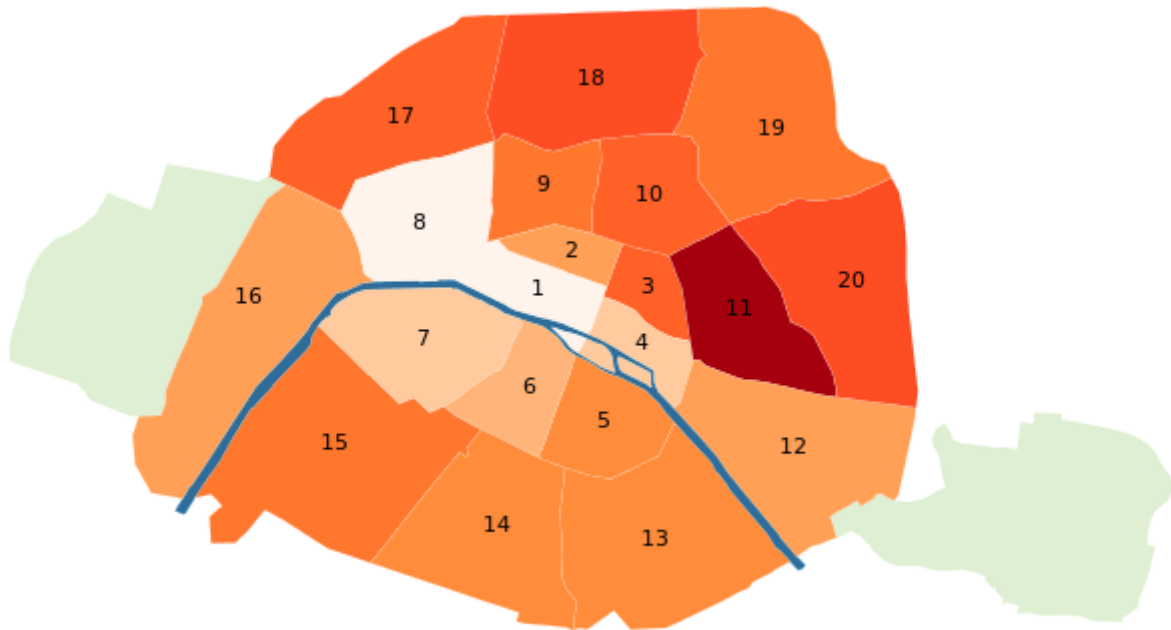


INTRODUCTION

The city of Paris is divided into 20 administrative districts called Arrondissements. The number of the arrondissements is indicated as the last two digits of the postal code. They are divided in a spiral clockwise pattern starting from the middle of the city going outward. In French, the number is noted in Roman numerals, ie The Eiffel Tower belongs to the *VII^e arrondissement*.



https://upload.wikimedia.org/wikipedia/commons/thumb/3/3d/Population_density_map_of_Paris_in_2012.svg/600px-Population_density_map_of_Paris_in_2012.svg.png

The client is interested in the return and pickup patterns of Bluecars in the *XVI^e arrondissement* (Vaugirard) and *XVII^e arrondissement* (Batignolles-Monceau)

PROBLEM STATEMENT

1. Practical Problem

The claim to be investigated is whether the rate of return for Bluecars is higher in Vaugirard than in Batignolles-Monceau.

2. Statistical problem

The hypotheses to be tested are:

Null hypothesis: The rate of return for Bluecars is higher in area 75015 than in area 75017.

Alternate hypothesis: The rate of return for Bluecars is the same in both areas 75015 and 75017.

DATA DESCRIPTION

The data for this has been provided on the given link: <http://bit.ly/DSCoreAutolibDataset>

The variable description has been provided on the link:

<http://bit.ly/DSCoreAutolibDatasetGlossary>

The data has 13 variables and 16085 observations. The dataset is a daily aggregation, by date and postal code, of the number of events on the Autolib network (car-sharing and recharging). The car types are: Bluecars, Utilib, and Utilib_14. The variable description is given below:

<u>VARIABLE</u>	<u>VARIABLE DESCRIPTION</u>
Postal code	Postal code of area in Paris
date	Date of row aggregation
n_daily_data_points	Number of data points that were available
dayOfWeek	identifier of weekday (0: Monday -> 6: Sunday)
day_type	Weekday or weekend
Bluecars_taken_sum	Number of bluecars taken that date in that area
Bluecar_returned_sum	Number of bluecars returned that date in that area
Utilib_taken_sum	Number of Utilib cars taken that date in that area
Utilib_returned_sum	Number of Utilib cars returned that date in that area
Utilib_14_taken_sum	Number of Utilib_14 cars taken that date in that area
Utilib_14_returned_sum	Number of Utilib_14 cars returned that date in that area
Slots_freed_sum	Number of charging slots freed that date in that area
Slots_taken_sum	Number of charging slots taken that date in that area

Day of the week and day type variables are categorical data.

HYPOTHESIS TESTING PROCEDURE

The main focus was on the BlueCars column as the test is interested in the rate of return of only BlueCars. During data cleaning, the Utilib columns were dropped to concentrate only on the BlueCars data.

The records associated with the returned BlueCars for both 75015 and 75017 were 2 records for each. The sample size is small and with the assumption that the variances are homogenous because the standard deviations are almost equal, the T statistic will be employed.

The level of significance was set at 1% therefore making
 $\alpha = 0.01$

HYPOTHESIS TESTING RESULTS

The T statistic is 1.067906.

The p-value is 0.397386

$0.397386 > 0.01$

As the p-value is greater than the value of α we, therefore, have insufficient evidence to reject the null hypothesis. More investigation will have to be carried out on the data to investigate the claim.

DISCUSSION OF TEST SENSIBILITY

The data for this test was very small and hence not sufficient to make inferences about the whole population. As the variables were not normally distributed hence making inferences using the mean increases the probability for a Type II error as we failed to reject the null hypothesis.

SUMMARY AND CONCLUSION

In conclusion, there is need for more investigation of the claim as there is insufficient evidence to reject the null.