# **Problem Statement**

The data that was used was very specifically filtered to contain entries from the weekend and particularly from the Postal Code 75015. The random variable was the mean number of BlueCars returned and borrowed.

* *Null hypothesis*: The average number of BlueCars borrowed and those returned is the same on weekends.
* *Alternate hypothesis:* On average, more BlueCars are borrowed than returned on weekends in the most popular postal code.

The interest behind this investigation was to see whether there was a temporary shortage over the weekend due to people borrowing more than they returned. Supposing the Alternate hypothesis were true, then there would be a deficiency in supply that would have to be addressed by the company in order to keep their consumers satisfied with their services.

# **Data Description**

The data that was used in the analysis was from Autolib electric car-sharing service company that contained columns with the following information:

* The postal code from which a car was either borrowed or returned on a particular date
* The date of when a car was either borrowed or returned.
* Day of the week a car was either borrowed or returned
* Type of day: either a weekday or weekend.
* Number of BlueCars borrowed on that particular day.
* Number of BlueCars returned on that particular day.

The descriptive analysis obtained from the data are the following:

1. Weekends saw the highest number of cars being borrowed and surprisingly enough, the highest number of returned cars occurred over the weekends as well.
2. The fifth day of the week had the highest number of cars taken and returned in general.
3. There was a strong linear relationship between the number of BlueCars taken and the ones returned.

# **Hypothesis Testing Procedure**

Since a sample size of 384 was used, it was only logical to use a z test over a t-test. The confidence level used in the test was 5%. The alternate hypothesis came from the idea that with a high number of vehicles being rented, especially from the most popular postal code, there might be a subsequent shortage of available cars if the number of vehicles returned is significantly lower than that of the rented ones. The alternate hypothesis therefore aims to anticipate a market problem before it even exists.

The test will be conducted as follows:

A sample will be drawn from the returned cars with respect to the significance level and the size of the population. The sample mean will then be calculated to give a point estimate of the average number of cars being returned. Thereafter, this figure will be compared to the average number of cars taken using a z-test to see whether the null hypothesis will be rejected or not.

# **Hypothesis Testing Results**

The p-value of the Z test conducted was revealed to be 0.5184. This therefore led to the conclusion that the alternate hypothesis cannot be accepted since this value did not fall between 0.95 and 1.0 which define the rejection zone. The point estimate for the parameter was:

* BlueCars returned = 179.487
* BlueCars taken = 178.179

The confidence interval for the difference of means at 95% confidence was:

(-1.1256132769515264, -0.49828528311724063)

# **Discussion of Test Sensitivity**

The power of the test was calculated from the samples drawn from the cars returned and those borrowed and are as follows:

* Sample of cars returned = 0.9999996621821322
* Sample of cars borrowed = 0.9999996621821322

Increasing the sizes of any of the sample sizes would lead to inaccuracy as the p-value obtained in either one of them would be too small to reflect any significant results when there could actually have been a difference between the two means under investigation. Doing the opposite would have led to the p-value being to large and possible pointing out there was a difference when there actually wasn’t any.

# **Summary and Conclusions**

The summary of the process entails checking for anomalies within the data, plotting univariate and bivariate summaries and filtering out the dataset to remain only with the postal code of interest within the days of interest as well. The means of BlueCars taken and those returned were then compared to see whether they were approximately the same or there was a significant difference between the two averages worth exploration.

In conclusion, there was not enough evidence to reject the null hypothesis thereby creating a supply gap that needs to be addressed by Autolib Company.