Blake Conrad

CSCI-49500 – Capstone Research

Project Advisor: Murat Dundar

FBI City Crime Data Prediction Analysis Via Unsupervised Learning

Introduction:

* What is my project?
* What was the goal of my project?
* Why is this project interesting?
* How is this project actually useful?
* In what ways could this project scale?
* Thesis: Each stepping stone along the way, successful or not, list each in a long comma separated sentence, then dedicate 1 paragraph to each approach and images if necessary.

What are good and bad cities?

* My original sources <http://www.usatoday.com> was skewed by population, the corrected sources were at <http://time.com/money/collection/best-places-to-live/> and <http://247wallst.com/special-report/2016/06/28/the-worst-cities-to-live-in/11/> because they are not population dependent, so small and large cities can be good and bad.

How many crime patterns are in the data? (i.e., What’s a good K?)

* My original approach used exhaustive counting of all practical options. The corrected solution used purity scores from my clustering algorithm from 5,10,15,20, and 25 to find which shows the highest purity amongst my 50 best and 50 worst cities from the above sources. The conclusion was K=10 showed the highest purity. This also showed clusters 3 and 5 were considered the best, while all others were associated with the worst cities.

How can we logically group cities into the categories of increasing, decreasing, or stagnant?

* My original approach I was considering using raw numbers, such as above 50K in 2014 but below 10K in 1979, but this had some holes in it. The updated solution is to use the Average Yearly Population Increase (AYPI) of a city looking backward from 2014 to 1979 to see how it moved on average with respect to population. This allowed me to separate them into increasing and decreasing cities.

How do certain groups of cities behave as population changes? Can we determine which crimes make them less favorable or more favorable as a city grows or shrinks in size?

* By looking at our population increasing cities, we can look at how each cluster on average per year changed per crime and plot these over time to look at which crimes were actually associated with the worst cities. We can also see which types of crime the good crime pattern based cities had and didn’t have with respect to the cities in the bad crime pattern cities. This showed that assault crime (total Assault, officers Assaulted, and rape) was most heavily associated with the bad crime based cities upon increase in population and decrease in population. An interesting find from the data is that the main crime a city wants to avoid, whether increasing or decreasing, is assault and rape. This was relatively unintuitive for me, because I would see life and death as the most important issue regarding cities and their desire for people to want to live there, however the study shows murder was relatively consistent between amongst cities, the assault and rape was the determining factor to becoming one of the worst cities in the U.S.

Conclusion:

* The after thoughts
* Struggles in the project
* Eye opening to how research actually looks
* Eye opening to how scientists work in their analysis
* Eye opening the difficulty of good plots
* Level of creativity involved in the project
* Where this project could go in the future