Investigating the Causal Relationships between Public Health and Violence/Crime in Chicago

Why Mental Health ?

1. Mental Health Affects Everyone
2. Mental Health and Physical Health are Interconnected
3. Mental Health Impacts Relationships
4. Mental Health and Productivity
5. Mental Health Stigma and Advocacy

Why Mental Health in Chicago? Let's look at some statistics

- Approximately **1 in 5 adults** in Chicago State **experiences a mental health disorder** each year (Source: Chicago Department of Public Health).
- Among youth aged 12-17 in Chicago State, the **prevalence of depression is 17.9%**, surpassing the national average (Source: Healthy Chicago Survey).
- In 2020, there were **581 reported suicides** in Cook County, which includes Chicago State, highlighting the urgency to address mental health issues (Source: Cook County Medical Examiner's Office).
- The rate of mental health-related emergency department visits in Chicago State is **higher compared to the national average**, indicating the need for accessible and effective mental health services (Source: Illinois Department of Public Health).

Our Study

- In this study, we aim to **investigate the causal relationships** between various factors that affect public health, violence/crime, and transportation in Chicago.
- By investigating the causal relationships between these elements, we aim to **gain a deeper understanding** of their interconnectedness and inform evidence-based strategies for enhancing public well-being and safety.
- The project will contain two components: **causal discovery** to find the causal structure and **causal inference** to estimate the causal effect.

What is Causality? Causal Discovery ? Causal Inference ?

- **Causality** refers to the relationship between cause and effect, where a cause is an event or condition that leads to an effect.
- **Causal Discovery** is the process of identifying causal relationships from observational or experimental data.It involves determining whether a particular variable or event influences another variable or event.
- **Causal Inference,** on the other hand, is a statistical and analytical approach used to draw conclusions about causal relationships from observational data. It involves making inferences about cause and effect relationships in situations where it is not feasible or ethical to conduct controlled experiments.

Datasets

Here are the data sets we will be looking at:

1. Clinics with zip codes and if they offer certain types of services

2. Types of health services provided

3. Public health indicators per neighborhood - we can make this at a zipcode level

4. Locations of affordable housing

5. Trust in police based on different demographics

6. School lead levels

7. Park lead levels

8. Chicago Crime data set

Exploration

Here’s a general rule we followed (given by the professor herself) :-

“ For the first step, carefully check the data properties, including whether the data is **continuous or discrete**, whether the data is **Gaussian or not**, whether the **causal relationships are very nonlinear or can be captured by linear relations**, whether there are **missing data**, whether there are **hidden confounders** according to the background knowledge, and whether there is **selection bias** in the data. ”

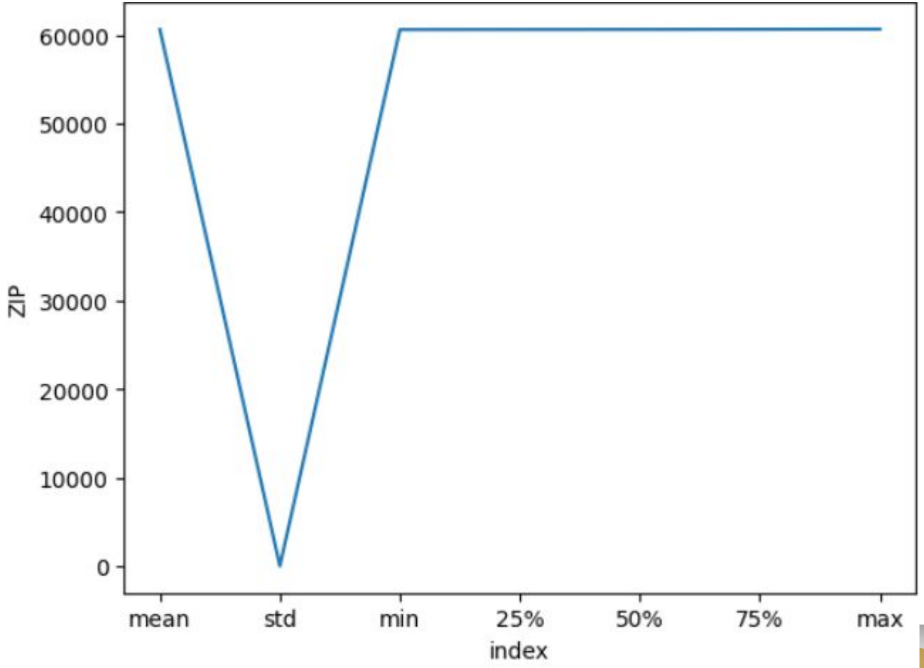
Datasets	Continuous/Discrete	Gaussian or not	Causal relations - similar for all	Missing data	Hidden Confounders Similar for all	Selection bias - same for all
Clinics	Zipcode - Discrete	Approx	Non-linear	Yes	Genetic factors, Lifestyle factors	Tending towards the “privileged”
Health service	All are discrete(binary) except zipcode	Approx		Yes		
Public health indicators	All are continuous	Approx		Yes		
Affordable housing	Community Area Number , zipcode - continuous	Yes		Yes		
Trust in police	Safety, Trust are continuous	Yes		Yes		
School lead levels	Score, num_fixtures	Yes		Yes		

Park lead levels	Score, num_fixtures	Approx		Yes		
Chicago crime	Discrete data	Approx		Yes		

