Citibike Analysis - Mini Project

Anupama Santhosh

Introduction

Citibike, New York's bike sharing system, releases public data. Citi Bike makes data available for every individual trip in the system. This is an article about CitiBike analysis. The report includes four parts Abstract, Data, Analysis and Result, which will be described in the following parts.

Abstract

The idea behind the analysis is to test whether young or old people commute more. Commuters are the crowd who use the bike share system over the weekdays. The Null hypothesis can be formulated as: "The ratio of young people(age $<$ 35) biking on weekends over young people biking on weekdays is greater than or equal to the ratio of old people(age $>=$ 35) biking over weekends to old people biking on weekdays"

$H_0 : \frac{Oweekend}{Oweek} \leq \frac{Y weekend}{Y week}$
$H_1 : \frac{Oweekend}{Oweek} > \frac{Y weekend}{Y week}$
significance level: $\alpha = 0.05$

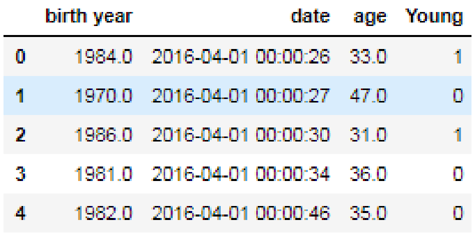
Data

In this analysis, the data is based on the citibike usage in April. 2016. There are several attributes in dataset. Each trip record includes:

- Station locations for where the ride started and ended
- Timestamps for when the ride started and ended
- Rider gender
- Rider birth year
- Whether the rider is an annual Citi Bike subscriber or a short-term customer
- A unique identifier for the bike used

However, "Birth year and "date" are the only two attributes required in the analysis. Age was calculated from the birth year and a binary variable "Young" was created which indicates if the rider is young or not. Note that riders below the age of 35 are classified as Young. The final dataframe is shown in Fig.1

As shown in Fig.1, the format of date should be transformed to weekend and weekdays. Therefore, the first step is to show the number of rides from Monday to Sunday, then normalized the data so we can compare the distribution of rider by age, which is visualized in Fig 2 and 3.



	birth year	date	age	Young
0	1984.0	2016-04-01 00:00:26	33.0	1
1	1970.0	2016-04-01 00:00:27	47.0	0
2	1986.0	2016-04-01 00:00:30	31.0	1
3	1981.0	2016-04-01 00:00:34	36.0	0
4	1982.0	2016-04-01 00:00:46	35.0	0

Figure 1. Final dataframe for analysis

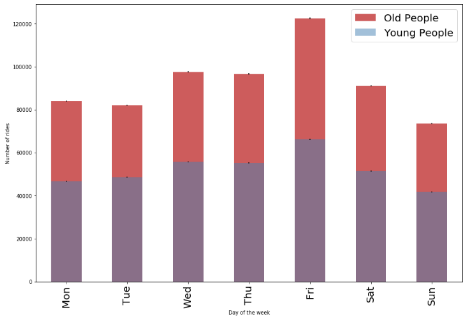


Figure 2. Distribution of Citibike riders by age group with statistical errors

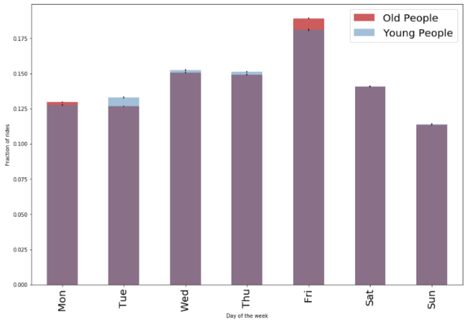


Figure 3. Distribution of Citibike riders by age group normalized, with statistical errors

Analysis

The data can be separated into weekend and weekdays and the error bar is shown in Fig 4. As shown in Fig 4, old people have higher proportion over weekdays than young people.

Results

Result Accordingly, the zscore based on the above analysis is -0.6 and P-value is less than 0.05, which means that the null

Anupama Santhosh is with New York University

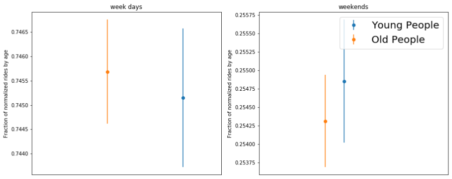


Figure 4. Fraction of Citibike riders by age for week days (left) and weekends (right)

hypothesis cannot be rejected. As a result,it is likely that Old people commute more than young people

: