**HYPOTHESIS REPORT**

This report is a rigorous study on an electric car sharing company in France.

# **PROBLEM STATEMENT**

This fact finding study sets out to investigate a claim about the blue cars from the provided Autolib dataset. In a bid to do this, I will go ahead and identify some areas and periods of interest via sampling stating the reason to the choice of method, then perform hypothesis via google colab, testing with regards to the claim that we will have made then document everything in this report. For the periods of interest, I have been given the choice to either assess the weekdays or weekends, I have chosen to assess the weekends. As for the area of interests I am at liberty to work with either postal code 75015 or 75017

**I. Null hypothesis**

The average number of cars taken during the weekends equals or is greater than 80

**H0 : 80 ≥ µ**

**H:BlueCars\_taken\_sum≥ 80**

**II. Alternate hypothesis**

The average number of Blue cars taken during the weekends is greater 100

**H1 : µ > 80**

H**:BlueCars\_taken\_sum> 80(Claim)**

This hypothesis will provide us with a clear and concise insight into the use of blue cars at Autolib Electric car sharing company.

# **DATA DESCRIPTION**

Below is a brief description of each column in the dataset in accordance to the glossary provided.

|  |  |
| --- | --- |
| Column name | Description |
| Postal code | Postal code of the area (in Paris) |
| Date | Postal code of the area (in Paris) |
| N\_daily\_data\_points | Number of daily data points that were available |
| Day of Week | Identifier of weekday (0: Monday -> 6: Sunday) |
| Day type | Number of blue cars taken that date in that area |
| Blue cars taken sum | Number of blue cars returned that date in that |
| Blue cars returned sum | Number of Utilib taken on that date in that area |
| Utilib cars taken sum | Number of Utilib returned on that date in that area |
| Utilib cars returned sum | Number of Utilib returned on that date in that area |
| Utilib 1.4 cars taken sum | Number of Utilib 1.4 taken on that date in that area |
| Utilib 1.4 cars returned sum | Number of Utilib 1.4 returned on that date in that area |
| Slots freed sum | Number of recharging slots released on that date in that area |
| Slots taken sum | Number of recharging slots taken on that date in that area |

# **HYPOTHESIS TESTING PROCEDURE**

# determining the population shape

autolibcopy.shape

# determine the margin of error or confidence level which determines how close the sample results will be to the true value of the overall population in the study

margin of error = +/- 5%

# determine the confidence level which will measures your degree of certainty in regards to how well a sample represents the overall population within the chosen margin of error

because the margin of error = plus/minus 5, the confidence level = 95%

# setting the standard deviation so as to show teh variance in the dataset

because the datasetis skewed then the standard deviation will be set to .5 (50%)

# determine the z-score

this will determine the number of standard deviations between any selected value & the mean of the population.

95% confidence => 1.96 z-score

# determining the sample size

Sample Size = [z2 \* p(1-p)] / e2 / 1 + [z2 \* p(1-p)] / e2 \* N] N = population size z = z-score e = margin of error p = standard deviation Sample Size = [2.582 \* 0.5(1-0.5)] / 0.05 Squared / 1 + [2.582 \* 0.5(1-0.5)] / 0.05 Squared \* 16085] = [6.6564 \* 0.25] / 0.0025 / 1 + [6.6564 \* 0.25] / 40.2125] = 665 / 1.041 =638

# **DISCUSSION OF TEST SENSITIVITY**

The sensitivity value that will be put into use for the alternate hypothesis is expected to be higher therefore at least 70% will be used to confirm that the study is fairly sensitive.

# **SUMMARY AND CONCLUSIONS**

The null hypothesis was accepted since it was true. The level of Acceptance is between 0.734 and 1, i.e. 80% falls in the region of acceptance and falls outside the region of rejection on the left hand side.