Capstone Project Proposal

The primary problem I will be focusing on is: **What attributes to bank failure?**

I want to solve if there are specifically regional and population variables that are correlated with bank failure and, if so, if there is a way to predict which banks would be susceptible to failure under FDIC regulations. Essentially, if any correlation is found, I would like to see if there is a regression that would accurately predict if the bank would fail.

My client would be larger banks interested in acquiring or merging, and would also be someone who was interested in economic indicators for the financial health of the country. Bank closure signals insufficient funds, so understanding potential causes or ‘breeding grounds’ for a bank closings could potentially lead to a better understanding of what makes a healthy financial institution in today’s world, and how to avoid unproductive situations.

The data set I will be using is titled “FDIC Failed Bank List”. This data is a collection of all banks that have failed since Oct 1, 2000. This data is obtained by going onto data.gov , a federal site with credible data sets ranging in topics. The original source for this data set came from: <http://www.fdic.gov/bank/individual/failed/index.html>

This data set contains 7 variables, with two of them being city and state, which directly helps with region. In dealing with population, I need to gain access to a reputable data set with city populations, and include the cities that are in the FDIC list. This way I could easily add a population variable along side the city variable. It is worth noting that I plan to use the population of the town on the date that the bank closed, which that information is disclosed in the initial data set.

For the dates of banks that closed from 2000-2010, I will be using a data set that lists cities and population from 2000-2010 to get the population figure. This was also obtained from data.gov , here is the link :

<https://catalog.data.gov/harvest/object/c1610a4c-dc1e-4489-9e3e-2efaab51e9b2>

I have been having a harder time finding a data set for date 2010 – 2016 and their populations. While I strongly believe I can find this data, there is a chance I will not find a reliable data set. If this is the case, I will most likely reshape my initial question and strictly focus on the years 2000-2010.

**Solving the problem**

1. Arranging the data so that the failed banks are grouped together by state
2. Arranging the data so that the failed banks are grouped together by national region (i.e. Midwest etc.)
3. Arranging the data so that the failed banks are grouped in strata of population density

I am interested in determining if some states have unusually higher or lower numbers of banks that failed relative to the others. If there are any statistically significant differences in the amount that have failed (or lack of), I plan to do research on that state in terms of population and dominant industries that may affect these numbers.

With the data collected on the states, I then plan to group them by region and see if the collective states per region are correlated by their susceptibility to have higher (or lower) numbers of bank closures.

For the years 2000-2010, I plan to create strata that categorize the cities by population, regardless of region. From there I would like to explore any correlation with population and numbers of bank closures. After that, I would like to analyze population and region (e.g. HIGH POP/LOW POP paired with region) to see if there are any significant correlations between region and population.

From there, I would use independent research again to find if there were any significant changes in the area or notable economic trends within certain regions that may affect the number of bank closures.

**My Deliverables**

I would include code showing how I organized the data, and which regressions and analysis were used on specific variables that would lead me to believe there are any causation with certain variables and bank closures.

I would include slides demonstrating my own research regarding particular regions and cities in relation to their economic conditions, as well as regression charts showing my reasoning behind any conclusions of causation (whether I determine any causation or not).

Ultimately my paper would explain my rational of using any found statistical information to guide my own research and conducting mini-case studies on cities and areas affected by a bank closing.