

Classy

Location Analysis: Let's do a pseudo-location analysis, by finding and downloading data about USA state populations, generating data, doing some simple analyses on the generated data, then visualizing the data.

We are looking for you to show that you know basic methods for loading, manipulating, visualizing, and modeling data. We also want to see that you can research, select, access, and understand new high-quality datasets from the Internet.

For this exercise, please:

1. **Access a data source:** Different states have different populations. Find a reliable data source online, download it, load it in your program, and verify that the data loaded properly.
2. **Generate raw data:** Generate 50,000 rows of example data into a CSV file. Each row should contain: an age, gender, gift size, billing state. Use the relative populations of each state to generate appropriate samples per state.
3. **Summarize the data:** Print the average gift size by state, the average per capita gift size, and the average gift size by gender. Show histograms of gift size and age. What are the shapes of the histograms, and why?
4. **Create visualizations:** Plot the average gift size by state, and number of donations per state, to two separate maps of the USA.
5. **Build a model:** Compute a linear regression of the number of donations per state, vs. the state population. Overlay the regression line on a scatter plot. What is the variance explained by this regression line?

Here's a bit more guidance:

- Get as far as you can get in a **couple of hours**. To fully solve this problem, it would take many people > 1 day.
- Focus on showing that you *could* execute on a project like this, from beginning to end, and showing your ability to justify your decisions and communicate about results.
- With that said, this is a technical position. Choose a language for each part. Make sure you choose appropriate libraries / methods, and show that you know how to work with data within that language. Make us believe that, with sufficient time, you could easily polish the code to handle any missing parts.
- Make sure to add comments about what decisions you're making and why.