**1 . Business Understanding**

**Business Objectives**

**Business Overview**

The electric car-sharing company in Paris wants me as a data scientist to process station data to understand electric car usage over time. In particular, the company would like to know 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. Upon looking at the dataset, I found out is for a period of 9 days.

**Business Objective**

The objective of this business is to help the electric car-sharing company to understand electric car usage in the period of 9 days.The analysis will help the company to strategize and make better business plans

**Business Success Criteria**

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.

**Assessing the situation**

1. **Resources**
2. **Personnel**(Data Mining Expert)
3. **The Project Datasets**( the dataset <http://bit.ly/autolib_dataset>])
4. **Computing resources.**
5. **Software(**Colaboratory, Github)

**b. Assumptions**

1. Data Sampled will be an accurate representation of the dataset.

**c . Risk and contaganceis**

We are provided with data that would lack vital information

**Data Mining Goals**

My data mining goal 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.

Potential Questions Include:

* What is the most popular hour for returning cars?
* What station is the most popular? (It can be the most popular station in Paris)
* What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?
* Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

**2. Data Understanding**

**Data Understanding Overview**

We will be working with only one dataset for our analysis which is

[<http://bit.ly/autolib_dataset>]

**Where The Data was Sourced**

This Data was sourced from the company database. It is a sample data that contains data for a period of 9 days.

**Brief Description of the data**

[<http://bit.ly/autolib_dataset>] == This dataset contains data of the company for just a week and 2 days. The dataset contains objects and integers together with null values. The dataset also contains columns that are unnecessary which I dropped for better analysis.

**Verifying Data Quality**

As stated above the dataset contains null values at the Scheduled At Column. I had to drop the column since there were 4953 null values upon doing the summation.

**3.Data Preparation**

These were the steps that were taken for the preparation:

1. Selecting the data

This was the dataset that was used during the analysis [<http://bit.ly/autolib_dataset>]

I loaded the URL using read\_csv to take first see how the dataset looks like

1. Cleaning the Data

* First and foremost I checked the column names of the dataset and noticed there were unnecessary columns like Geo Point, Displayed Comment, Subscription Status. So I dropped the column names.
* I then checked the shape of the dataset using the shape() function and noticed there were 5000 rows and 17 columns
* I later checked for null values in the dataset and found that the Scheduled At Column had Null values so I proceeded to drop the column

**4.Analysis**

During my analysis, I found out that the utilib counter was 0:

|  |  |  |
| --- | --- | --- |
|  |  | **utilib counter** |
| **public name** | **hour** |  |
| **Paris/25 AoÃ»t 1944/8** | **3** | 0 |
| **5** | 0 |
| **7** | 0 |
| **12** | 0 |
| **18** | 0 |
| **...** | **...** | ... |
| **Paris/Ãtienne Marcel/36** | **15** | 0 |
| **20** | 0 |
| **23** | 0 |
| **Paris/Ãvangile/45** | **4** | 0 |
| **9** | 0 |

2580 rows × 1 columns

On the hours