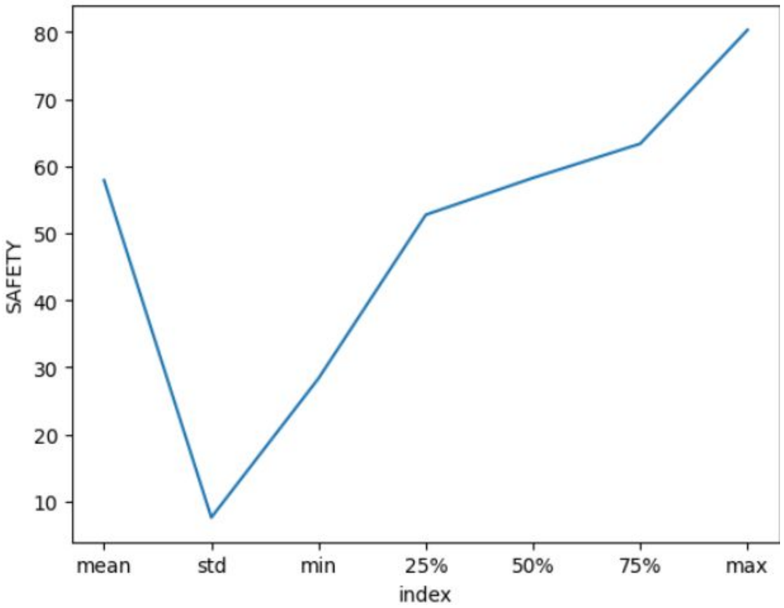
Here are some data exploration results and visuals

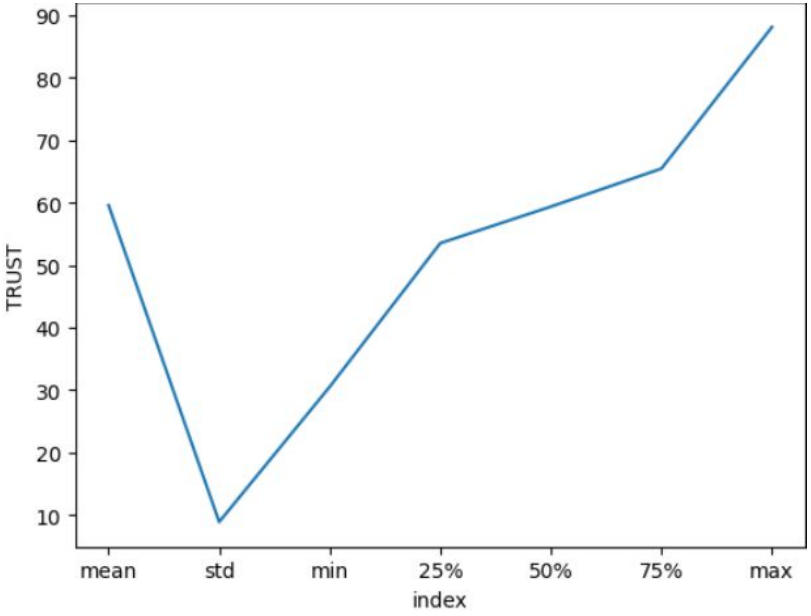
Summary Statistics visualized of the “important” columns:-

