Table of Contents

[Hypothesis Report 1](#_Toc40650426)

[Problem Statement 1](#_Toc40650427)

[Data Description 1](#_Toc40650428)

[Hypothesis Testing Procedure 1](#_Toc40650429)

[Hypothesis Testing Results 1](#_Toc40650430)

[Summary and Conclusions 2](#_Toc40650431)

# Hypothesis Report

## Problem Statement

Autolib electric car-sharing service company have tasked me to investigate a claim made about the blue cars .The provided Autolib dataset is [autolib\_daily\_events\_postal\_code.csv](file:///C:\Users\SPECTRE\Downloads\autolib_daily_events_postal_code.csv).The variable being investigated is the blue cars. The claim is that the postal codes differences in mean are equal to zero.

The null and the alternative hypothesis are as follows:

Ho = the postal code chosen represents that the difference in the means of blue cars taken in postal code 75020 and 94300 is 0.

H1!= The postal code chosen represents that the difference in the means of blue cars taken in postal code 75020 and 94300 is 0.

## Data Description

The data to be used in the hypothesis testing will be the difference in mean of two postal codes that is postal codes 75020 and 94300.

The random variable being tested is a discrete variable in the form of the mean of blue cars postal codes taken during a specific day throughout the month in question.

The descriptive statistics are as follows:

## Hypothesis Testing Procedure

* The first step was loading the dataset, cleaning and dealing with outliers. In my case I did not remove the outliers as I felt I will need them for the analysis
* Step two will involve selecting the variables of interest in the study which is the postal codes
* Step 3 was creating a data frame containing our simple random sample which had sample size =5000.I will use this to compare whether I should reject the null hypothesis or not
* I calculated the test statistics using the Z – test assuming the level of significance = 0.5 because I was using the difference in mean of two postal codes.

.

## Hypothesis Testing Results

* if pval1 < 0.05:

    print ("Failed to accept null hypothesis")

else:

    print ("Failed to reject null hypothesis")

* The results shows that the P\_ value is 1.0105
* A P\_ value higher than 0.05 (>0.05) is not statistically significant.
* We fail to accept null hypothesis.

.

## Summary and Conclusions

* The summary of the procedures to conduct the hypothesis testing were:
  + Importing the data-set and doing the necessary data cleaning.
  + Doing EDA.
  + Preparing a sample from the data-frames of interest
  + Conducting a z-test
  + Finding the p-value of the result got from the t-test

According to our hypothesis test on the p-value is 1.01529. A p-value higher than 0.05 (> 0.05) is not statistically significant and indicates weak evidence against the null hypothesis.

Therefore, we fail to accept the null hypothesis.