate effectively and thus increase production to the customers.

**Business success criteria**

To enable the company to monitor the business well thus increasing the income and also operate well to their customers.

Best of my knowledge the analysis is the best.and can be used to implement the project or monitor the project.

**Assessing the situation.**

**Requirements, Assumptions and Constraints**

**Resources:**

1. Personnel -data mining expert.
2. Computing resources
3. Softwares (collaboratory,github)
4. Project Datasets.

* Autolib dataset- <http://bit.ly/autolib_dataset>
* Data description: <https://bit.ly/3wnjnqj>

**Assumptions:**

1. Data samples represent the entire dataset of these projects.

**Constraints:**

1. No constraints at all.

**Data mining Goals**:

Our data mining goal is to identify the most popular hour of the day for picking up a shared electric car (Bluecar) in the city of Paris over the month of April 2018.

Potentials questions for consideration:

* What is the 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 you consider Utilib and Utilib 1.4 instead of Blue cars?

**Project plan.**

In regard to this project in followed the CRISP DM methodology to do my analysis and research concerning this project that is:

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Time | Resources | Risks |
| 1. Business understanding | 1 hour | Datasets  Data Analyst | Insufficient business information. |
| 1. Data understanding | 2 hours | Datasets  Data Analyst | Insufficient dataset description. |
| 1. Data preparation | 2 hours | Datasets  Data Analyst | Some of the columns and were null(no values) |
| 1. Analysis | 3 hours | Datasets  Data Analyst | Missing data could lower the accuracy of the Analysis |
| 1. Recommendation | 1 hour | Datasets  Data Analyst | To check on the dataset inorder to improve the accuracy for the next projects. |
| 1. Evaluation | 1 hour | Datasets  Data Analyst |  |

**2.DATA UNDERSTANDING**

**Data understanding overview**

I used the Autolib dataset to do my project,links for my data are listed below.

Autolib Dataset: <http://bit.ly/autolib_dataset>

Autolib description: <https://bit.ly/3wnjnqj>

**Collecting initial data:**

The data collected was sourced from the company's database infrastructure. This data is a sample from the original company's dataset.

**Describing and explaining the data.**

There are two data sets available . The two are descriptive of the column and the values and the location details of the users and the cities for some specific days.

Autolib dataset:

The dataset outlines the description of the column names and the format. No null values

Some of the columns were not necessary.

Autolib description:

It just describes how the dataset should be analysed and the meanings of the columns.

**Data quality.**

In terms of completeness I can say the data was complete(approximately 98% complete),for sure some of the values were missing and others were repeated(duplicates).

but all the data were correct for this project(mend for this project)

To achieve a better quality in future projects I will recommend the collection of data for areas that were missing and also duplicated areas.

**3 .DATA PREPARATIONS**

I can say this is the stage that matters the success of any project. For my case I used many techniques and steps to clean my data.

**3.1 loading my data.**

I first decided to load all my dataset so that I can easily identify the errors.in the dataset (Autolib dataset) and display them in dataframes .

**3.2 cleaning my data.**

Data cleaning procedures performed :

* Dropping of the irrelevant columns
* Check the null values .
* Creating a datetime column and converting it to the date time format.

**4.DATA ANALYSIS**

During the analysis the following questions were answered:

1. Identify the most popular hour of the day for picking up a shared electric car (Bluecar) in the city of Paris over the month of April 2018.

* Timestamp('2018-04-08 09:42:00')

Bonus Questions (Optional)

1.What is the most popular hour for returning cars?

* Timestamp('2018-04-06 07:02:00')

2.What station is the most popular?

* + Overall?
* Paris/Olier/15
  + At the most popular picking hour?
* Paris/EugÃ¨ne OudinÃ©/51
* What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?
  + Overall?
* 75015
  + At the most popular picking hour?
* 75017
* Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

YES -I did the analysis with both the utilib and utilib 1.4.

**5.RECOMMENDATION**

Following the analysis i provide these recommendations :

1. The stations with the maximum overall pick should be upgraded to enable many customers meet their needs
2. The stations that are not operational but rent vehicles to those that are operational during their highest picking hour.
3. Those stations that are operational but their charging status are not operational should be repaired so as to meet all the needs of the customers.

**6.EVALUATION.**

I have done a brief analysis of the of the three that is blue cars counter, utilib counter and the utilib 1.4 counter and concluded that there is one relationship between the blue cars and utilib returning hour. That is the time is the same.

Also the highest picking hours are morning hours.

**GIT-HUB LINK-** <https://github.com/RaymondKiplangat/WEEK-4-IP.git>

.