1. Chicago Department of Public Health clinics

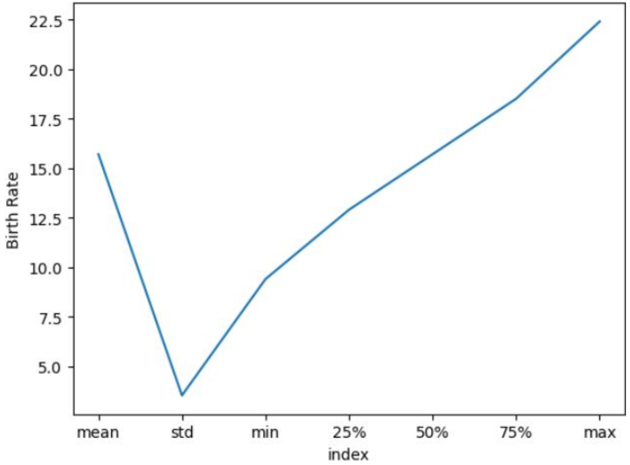


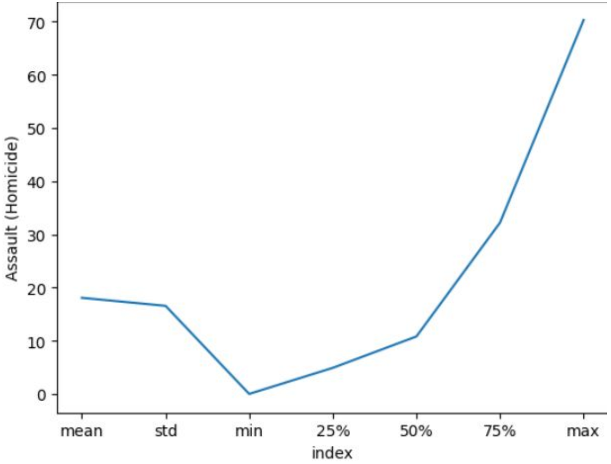
2. Police Sentiment Scores

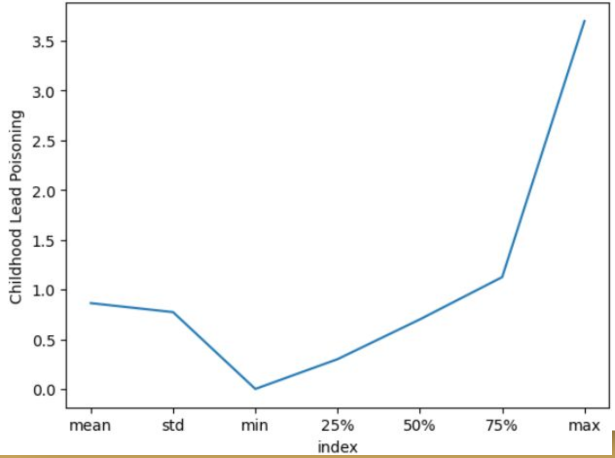


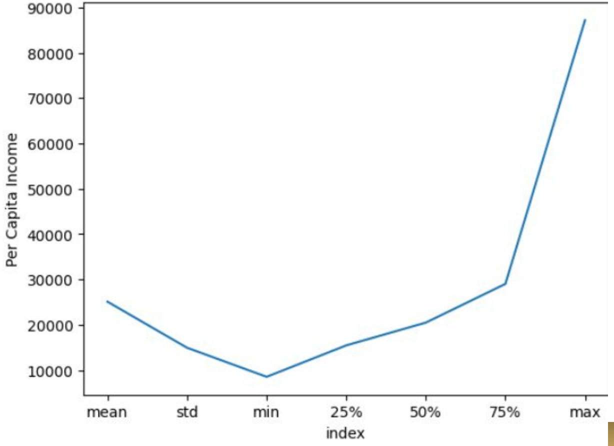


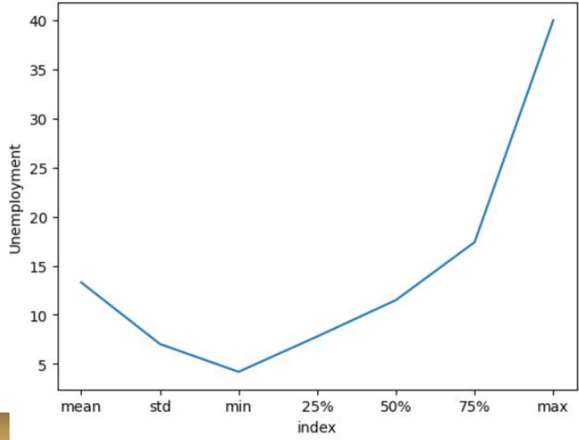
3. Public health indicators by Chicago community area



