**Elizabeth Josephine**

**Data report for an electric car-sharing service company**

**Embedded links**

* [**Git Repository**](https://github.com/alfafimel/ELIZABETHJOSEPHINE-IPWK4)
* [**JIRA Kanban Board**](https://elizabethjosephine.atlassian.net/secure/RapidBoard.jspa?projectKey=MFP&rapidView=1)
  + - [**Dataset for this project**](http://bit.ly/autolib_dataset)

The python codes used for analysis of this dataset are found in the link to my git repository above. Understandably these are the codes from which this report is obtained. As a data scientist working for an electric share company, I have been tasked with processing the station's data in a bid to understand the electric car usage over time. Notably, the research question in this task is to identify an hour most popular in the day for which a shared blue electric car was picked up in the city of paris over the month of April 2018.

1. **Business Understanding**

An understanding of the key objectives of a project is very essential for the success of the whole thing. The understanding phase of any project lifecycle determines whether or not the accomplishments are understood and from a perspective relevant to the business. For this project, it is crucial that I understand what it is that I need to do with the dataset provided along with the relevant strategies. This project requires estimations from my application of my data science knowledge to process the data of the stations. This is to enable me identify an hour most popular in the day for which a shared blue electric car was picked up in the city of paris over the month of April 2018.

1. **Data Understanding**

The data understanding phase involves the exploration, description verification and making of a report form analysis of the data provided. This phase of the project lifecycle actualises the objectives in the first step while also helping in the understanding of what the attributes of the data mean. Data is just a collection of useless numbers, until someone makes them useful. As such, a proper understanding of the datasets is essential. From the given data set, there are 25 columns namely: **['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'].**

The address refers to the address of the station while the Cars is for the Number of cars available at the station - redundant with Bluecar counter, always the same value. Bluecar counter is the Number of Bluecars available at the station, Utilib counter is the Number of Utilibs available at the station. Utilib 1.4 counter is the Number of Utilib 1.4 available at the station, Charge Slots refers to the Number of Charging slots available at the station. Charging Status can be "nonexistent","operational","broken",or even the "future". These are just but a few of the listed columns above.

1. **Data Preparation**

Each of the columns in this dataset contains data which is essential to this project. There is however the need to clean the data and remove duplicates along with missing values/numbers it is only after data cleaning that data exploration is possible. With data preparation, all the possible solutions to the questions about this dataset will be answered. It is at this stage that i get to select,integrate, after cleaning and finally construct the appropriate data for a complete and accurate analysis.

1. **Analysis**

At this stage of the project, the identification of a necessary and applicable model is significant. This is because this model will be needed in the determination process of a proper strategy. This can only be done when all the information is presented and the data to be used has already been cleaned and prepared. Notably, analysis takes place with the imported libraries being recalled and the dataset being monitored.

1. **Evaluation**

At this stage of the project, the data is checked for accuracy and essentiality as part of the final data investigation steps. It acts as a confirmation of whether or not the results gotten from the data does meet the objectives as expected by the research question. The research question in this project is to identify an hour most popular in the day for which a shared blue electric car was picked up in the city of paris over the month of April 2018.

1. **Recommendation**

Recommendations are essential for any research done. For this, I would recommend the use of data which is not highly corrupted and with numerous issues.