**HYPOTHESIS TESTING REPORT**

1. **Problem statement:**

This dataset contains details about the operation of cars within Paris. It showed a compilation of dates when the blue cars were picked from and returned to the addresses. The claim being investigated was whether the average number or blue cars taken was different from the average number of blue cars returned during that period.

**Null hypothesis**: The average number of blue cars taken is not different from that of the blue cars returned.

**Alternative hypothesis:** The average number of blue cars taken is different from that of the blue cars returned.

**2.** **Data description**

The dataset contains 13 columns.

The dates ranged between January and July of 2018 with the number of daily data points that were available for aggregation on the days of aggregation within the specified time periods. The days of the week were the usual Monday to Friday with the specifications and special assignments of days. Weekday or weekend the dataset had the specific days within the period. The blue cars that were taken and returned, the Utilib data and the slots set of data were also contained in the dataset.

**Hypothesis testing procedure**

The dataset is large with 16, 085 \* 13 entries. I picked a sample from it using a stratified sampling method. I used this method of sampling because with it, the sample would be unbiased. Stratified sampling required me to group the data into different sets of strata and then randomly pick a sample from the strata. The dataset only had count and dates of picked and taken bluecars, I used those to determine my hypothesis. I decided to work with an average to find out whether the average number of cars that were picked in a day was like the average number of cars that were returned on that very day. From my stratified sampling, I got a sample of 296 \* 4 entries. This meant that n > 30. As such, I used the z-score to determine the p-value. I did perform normality tests on the data since the statistic I was using demanded so. The alpha level of significance that I used was 0.05.

**4.** **Hypothesis testing results**

From the hypothesis test, I found that there was not sufficient evidence to prove that the average means of the blue cars taken and the bluecars returned are not equal. The z-score was 1.2541048695208732 and as a result, the null hypothesis was not rejected. The z-critical value was 1.959963984540054

with the confidence interval being:

Confidence interval:

(35.33555372011692, 77.583365198802)

The p-value was 0.10490194533288694

which was greater than the significance level.

**5.** **Summary and conclusions**

I failed to reject the null hypothesis because there was not enough evidence for me to reject the null hypothesis.

**Null hypothesis**: The average number of blue cars taken is not different from that of the blue cars returned.

**Alternative hypothesis:** The average number of blue cars taken is different from that of the blue cars returned.