1. **Business Understanding**

**Business Objectives**

**Business Overview**

Autolib is one of car sharing service company in Paris, France and would like to process stations data to understand electric car usage over time for its subscribers in France.

Data scientists for the company , we are required to study and analyze the Autolib dataset, aiming to aid in understanding electric car usage.

**Business Objective**

The objective of this project is to draw insight from a dataset sampled over 9 days across a city of Paris in France.These insight will act as an upper hand in understanding the rate at which electric cars are preferred.

**Business Success Criteria**

To determine the performance of an Electric car over the 9 days period.

**Assessing the situation**

**Requirements, Assumptions, and Constraints**

1. **Resources**

i . Personnel(Technical support, Data mining experts)

ii. Project Datasets

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

iii. Computing resources

Iv. Software(Collaboratory,Google Suite, Github, etc)

**b.Assumptions**

i. Data sampled will be an accurate representation of the entire dataset

ii. The data given represents a uniform product consumption trend users in the long run

**C. Constraints**

I. The constraint is that the dataset used is large in size.This could make the task stall more than expected.

**d. Risk and Contingencies**

i.We are provided with sample data that could lack vital information.

ii.There may be bias in our sampled data.

**e. Cost Benefit Analysis**

The cost associated with consultancy and technology infrastructure implementation cost

will be minimal to the potential revenues of the projects.

However, this will be independent on factors such as funding and affiliating with right pa

Partners.

**Data Mining Goals**

Our data mining goal for this project is to determine how different customers are subscribing to use

Electric car.

Potential questions for consideration include:

* Which time in Paris does a company receive more incoming requests for a shared electric car (Bluecar)?
* 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 Utility 1.4 instead of Blue cars?

**Project Plan**

The Cross\_industry Standard Process for Data mining (CRISP-DM) will be used as a guideline for conducting this research.Below is an overview plan for this study:

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Time | Resources | Risks |
| Business Understanding | 2 hours | Project Datasets  Data Analyst/Scientist |  |
| Data Understanding | 2 hours | Project Datasets  Data Analyst/Scientific | Insufficient datasets |
| Data Preparation | 3 hours | Project Datasets  Data Analyst/Scientist | Insufficient datasets |

**2 .Data Understanding Overview**

The existing dataset files we have comprises sampled data collected from car service subscribers.The data files that we will need for this project will be:

1.Autolib dataset [http://bit.ly/autolib\_dataset]

**Collecting initial data**

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

**Describing and exploring data**

There is only one dataset available for data analysis.

A further description of the provided dataset is as follows:

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

**3.Data preparation**

Steps were taken during data preparation are as follows:

**Selecting Data.**

The following dataset was used for analysis:

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

Used a DataFrame to load data from a csv file, examine basic statistics of the data , change some values and finally output the result.

**Cleaning Data**

Data cleaning procedures performed during analysis include:

* Drop column cars since it has the same values with the column for Blue cars.
* Drop column Address since we can get it through postal code.
* Drop column for Displayed comment since it has got alot of missing value

**Constructing new data**

New data created upon splitting city names. This has been described in the integration section.

**Integrating and Formatting Data**

Splitted cities are now grouped together based on hour and type of cars.

Changed the date and time column data type to date time for us to work with the column date time which would give results when doing analysis.

**4. Analysis**

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

hour

3 49

12 48

10 48

6 46

8 44

21 43

11 42

19 42

15 42

18 40

9 40

0 39

22 38

2 38

1 38

16 37

5 37

23 35

13 33

7 33

17 33

4 31

20 27

14 25

2. What is the most popular hour for borrowing a car

hour

21 80

16 68

7 64

13 62

5 62

23 61

20 61

3 60

8 59

4 58

18 57

9 57

17 55

19 55

22 55

14 54

12 54

0 52

10 52

2 51

6 50

15 49

11 47

1 43

Name: hour, dtype: int64

**5.Recommendation**

Following our analysis the following recommendation were made:

1. Blue cars are mostly used types of cars due to its high performance.
2. Utilib 1.4 counters and Utilib are not mostly preferred compared to Electric cars. The company should restructure its budget by reducing the cost of resources which are not mostly used and invest more on those Blue cars.
3. It's clear from the analysis result that Blue cars bring in more revenue to the company, maybe because its performance is very high compared to the rest of the cars.

**6. Evaluation**

From our business success criteria, we have been able to determine the most used type of Electric car based on insight from our analysis in our recommendations. Our approach determining the questions would help determine the most revenue potential cars and its performance which can be verified by use of different samples using the same assumptions.

For purpose of further investigations , we have also provided further questions that can provide more insights while performing

.