**ELECTRIC CAR SHARING SERVICE COMPANY DATA REPORT.**

***1. Business Understanding***

***Determining Business Objectives***

Autolib' was an electric car sharing service which was inaugurated in Paris, France, in December 2011. The Autolib' scheme maintained a fleet of all electric Bluecars for public use on a paid subscription basis, employing a citywide network of parking and charging stations.

We would like to process stations data to understand electric car usage over time by understanding which is the most popular hour for picking up a Bluecar in certain stations in Paris.

***Defining Business Objectives***

We would like to know the most popular hour for picking up a shared electric car as well as the most popular stations so that we can understand the usage of these cars over time. We would also like to explore more on utilib and utilib 1.4 as compared to Bluecars.

***Assessing the Situation***

We have two datasets that contain so much information on electrical car sharing. We have details on the number of cars available, postal addresses,stations, rental status, subscription status and time. From this dataset, we get to know when cars are picked up and which stations they come from. This information is from 9 days, this is enough time to come up with a conclusion in analyzing the stations.

***2. Data Understanding***

***Collecting Initial Data***

From the datasets we have, there are a total of 25 columns and 5001 rows. Of the columns, the ones that seem most promising in giving results are: Bluecar counter, Utlib counter, Utilib 1.4 counter, postal code and time. The irrelevant columns are: cars, displayed comments,station type,year and month. We will be merging the hour and minute columns to come up with a column time, for an easy analyzation of most popular hour. We will then drop the missing values so that we work with clean data.

***Describing Data***

We have a large dataset with the value types as integers and strings. Our data is represented by the following column descriptions:

|  |  |  |  |
| --- | --- | --- | --- |
| **Column name** | **Type** | **Values** | **Comments** |
| **Address** | **String** |  | **address of the station** |
| **Cars** | **Number** | **[0-7]** | **Number of cars available at the station - redundant with Bluecar counter, always the same value** |
| **Bluecar counter** | **Number** | **[0-7]** | **Number of Bluecars available at the station** |
| **Utilib counter** | **Number** | **(0-4]** | **Number of Utilibs available at the station** |
| **Utilib 1.4 counter** | **Number** | **[0-5]** | **Number of Utilib 1.4 available at the station** |
| **Charge Slots** | **Number** | **[0-3]** | **Number of Charging slots available at the station** |
| **Charging Status** | **String** | **{"nonexistent","operational","broken","future", some typos}** | **Whether the station is operational for recharging. Mainly "nonexistent", "operational" or "broken": charge slots can only be greater than 0 when "operational"; slots and vehicles can be available in all situations (except future stations that have 0 resources)** |
| **City** | **String** |  | **City** |
| **Displayed comment** | **String** |  | **Some comments like "station within parking, access through …"** |
| **ID** | **String** |  | **ID of the station** |
| **Kind** | **String** | **{"STATION","SPACE,"PARKING",CENTER"}** | **"CENTER" have no resources at all; "PARKING" do not have charge slots, but can have bluecars and utilib; "STATION" and "SPACE" can have all resources** |
| **Geo point** | **String** |  | **GPS coordinates of the station** |
| **Postal code** | **Number** |  | **Postal code of the station** |
| **Public name** | **String** |  | **Name of the station** |
| **Rental status** | **String** | **{"nonexistent","operational","broken","future", a few empty}** | **Whether the station is available for renting vehicles. Resources are only available when "operational", except for "broken" which can have Slots, but none of the other resources (Bluecars, utilib or charging slots).** |
| **Scheduled at** | **String** | **datetime** | **Planned opening date: non empty values only for stations that have "future" in one of the statuses.** |
| **Slots** | **Number** | **[0-7]** | **Number of parking slots available at the station?** |
| **Station type** | **String** | **{"station","full\_station","subs\_center"}** | **No resources available for "subs\_center" - which is just one location. Was that actually a selling point for Autolib subscriptions?** |
| **Status** | **String** | **{"ok","closed","scheduled"}** | **No resources available for "scheduled", which is the status if there is a "scheduled at" date. Yet there can be resources associated with "closed" stations** |
| **Subscription status** | **String** | **{"nonexistent","operational","broken","future"}** | **Whether it is possible to subscribe to the autolib service in that station? No resources available when "future", but other values can have resources** |

***Verifying Data Quality***

The data has several null values which we will delete. We also have columns which have a constant value such as year which only has 2018 and month only contains the value 4. A column such as cars has values similar to Bluecar counter. We will drop these columns as we will not need them in our final data for analysis.

***4. Data Preparation***

***Data Preparation Overview***

We will merge several columns like year, month and day to come up with a new column for date, merge hour and minute to come up with time...This will help us in coming up with a solid conclusion on what hour of the day is the most popular.