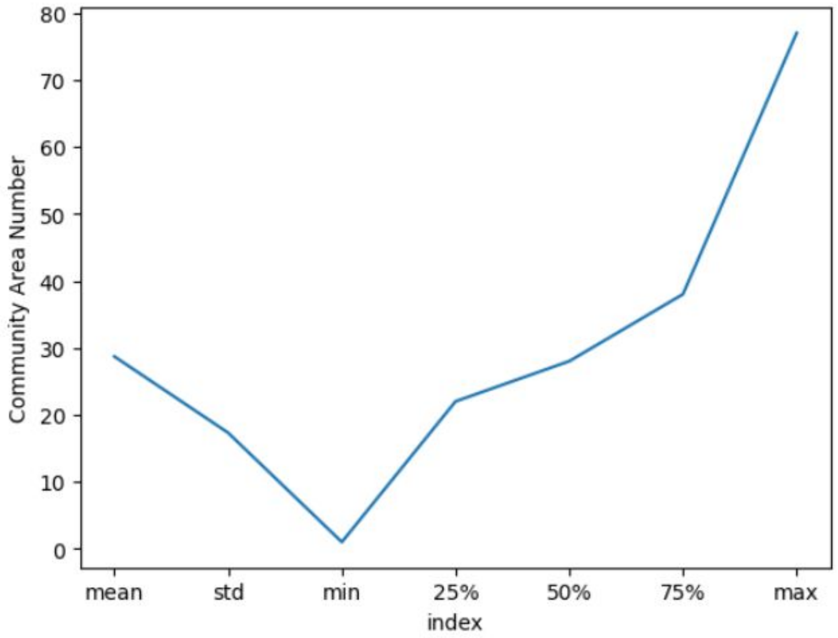


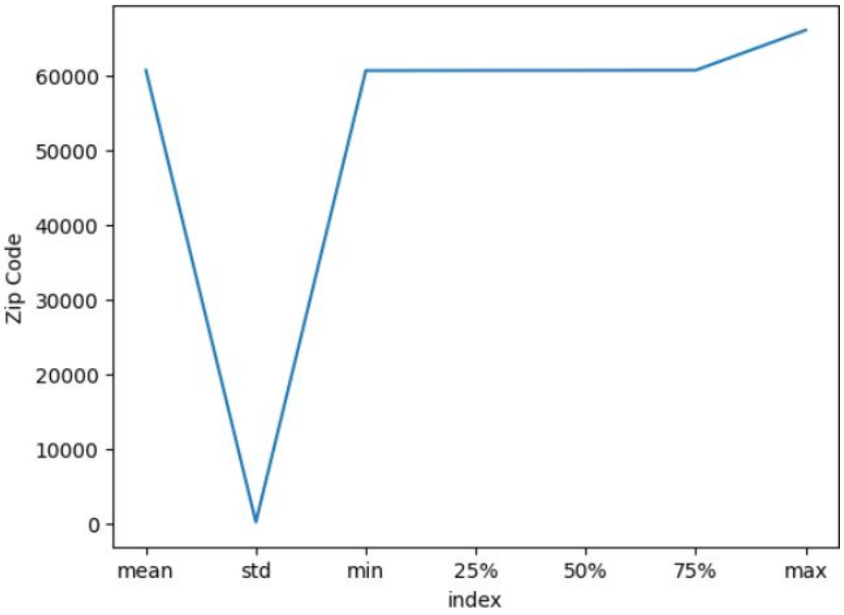




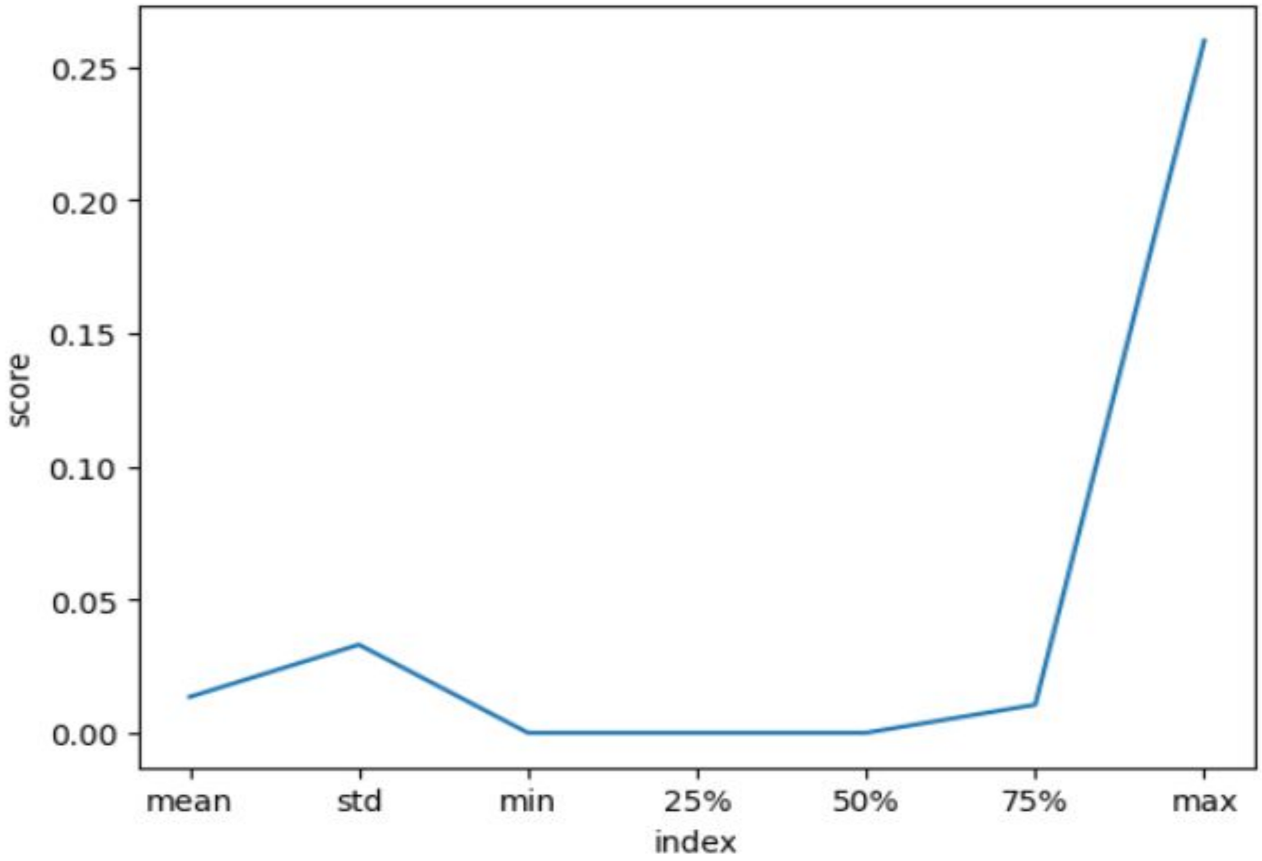


4. Affordable housing developments

