AUTOLIB SERVICES DATA REPORT

# BUSINESS UNDERSTANDING

## BUSINESS OVERVIEW

Autolib is a company based in France.It offers electric car sharing services in various cities worldwide.Autolib possesses a fleet of different cars i.e Blue cars, utilib ,utilib 1.4 .Each of these cars can be borrowed and used within the given city.Each City has different types of stations where its citizens may borrow these cars.

In this report we shall be focusing on the company’s operations in France and how these stations can be maximised to improve the overall success of Autolib as a company.

## BUSINESS OBJECTIVE

Our main objective would be to 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.

## 

## BUSINESS SUCCESS CRITERIA

Ensure more vehicles are available at the more popular stations at the most popular picking hours of the day.

## SITUATION ASSESSMENT

## RESOURCE INVENTORY

1. Datasets

Autolib main dataset : <http://bit.ly/autolib_dataset>

Dalberg insights : [[Link]](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)

1. Software

Google Collaboratory(Python Notebook)

Jira Kanban board

GitHub Repository

## ASSUMPTIONS

The data sampled between April 1 to April 9,2018 is correct and up to date.

## CONSTRAINTS

There are no constraints for this set of data

## DATA MINING GOALS

Our data mining goals exist as follows:

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?Overall?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?Overall?At the most popular picking hour?

5. Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

## DATA MINING SUCCESS CRITERIA

Our data success criteria will be based on identifying the top three most popular stations at the most popular picking hours of the day.

# DATA UNDERSTANDING

## DATA UNDERSTANDING OVERVIEW

We shall be focusing on one specific dataset which contains all pertinent information regarding the stations data in France from April 1 to April 9,2018

Autolib main dataset : <http://bit.ly/autolib_dataset>

## DATA DESCRIPTION

Autolib Dataset- It mainly constitutes Station information such as ‘ID’ to identify the station, ‘GEO LOCATION’ to locate the station ‘KIND’ to identify the type of station,’ADDRESS’ to identify the region of the station.The Information provided by the dataset may be relevant or irrelevant to our study.

## QUALITY OF DATA

The data in the dataset had some missing values.It also contained outliers which eventually had to be removed to enhance the quality of data.

# DATA PREPARATION

The data collected in the dataset was prepared in the following sequence:

## Importing of Libraries

Import Pandas-Manipulation of data

Import Numpy - Manipulation of data

Import Matplotlib- Plotting of desired graphs

All imports are done in the Google Collaboratory environment.

## Loading Dataset

Load the Autolib csv file into Google Collaboratory environment.

## Cleaning of Data

Autolib Dataset

Dropping of columns that provided additional information that was not necessary for our objectives :['Geo point','Cars','Charging Status','Charge Slots','Subscription status','Slots']

Merging of Date time columns:['year','month','day','hour','minute'] to [‘Date\_Time’] to improve functionality of the of the dataset

Handling outliers:Error in computation

# ANALYSIS

The dataset above was analysed using Python programming language in a Python notebook.

The full analysis can be obtained from the following notebook in the Git Repository.

<https://github.com/SydneyTsuma/Week_4_Autolib_Project.git>

# RECOMMENDATIONS

Ensure enough cars are available at 2018-04-06 14:48:00 at Place Montgolfier - Parking Montgolfier.

# EVALUATIONS

1.The process of evaluating our data using CRISP-DM was successful

2.Deployment to the data to the GIT repository was successful.

3.There was a setback in the data analytics, hence the inconclusive results.