**DATA REPORT**

**1. Business Understanding**

**Business Overview**

Autolib-Car-Sharing, a car sharing company wants to understand its electricity usage. It has selected Paris to be the city to study. As their Data Scientist, I have been tasked to study and analyse the autolib car sharing dataset, aiming to understand electric car usage over time.

**Business Objectives**

The objective of this report is to draw insights from the  9 day dataset across the various stations in Paris City. The insights  will assist to understand electric car usage over time of their different cars types.

**Business Success Criteria**

To determine how efficient their cars are in usage of electricity.

**Assessing the Situation**

**Requirements, Assumptions, and Constraints**

1. **Resources**
2. Personnel(Technical Support, and data mining experts)
3. Projects Datasets: [<http://bit.ly/autolib_dataset>]
4. Computing Resources
5. Softwares( Colaboratory, Github, JIRA)
6. **Assumption**

* That the data sampled was the accurate representation of the entire datasets

1. **Constraints**

The only constraint is that the dataset is large making it stall more than expected.

1. **Risks**

i. Sampled data may contain vital information

ii. There might be bias in the sampled data

1. **Cost/Benefit Analysis**

The cost associated with the study will be minimal. Insights from this investigation will be critical to helping the company manufacturing sector come up with cars that are cost effective cars that use less electricity.

**Data Mining Goals**

The data mining goals for this project is to determine the electricity usage of the Autolib-Car-Sharing cars across various stations in Paris.

The research questions is:

* 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.

Other questions to be considered in the investigation include:

1. What is the most popular hour for returning cars?
2. What station is the most popular?
   1. Overall?
   2. At the most popular picking hour?
3. What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?
   1. Overall?
   2. At the most popular picking hour?
4. Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

**Project Plan**

The **Cross-Industry Standard Process for Data Mining** (CRISP-DM) will be used as the guideline for conducting the research.

**2. Data Understanding**

**Overview**

The existing dataset file is a sampled dataset for cars rented in nine days in april 2018. A description of the dataset columns have been provided for easy reference.

**Description and Exploration**

The dataset contains 25 columns with each having 5000 columns. Only  Displayed\_comments and Scheduled\_at\_columns had null values. There are no duplicated entries in the dataset.

**3. Data Understanding**

These are the steps followed in preparation of the data:

1. Loaded the datasets

Loaded the datasets from both CSV and Excel using pandas.

1. Cleaning the Data

The data cleaning procedure:

* Dropping all columns that seem irrelevant for the analysis.
* Checking for outliers. Although our datasets have outliers we shall not drop them to avoid losing vital data for the study.
* Changing the column names to be in lowercase and replacing the spacing with hyphen to have uniformity
* Creating a new dataset that only contains Paris city
* Creating new columns that showed the difference between the car types; bluecars, utillis, and utillis 1.4.

**3. Analysis**

During the analysis the following questions were answer:

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.

*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*

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*

1. What station is the most popular overall?

*public\_name*

*Paris/Porte de Montrouge/8         13*

*Paris/Voltaire/182                 11*

*Paris/Courcelles/69                11*

*Paris/Mathis/35                    11*

*Paris/Philippe Auguste/126         11*

*..*

*Paris/PÃ©guy/2                      1*

*Paris/Faubourg-Saint-Martin/168     1*

*Paris/Bobillot/16                   1*

*Paris/FÃ©dÃ©ration/10               1*

*Paris/Murat/157                     1*

1. # What station is the most popular at the most popular picking hour(2100 hrs)?

*Paris/Tronchet/19               4*

*Paris/Voltaire/182              3*

*Paris/Ãmile Reynaud/4          2*

*Paris/Danton/8                  2*

*Paris/Boulard/1                 2*

*..*

*Paris/Pelleport/86              1*

*Paris/Pereire/170               1*

*Paris/Pompe/137                 1*

*Paris/Pont Louis-Philippe/24    1*

*Paris/25 AoÃ»t 1944/8*

1. # What postal code is the most popular for picking up Bluecars Overall?

*postal\_code*

*75015    220*

*75016    197*

*75017    180*

*75011    155*

*75013    155*

*75008    150*

*75020    149*

*75012    138*

*75019    134*

*75014    128*

*75018    120*

*75007     99*

*75010     98*

*75009     93*

*75005     85*

*75006     74*

*75004     58*

*75003     55*

*75002     54*

*75001     43*

*75112      6*

*75116      5*

1. What postal code is the most popular for picking up Bluecars at the most popular picking hour?? Does the most popular station belong to that postal code?

*postal\_code*

*75014    15*

*75019    12*

*75012    11*

*75008    11*

*75016    10*

*75013     9*

*75011     8*

*75015     7*

*75020     5*

*75017     5*

*75007     5*

*75009     4*

*75006     3*

*75003     3*

*75010     3*

*75018     3*

*75005     2*

*75004     2*

*75001     2*

*75002     1*

The above analysis was done using the python programming language. This analysis can be found with the python notebook. [[Week 4 IP](https://drive.google.com/open?id=1TZWJutwXLz8oeGvkij-AxjjzWjB7waJ3)]

**4. Recommendations**

Following the above analysis, the following recommendations were provided:

1. More Cars to be added to the *Paris/Tronchet/19* to avoid traffic during picking hours. Also more cars to be added to *Paris/Porte de Montrouge/8* which is the most popular station in the city. This will lead to more revenue for the company.