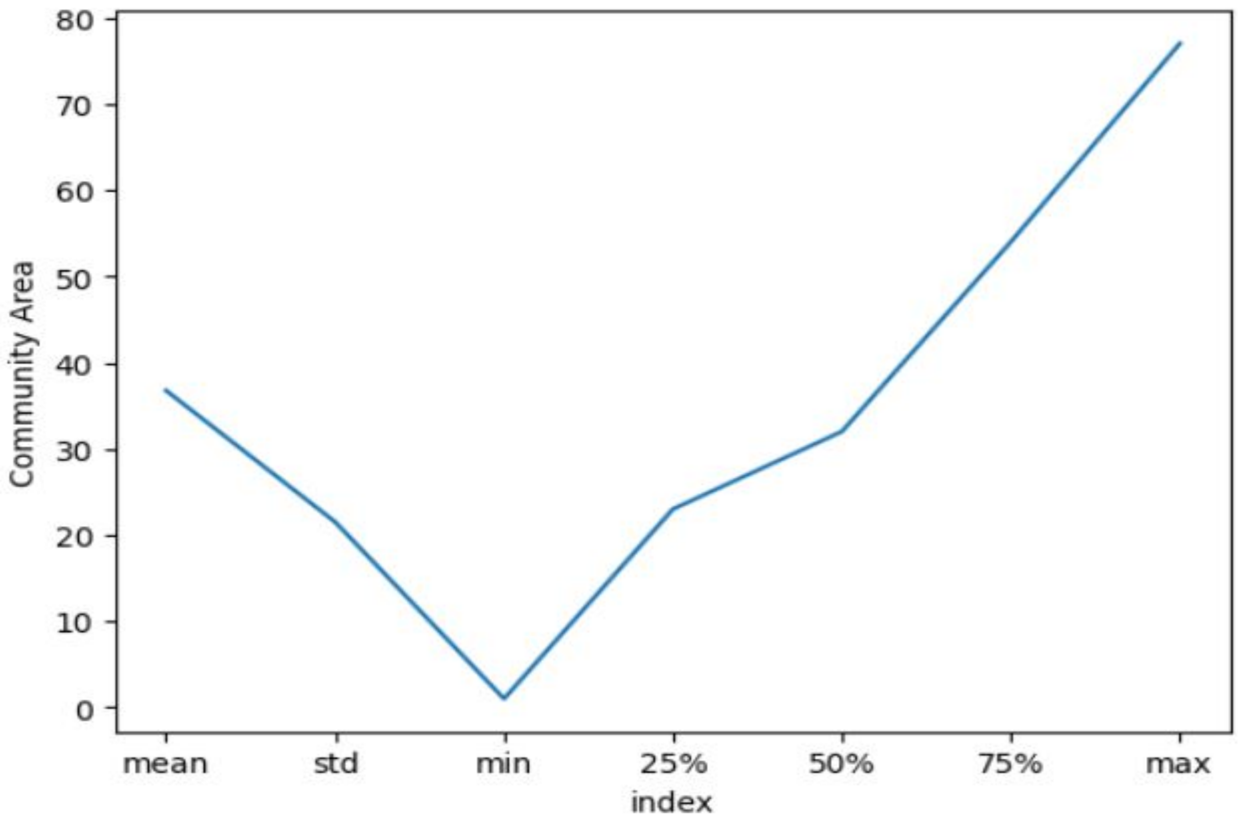




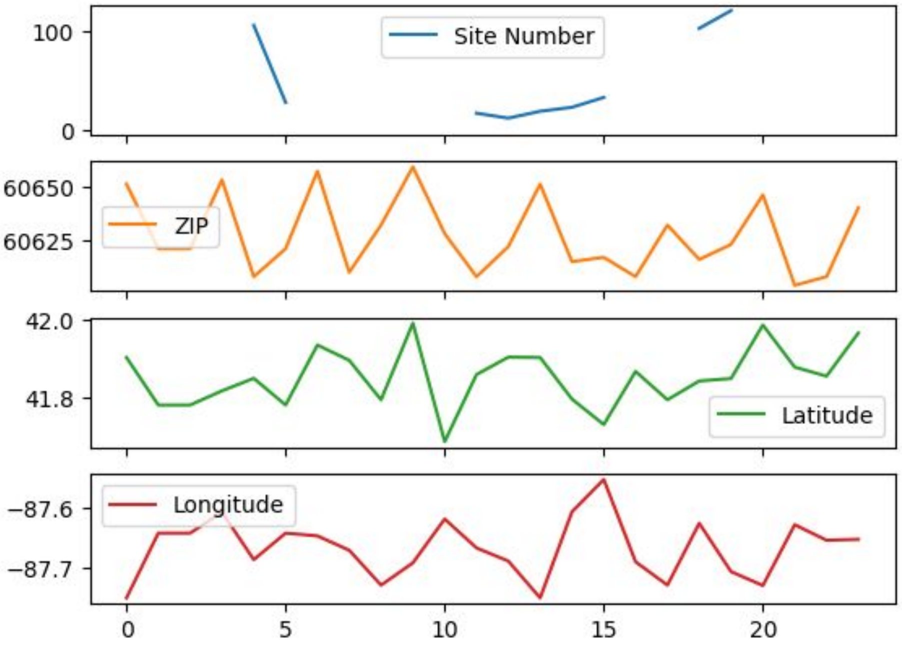
Lead Scores

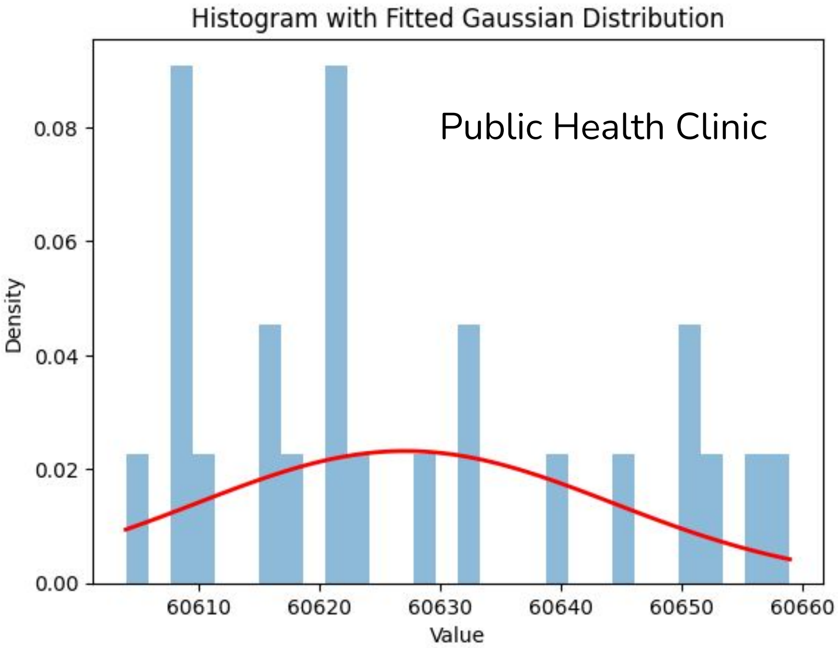


