Tableau Assignment - Citi Bike Analytics

Background

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 to build a data dashboard, story, or report. 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, especially if you are working with merged datasets. Do not limit yourself to these questions; they are suggestions for a starting point. Be creative!

- 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?

Additionally, city officials would like to see the following visualizations:

- 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.
- If you're working with a merged dataset: a dynamic map that shows how each station's popularity changes over time (by month and year) -- with commentary pointing to any interesting events that may be behind these phenomena.

Lastly, as a chronic over-achiever:

- Find at least two unexpected phenomena in the data and provide a visualization and analysis to document their presence.

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.

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 dashboards 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 phenomena is happening, 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.

- The final "format" of your deliverable is up to you. It can be an embedded Tableau dashboard, a Tableau Story, a Tableau visualization + PDF -- you name it. The bottom line is: This is your story to tell. Use the medium you deem most effective. (But you should definitely be using Tableau in some way!)

- 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