**HYPOTHESIS TESTING REPORT.**

1. **Problem Statement**

An electric car company has provided a data set on the electric cars and their stations. I will be investigating a random variable (dIff), which represents electric car activity at the stations (this is whether an electric car arrived at the station or left).

The problem statement is to investigate whether it is true to say that the average blue car activity per station in Paris is greater in the morning hours than in the afternoon hours.

Null Hypothesis

Ho: The average blue car activity per station in Paris is no greater in the morning hours(6-12) than in the afternoon hours(12-18). Ho: U1=u2

Alternate Hypothesis

The average blue car activity per station in Paris is greater in the morning hours (6-12) than in the afternoon hours (12-18). Ho: U1>u2

1. **Data description.**

The data set to analyze is provided by the Autolib car sharing company. It contains a daily aggregation, by date and postal code, of the number of events on the Autolib network.

The specific random variable variable to be analysed is( diff ) which represents electric car activity at the stations (this is whether an electric car arrived at the station or left). The variable is created through feature engineering using station ID, Date-Time and the Blue car count.

The new variable is used to answer the business question.

1. **Hypothesis Testing Procedure**

The hypothesis testing is to be carried out to help the Autolib Company know how to get more customers and ensure availability of the electric cars when they are most needed. This was one of the concerns raised during a marketing pitch meeting and thus the research need be done.

The test statistic to be used is the t-statistic. We will be comparing two means from two samples. The sample data sets are normally distributed (shown by a histogram) and the sample size choosen is less than 30. This further confirms the appropriateness of the t-test

The significance level chosen is 0.05. Research allows a 5% error This means there is a 5% risk that we will be rejecting null when it’s true.

1. **Hypothesis Testing Results**

The test of the results include.

T-test statistic- (-0.898)

P-Value – (0.373).

The P-Value is greater than the level of significance. The null hypothesis is not rejected

1. **Discussion of Test Sensitivity.**

The power of the test is 1-0.05=0.95.

When the sample size is increased, the statistical power of the hypothesis test (1.00 minus Beta) also increases. This means with a bigger data set we get more accurate results. This on the contrary would increase probability of committing a type 2 error (not rejecting null when its false).

1. **Summary and Conclusions**

From our hypothesis test there is no significance difference between the two means. Blue car activity in Paris during the morning hours is not different from Blue car activity in Paris during the afternoon hours.

For the Autolib company, cars should equally be made available in stations both in the morning hours and the afternoon hours.