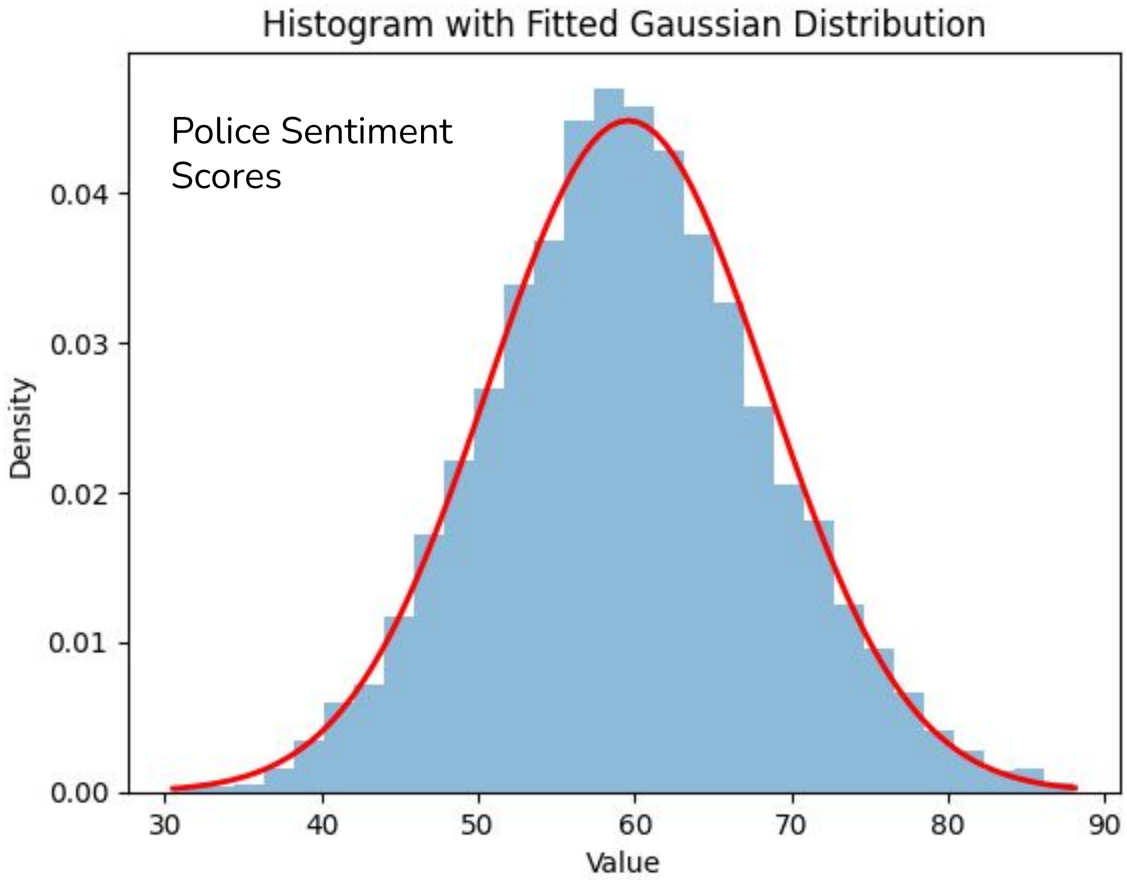
Chicago Crime dataset



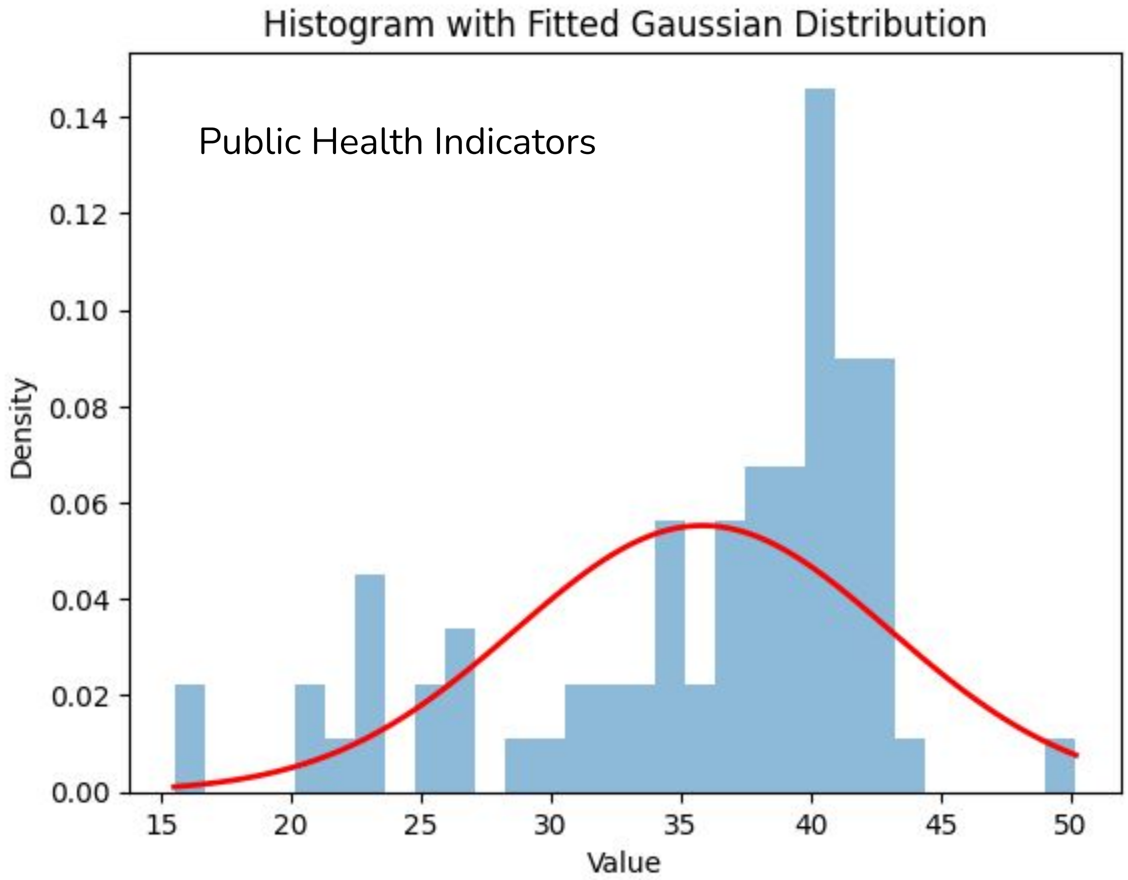
Some more visuals that helped us judge the data properties



