Data Report

Peter Kiragu

Moringa School

5th April 2020

Data Report

**Business Understanding**

In this project, the client is Autolib, a French electric car sharing service. Autolib maintains a fleet of electric cars for public use on paid subscription. The company utilizes a network of parking and charging stations throughout the city. The goal of this project is to help Autolib gain some insights from the data collected between April 1 and April 9 2018. One of the most important questions that Autolibs wants us to help them answer is: what is the most popular hour of the day for picking up a shared electric car in Paris over the month of April. This information could help the company optimize its services so that it is able serve the customers better.

The Business also wants to understand the popular hour for returning cars, the most popular station, and postal code that is most important for picking up Blue cars.

Answering these important questions will help Autolib understand how the business is performing at different points of the day and different locations. This information can be used to adjust prices to ensure that there is a match between demand and supply of electric cars. The insights gathered from this analysis can also be used to understand which locations are most profitable and why this is so.

**Data Understanding**

The data provided is for nine days of operations. The data has 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 and minute. These fields are very important in aiding in the analysis.

The data is of integer and object type. The entire data set has 5,000 entries for each field. The data is verified to be from an accurate source. In the data cleaning and preparation process, we will establish whether there are missing value or errors and what to do about these issues.

**Data Preparation**

At this stage, we must decide the data that we need to use to make decisions. The dataset will be loaded onto the notebook in its current form and further cleaning and analysis will be conducted. Not all fields will be included in the analysis. The fields of Scheduled at and Displayed comment will be dropped because they have numerous missing values and they do not play any role in our analysis.

The data set has no missing values and no action is required to make the data ready for analysis. The data passes all checks for validity, consistency, accuracy and completeness. Throughout the analysis, we will generate new records by filtering the data based on the outcome we want to achieve. This process is important because it allows effective analysis to be conducted.

**Analysis**

Python will be used for the analysis section of this project. The modelling technique used to help our client understand the most popular hour to pick a car in Paris is filtering the records to get entries where Blue Counter is greater than zero, and the city is Paris and status is operational. Filtering the records based on this criterion allows us to find the number of times each hour is repeated in the filtered entries. The hour with the highest count is the most popular hour. This modelling technique is simple and straightforward.

The assumption we are making during the analysis is that in instances when Blue counter was zero, there was no car to pick. Removing these entries allows us to get the values when there were commuters actively looking for an electric car to use.

**Recommendation**

Based on the analysis conducted. 21 was the most popular hour to pick an electric car, followed by 12, 6 and 9. Further investigation needs to be made on why these periods are very popular.

Actions that can be taken based on this analysis is to ensure that the number of cars available at the Blue counter are increased. This will ensure that Autolib is able to meet demand for electric car sharing service at these points. Various maintenance and repair services can be rescheduled for these periods to ensure that the company does not lose revenue because of being unable to meet demand.

**Evaluation**

I believe that the model created is effective in meeting the general business objectives. The model is efficient because it provides the exact solution as to help the organization meet demand. Other follow up questions need to be made to help Autolib understand the users and the market better. That said, the model will be sufficient for use in this project.

Link to Jira Kanban Board

<https://cresent.atlassian.net/jira/software/projects/IPW4/boards/2/roadmap?selectedIssue=IPW4-14&shared=&atlOrigin=eyJpIjoiOGFkM2ZjMmYyYjhkNDBlZTllNGI1NzA3NzAyYzA4MTgiLCJwIjoiaiJ9>

Link to Git Repository

https://github.com/peterzk/Indipedent-Project-Week-4.git