

STA 141C project proposal - Winter 2019

Version 1.0

Form a group of 2-3 students in this class and write a half page typed proposal that addresses the following:

- Come up with a question you can ask using one of the data sets at <http://anson.ucdavis.edu/~clarkf/sta141c/>.
- Describe briefly why this question is interesting, and how you plan to answer it.

We (your instructor and TA) will evaluate the final projects based on the course rubric. We will also consider creativity, ambition, and the technical merit of the solution. A panel of experts from the Data Science Initiative will look at the top 5 or so projects and choose 1 or 2 to submit to usaspending.gov to be displayed at their Student Innovator's page, subject to approval by their External Affairs team.

Data

We have around 60 GB of US federal government transaction records that contain:

- date
- amount
- description, free text and discrete categories
- industry naics code
- hierarchy of funding agency names and ID's
- recipient names and ID's
- business categories of recipients, i.e. `small_business`, `us_owned_business`

You may supplement the data with external data sets.

Example Questions

These are some questions to get you thinking. You can use one of these, or come up with your own.

Prediction

- Given the transaction description, what is the NAICS code?
- How much will an organization spend in the next fiscal year based on the previous spending patterns?
- How much will a contractor receive in the next fiscal year?

Anomaly Detection

- What are the ‘most unusual’ transactions or recipients?
- Do some descriptions use words that are not found in any other descriptions?
- Are there large numerical outliers in some industry categories? These could be identified by the residuals from a model.

Politics

- Does the government party that is in power influence how the money is spent? For example, spending in defense, education, and social welfare may be correlated with the party that is in power.
- How well is the government fulfilling its mandates to support various business categories?

Historical Trends

How have US government purchasing patterns changed over time? This question has many possible facets:

- Are some geographic areas (countries, states, counties) receiving more or less money over time?
- Which industries are growing or declining?