PROBLEM STATEMENT

The data used for this project is an autolib dataset. Autolib is a company for an electric car-sharing service scheme in which people can rent an electric car for a short period of time. The claim under investigation is whether the average number of blue cars being picked in a particular area is different from the average number of blue cars being picked in another area during weekdays.

The null and alternative hypotheses of this study are:

- The average number of blue cars taken from postal code 75015 is the same as in postal code 75017 during weekdays.
- The average number of blue cars taken from postal code 75015 is not the same as in postal code 75017 during weekdays.

From the hypotheses above, we can determine which area(postal code) to ensure that there are more cars available for pick up during weekdays.

DATA DESCRIPTION

The dataset has a lot of information relating to blue cars and also not relating to blue cars. To conduct the study successfully, all information not relating to blue cars was dropped so that the only remaining columns were postal code, date, type of day(weekday or weekend), the sum of blue cars taken, the sum of blue cars returned, daily data points, slots taken, and slots freed. A descriptive summary of the dataset information can be found here.

The data was already provided so there was no need source for it.

HYPOTHESIS TESTING PROCEDURE

A simple sampling technique was used to pick the samples. This method was used because random samples are the best method of selecting a sample from the population of interest. The advantage is that the sample represents the target population and eliminates sampling bias.

It is interesting to note the average number of blue cars being picked in a particular area during certain periods of the week, that is either weekend or weekday. This gives us a chance to test and certainly prove that the average number of blue cars being picked in area 75015 is not the same as the average number of blue cars being picked in area 75017. When the stated null hypothesis is rejected, the company will be sure to increase the availability of blue cars where they are being picked most during weekdays for a revenue boost.

The test statistic used to test the hypothesis is the t-test because we are comparing the means of 2 groups, the population, and the sample.

The confidence level is set at 95% and the alpha value 0.05.

HYPOTHESIS TESTING RESULTS

From the results, there is sufficient evidence that the average number of blue cars picked from area 75015 is the same as those picked from area 75017.
The value of the test statistic is -0.0569 and we fail to reject the null hypothesis that the average number of blue cars taken from postal code 75015 is the same as in postal code 75017 during weekdays.
The p-value of the test statistic is 0.4773

TEST SENSITIVITY

Sensitivity in a statistical test is the measure of performance of a binary classification test. It measures the proportion of the actual positive i.e. the probability of a null hypothesis being true. In this case, the sensitivity was 91%.

SUMMARY AND CONCLUSIONS

The project was comprehensive. Exploratory data analysis was performed with hypothesis testing as its implementation. Conclusively, the null hypothesis was not rejected because there was not sufficient evidence to reject it.