Citi Bike Trips Distance Analysis for Men and Women

Ruben Hambardzumyan[1]

[1]NYU Center for Urban Science & Progress

November 8, 2017

Abstract

Citi Bike is one of the most commonly preferred options for public transportation in New York. As such, it continuously collects millions of rows of data about its customers and subscribers, that contain significant insights reflecting various behavioural patterns of Citi Bike's users. Particularly, considering that men in average are physically more developed, do they cycle longer distances than women? In this article, a research of the platform's data has been conducted to discuss the issue and the results have been discussed.

Introduction

Launched in May 2013, Citi Bike quickly became one of the most favorite menas of transportations for people in New York. Nowadays, the platform utilizies more than ten thousand bikes and manages over fifty thousand rides per day adding millions of rows of data every month. Such amounts of data is interesting in terms of statistical analyses to determine behavioural patterns of the platform's users and the context behind the data.

Gender-based analyses are interesting in terms of determining the factors that may affect any differences in between the usage patterns of males and females. For this research, the idea was: do men cycle longer distances than women? To try to analyse the idea, the author developed the following null hypothesis: the average trip distance cycled by women in the month of October, 2016 is the same or less than that of men.

(?)

Data

To tacle the issue, the Citi Bike usage data for the month of October 2016 was parsed from the platform's website. The author chose October based on the premise that during that month the climate is mind to choose cycling for commuting or traveling to a park.

the following data was needed:

- Start station latitude
- Start station longitude
- End station latitude
- End station longitude
- User type
- Gender
- Travel distance

Out of the mentioned data, only travel distance needs to be calculated by using the given geographical coordinates (described in the methodology section). Note, that for the research only Subscribers of Citi Bike were used. The figure below shows the distributions of calculated distances and observed frequences for men and women.

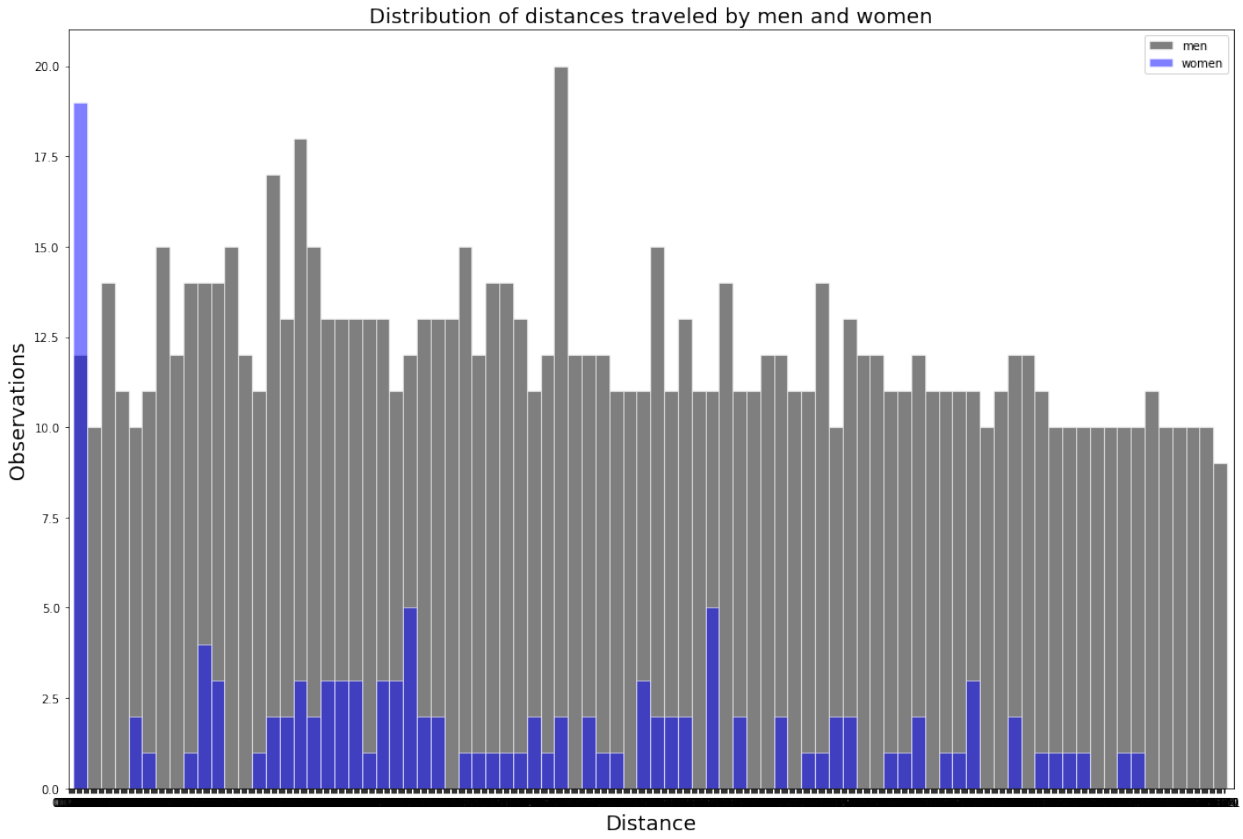

Figure 1: The figure shows the distances traveled by men and women and the frequencies of observed distances. It is clearly visible from the histograms that men travel longer distances.

Methodology

To reject the null hypothesis, the Student's t test to compare the means of two groups of the same population was used. The traveled distances was calculated by writing a function in Python that gets the geographical coordinates as values and, taking into account the spherical shape of the Earth, calculates the distance in kilometers in between the two points as an arc.

Conclusions

Student's t test to compare the means of two groups of the same population has given 0.118 as the p-value. Since our significance threshold for the research was specified as 0.05 (95% confidence intervals), we conclude that the null hypothesis holds, meaning that men did take longer rides than women during the month of October 2016.