Explore-US-Bikeshare-Data

Udacity Data Analyst Degree - Project II

Overview

In this project, I will explore data related to bike share systems for three major cities in the United States—Chicago, New York City, and Washington. Using Python, I will write a code import the data and answer interesting questions about it by computing **descriptive statistics**. I will also write a script that takes in raw input to create an interactive experience in the terminal to present these statistics.

What Software Do I Need?

To complete this project, i'll require the following softwares:

- Python
- A text editor, like Sublime or Atom
- A terminal application

The Datasets

The datasets provided by Motivate, a bike share system provider in the US, contains randomly selected data for the first six months of 2017 for all three cities. The data files for all three cities contain the same core six columns:

- Start Time
- End Time
- Trip Duration
- Start Station
- End Station
- User Type

The Chicago and New York City files also have the following two columns:

- Gender
- Birth Year

Answering Questions

Through writing code, I'm able to answer the following questions about the bike share data:

- What month occurs most often in the start time?
- What day of the week (Monday, Tuesday, etc.) occurs most often in the start time?
- What hour of the day occurs most often in the start time?
- What is the total trip duration and average trip duration?
- What is the most frequently used start station and most frequently used end station?
- What is the most common trip (i.e., the combination of start station and end station that occurs the most often)?
- What are the counts of each user type?
- What are the counts of gender?

- What is the earliest birth year (when the oldest person was born), most recent birth year, and most common birth year?

```
23
24      while True:
25          city = input("\nWhich city would you like to filter by? New York City, Chicago or Washington?\n")
26          if city not in ('New York City', 'Chicago', 'Washington'):
27              print("Sorry, I didn't catch that. Try again.")
28              continue
29          else:
30              break
```

The Interactive Experience

For a more interactive experience, I wrote a script that took raw input in the terminal to present the statistics. Below screenshots show the results after requesting for information in Chicago for the month of January on Sunday:

```
Hello! Let's explore some US bikeshare data!

Which city would you like to filter by? New York City, Chicago or Washington?
Chicago

Which month would you like to filter by? January, February, March, April, May, June or type 'all' if you do not have any preference
?
January

Are you looking for a particular day? If so, kindly enter the day as follows: Sunday, Monday, Tuesday, Wednesday, Thursday, Friday,
 Saturday or type 'all' if you do not have any preference.
Sunday
```

```
Calculating User Stats...

User Types:
 Subscriber    1643
Customer       232
Name: User Type, dtype: int64

Gender Types:
 Male      1250
Female     393
Name: Gender, dtype: int64

Earliest Year: 1945.0
```