**Data report for Autolib car-sharing service company.**

1. **Business Understanding**

Autolib is a car sharing service based in France Paris which has a fleet of all-electric Bollore Blue-cars for public use on subscription basis. The company also runs electric car stations, parking and charging stations for electric cars.

* **Business objective**

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.

* **Project plan**

The following steps will be taken to ensure that the objective of this research is met.

1. Find out whether data is enough to make an informed conclusion.
2. Use resources available to analyse the data and give an informed decision.
3. Submit a conclusive report.

The success criteria will be if we are able to identify the most popular hour for picking up a shared electric car in the city of paris over the month of April 2018.

**Requirements , assumptions and constraints:**

* Deadline of submission 14th Aug 2021 at 6:00PM.
* I have checked to find outliers in the dataset thus it is a reliable source.
* I have assumed that the data will be enough for the study.
* I have assumed that the data is real and will be consistent with real world scenarios for the company.
* Resource constraint- more data could inform the study better.

**Our data mining goals:**

* Determining the most popular hour for returning cars.
* Finding the station that is most popular.
* Overall.
* At the most popular picking hour.
* Determining what postal code is the most popular for picking up Blue cars
* whether the most popular station belongs to that postal code.

1. Overall
2. At the most popular picking hour
3. **Data Understanding**

The data is in the form of a csv file named autolib\_dataset. The data was extracted from opendataparis.com where Autolib data was available without constraints. Found in the document below:

Source document: [Autolib\_DDI\_DB\_description\_MoringaSchool\_w4.docx](https://drive.google.com/a/moringaschool.com/file/d/13DXF2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing)

* Content in the data

The data is contained in a csv file with 5,000 entries and 25 columns.

Here’s a list of the columns and the datatypes they contain.

RangeIndex: 5000 entries, 0 to 4999

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Address 5000 non-null object

1 Cars 5000 non-null int64

2 Bluecar counter 5000 non-null int64

3 Utilib counter 5000 non-null int64

4 Utilib 1.4 counter 5000 non-null int64

5 Charge Slots 5000 non-null int64

6 Charging Status 5000 non-null object

7 City 5000 non-null object

8 Displayed comment 111 non-null object

9 ID 5000 non-null object

10 Kind 5000 non-null object

11 Geo point 5000 non-null object

12 Postal code 5000 non-null int64

13 Public name 5000 non-null object

14 Rental status 5000 non-null object

15 Scheduled at 47 non-null object

16 Slots 5000 non-null int64

17 Station type 5000 non-null object

18 Status 5000 non-null object

19 Subscription status 5000 non-null object

20 year 5000 non-null int64

21 month 5000 non-null int64

22 day 5000 non-null int64

23 hour 5000 non-null int64

24 minute 5000 non-null int64

1. **Data Preparation**

This stage involved the following steps:

* Loading a Jupyter notebook into our visual studio code Environment.
* Importing Pandas, numpy to enable us to perform python queries on our dataset.
* Loading our data set from the link in the document where it was stored as a CSV file.
* Describing the data to view the type they've been stored in.

1. **Analysis**

This stage started with a review of the Cars column and the Bluecar counter column, since both are identical, the cars column was deleted to remove the redundancy.

Second, we found the most popular city for picking up the cars which was Paris as it has the largest sum of Bluecars. Below is a list of the top 5 Cities

City

Paris 2816

Nanterre 127

Issy-les-Moulineaux 101

CrÃÂ©teil 88

Rueil-Malmaison 88

The next step was to find the most popular hour where a new column was formed to contain all the date and time. By grouping this column, we are able to see which hours had the most pickup. Assuming that cars in parking are the ones dropped off, we can make the execution the following time with the highest number means many vehicles were dropped off while with the least number of slots means that it is the least drop off time.

While investigating using utilib and utilib 1.4 we can conclude that the same time and conclusion is arrived at. That the two columns can be used interchangeably.

1. **Review of the objectives**

Determining the most popular hour for returning cars.

* Finding the station that is most popular.

From the analysis the station was found by assuming the id is the station marker.

The station ID therefore was paris-portedemontrouge-8 13

* Overall.
* At the most popular picking hour.
* Determining what postal code is the most popular for picking up Blue cars

The most popular address is 8 Avenue de la Porte de Montrouge 13

* whether the most popular station belongs to that postal code.

Overall

At the most popular picking hour

1. **Evaluation**

We have been able to identify the most popular pickup day for Paris which is the 6th of April.