Milestone: - Project Proposal

Uncovering Crime Patterns in Communities: A Data-Driven Approach

Group 87

Email:gupta.vishe@northeastern.edu

Signature of Student : Vishesh Gupta

Submission Date: **January 27th 2023**

Problem Setting: Urban communities are facing an alarming rise in crime rates, which poses a significant threat to the safety and well-being of residents. The increasing crime not only affects the physical security of individuals but also leads to decreased property values, slowed economic development, and a negative impact on the overall quality of life in a community. Despite these negative impacts, many urban communities struggle to effectively address crime and make their neighborhoods safer due to lack of relevant data and proper analysis. So, each records of the data involve demographic information, economic information, and crime statistics to identify patterns and trends that may be associated with higher crime rates.

Problem Definition: The problem of increasing crime rates in urban communities is a complex issue that requires a multifaceted approach. To effectively address this problem, a thorough analysis is needed to identify patterns and trends that may be associated with higher crime rates. This information can then be used to develop targeted strategies to reduce crime and improve safety in these communities. Additionally, this project aims to provide insights into the root causes of crime in urban communities, which can inform policies and programs to prevent crime and promote community safety.

Key objectives of the project:

- Collect and clean crime data from various sources, including local law enforcement agencies and government databases.
- Analyze crime data at the neighborhood level to identify patterns and trends in crime incidents.
- Investigate how various factors, such as population density, socio-economic status, and the presence of amenities or infrastructure, are related to crime rates in different communities.
- Develop a model to classify crimes or predict crime rates in neighborhoods based on the identified factors.
- Use the findings from the analysis to develop recommendations for reducing crime and improving community safety.

Key Deliverables:

- A detailed analysis of crime patterns and trends in the target communities.

Data Sources:

Website link: https://archive.ics.uci.edu/ml/datasets/communities+and+crime

Data agency: University of California, Irvine (UCI)

Dataset owner: Center for Machine Learning and Intelligent Systems (CMLIS)

Type of data: Community demographic, economic, and crime statistics

Number of records: 2100

Number of attributes: 128

Additional information: The data was collected by the FBI's Uniform Crime Reporting Program and the U.S. Census Bureau's. It includes information on various community characteristics such as population, race, education level, and income, as well as crime statistics such as number of reported violent crimes and property crimes. We will be utilizing dimension reduction techniques to enhance the effectiveness of the dataset.

Sample Dataset Representation:

8	?	?	Lakewood	1	0.19	0.33	0.02	0.9	0.12	0.17	0.34	0.47	0.29	0.32	0.2	1	0.37	0.72	0.34
53	?	?	Tukwilacit	1	0	0.16	0.12	0.74	0.45	0.07	0.26	0.59	0.35	0.27	0.02	1	0.31	0.72	0.11
24	?	?	Aberdeent	1	0	0.42	0.49	0.56	0.17	0.04	0.39	0.47	0.28	0.32	0	0	0.3	0.58	0.19
34	5	81440	Willingbor	1	0.04	0.77	1	0.08	0.12	0.1	0.51	0.5	0.34	0.21	0.06	1	0.58	0.89	0.21
42	95	6096	Bethlehem	1	0.01	0.55	0.02	0.95	0.09	0.05	0.38	0.38	0.23	0.36	0.02	0.9	0.5	0.72	0.16
6	?	?	SouthPasa	1	0.02	0.28	0.06	0.54	1	0.25	0.31	0.48	0.27	0.37	0.04	1	0.52	0.68	0.2
44	7	41500	Lincolntow	1	0.01	0.39	0	0.98	0.06	0.02	0.3	0.37	0.23	0.6	0.02	0.81	0.42	0.5	0.23
6	?	?	Selmacity	1	0.01	0.74	0.03	0.46	0.2	1	0.52	0.55	0.36	0.35	0	0	0.16	0.44	1
21	?	?	Hendersor	1	0.03	0.34	0.2	0.84	0.02	0	0.38	0.45	0.28	0.48	0.04	1	0.17	0.47	0.36
29	?	?	Claytoncit	1	0.01	0.4	0.06	0.87	0.3	0.03	0.9	0.82	0.8	0.39	0.02	1	0.54	0.59	0.22
6	?	?	DalyCitycit	1	0.13	0.71	0.15	0.07	1	0.41	0.4	0.52	0.35	0.33	0.15	1	0.49	0.71	0.16
36	?	?	RockvilleC	1	0.02	0.46	0.08	0.91	0.07	0.1	0.34	0.36	0.22	0.57	0.04	1	0.72	0.53	0.23
25	21	44105	Needhamt	1	0.03	0.47	0.01	0.96	0.13	0.02	0.29	0.32	0.2	0.52	0.04	1	0.8	0.55	0.18
55	87	30075	GrandChut	1	0.01	0.44	0	0.98	0.04	0.01	0.35	0.53	0.32	0.23	0.02	0.77	0.46	0.77	0.41
6	?	?	DanaPoint	1	0.04	0.36	0.01	0.85	0.14	0.26	0.32	0.46	0.3	0.31	0.05	1	0.71	0.67	0.42
19	187	91370	FortDodge	1	0.03	0.34	0.06	0.93	0.03	0.03	0.39	0.41	0.28	0.58	0	0	0.18	0.42	0.81
36	1	1000	Albanycity	1	0.15	0.31	0.4	0.63	0.14	0.06	0.58	0.72	0.65	0.47	0.16	1	0.22	0.52	0.1
34	27	17650	Denvilleto	1	0.01	0.53	0.01	0.94	0.2	0.03	0.34	0.39	0.27	0.36	0.02	0.76	0.79	0.77	0.13
18	?	?	Valparaiso	1	0.02	0.47	0.01	0.97	0.07	0.02	0.7	0.67	0.63	0.37	0	0	0.33	0.56	0.28
42	129	66376	Rostravert	1	0	0.41	0.05	0.96	0.01	0.01	0.37	0.37	0.24	0.55	0.01	0.58	0.23	0.34	0.33
6	?	?	Modestoci	1	0.25	0.54	0.05	0.71	0.48	0.3	0.42	0.48	0.28	0.32	0.26	1	0.33	0.55	0.37
12	31	?	Jacksonvill	1	1	0.42	0.47	0.59	0.12	0.05	0.41	0.53	0.34	0.33	1	0.99	0.28	0.62	0.16
41	?	?	KlamathFa	1	0.01	0.34	0.02	0.87	0.07	0.11	0.49	0.56	0.43	0.47	0	0	0.13	0.4	0.26
19	193	93926	SiouxCityc	1	0.11	0.43	0.04	0.89	0.09	0.06	0.45	0.48	0.31	0.46	0.13	1	0.22	0.52	0.44
6	?	?	Delanocity	1	0.02	0.96	0.05	0	1	1	0.54	0.58	0.39	0.33	0	0	0.16	0.61	0.41

The dataset will include the most important attributes as shown in the sample below, as well as a few other attributes that will be countered later on as per analysis:

Attributes	Attributes Type	Illustration
Community_Name	Str/Text	Community name for information only
State_Name/Code	Str/Text	Code for State_Name
Race_Asian_Pct	Numeral-Decimal	Percentage of population that is of asian heritage
Race_White_Pct	Numeral-Decimal	Percentage of population that is caucasian
Auto_Theft_Per_Pop	Numeral-Decimal	Theft cases per 100K population
House_Hold_Size	Numeral-Decimal	Mean people per household
Violent_Crimes_Per_Pop	Numeral-Decimal	Total number of violent crimes per 100K population GOAL attribute (to be predicted)