## ***HYPOTHESIS TESTING REPORT***

## Business Understanding

### **Business Overview**

Electric car-sharing companies offer car rental services. They make it possible for people to rent electric cars for short periods of time. Electric car-sharing is an instrument which provides the local community with an alternative mobility service for their daily or occasional commuting, and it reduces the ownership of private vehicles circulating in the area.

### **Problem Statement**

To investigate a claim about blue cars from the dataset provided. We shall investigate the claim of our choice using a hypothesis test. Hypothesis testing is important here because it evaluates two mutually exclusive statements i.e., the null and alternative hypothesis about a population to determine which statement is best supported by the sample data

H0: The average usage of blue cars during weekdays is equal to the average usage of blue cars over the weekend

Ha: The average usage of blue cars during weekdays is different from the average usage of blue cars over the weekend

Our study will provide an analysis on when the electric car-sharing service is popularly used and with this information our client is able to make informed decisions such as:

* Which day type is best to be open for business (Weekdays or Weekends)?
* Whether to increase work hours and decrease work hours on lower usage day type

## Data Understanding

### Data Understanding Overview

For this project, we are using the availed dataset by the company.

* Autolib Dataset [[Link]](http://bit.ly/DSCoreAutolibDataset) - This dataset contains information on the sum of blue cars taken and blue cars returned for each day of the week.

## ***Data Description***

The dataset was an open dataset about cars in Paris. It contains variables like the postal code of the area in Paris where the stations are located, the dates of data collection. The dates ranged between 1st of January 2018 and 19th June 2018. The dataset also had the number of daily data points that were available for aggregation on the particular days of aggregation within the specified time periods. The days of the week were listed as either weekday or weekend. The blue cars that were taken and returned, the utilib data and the slots set of data were also contained in the dataset. The problem under investigation was on the averages which would make the null and alternative hypotheses.

It was a set of data that was already collected. Using the data, we shall investigate the data in details and carry out data cleaning to ensure the completeness, validity and reliability of the data for our analysis.

## ***Data Analysis***

**Sampling Method**

From the data set, we used the clustering method to group the data into two categories i.e., Weekday dataset and Weekend dataset.

After clustering, we used random sampling to pick 100 records as our sample from both datasets. This will be our sample dataset and has the following features:

|  |  |  |
| --- | --- | --- |
|  | Weekday Sample | Weekend Sample |
| count | 100.000000 | 100.000000 |
| mean | 56.220000 | 62.000000 |
| std | 64.934712 | 57.922449 |
| min | 0.000000 | 2.000000 |
| 25% | 14.000000 | 20.750000 |
| 50% | 35.000000 | 44.500000 |
| 75% | 69.000000 | 92.500000 |
| max | 290.000000 | 299.000000 |

## ***Hypothesis testing procedure***

* Step 1: Specify the Null Hypothesis

H0: Blue cars usage during weekdays is equal to usage during the weekends

* Step 2: Specify the Alternate Hypothesis

Ha: Blue cars usage during weekdays is different from usage during the weekends

* Step 3: Set the Significance Level

The significance level that is denoted by the alpha Greek letter shall be set at 5%.

Alpha = 0.05

* Step 3: Calculating the p-value

Two simple random samples were used and the corresponding p-value calculated was 0.235.

* Step 4: Drawing Conclusion

Since the p-value (0.235) > alpha (0.05), we fail to reject the null hypothesis

## ***Hypothesis testing results***

From the hypothesis test, we found that there was not sufficient evidence to prove that there is a difference in the average number of blue cars taken on weekdays to that of the blue cars taken over the weekends. The significance level used was 0.05 and the p-value calculated was 0.235.

The p-value is greater than the level of significance, hence we fail to reject the null hypothesis. Our conclusion thus is blue cars taken during the weekday are significantly equal to blue cars taken over the weekends

## ***Discussion of test sensitivity.***

Since we fail to reject the null hypothesis with a significance level of 0.05; there is a 95% chance that this finding is true. In the case we increase the size of the dataset, this confidence level (95%) might decrease leading to the decision of rejecting the null hypothesis.

## ***Summary and conclusion.***

In order to investigate the claim, we made on blue cars, it was necessary to first understand the dataset provided, then tidy up the data to remain with a complete dataset with the fields of interest free of any anomalies. Once we have the clean data, we carry out data analysis: An analysis on the blue cars taken. From our analysis and the hypothesis formulated we carry out a statistical test to enable us decide on whether to Accept or Reject the Null Hypothesis. As per our results we made the decision not to reject our Null Hypothesis. We conclude that blue cars taken during the weekdays are significantly equal to blue cars taken over the weekends.