IBM DATA SCIENCE CAPSTONE PROJECT

DOES POST-SECONDARY EDUCATION AFFECT THE RATE OF CRIME IN CITIES?

An investigation of the relation between the increase in the number of academic centers and libraries and a decrease in the rate of crime in American cities.

Iman Soltani

OUTLINE

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

1.INTRODUCTION (BUSINESS PROBLEM)

- After a significant decrease in 2020 the crime index of the United States has been determined as 47.7.
- In 2012 United states was considered as one of the creepiest countries in the world, with a crime index of 64.93.
- The crimes in the cities are divided into two main groups of violent crime and property crime.
- violent crime : offender uses or intimidates to use force on the victim. E.g. robbery, hijacking, carjacking, rape, kidnapping, shooting, torture,…
- Property crime : committed to obtaining money or any other similar benefit, e.g. theft, shoplifting, vandalism, larceny, and burglary.

https://www.numbeo.com/crime/rankings_by_country.jsp?title=2016
https://worldpopulationreview.com/country-rankings/crime-rate-by-country
https://en.wikipedia.org/wiki/Violent_crime
https://en.wikipedia.org/wiki/Property_crime

1.INTRODUCTION (BUSINESS PROBLEM)

- **Possible reasons:**
- The economic issues that the people under the poverty line would struggle with can be considered as the major motivation.
- Gun freedom is considered as one of the reasons.
- It is commonly believed that an increase in the level of education of people may play a major effect on decreasing the cities' crime rates.

- This study endeavored to scrutinize the existence of such a relation.
- One may consider the number of academic and educational institutions per capita as a possible measure of the education level in a city.

- To make it more specific we can account for the number of pos-secondary academic institutions as a measure of the education level of people in the vicinity.

2.DATA

2.1. Crime rate in American cities

- To determine the rate of crime in American cities, a Wikipedia article on the list of a selected United States cities by crime rates. Was used
- In this article, there is a table containing about 90 most populous American cities along with their crime rates per 100,000 people based on Federal Bureau of Investigation (FBI) Uniform Crime Reports (UCR) statistics from 2017.
- Both violent crime data including murder and nonnegligent manslaughter, rape, robbery, aggravated assault, as well as property crime including burglary, larceny-theft, motor vehicle theft, and arson.

https://en.wikipedia.org/wiki/List_of_United_States_cities_by_crime_rate

2.DATA

2.2.Geographic coordinates of the selected most populous American cities.

- To determine the geographic coordinates of the selected most populous American cities, we used the geopy.geocoder library to convert the name of the cities into their corresponding coordinates.

2.3.Number of post-secondary schools in American cities

- To get detailed information about the cities post-secondary schools we used Foursquare, which is a location data provider website. So, we constructed some URL to make Foursquare API calls to search the cities for universities and colleges. Then, we received the json file of the corresponding detailed information upon calling get request.

3.METHODOLOGY

3.1.Information scrapping from the crime data table in American cities.

- Imported Pandas and NumPy libraries,
- Used read_html function from Pandas library to read the webpage table data into a data frame.
- Performed data cleaning steps on the table of data.

3.2.Determining the geographic coordinate of the cities.

- We used geopy.geocoder and imported the Nominatim function to do the geocoding of the cities' names into geographic coordinates.

3.METHODOLOGY

3.3. Creating a map of American cities using Folium library.

- Matplotlib library with cm and colors modules used to plot the map. Folium graphical map rendering library was used.
- Geopy geocoder to determine the geographic coordinate of the United States.
- A map of the United States with the selected cities superimposed on top.

3.4. Exploring the American cities using Foursquare API

- Foursquare API calls (upon registration) was used to explore the selected American cities and cluster them.
- Data cleaning was performed on the json detailed information file (received upon calling get request) to result in a desired table of appropriate data .

3.METHODOLOGY

3.5. Clustering cities based on their post-secondary schools

- Clustering (an unsupervised machine learning method)
- KMeans algorithm (from Scikit Learn library) was used to do the clustering of the cities into 3 groups based on their number of post-secondary education centers.

3.6. Clustering cities based on their total crime rates

- KMeans algorithm (from Scikit Learn library) was used to do the clustering of the cities into 3 groups based on their crimes rates.

4.RESULTS

- The initial map shows the United States and selected most populous cities superimposed on top of it that were plotted using folium library.

