Automotive Car sharing usage analysis Report

Name: Ian Mugambi

[**Project Links**](#_5086vp6puhqi) **2**

[1. Business Understanding](#_o35tu6eilg90) 3

[Business Objective](#_uvamsid3e42x) 3

[Business Success Criteria](#_guyg2zb3gn4f) 4

[Assessing the situation](#_qz1il4fzqbmq) 4

[Resource Inventory](#_u2y502are2ox) 4

[Assumptions](#_vstkkdh1q0nn) 5

[Constraints](#_izmn8c988krz) 5

[**2. Data Understanding**](#_nnjc7fjw1sm3) **5**

[Data Understanding overview](#_4cbmlwdj6b82) 5

[Data Description](#_twebsqngnykv) 5

[**3. Data Preparation**](#_ortveatupzma) **6**

[4. Hypothesis testing](#_haobrk9d0eg7) 6

[Hypothesis Testing Procedure](#_vqzb4p3qf5wl) 6

[Hypothesis Testing Results](#_ozgy3qual2cb) 7

[3. Summary and Conclusions](#_cuyzh2c08qdk) 7

# 

# Project Links

|  |  |  |
| --- | --- | --- |
| **Item** | **Description** | **URL** |
| Github Reposiroty | URL to Github respiratory | https://github.com/mugambiian/C-coding/blob/main/Moringa\_Core\_Week4\_IP\_IanMugambi.ipynb |

## **1. Business Understanding**

Business Overview

Autolib, an electric car hire company, would like to understand the car usage patterns in various locations in France so as to inform its strategy and give it a competitive advantage. The data set categorises the different areas using postal codes with the first two digits representing the district, and the last three digits indicating the specific town or neighborhood.

#### Business Objective

The choice of the two postal areas in Paris i.e 19ème Arrondissement Paris and 13ème Arrondissement Paris were due to the fact that we needed to investigate if the demand for cars remains the same within Paris, which is the most populous city in Northern France. By comparing Blue car utilisation over weekdays and weekends in the two postal codes, it will lead to insights on whether Paris needs to be treated as a city or if further intervention is needed on a case by case basis for the different postal codes. This will be useful in determining the staff and resource allocation for Paris i.e to treat as a single administrative unit or increase staff in specific areas. The staff deployed has a direct bearing of the company's running costs and can be avoided if not strictly needed.

#### Business Success Criteria

Hypothesis

In our research , we will seek to answer the following questions . Is the usage of blue cars in 19ème Arrondissement Paris the same as the blue car usage in 13ème Arrondissement Paris Paris

Hypothesis

**Null Hypothesis:** On weekdays,the number of Blue Cars taken from Hauts-de-Seine is the same as in Paris

***H0:U1 = U2***

**Alternative Hypothesis**: On Weekdays, the number of Blue Cars taken from Paris is not the same as in Essonne.

***Ha: m1 ≠ m2***

### Assessing the situation

#### Resource Inventory

a)Datasets:

* Autolibd dataset description
* Autolib dataset

b)Software:

● GitHub repository

● Python

● Numpy

● Pandas

● Seaborn

● Scipy

● Matplotlib

● JIRA software

● Google Colab

● Google Docs.

#### Assumptions

#### Constraints

There are no constraints

# **2. Data Understanding**

### Data Understanding overview

We will use two datasets provided by the client that had been collected by the client as below

* Autolib\_daily\_events\_postal\_code
* Columns\_explanation

### Data Description

We have two datasets for use on this project. One describes the columns found on the second dataset which is a large file that contains the data collected regarding the electric cars, stations and car details.

