Table of Contents

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Analysis
5. Recommendation
6. Evaluation

Resources and GitHub Link

[Github W4 IP Link](https://github.com/emunyala/W4-IP)

## 

## BUSINESS UNDERSTANDING

##### **1.1 Business Objective:**

Business Overview

We are to work for an electric car-sharing service company to process stations data to understand electric car usage over time by solving for the following 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.
  2. What is the most popular hour for returning cars?
  3. What station is the most popular?
     1. Overall?
     2. At the most popular picking hour?
  4. 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?
  5. Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

#### **1.2 Accessing the Situation:**

1. Requirements

Data Scientist

1. Assumptions

The data given is not conclusive or wholesome description of the ins and out of the business

1. Constraints:

The constraint with the data provided is that it's large, making handling somewhat difficult and more time consuming.The dataset contains data collected for a period of 9 days. The dataset may take a bit of time to load [~ 10 minutes].

**1.3 Data Mining Goals**

Our data mining goals for this project are to determine….

Potential questions for consideration include:

* Identify the most popular hour of the day for picking up a shared electric
* 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 Utilib 1.4 instead of Blue cars?

Other potential questions that can be answered include:

**1.4 Project Plan**

The **Cross Industry Standard Process of Data Mining ( CRISP-DM)** will be our guideline for research.Below is an overview plan.

| Phase | Time | Resources | Risk |
| --- | --- | --- | --- |
| Business Understanding | 1 hour | Data analyst/Scientist  Project data set | Insufficient wholesum business information ins and outs |
| Data Understanding | 2 hours | Data analyst/Scientist  Project data set | Insufficient data set only ranges 9 days |
| Data Preparation | 2 hours | Data analyst/Scientist  Project data set |  |

| Data modelling | 2 hours | Data analyst/Scientist  Project data set |  |
| --- | --- | --- | --- |
| Evaluation | 1 hour | Data analyst/Scientist  Project data set |  |

## DATA UNDERSTANDING

**Data Understanding Overview**

The data we have is sample data was collected from a period of 9 days

The data files links:

1. [Autolib\_DDI\_DB\_description\_MoringaSchool\_w4.docx](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)
2. <http://bit.ly/autolib_dataset>

**Collecting Initial Data**

Data collected was extracted from opendataparis.com, where the Autolib availability information was available in real-time. The accessed database has the following:

Name: Stations Autolib: Disponibilité en temps réel

Producer : Autolib

Date : see below

License : Open Database License

**Describing and Exploring Data**

We have one data set available collected over 9 days

This sample contains data from April 1 to April 9, 2018.

Overall, Dalberg Data Insights had collected data at the following times:

1. First pilot: every 5 minutes from October 6, 2017, 11:13 AM to October 8, 2017, 10:21 AM
2. Second Pilot: every minute from October 9, 2017, 15:53 PM to October 10, 2017, 15:31 PM
3. Production: every minute from October 30, 16:59 PM to July 31, 2018, 23:59 PM

(date of the end of the Autolib services, although our automatic downloads went on after that)

## DATA UNDERSTANDING

Steps that were taken during data exploration:

1. Selecting Data

We well select our data which is in CSV format and convert it to either SQL or

Just work on it on python with pandas and Numpy

1. Cleaning Data

We will check for null values, duplicated columns

We will then do type conversions on dates if necessary

1. Integrating and Formatting the data

We will concatenate where needed and drop any unnecessary columns that are not needed for our research

## Analysis

During our analysis we were able to determine

1. The data provided hall two columns with Null values that we dropped off
2. Most popular hour for picking up cars was at the 23 hour
3. Most popular station was Quai de SÃ¨vres

## Recommendations

There were no changes if you consider Utilib and Utilib 1.4 instead of Blue cars.

## 