**Name: Paul Ndirangu**

**Topic: Hypothesis Testing Report**

**Problem Statement**

The hypothesis test was performed for an Autolib electric car-sharing service dataset.Theaim of the test was to investigate a claim about blue cars from the provided Autolib dataset. The hypothesis question that whose test we performed was; Was the number of Blue Cars taken, different from the number of Blue Cars returned, on a weekday?

Ho = Number of Blue Cars taken on a weekday is equal to the number of Blue Cars returned.

Ha = Number of Blue Cars taken on a weekday is not equal to the number of Bluecars returned on a weekday.

This particular hypothesis was important because it allows us to get a better understanding of whether there is a disparity in the number of cars taken and those returned on weekdays or not. This helps us to determine whether the customers stay with the cars for a longer time and whether the blue cars are available on weekdays.

**Data Description**

The data provided has 13 columns which contain data about the cars offered by the company and how their customers rent them. These columns are:

'Postal code',

'Date',

'n\_daily\_data\_points',

'dayOfWeek',

'day\_type',

'BlueCars\_taken\_sum',

'BlueCars\_returned\_sum',

'Utilib\_taken\_sum',

'Utilib\_returned\_sum',

'Utilib\_14\_taken\_sum',

'Utilib\_14\_returned\_sum',

'Slots\_freed\_sum',

'Slots\_taken\_sum'

The company provide 3 types of cars which are; Blue cars, Utilib cars and the Utilib 14 cars.

The dataset and glossary to use for this project can be found here [http://bit.ly/DSCoreAutolibDataset] and here [Link] respectively.

The provided dataset is a daily aggregation, by date and postal code, of the number of events on the Autolib network (car-sharing and recharging).

**Hypothesis Testing Procedure**

The hypothesis test was done with the help of Python libraries and the test performed on a Python notebook.

The hypothesis was carried out using a t-test statistical test with the aim of accepting or rejecting the null Hypothesis.

The logic behind the null hypothesis was due to the assumption that the number of blue cars taken was equal to the number of blue cars returned on a weekday since we assume that customers return the vehicles on the same day they take them.

We performed exploratory data analysis on our dataset and realized that the data followed a normal distribution. This prompted us to perform a hypothesis test since the data was normally distributed.

We performed a t-test due to the following reasons:

* The number of samples was less than 30. (Samples were 25).
* The alternative hypothesis required us to determine whether the number of blue cars taken is greater than or less than the number of cars returned.

We set the significance level for our test as 0.05.

**Hypothesis Testing Results**

We calculated the mean and the standard deviation of the number of cars taken and the number of cars returned and the following were the results;

* Mean of taken cars: 185.12
* Mean of returned cars: 186.52
* Standard Deviation of Taken cars: 222.4418126162435
* Standard Deviation of Returned cars:223.8906206164072

The number of samples was constant for both cases, and it was determined to be 25 .

The results of our hypothesis test were as follows;

The test statistic is: -0.022179528185530838

The p-value is: 0.9823966943294226

Hence we failed to reject the null hypothesis because the p-value was greater than the level of significance for the test which was set to 0.05.

Thus we did not get enough evidence to accept our alternative hypothesis which claimed that the number of cars taken during weekdays is not equal to the number of cars returned.

**Summary and Conclusions**

We successfully performed our hypothesis test and were able to come up with concrete evidence that prompted us to accept the null hypothesis and reject the alternative hypothesis. The