Hypothesis Testing Report

# Problem statement

To show how the variable of blue cars taken is affected by the day selection type and whether there is a difference by day type

H0: There is no difference in the sum blue cars taken between weekdays and weekends

H1: There are more blue cars taken in the weekend compared to the weekdays

This hypothesis can help the company to prepare more cars if there is higher demand for blue cars during the weekend

# Data description

The data is derived from the autolib dataset from Autolib car sharing company

It contains:

* Postal -code postal code of the area (in Paris)
* date -date of the row aggregation
* n\_daily\_data\_points -number of daily data points that were available for aggregation, that day
* dayOfWeek -identifier of weekday (0: Monday -> 6: Sunday)
* day\_type -weekday or weekend
* BlueCars\_taken\_sum -Number of bluecars taken that date in that area
* BlueCars\_returned\_sum -Number of bluecars returned that date in that area
* Utilib\_taken\_sum -Number of Utilib taken that date in that area
* Utilib\_returned\_sum -Number of Utilib returned that date in that area
* Utilib\_14\_taken\_sum -Number of Utilib 1.4 taken that date in that area
* Utilib\_14\_returned\_sum -Number of Utilib 1.4 returned that date in that area
* Slots\_freed\_sum -Number of recharging slots released that date in that area
* Slots\_taken\_sum -Number of recharging slots taken that date in that area

# Hypothesis testing procedure

The hypothesis testing procedure is as follows:

1. Identifying the data
2. Coming up with a null and alternate hypothesis and setting a significance level (0.05)
3. Selecting a sample chosen randomly and with independent variables (variables are assumed to be dependent as 10% of the population is sampled)
4. Selecting the hypothesis testing method ( 1 sample z test for proportions) in accordance with the assumptions of a
   * normal distribution
   * Independence of the variables selected
   * Data is continuous
5. Calculating the test statistic using google colab and getting the p-value (0.0). As the p value is below the level of significance we reject the null hypothesis and the Z score

# Hypothesis testing result

The result of the test was that there are more blue cars taken during the weekend compared to the weekdays

The value of the test statistic was a z score of 47.458

The null hypothesis was rejected and the alternate hypothesis was considered

The p value of the test was 0.00

The confidence interval was between 454.0220017736116 and 454.02200177361163. It was noted that as the sample size increased the confidence level decreased

# Test sensitivity

For the test as the p value was so low there is a very tiny chance for a Type 2 error and the power level is practically to 1

The bigger the sample size the closer the power level moves to 1

# Summary and conclusions

In summary the hypothesis test was conducted successfully and meaningful insights could be drawn from the results

The hypothesis was to test for an increase in blue car usage for the weekends

The result concluded more blue cars were taken during the weekends

The test sensitivity was very low as the sample size was relatively large