**MORINGA SCHOOL PREP INDEPENDENT PROJECT WEEK 4.**

**1 BUSINESS UNDERSTANDING.**

As the data scientist for an electric car sharing company i am seeking to process stations data to understand electric car usage over time by solving for the following research objectives.

So the main objective of this company 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 objectives are to determine:

* 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?

**2 DATA UNDERSTANDING.**

So the data we are going to use for the work below is contained in the link below which will be read once in the python.

Url = [<http://bit.ly/autolib_dataset>]

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

The dataset below contains a number of field which have been explained in details in the link below

Link for Autolib sample data

;<https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing>

**3.DATA PREPARATIONS**

The following data had one or two problems which we needed to solve before proceeding to analyse data.

These are the steps I undertook to prepare my data for analysis.

**Step1** : Read my data in python collab notebook

**Step2** : Check the validity of the data

So in this step I focused on checking for validity by removing all the rows and the column that I will no longer need.

Therefore i removed the following columns

* Displayed comments -This is because we will not be needing this for analysis
* Year- This is because all the data had the same year that is 2018
* Month -This is because most of the data shared the same month.
* Car- This is because the car column matched the Bluecar counter.
* Scheduled at - This is because it had many null values and we were also not using it during analysis.

Step 3:Consistency

Here we check if the data is consistent and that there are no duplicates in the dataset

Step4: Completeness

Lastly we check if the data is complete and that there are no null values.

After following the following steps my data was ready to be exported for data analysis.

**3 DATA ANALYSIS**

In data analysis we used python to answer the questions on the objectives.We used a set of codes .

So for the first objective which is to find 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, we use the codes below to compute that.

We first had to compute when the Blue Car == 0 because that's when it has been picked

We the compute when the city is paris from the data in the first step

And lastly we compute the hour with the most values of zeros in the Bluecar,

So in our results we found that the most popular hour was 2 where most blue cars were picked.

So the next one was to find the most popular station overall so here we found out that it was limeil brevannes-lavoisier-12

The last analysis we did was to find if results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

Here we found utilibs and blue cars and compared them side by side.

TABLE FOR SUM OF UTILIB AND SUB UTILIB

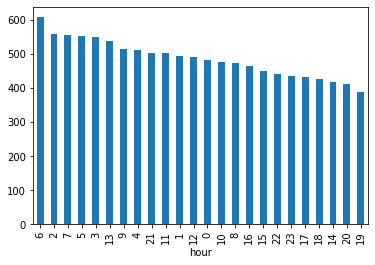
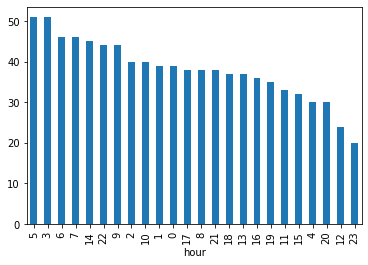


TABLE FOR BLUE CAR

As you can see the combination of utilib and utilib 1.4 still is small compared to blue car, so the results would change by decreasing

And lastly we found the most popular postal code for picking blue cars as 75008 94.

**RECOMMENDATION**

So my recommendation would be to focus more of our resources on the stations which most cars were picked from.That is limeil brevannes-lavoisier-12 and the postal code 75008.Also we need to find out why most Blue cars were picked at that particular time and make our stations operational at that time.

**EVALUATION**

According to my evaluation for this analysis the data was explained well and that made it easy for us as the data science team to be able to work on it with much accuracy.

Below are the relevant links for my projects

JIRA LINK -<https://norahip.atlassian.net/jira/software/projects/MIP/boards/1>