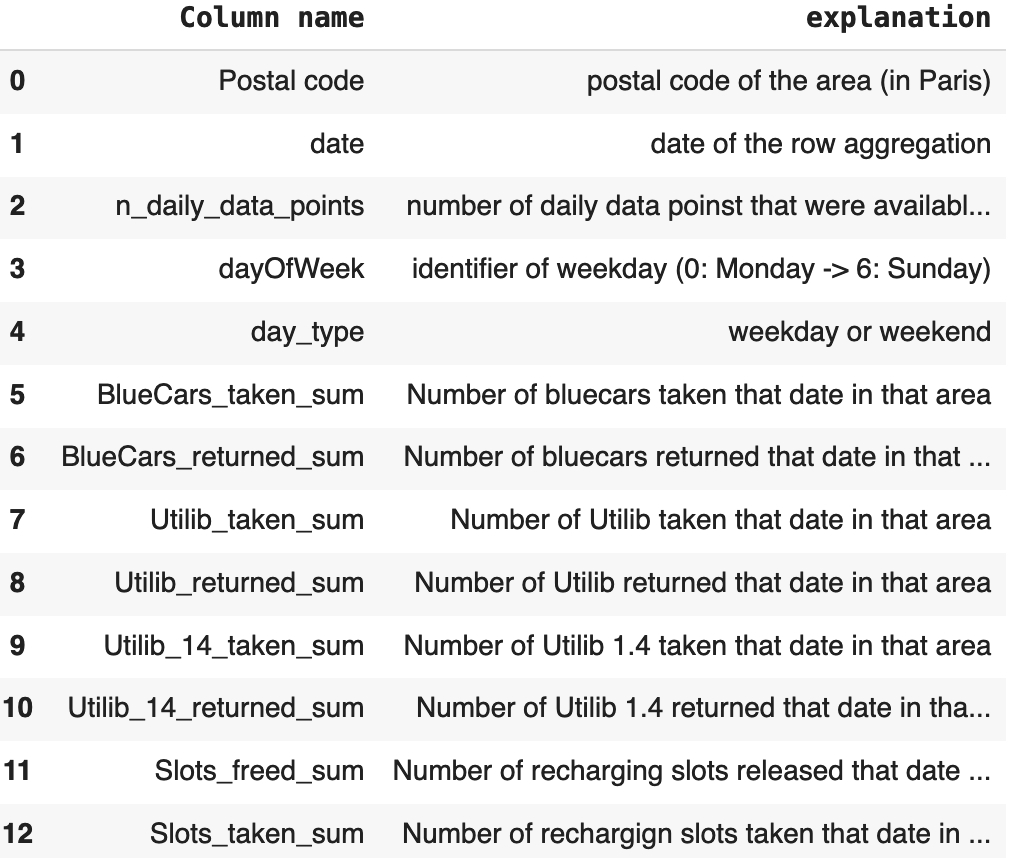
A further description of the datasets is as below;

* Columns\_explanation

This is a data collection report which gives an explanation of the columns found in the Autolib dataset. It is a report that describes the different cars used by clients, their postal code, day of the week and slots utilized.

* Autolib Dataset

This data set lists data about electric cars, their rental status and Postal codes . It has 13 columns and 16085 rows.

****

# **3. Data Preparation**

The steps below were followed in preparing the data

*Data loading and cleanup*

The data was loaded onto google colab using pandas. The data was checked for duplicates and missing values

*EDA*  
 Exploratory Data Analysis was performed to get a better understanding of the data. Measures of central distribution, skewness as well as checking for null values, the various data types to ensure that they can be manipulated appropriately.

## **4. Hypothesis testing**

### Hypothesis Testing Procedure

The dataset was obtained and required columns extracted

The Hypothesis test was then specified as below

* Significant level of 0.05 % was adopted
* Z test was adopted as the records count is greater than 30.
* The z-score and p-value was calculated using the mean and standard deviation of the two samples.
* The p\_value was compared against the significance level

### Hypothesis Testing Results

Outcome: The Null Hypothesis: On weekdays,the number of Bluecars taken from postal code 75019 is the same as in postal code 75013 is not rejected with a p value of 0.5

## **3. Summary and Conclusions**

There was no significant difference in the utilization of the blue cars between the two postal codes over weekdays and weekends as well.As such there is no need for smaller business units with more staff and Paris can be treated as a single administrative business unit.

It may be worth increasing the sample size to see if the hypothesis holds for a larger data set