# Tableau Homework - Citi Bike Analytics

# ## Before You Begin

\* This assignment will be saved to your Tableau Public account rather than GitHub.

\* If you haven't already, be sure to create a Tableau Public account [here](https://public.tableau.com/s/).

\* The free tier of Tableau only lets you save to their public server. This means that each time you save your file it will be uploaded to your Tableau Public profile.

\* You are able to load and continue working on the same workbook.

\* When you are finished with your assignment, you will turn in the URL to your Tableau Public workbook along with any additional files used for your analysis.

# ## Background

![Citi-Bikes](Images/citi-bike-station-bikes.jpg)

Congratulations on your new job! As the new lead analyst for the [New York Citi Bike](https://en.wikipedia.org/wiki/Citi\_Bike) Program, you are now responsible for overseeing the largest bike sharing program in the United States. In your new role, you will be expected to generate regular reports for city officials looking to publicize and improve the city program.

Since 2013, the Citi Bike Program has implemented a robust infrastructure for collecting data on the program's utilization. Through the team's efforts, each month bike data is collected, organized, and made public on the [Citi Bike Data](https://www.citibikenyc.com/system-data) webpage.

However, while the data has been regularly updated, the team has yet to implement a dashboard or sophisticated reporting process. City officials have a number of questions on the program, so your first task on the job is to build a set of data reports to provide the answers.

# ## Task

## \*\*Your task

in this assignment is to aggregate the data found in the **Citi Bike Trip History Logs** and **find two unexpected phenomena**.\*\*

## \*\*Design 2-5 visualizations

for each discovered phenomena (4-10 total). You may work with a timespan of your choosing. Optionally, you may merge multiple datasets from different periods.\*\*

## \*\*The following are some questions

you may wish to tackle. Do not limit yourself to these questions; they are suggestions for a starting point. Be creative!\*\*

### Bar chart Plot duration of trip vs Station start

### Bar chart Plot duration of trip vs month

### Duration of trip Male vs Female vs undefined

### Top 10 station to start trip + Bottom 10 stations 🡺 why

### Top 10 station to end trip + Bottom 10 stations 🡺 why

### Average trip duration by age

### Average trip duration by sex

### Average trip duration by BikeID

### Average trip duration by User Type

### Number of trips by Hours of the day

### Scatter plot duration VS Distance

\* How many trips have been recorded total during the chosen period?

\* By what percentage has total ridership grown?

\* How has the proportion of short-term customers and annual subscribers changed?

\* What are the peak hours in which bikes are used during summer months?

\* What are the peak hours in which bikes are used during winter months?

\* Today, what are the top 10 stations in the city for starting a journey? (Based on data, why do you hypothesize these are the top locations?)

\* Today, what are the top 10 stations in the city for ending a journey? (Based on data, why?)

\* Today, what are the bottom 10 stations in the city for starting a journey? (Based on data, why?)

\* Today, what are the bottom 10 stations in the city for ending a journey (Based on data, why?)

\* Today, what is the gender breakdown of active participants (Male v. Female)?

\* How effective has gender outreach been in increasing female ridership over the timespan?

\* How does the average trip duration change by age?

\* What is the average distance in miles that a bike is ridden?

\* Which bikes (by ID) are most likely due for repair or inspection in the timespan?

\* How variable is the utilization by bike ID?

# Next, as a chronic over-achiever:\*\*

\* Use your visualizations (does not have to be all of them) to design a **dashboard for each phenomena.**

\* The dashboards should be accompanied with an **analysis explaining why the phenomena may be occuring**.

\*\*City officials would also like to see **one of the following** visualizations:\*\*

## \*\*Basic:\*\*

A static map that plots all bike stations with a visual indication of the most popular locations to **start and end** a journey with zip code data overlaid on top.

## \*\*Advanced:\*\*

A dynamic map that shows how each station's popularity changes over time (by month and year). Again, with zip code data overlaid on the map.

\* The map you choose should also be accompanied by a write-up unveiling any trends that were noticed during your analysis.

## \*\*Finally, create your final presentation\*\*

\* Create a Tableau story that brings together the visualizations, requested maps, and dashboards.

\* This is what will be presented to the officials, so be sure to make it professional, logical, and visually appealing.

# ## **Considerations**

Remember, the people reading your analysis will \*\*NOT\*\* be data analysts. Your audience will be city officials, public administrators, and heads of New York City departments. Your data and analysis needs to be presented in a way that is focused, concise, easy-to-understand, and visually compelling. Your visualizations should be colorful enough to be included in press releases, and your analysis should be thoughtful enough for dictating programmatic changes.

# ## Submission

Your final submission should include:

## \* A link to your Tableau Public workbook that includes:

### \* 4-10 Total "Phenomenon" Visualizations

### \* 2 Dashboards

### \* 1 City Official Map

### \* 1 Story

## \* A text or markdown file with your analysis on the phenomenons you uncovered from the data.

## Sharing Your Work

In order to share your work, we are asking that you will save your workbook as a .twbx file so that your TA's can grade them.

To save your workbook as a .twbx file, you will just need to select "Save As..." from the "File" dropdown. Then, select the .twbx option.

# ## Assessment

Your final product will be assessed on the following metrics:

\* Analytic Rigor

\* Readability

\* Visual Attraction

# ## Hints

\* You may need to get creative in how you combine each of the CSV files. Don't just assume Tableau is the right tool for the job. At this point, you have a wealth of technical skills and research abilities. Dig for an approach that works and just go with it.

\* Don't just assume the CSV format hasn't changed since 2013. Subtle changes to the formats in any of your columns can blockade your analysis. Ensure your data is consistent and clean throughout your analysis. (Hint: Start and End Time change at some point in the history logs).

\* Consider building your visualizations with small extracts of the data (i.e. single files) before attempting to import the whole thing. What you will find is that importing all 20+ million records of data will create performance issues quickly. Welcome to "Big Data."

\* While utilizing all of the data may seem like a nice power play, consider the time-course in making your analysis. Is data from 2013 the most relevant for making bike replacement decisions today? Probably not. Don't let overwhelming data fool you. Ground your analysis in common sense.

\* Remember, data alone doesn't "answer" anything. You will need to accompany your data visualizations with clear and directed answers and analysis.

\* As is often the case, your clients are asking for a LOT of answers. Be considerate about their need-to-know and the importance of not "cramming in everything". Of course, answer each question, but do so in a way that is organized and presentable.

\* Since this is a project for the city, spend the appropriate time thinking through decisions on color schemes, fonts, and visual story-telling. The Citi Bike program has a clear visual footprint. As a suggestion, look for ways to have your data visualizations match their aesthetic tones.

\* Pay attention to labels. What exactly is "time duration"? What's the value of "age of birth"? You will almost certainly need calculated fields to get what you need.

\* Keep a close eye for obvious outliers or false data. Not everyone who signs up for the program is answering honestly.

\* In answering the question of "why" a phenomenon is occurring, consider adding other pieces of information on socioeconomic or other geographic data. Tableau has a map "layer" feature that you may find handy.

\* Don't be afraid to manipulate your data and play with settings in Tableau. Tableau is meant to be explored. We haven't covered all that you need -- so you will need to keep an eye out for new tricks.

\* Treat this as a serious endeavor! This is an opportunity to show future employers that you have what it takes to be a top-notch analyst.

\* Good luck!

### Copyright - Data Boot Camp © 2019. All Rights Reserved.

# Grading rubric