4.RESULTS

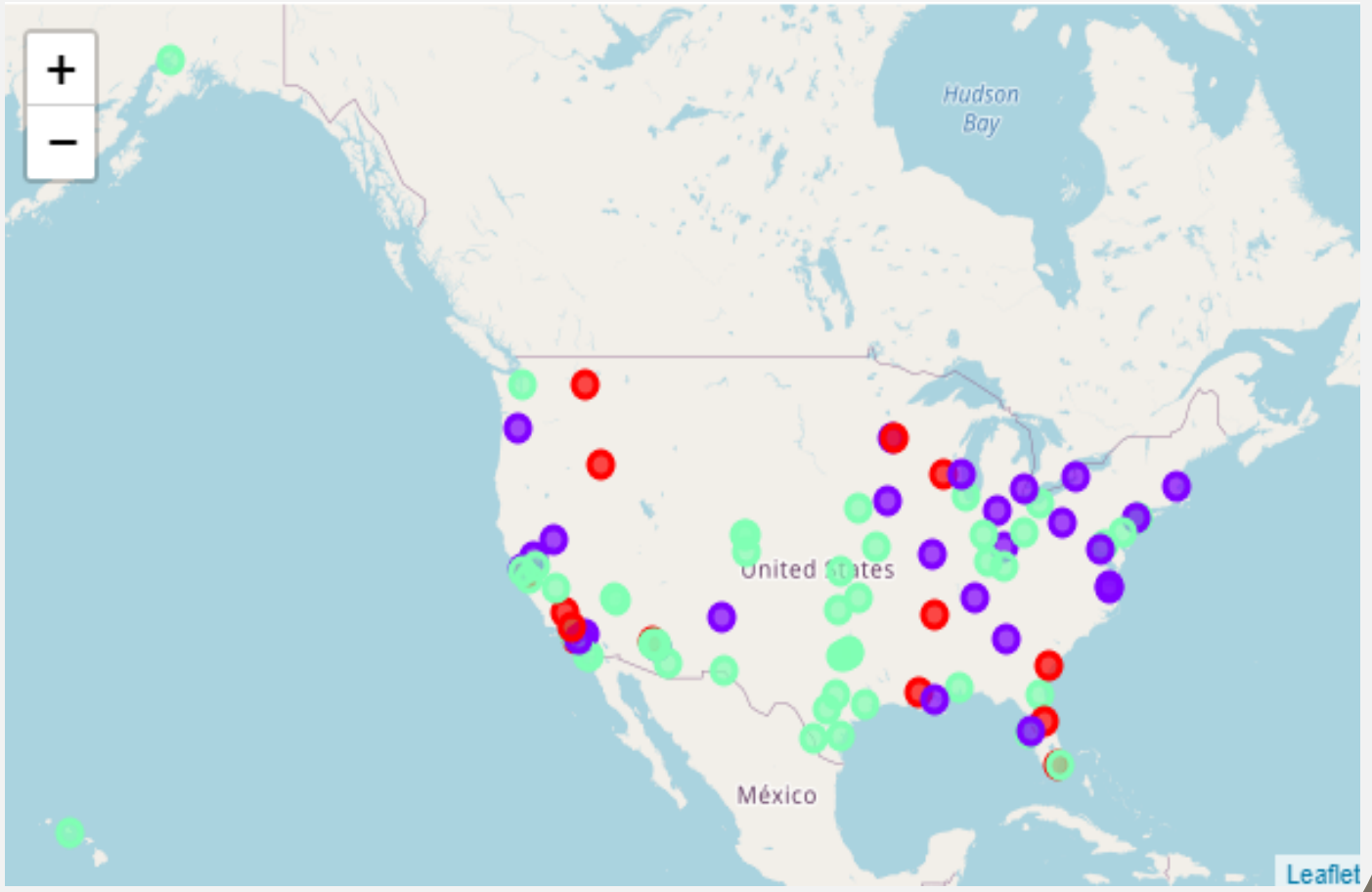
4.1. Clustering cities based on their post-secondary schools

- Upon making Foursquare API calls and data cleaning we found 558 postsecondary academic centers in 91 most populated American cities.
- The following table shows the first 5 rows of the resulting data.

