PROBLEM STATEMENT

In this week's independent project, you will be working as a data scientist working for an electric car-sharing service company. You have been tasked to process stations data to understand electric car usage over time by solving for the following research question;

Research Question

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.

Bonus Questions (Optional)

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

3.What station is the most popular?

4.Overall?

5At the most popular picking hour?

6.What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?

7Overall?

8.At the most popular picking hour?

9.Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

REPORT

Summary

The problem question will involve an Electric Car sharing Company. The task revolves lending the cars and returning them . The company wants to establish a profitable way of doing its business by understanding the customers preferences according to time lended and time the vehicle was returned by the clients. Through this a proper model can be devised and recommended to ensure the clients get their preferences. Using Data Science techniques, we will be scheming and planning towards a strategy to achieve reliable results.

CRISP DM utilisation

This is the standard and best routine to follow when taking up data Science projects. It is the comprehensive and structured methodology with a clear path of conducting a Data Science project.

THe following are the methods which are followed in this order.

1. Business Understanding
This is the first step of the process. It involves understanding the objectives of the business and the relation with the study.
The objectives of the study is

1. To get a clear and propagated look at the customer behaviour and trends revolving around the car lending business
2. To know which hours are the best and most convenient for clients to pick and return the lended vehicles.
3. To establish a sustainable and longevity that will ensure most clients preferences are met.

2. *Data understanding*

This is the second stage of the process. The initial collection involves loading the necessary dataset(s). For our case here, we have one dataset that we will be using; The Autolib Dataset.The time frame of this dataset was one month ,which was the month of April.

In this case we will also not be using the conventional Python offline platforms but Google Collaboratory with a python extension. In this stage we will also verify the quality of data that we have. Having to use this as secondary data we have to make sure that it is precise, legitimate, complete, unique , comprehensive, relevant and valid.

3. *Data Preparation*

Involves three tasks namely:

    a. Load the dataset(s)
    b. Cleaning the data
    c. Data manipulations

To ensure our data is ready for analysis . We need to ensure that all our datasets for This project is viable and accurate and gives reliable results and outcomes.

    a. Loading the datasets

I loaded the datasets that are ready for data analysis. I loaded the datasets using the read_csv to read data that is in the format of Comma Separated values. I also imported the Python Library Pandas and numpy that would assist in mathematical manipulations.

    b. Cleaning the data

To ensure our datasets are ready for analysis, we need to ensure our dataset for this project is viable for accurate and realistic outcomes.Data cleaning is the process of detecting and correcting corrupt and inaccurate records. My approach was to check for the following attributes so as to ensure my data is ready for analysis.

  i). Validity- This involves removing all columns that are not helpful in the process of

Analysis. I hence dropped three columns which were ' status, displayed_comments and Sheduled_at. This ensured I got rid of the columns I did not need.

ii) Accuracy -I implemented this by checking for null values.Fortunately there were no null values.

iii) Consistency - This was done by checking for prefixes and suffixes that may make our dataset not consistent. Fortunately also I did not notice any in this dataset.

iv) Uniformity - I achieved this in the dataset by changing all the data that was in uppercase to lowercase. I did this for all columns too.

v)Completeness- I checked for duplicates. Fortunately the data was clean from duplicates.

4.*Analysis*

Finding the most of the data we will find the analysis of the dataset. Using the time Columns,we will check how the vehicles are taken and returned by the clients. We will also look at the most popular picking hour and the popular station.

From the question   .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. THe resulta are as below

0   21
dtype: int64

	hour
21	95
12	90
6	81
9	80
13	80

10	80
15	79
8	79
17	78
16	78
19	75
18	75
5	74
7	73
3	73
11	70
22	69
2	69
4	68
0	68
23	67
1	64
20	63
14	59

This shows most vehicles were taken at around 9 PM which is 2100hrs.

To find out the most used station, the results were as follows;

```
address
```

Recommendations

Our recommendations include knowing the most popular time to know which is the most for pick time. The analysis shows that 2100 hrs and around those hours are the most the vehicles are picked. This advocates for all stations to be opened at this time

Also knowing the most popular Station should make sure that there are enough vehicles at that station.Another recommendation is to ensure that there are enough personnel at that station. Also a good idea would be to set up near stations to those having most pick ups to ensure that clients do not miss vehicles.