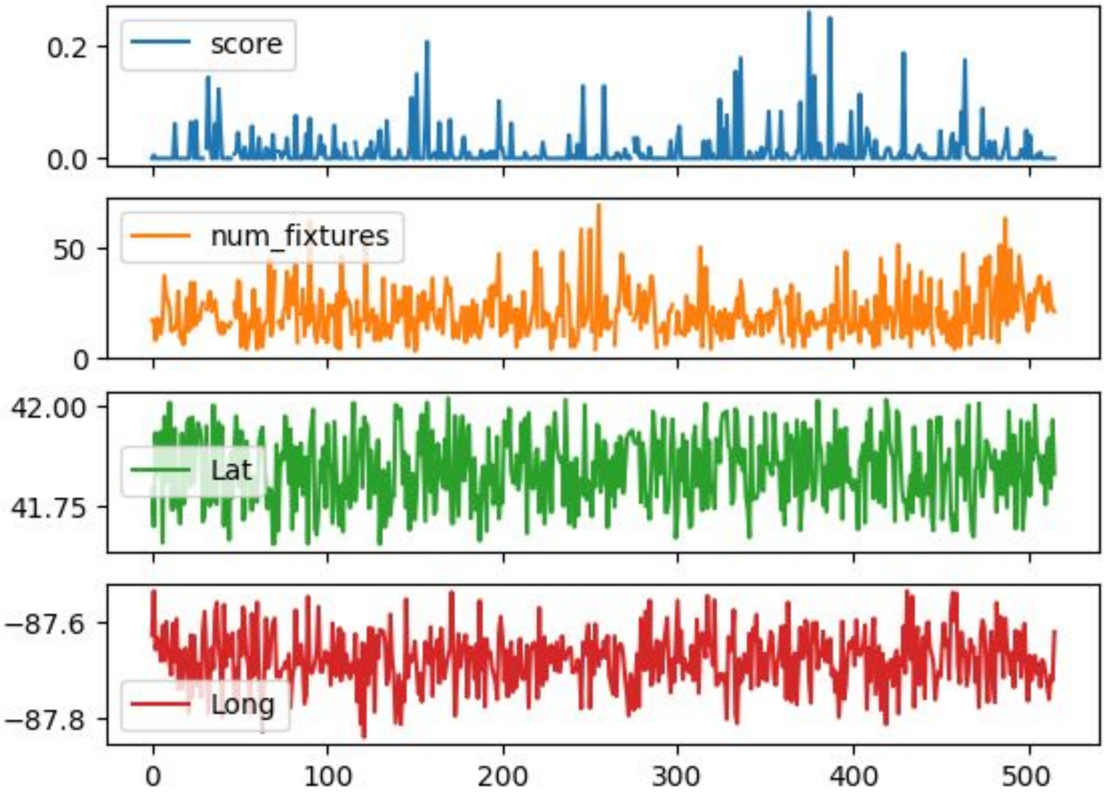




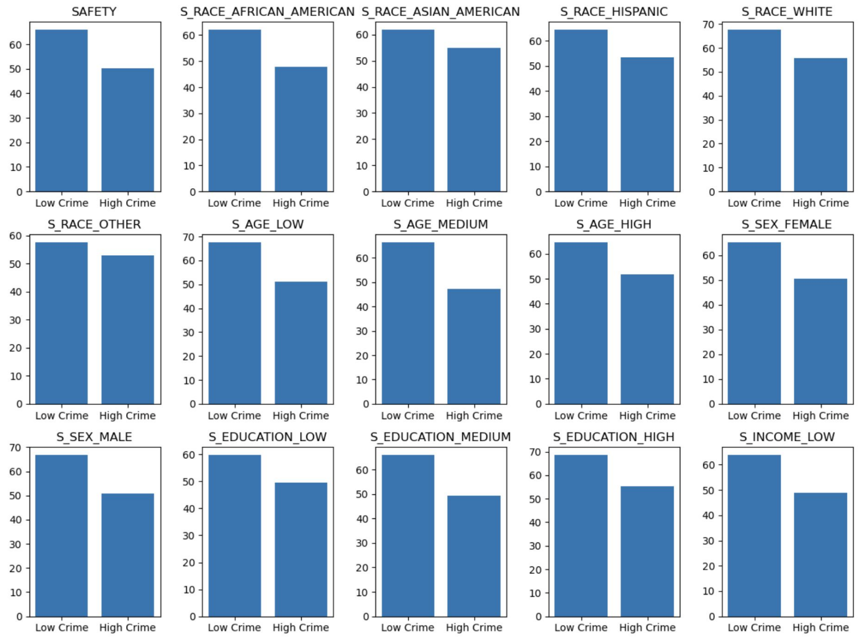
Histogram with Fitted Gaussian Distribution
Police Sentiment Scores
Density
0.00
0.01
0.02
0.03
0.04
30
40
50
60
70
80
90
Value


Histogram with Fitted Gaussian Distribution
Public Health Indicators
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
15
20
25
30
35
40
45
50
Value

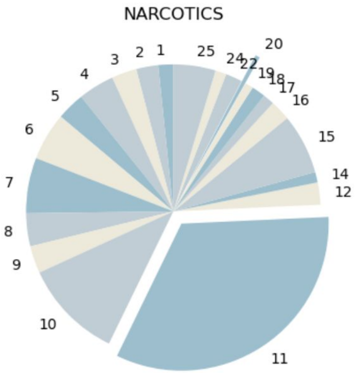
Lead Scores-Levels

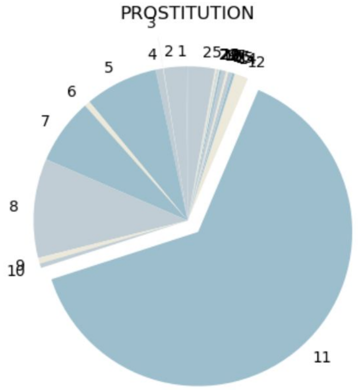


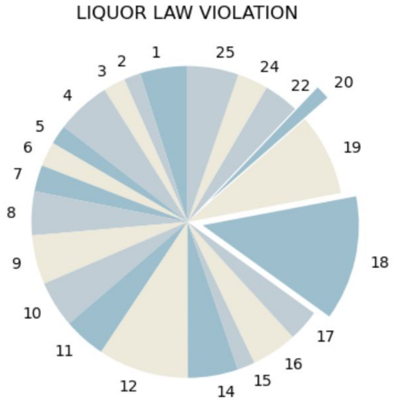
Bar Charts comparing the highest crime district to the lowest crime district (highest being district 11 and lowest being district 20)

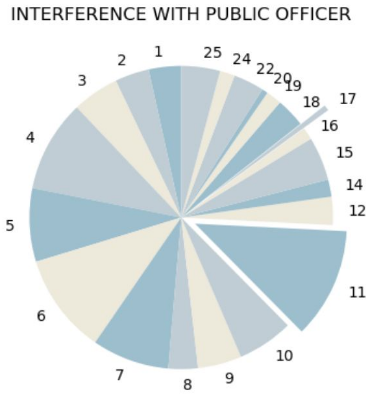


Noteworthy Pie Charts









Sentiment comparison for highest/lowest crime districts (11 and 20 respectively)

Sentiment analysis of the dataset, see for further explanation of category meanings.