	Edu Cluster Labels	City	University per capita	College & University per capita	General College & University per capita	Latitude	Longitude
0	1	Albuquerque	0.000011	0.0	0.000004	35.084103	-106.650985
1	1	Anaheim	0.000011	0.0	0.000006	33.834752	-117.911732
2	2	Anchorage	0.000003	0.0	0.000003	61.216313	-149.894852
3	2	Arlington	0.000008	0.0	0.000005	32.701939	-97.105624
4	1	Atlanta	0.000010	0.0	0.000004	33.748992	-84.390264

4.RESULTS

- Cluster 1 (violet color) shows the cities with the highest number of universities per capita.
- Cluster 0 (light green) shows a relatively high number of General College and University per capita,
- Cluster 2 (red color) shows the lowest number of General College and University per capita.
- The number of College and University per capita is low for all 3 clusters.



4.RESULTS

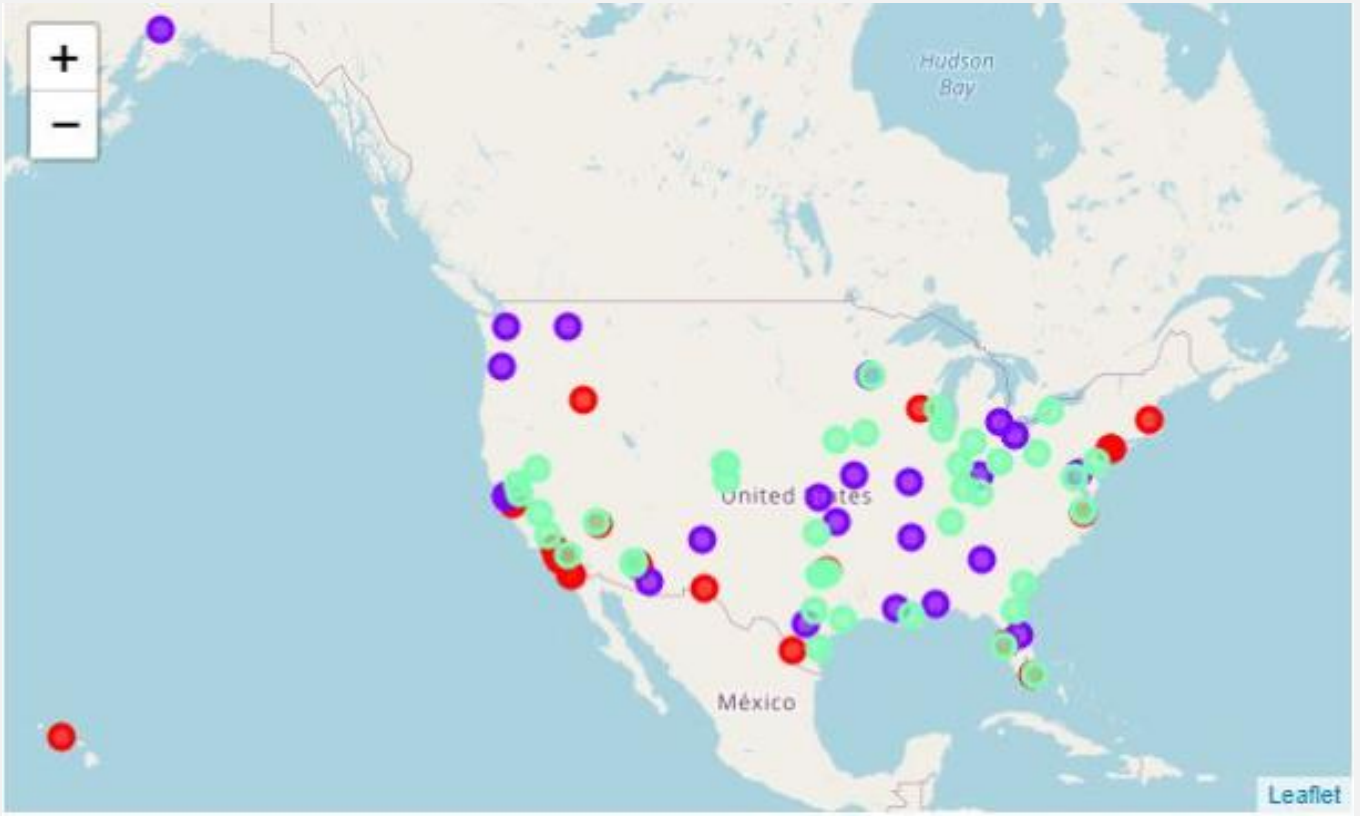
4.2. Clustering cities based on their crime rates

- Clustering of the cities based on their crime rates using the K-Means algorithm resulted in 3 groups.
- The following image shows the upper part of the resulting table.

