

Project Description

In this project, you'll create visualizations to reveal insights from a data set. You will create data visualizations that tell a story or highlight patterns in the data set. Your work should be a reflection of the theory and practice of data visualization, such as visual encodings, design principles, and effective communication.

There are 3 different data sets you can choose from.

- Flight Delays and Cancellations
- US Census Demographic Data
- Youtube Data from the US

1) Flight Delays and Cancellations

This data comes from a Kaggle dataset, it tracks the on-time performance of US domestic flights operated by large air carriers in 2015. You can find the dataset in supporting materials at the bottom of this page.

The file you must use in creating your data visualizations is the `flights.csv` file. The other two provided files may be used in conjunction with the `flights.csv` file, but should not be used alone.

You are required to create three visualizations. Some questions you may attempt to answer include those pertaining to the following areas:

1. Which airlines or airports have the worst delays?

Determine which destinations and arrival destinations have the most delays? Doing this using maps is actually pretty difficult, but you may

choose an alternative visual to provide this information. Think about what kind of aggregates might work best to determine which airlines and airports are the best and worst in terms of delays.

2. What causes delays?

Think about if you work at an airline and you want to decrease delays. What part of the flight causes the most delays? Do these causes vary by airport or time of year?

3. You can also come up with your own question!

As you work with the data, come up with a question you're curious about and can be answered from the data. Build a dashboard or story to answer your question and lead viewers to that answer.

2) US Census Demographic Data

This data comes from a Kaggle dataset, it includes the census data for all counties in 2015. You can find the dataset in supporting materials at the bottom of this page. Required dashboards

You are required to create three visualizations. Some questions you may attempt to answer include those pertaining to the following areas:

1. Which states have the best transportation?

This is a fairly subjective question, so your first job is to define what the best transportation is. Is it highest percentage of transit use? Is it lowest mean commute times. Then you need to determine how to aggregate the data from the county level to the state. Are there outlier

counties affecting the data? How should you aggregate all the data from the counties to represent the state effectively? Please provide your reasoning in your report.

2. How does income and poverty look across America?

Think about how best to contrast this data to show an interesting finding. You can look across many of the different fields to show interesting findings. Do counties with more construction experience more or less poverty? Do counties near the coast experience more or less income? Remember this is all correlation and not causation so we cannot say any one thing causes it but we can report descriptive statistics.

3. You can also come up with your own question!

As you work with the data, come up with a question you're curious about and can be answered from the data. Build a dashboard or story to answer your question and lead viewers to that answer.

3) Youtube Data US

This data comes from a Kaggle dataset, it includes a bunch of information for videos that were trending for at least one day. This data set will require some cleaning in excel prior to Tableau. You can find the dataset in supporting materials at the bottom of this page.

You are required to create three visualizations. Some questions you may attempt to answer include those pertaining to the following areas:

- 1.** What tags have grow in popularity over time? (The tag field is very hard to work with, consider asking a different question. Reminder all of the questions here are suggestions, they are not required. There is an additional file with the tags transposed into a a single column that you can make use of as well. To join it in you will want to join it on channel title and title of the video)

You will need to first figure out how to clean the data from having bunch of tag words all in one cell to a format where you can count the number of times the individual tag word is used each day, month, or year depending on how you are answering this question

- 2.** What categories are the most liked and disliked?

To answer this question you will need to replace the numeric category IDs with the actual names which are in a separate file. Then you can looks at this question deeper. Perhaps there is a time of year where one category is preferred? Are there certain channels with an outlier amount of likes or dislikes?

- 3.** You can also come up with your own question!

As you work with the data, come up with a question you're curious about and can be answered from the data. Build a dashboard or story to answer your question and lead viewers to that answer.