**DATA SCIENCE REPORT**

**1.BUSINESS UNDERSTANDING**

**Business Objectives**

**Business Overview**

In this research , we are expected to analyze the given dataset to understand electric car usage over time. We are expected to find out the popular hour for picking up a shared electric car across the different cities and also the most popular hours for returning the cars.

**Business Objective**

The main objective of the report is to find out the insights from the dataset of the usage of cars in the cities, the most popular hours and the most popular stations.

**Business Success Criteria**

To determine the most popular stations, the most popular postal codes and the popular hour.

**Assessing the Situation**

**Requirements, Assumptions and Constraints**

1. **Resources**

**i)** Personnel (Data Cleaning Expert, Data Mining Expert)

**ii)** Project Dataset link

**iii)** Computing resources

**iv)** Software (Google Python Colaboratory Notebook, GitHUB, Jira etc)

1. **Assumptions**

i) Data taken for the analysis is an accurate representation of the exact findings.

ii) The data that is available shows a uniform usage trend.

1. **Constraints**

No constraints were there in the dataset

1. **Risks and Contingencies**

The data may be biased.

1. **Cost/ Benefit Analysis**

Knowing the popular stations and the popular hour is crucial in insights on the potential revenues or marketing strategies. This is however dependent on a number of factors such as the resources allocated.

More investigation can be done and a cost benefit analysis can be produced in order to provide detailed and precise information on the cost-benefit analysis of the project.

**DATA MINING GOALS**

Our data mining goal is to determine the usage of electric cars across all cities, the popular hours and the areas postal codes that register the highest usages.

Potential questions for our analysis is:

* The most popular hour for returning cars
* The most popular station
* The most popular station and the postal code it belongs to

**Project Plan**

|  |  |  |  |
| --- | --- | --- | --- |
| PHASE | TIME | RESOURCES | RISKS |
| BUSINESS UNDERSTANDING | 1 HOUR | The Project Dataset/ Data Scientist |  |
| DATA UNDERSTANDING | 2 HOURS | The Project Dataset/ Data Scientist |  |
| DATA PREPARATION | 2 HOURS | The Project Dataset/ Data Scientist | Some assumptions |
| DATA MODELLING | 2 HOURS | The Project Dataset/ Data Scientist | Some assumptions |
| EVALUATION | 1 HOUR | The Project Dataset/ Data Scientist |  |

**2. DATA UNDERSTANDING**

**Data Understanding Overview**

The below dataset comprises of the sample data collected from electric car companies.

The following dataset will be in our analysis. [Autolib Dataset](http://bit.ly/autolib_dataset)

**Collecting Initial Data**

The data collected was sourced from the electric car sharing service company.

**Describing and exploring Data**

The dataset gives information on the electric car usage, the location, their corresponding postal codes, the day and the month.

The data exploration report can be written for further investigation.

**Verification of the Data Quality**

The dataset does not contain alot of missing values. The two columns that contained missing values were the displayed comment and scheduled at which I dropped. There was also no duplicated columns.

**3. DATA PREPARATION**

Steps taken during the data exploration are as follows:

**Selecting Data**

I used the dataset provided without any external resources.

I used dataframes to load data from a file, examine the basic information of the data, standardizing the column names.

**Cleaning the Data**

The data cleaning procedures that I undertook were:

* I chose the relevant columns that will give tangible results
* I also standardized the column names to lower case, removed the spaces between words.
* I also did a background check of duplicated rows which was not found and also a null items check on our newly formed dataframe.

**4. ANALYSIS**

During our analysis, the following questions were answered:

* Identifying the most popular hour of the day for picking up a shared electric car(Blue car) in the city of Paris over the month of April 2018.

hour

21 -177.0

7 -173.0

20 -149.0

13 -142.0

4 -133.0

16 -132.0

6 -129.0

2 -126.0

8 -125.0

3 -122.0

5 -121.0

18 -120.0

17 -118.0

9 -116.0

19 -115.0

1 -109.0

22 -108.0

23 -108.0

14 -105.0

15 -101.0

11 -95.0

10 -93.0

0 -92.0

12 -87.0

The most popular hour for returning the cars:

hour

6 313.0

5 287.0

12 279.0

2 276.0

3 269.0

1 261.0

13 251.0

4 243.0

0 242.0

10 240.0

7 240.0

8 237.0

18 236.0

11 234.0

23 231.0

22 229.0

9 227.0

15 222.0

20 206.0

21 201.0

17 200.0

14 189.0

19 183.0

16 182.0

This was the most frequent postal code for picking up cars overall:

postal\_code

75008 -314.0

75015 -290.0

75012 -280.0

75017 -249.0

75016 -226.0

The above analysis was done in python programming language and can be found in [this](https://drive.google.com/open?id=1QwqyDrMFrjgfueEgCorHyxYk6ejUteCt) notebook.

**5.RECOMMENDATION**

The following were the popular hours:

At 2100 hour was the most popular for bluecar electric cars being picked in the city of Paris with 177 cars being picked.

At 0600 hours was the most popular hour for cars being returned for bluecar electric cars with 313 cars being returned.

75008 was the most popular postal code overall for picking up bluecar electric cars.

**6.EVALUATION**

From out business success criteria, we have far been able to find out popular hours in our dataset and the popular postal codes for both picking and returning of the cars. This can be used in marketing strategy and upgrade of technology and infrastructure.