Crime Cluster Labels	Edu Cluster Labels	City	University per capita	College & University per capita	General College & University per capita	Total Violent Crime	Total Property crime	Latitude	Longitude
0	1	Irvine	1.810840e-05	0.000000e+00	7.243359e-06	61.21	1316.48	33.685697	-117.825982
0	2	Gilbert	8.261390e-06	0.000000e+00	0.000000e+00	85.51	1385.85	33.352763	-111.789037
0	2	Plano	3.443372e-06	0.000000e+00	0.000000e+00	149.79	1733.74	33.013676	-96.692510
0	0	Santa Clarita	4.622140e-06	9.244280e-06	1.386642e-05	162.70	1424.08	34.391664	-118.542586
0	0	Fremont	0.000000e+00	4.230691e-06	1.269207e-05	182.34	2150.46	37.548270	-121.988572
0	2	Henderson	3.341297e-06	3.341297e-06	3.341297e-06	185.11	1833.04	36.030113	-114.982619
0	0	Hialeah	8.394191e-06	0.000000e+00	2.937967e-05	198.52	2213.55	25.857596	-80.278106
0	2	Irving	8.262346e-06	0.000000e+00	0.000000e+00	226.80	2539.43	32.829518	-96.944218
0	2	Honolulu	2.019419e-06	2.019419e-06	3.029128e-06	246.37	2774.38	21.304547	-157.855676

4.RESULTS

- Cluster 0 (red color) showed the cities with the lowest Total Property Crime rate and the lowest Total Violent Crime rate.
- Cluster 2 (light green color) showed the cities with the mid-level of the Total Property Crime rate and Total Violent Crime rate.
- Cluster 1 (violet color) demonstrated the highest rate of this type of crime.



5.DISCUSSION

- Comparing both cities post-secondary educational centers numbers and their rates of crime clustering trends showed some level of correlation between them.
- However, there were exceptional cases as well, where despite good academic levels, the rates of crimes displayed showed very high levels.
- The particular exceptions were the cities like Albuquerque, Oakland, St Louis, Memphis, Baton Rouge, and Detroit,
- It may be argued that other factors are affecting the rate of crime as well.
- For example, the economic condition of a city can affect the rate of crime as we see in the cities like Baton Rouge, Memphis, and St Louis.

5.DISCUSSION

- Another exception is the city of Oakland, CA, a short distance from Berkeley CA,
- Since we set for the radius of 5000m to count for the post-secondary schools near each city, Berkeley's academic centers would increase the academic grade around Oakland.
- However, Oakland has been historically considered as a poor region in the San-Francisco Bay Area that affects the rate of crime in this city.

6.CONCLUSION

- In this study it was attempted to investigate the rate of crime in most populated American cities, as well as their level of post-secondary education.
- For the rate of crime, both violent crime rates and property crime rates data were scrapped from the relevant websites.
- After data cleaning, the K-Means clustering as an unsupervised machine learning algorithm was performed on them.
- The clusters were visualized into a map using the Folium library. Accordingly, the cities were divided into 3 clusters based on their crime rates.
- The numbers of universities were determined by making Foursquare API calls.
- Data cleaning was performed on the query results, a json file containing the relevant detailed information about those cities.
- K-Means clustering was conducted on the prepared data to categories the cities based on the numbers of their post-secondary educational centers into 3 groups.
- Visualization was done for education-based clustering of the cities, using the Folium library.

6.CONCLUSION

- Despite observing some contradicting exceptions, many cases of correlations were observed between two different clustering results, pointing to a decrease in the rate of crime upon an increase in the number of universities in the cities.
- The results confirmed the assumption that an increase in the post-secondary education level of people may decrease their intentions towards committing the crime.
- On the other hand, for the cities with a good level of academic education and still a high rate of crimes, economic problems may be considered as the crime controlling factors.

REFERENCES

1. https://www.numbeo.com/crime/rankings_by_country.jsp?title=2016
https://worldpopulationreview.com/country-rankings/crime-rate-by-country

2. https://en.wikipedia.org/wiki/Violent_crime

3. https://en.wikipedia.org/wiki/Property_crime

4. List of United States cities by crime rate - Wikipedia:
https://en.wikipedia.org/wiki/List_of_United_States_cities_by_crime_rate