1.Business Understanding

1. Business Overview

Autolib, an electric car-sharing company based in Paris, France was founded in December 2011 and operated by Bollore Industry. Its complemented by the City’s Bike Sharing system as well known as the Velib. The Autolib services maintain a Fleet of cars all-electric and Bollore Blue Cars for public use on a subscription basis. This is is a research data report regarding Autolib’s request to establish 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.

1. Business objectives

Our objective in this report is to generate insights from a 9-day dataset collected automatically in real-time to help identify the most popular hour of the day during which electric blue cars were picked across Paris on April (from date 1 to 9)

1. Business Success Crieria.

To come up with the most effective strategy for profit turnover maximization by ensuring efficiency during the peak business times

Situation analysis

Requirements, Assumptions and constraints

a.Resources

i) Data mining and Technical experts from Dalberg Data Insights

II) Dataset Description handbook from Dalberg

III)Datasets (Autolib files Availability over time)-

Iv)Data analysis tools (Google Colab. Git, JIRA)

Assumptions.

i) the collected dataset is accurate to a considerable

ii) the sample set gives a rather true picture of the business operation in the long run

Constrains.

The data had missing files each month.

E. Cost-Benefit analysis- the cost-benefit analysis indicates the decision to invest on better data acquisition and storage mechanisms will help drive company sales by facilitating better insights on their business coverage.

Data mining Goal

The data mining goal for this project is to determine the usage patterns of the Electric Blue cars by the Customers of Autolib and derive further insights from the questions below

Potential questions

* What is 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?

Project plan

The CRISP-DM(Cross-industry Standard Process for Data Mining) shall be used as a guideline to carry out the research as outlined below.

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Time | Resources | Risks |
| Business understanding | 2 hours | Project Datasets | Missing data files caused by dome gaps |
| Data understanding | 2 hours | Project Datasets | Insufficient data files due to a download failure |
| Data preparation | 2 hours | Data Scientists |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Data Modelling | 2 Hours | Data Scientist/project datasets | Too many assumptions |
| Evaluation | 1hr | Data Scientist |  |

2. Data Understanding

The given datasets were extracted from the opendataparis.com since they had real-time data for the Stations Autolib the dataset

Overall, Dalberg Data Insights had collected data at the following times:

First pilot: every 5 minutes from October 6, 2017, 11:13 AM to October 8, 2017, 10:21 AM

Second Pilot: every minute from October 9, 2017, 15:53 PM to October 10, 2017, 15:31 PM

Production: every minute from October 30, 16:59 PM to July 31, 2018, 23:59 PM (date of

the end of the Autolib services, although our automatic downloads went on after that).

The data also contained 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-
* The data has 24 columns containing 5000 rows each

Verifying Data Quality

Two columns labeled displayed comment and Scheduled at 111 and 47 null values respectively. Our business understanding established that the missing data gaps were caused by download failure and Dome gaps.

1. Data Preparation

The dataset below was used for our analysis based on the relevance of the research objective and the quality of the data.

Data cleaning

We dropped 2 columns that contained null values /irrelevant for analysis so we can obtain quality results

During our analysis, the following questions were answered

1. What was the most popular hour for picking an electric blue car?

* - 6

1. What was the most popular hour for returning the cars?

-13

1. Whether the most popular station belongs to the postal code - 92140

-most popular station - Sur-Seine/Foch/53

1. Overall most popular postal code for picking cars

* 75011

Recommendation

Following this analysis, the following recommendation was provided

The business should invest in more efficient inventory management systems do that business can forecast future demand with higher accuracy.

Knowing the peak business hours can help in decision making as to how to run marketing campaigns and perhaps use the quiet times to focus on servicing the units and systems

Evaluation

The business success criteria have been able to determine peak business performance periods and most importantly the very popular destination addressed based on our insights and recommendations. This will help determine the most revenue potential areas.