Follow this link to access my GitHub repository:

<https://github.com/CindyMG/Week4IP>

I had added the T.M to my JIRA Kanban board: <https://mptestrun.atlassian.net/jira/software/projects/EC/boards/3>

**E-CAR CRISP-DM DATA REPORT**

1. **BUSINESS UNDERSTANDING**

**Business overview**

**E-Car** is an electric car-sharing service company based in the city of Paris in France. For years, this company has been providing shared electric cars, known as **Blue cars**, which the residents of Paris are free to use or rent out at any time of day to facilitate their daily activities. E-Car has without a doubt proven to the world to be one of the best, if not the best, electric car servicing companies. In addition to directing the world’s attention away from the use of fuel-driven cars which only cause air pollution and increase global warming, E-Car has also massively aided the population of Paris by providing fast, efficient, and environmentally friendly cars to conduct their daily activities.

As exceptional as the company is in terms of servicing electric cars for the public, there is always room for improvement and E-Car would love to do just that; improve its client service. As a data scientist here at E-Car, my team has been tasked with collecting, cleaning, and analyzing all possible datasets concerning client use of the Blue cars. In this way, we will be able to clearly identify the most popular hours of the day in which the Blue cars are picked up to ensure that that time is prioritized and that there is always an abundant number of Blue cars then. This will be done with the objective of improving our services to our Parisian clientele hence boosting company profits.

**Business objective.**

* Improve our clientele service by making sure that the Blue cars are readily available when they are needed by the public.

**Business success criteria.**

* Identifying the most popular hour when the Blue cars are picked up in order to improve client service.

**Assessing the situation.**

* Resources - The following datasets are available to our team for cleaning and analysis:
* [<http://bit.ly/autolib_dataset>]
* [[Link]](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing).
* Assumptions:

- The data collected is accurate.

* Constraints:

- There are no constraints.

**Data mining goals.**

* Identify the most popular hour of picking up Blue cars throughout the day.
* Identify the most popular hour of returning the Blue cars.
* Identify the most popular / most frequently used charging station.

1. **DATA UNDERSTANDING**

**Dataset overview and description**

* [<http://bit.ly/autolib_dataset>] : This is an Autolib dataset containing raw data concerning the use of the Blue cars by the Parisian public across the month of April 2018. It contains the following fields:
* *Address*
* *Cars*
* *Bluecar counter*
* *Utilib counter*
* *Utilib 1.4 counter*
* *Charge Slots*
* *Charging Status*
* *City*
* *Displayed comment*
* *ID*
* *Kind*
* *Geo point*
* *Postal code*
* *Public name*
* *Rental status*
* *Scheduled at*
* *Slots*
* *Station type*
* *Status*
* *Subscription status*
* *Year*
* *Month*
* *Day*
* *Hour*
* *Minute*
* [[Link]](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing). : This is an analysis containing a detailed description of the raw dataset linked above. It lists the flaws in the dataset including missing values, null values, outliers, and more.

**Verifying data quality.**

* Some of the columns had missing values, null values, and some outliers in the datasets were identified.

1. **DATA PREPARATION**

**Loading the data.**

After importing the pandas library into my python notebook, I created a data frame from the URL given above which contains the dataset.

**Inaccurate data.**

Some data values were inaccurate hence made it difficult to analyze the entire dataset. For example, under the field, Public name, there were abnormalities in the data such as ‘Paris/Eug√É¬®ne Oudin√É¬©/51’.

**Missing values.**

Columns such as ‘Scheduled at’ had missing values. These columns were dropped from the data frame.

1. **ANALYSIS**
2. **RECOMMENDATION**

* The most common time of picking up the Blue cars should be prioritized; in that, whenever any client might be in need of the product, they will be readily available to ensure that the public is able to conduct their daily activities with ease.

1. **EVALUATION**

* The importance of this project is to ensure that the Blue cars are readily available and accessible by the Parisian public at any time of the day, especially during the most popular pick-up time. This is to ensure that no client is left stranded and frustrated after realizing that all the Blue cars have already been picked up by earlier clients. In the end, everyone is satisfied and able to carry on with their daily activities without any major inconveniences.