Summary:

- Asian-Americans, those highly educated, and those who are younger tend to agree between the two districts, especially when it comes to the police listening to community members.
- Medium-aged individuals, especially those who are african american, tend to disagree between the two districts, especially when it comes to the police respecting community members.

T_RESPECT_AGE_LOW	2.550000
T_LISTEN_RACE_OTHER	2.731214
T_RACE_ASIAN_AMERICAN	3.588068
T_LISTEN_EDUCATION_HIGH	3.600571
T_LISTEN_RACE_ASIAN_AMERICAN	3.768857
	...
T_RACE_AFRICAN_AMERICAN	16.732652
T_RESPECT_AGE_MEDIUM	17.364643
T_RESPECT_RACE_AFRICAN_AMERICAN	17.596643
T_AGE_MEDIUM	19.114962
S_AGE_MEDIUM	19.174091

Methodology - Basic Idea

1. **Finding which columns were gaussian in our datasets**
2. **Cleaning data, One Hot Encoding**
3. **Performing ICA**
4. **Attempting PCA**

PCA (Principal Component Analysis) - Joined Public Health Statistics Data and Chicago Crime Data based on Community Area

Here are the results -

```
('Birth Rate', 'General Fertility Rate')
('No High School Diploma', 'Crowded Housing')
```

ICA - Independent Component Analysis

First we cross multiplied all values of location with all values of crime types, then aggregated this data.

	Crime	Location	Count
0	THEFT	STREET	147856
1	BATTERY	APARTMENT	128941
2	CRIMINAL DAMAGE	STREET	102442
3	MOTOR VEHICLE THEFT	STREET	94624
4	BATTERY	RESIDENCE	94580
...	...	...	...
6960	NARCOTICS	VEHICLE-COMMERCIAL - ENTERTAINMENT/PARTY BUS	0
6961	CRIM SEXUAL ASSAULT	CTA PARKING LOT / GARAGE / OTHER PROPERTY	0
6962	NARCOTICS	VEHICLE - OTHER RIDE SHARE SERVICE (LYFT, UBER...	0
6963	NARCOTICS	VEHICLE - COMMERCIAL: TROLLEY BUS	0
6964	WEAPONS VIOLATION	YMCA	0

6965 rows × 3 columns

Future Work

Exploration of Additional Data Sets:

In our ongoing project, we have successfully completed the data preprocessing and analysis phases for a subset of the available data sets. However, our work is not yet complete. Moving forward, we plan to extend our analysis to encompass the remaining data sets. These additional data sets hold valuable information that can provide a more comprehensive understanding of the underlying causal relationships within our domain.

Identifying Hidden Confounders and Theorizing Relationships:

One of our key objectives is to determine if there are any hidden confounders within the data sets. To achieve this, we will employ advanced techniques in causal inference and statistical analysis. By identifying these hidden confounders, we can unveil intricate relationships that may exist between variables, enabling us to refine our understanding of the causal mechanisms at play.

Analyzing Data Distribution and Variable Types:

In our future work, we will delve deeper into the data sets to analyze their distributions and ascertain whether the variables are continuous or discrete. This analysis will guide us in selecting appropriate methods and models for further exploration. By considering the specific characteristics of the data, we can employ tailored techniques that align with the underlying data structures.

Uncovering Independencies and Conditional Independencies:

We aim to identify independencies and conditional independencies within the data. These findings will aid in distinguishing between causally connected variables and those that are merely independent. By leveraging sophisticated statistical tools, we can uncover subtle dependencies and infer causal relationships that may have previously been hidden.

Utilizing the Do Operator and Average Causal Effect (ACE):

To further deepen our understanding of causal relationships, we will leverage the do operator and estimate the Average Causal Effect (ACE). This approach will enable us to evaluate the causal impact of interventions and shed light on the causal relationships between variables. By employing these techniques, we can move beyond observational associations and gain insights into causal mechanisms.

Applying the PC Algorithm to Infer Causal Graphs:

To approximate the "true" causal graph underlying the data, we plan to utilize the PC Algorithm. This algorithm effectively detects and constructs Directed Acyclic Graphs (DAGs) by capturing conditional independencies and exploiting graphical criteria. By applying the PC Algorithm to our data, we can infer causal relationships and establish a visual representation of the causal graph.

References

[1] Hamaker, E. L., Kuiper, R. M., \& Grasman, R. P. (2015). A Critique of the Cross-Lagged Panel Model. Psychological Methods, 20(1), 102-116.

[2] Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., \& Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry, 62(6), 593-602.

[3] Pearl, J. (2009). Causality: Models, Reasoning, and Inference. Cambridge University Press.

[4] Schuler, M. S., Putnam, K. M., \& Goldsmith, J. V. (2020). Understanding the Relationship between Physical Activity and Mental Health: A Causal Framework. Psychology of Sport and Exercise, 47, 101538.

[5] VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. Annual Review of Public Health, 37, 17-32.

[6] Office of Disease Prevention and Health Promotion (n.d.). Crime and Violence. Healthy People 2030. https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/crime-and-violence.

Thank you so much for listening

Comments? Feedback ?