

Overview

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

Bike Share Data

Over the past decade, bicycle-sharing systems have been growing in number and popularity in cities across the world. Bicycle-sharing systems allow users to rent bicycles on a very short-term basis for a price. This allows people to borrow a bike from point A and return it at point B, though they can also return it to the same location if they'd like to just go for a ride. Regardless, each bike can serve several users per day.

Thanks to the rise in information technologies, it is easy for a user of the system to access a dock within the system to unlock or return bicycles. These technologies also provide a wealth of data that can be used to explore how these bike-sharing systems are used.

In this project, you will use data provided by [Motivate](#), a bike share system provider for many major cities in the United States, to uncover bike share usage patterns. You will compare the system usage between three large cities: Chicago, New York City, and Washington, DC.

The Datasets

Randomly selected data for the first six months of 2017 are provided for all three cities. All three of the data files contain the same core **six (6)** columns:

- Start Time (e.g., 2017-01-01 00:07:57)
- End Time (e.g., 2017-01-01 00:20:53)
- Trip Duration (in seconds - e.g., 776)
- Start Station (e.g., Broadway & Barry Ave)
- End Station (e.g., Sedgwick St & North Ave)
- User Type (Subscriber or Customer)

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

- Gender
- Birth Year

	Start Time	End Time	Trip Duration	Start Station	End Station	User Type	Gender	Birth Year
0	2017-01-01 00:00:21	2017-01-01 00:11:41	680	W 82 St & Central Park West	Central Park West & W 72 St	Subscriber	Female	1965.0
1	2017-01-01 00:00:45	2017-01-01 00:22:08	1282	Cooper Square & E 7 St	Broadway & W 32 St	Subscriber	Female	1987.0
2	2017-01-01 00:00:57	2017-01-01 00:11:46	648	5 Ave & E 78 St	3 Ave & E 71 St	Customer	NaN	NaN
3	2017-01-01 00:01:10	2017-01-01 00:11:42	631	5 Ave & E 78 St	3 Ave & E 71 St	Customer	NaN	NaN
4	2017-01-01 00:01:25	2017-01-01 00:11:47	621	5 Ave & E 78 St	3 Ave & E 71 St	Customer	NaN	NaN
5	2017-01-01 00:01:51	2017-01-01 00:12:57	666	Central Park West & W 68 St	Central Park West & W 68 St	Subscriber	Male	2000.0
6	2017-01-01 00:05:00	2017-01-01 00:14:20	559	Broadway & W 60 St	9 Ave & W 45 St	Subscriber	Male	1973.0
7	2017-01-01 00:05:37	2017-01-01 00:19:24	826	Broadway & W 37 St	E 10 St & Avenue A	Subscriber	Female	1977.0
8	2017-01-01 00:05:47	2017-01-01 00:10:02	255	York St & Jay St	Carlton Ave & Flushing Ave	Subscriber	Male	1989.0
9	2017-01-01 00:07:34	2017-01-01 00:18:08	634	Central Park West & W 72 St	Columbus Ave & W 72 St	Subscriber	Male	1980.0

Data for the first 10 rides in the new_york_city.csv file

The original files are much larger and messier, and you don't need to download them, but they can be accessed here if you'd like to see them ([Chicago](#), [New York City](#), [Washington](#)). These files had more columns and they differed in format in many cases. Some [data wrangling](#) has been performed to condense these files to the above core six columns to make your analysis and the evaluation of your Python skills more straightforward. In the Data Wrangling course that comes later in the Data Analyst Nanodegree program, students learn how to wrangle the dirtiest, messiest datasets, so don't worry, you won't miss out on learning this important skill!

Statistics Computed

You will learn about bike share use in Chicago, New York City, and Washington by computing a variety of descriptive statistics. In this project, you'll write code to provide the following information:

#1 Popular times of travel (i.e., occurs most often in the start time)

- most common month
- most common day of week
- most common hour of day

#2 Popular stations and trip

- most common start station
- most common end station
- most common trip from start to end (i.e., most frequent combination of start station and end station)

#3 Trip duration

- total travel time
- average travel time

#4 User info

- counts of each user type
- counts of each gender (only available for NYC and Chicago)
- earliest, most recent, most common year of birth (only available for NYC and Chicago)

The Files

To answer these questions using Python, you will need to write a Python script. To help guide your work in this project, a template with helper code and comments is provided in a **bikeshare.py** file, and you will do your scripting in there also. You will need the three city dataset files too:

- **chicago.csv**
- **new_york_city.csv**
- **washington.csv**

All four of these files are zipped up in the **Bikeshare** file in the resource tab in the sidebar on the left side of this page. You may download and open up that zip file to do your project work on your local machine.

Some versions of this project also include a Project Workspace page in the classroom where the bikeshare.py file and the city dataset files are all included, and you can do all your work with them there.