Minneapolis Temperature and Crime Relation Analysis
(GROUP 4) - CSCI 5707

Akshara Madhu Suthanan
madhu047@umn.edu

Khushi Goyal
goyal136@umn.edu

Prakrit Sinha
sinha239@umn.edu

Jithendra Jagannatha Kagathi
jagan028@umn.edu

***Abstract*—To analyze if temperature has a causal relation on the rate of crime across different areas of Minneapolis, we will conduct a comprehensive study employing a rigorous research methodology. This research aims to uncover whether variations in temperature, across different seasons and locations in Minneapolis, can be identified as a potential causal factor influencing the occurrence of crime.**

I. Introduction

Crime in the United States exacts a substantial toll, with each instance carrying a significant financial burden. The repercussions extend far beyond immediate monetary costs, delving into the complex realms of mental and physical well-being. Consider the staggering figures associated with crime: a lone homicide rings up an economic invoice exceeding $8.9 million, a robbery demands over $42,000, and a burglary results in a price tag surpassing $6,000. However, the ramifications of criminal activities extend beyond the economic domain. Victims often find themselves grappling with enduring mental health challenges, with approximately 25% contending with post-traumatic stress disorder (PTSD). The psychological aftermath of crime paints a poignant picture of human suffering that transcends mere financial calculations. Moreover, the pervasive fear of crime casts a looming shadow over the general populace, inflicting additional harm on individuals' mental and physical health. This intangible cost is profound, influencing the daily lives and well-being of those who live under the specter of potential victimization. The psychological toll exacted by the fear of crime serves as a potent reminder that the impact of criminality extends far beyond the immediate victims. In the pursuit of understanding the roots of criminal behavior, researchers delve into a myriad of contributing factors. Traditionally explored dimensions include income, education, and intelligence. However, the investigative lens has widened to encompass less obvious connections, with scholars scrutinizing variables such as weather and temperature. This expanded scope underscores the multifaceted nature of crime, urging a comprehensive examination that considers both the overt and subtle influences shaping criminal behaviors in society.

II. Motivation

Efficient resource allocation for law enforcement can be achieved by leveraging weather-related crime patterns. By investigating the correlation between temperature and crime within the city, a more nuanced understanding of human decision-making in criminal activities can be attained. Analyzing crime data in relation to weather conditions allows for the identification of patterns and trends that can inform law enforcement strategies. This approach not only enhances the ability to allocate resources effectively but also contributes to a more comprehensive comprehension of the intricate dynamics underlying criminal behavior, thereby facilitating more informed and targeted crime prevention measures.

III. Dataset and tools used

- **Jupyter Notebook** - Data Preprocessing
- **Numpy** - Perform math and matrix operations
- **Pandas**- Cleaning and analyzing data
- **MySQL** - Database creation and answering questions
- **Seaborn** - Creating visualizations
- **R** - Hypothesis testing

A. Datasets

IV. Data Preprocessing

- Dropped irrelevant attributes from Crime Dataset, hence reducing columns from 24 to 10.
- Dropped null values. As they were very few in percentage it was better to remove them than using some impute technique.
- In the temperature dataset, there were no null values but for the 'Snowfall' attribute there was a value denoted by 'T' , which was making this numerical column behave as a string object. Upon further research we found out that 'T' denoted trace values: snowfall levels which were not 0, but some slight quantity not possible to measure with the devices. Hence we replaced all trace values into '0' and changed this column to type float.
- The date columns were of object type so we converted them to the appropriate date format (YYYY-MM-DD).

	Reported_Date	Occurred_Date	NIBRS_Group	NIBRS_Code	Offense_Category	Precinct	Ward	Neighborhood	Crime_Count	Time_To_Report
0	2019/01/10 13:00:00+00	2019/01/09 23:00:00+00	Non NIBRS Data	Non NIBRS Data	Subset of NIBRS Assault Offenses	4.0	5.0	Willard - Hay	1	0 days 14:00:00
1	2019/01/14 17:42:00+00	2019/01/14 17:18:00+00	Non NIBRS Data	Non NIBRS Data	Subset of NIBRS Assault Offenses	4.0	5.0	Willard - Hay	1	0 days 00:24:00
2	2019/01/26 14:15:00+00	2019/01/26 12:00:00+00	Non NIBRS Data	Non NIBRS Data	Subset of NIBRS Assault Offenses	5.0	10.0	Lowry Hill East	1	0 days 02:15:00
3	2019/02/01 16:39:00+00	2019/02/01 15:22:00+00	Non NIBRS Data	Non NIBRS Data	Subset of NIBRS Assault Offenses	4.0	5.0	Hawthorne	1	0 days 01:17:00
4	2019/02/03 23:12:00+00	2019/02/03 23:12:00+00	Non NIBRS Data	Non NIBRS Data	Subset of NIBRS Assault Offenses	3.0	11.0	Diamond Lake	1	0 days 00:00:00

Fig. 1: Minneapolis crime dataset

	Date	Maximum Temperature degrees (F)	Minimum Temperature degrees (F)	Precipitation (inches)	Snow (inches)	Snow Depth (inches)
0	2010-01-01	6.0	-9.0	T	T	9.00
1	2010-01-02	1.0	-15.0	0.00	0.00	9.00
2	2010-01-03	7.0	-14.0	0.00	0.00	9.00
3	2010-01-04	7.0	-10.0	0.00	0.00	9.00
4	2010-01-05	10.0	-9.0	0.00	0.00	9.00

Fig. 2: Minneapolis temperature dataset

Offense_Category	Precinct	Ward	Neighborhood	Date	Maximum Temperature degrees (F)	Minimum Temperature degrees (F)	Precipitation (inches)	Snow (inches)	Snow Depth (inches)	Month	reported_in	mean_temp	Year
Subset of NIBRS Assault Offenses	4	5	Willard - Hay	2019-01-09	16.0	6.0	0.0	0.0	0.0	January	0	11.0	2019
Subset of NIBRS Assault Offenses	3	9	Midtown Phillips	2019-01-09	16.0	6.0	0.0	0.0	0.0	January	0	11.0	2019
Subset of NIBRS Assault Offenses	4	4	McKinley	2019-01-09	16.0	6.0	0.0	0.0	0.0	January	0	11.0	2019
Counterfeiting/Forgery	1	7	Elliot Park	2019-01-09	16.0	6.0	0.0	0.0	0.0	January	0	11.0	2019
Assault Offenses	1	3	Downtown West	2019-01-09	16.0	6.0	0.0	0.0	0.0	January	0	11.0	2019

Fig. 3: Merged dataset

```
df_final.isnull().sum()

Reported_Date          0
Occurred_Date          1
NIBRS_Group            0
NIBRS_Code             0
Offense_Category       0
Precinct            1566
Ward                1783
Neighborhood        1647
Crime_Count            0
Time_To_Report         1
dtype: int64

df_final=df_final.dropna(axis=0) # dropped rows with null values
```

Fig. 4: Merged dataset

- After cleaning the two datasets we merged them on Date.
- We had a lot of data (from 2013), but our study interest was only for the most recent years (2019-2022). So we ordered our data and sliced the dataset for our target range.

```
df_merge_testing=pd.merge(df_final, df_temp, on='Date', how='inner')
```

```
df=df[(df['Date'] > pd.to_datetime('2019-01-01')) &(df['Date'] < pd.to_datetime('2023-01-01'))]
```

Fig. 5: Natural join code

- Furthermore, to efficiently write queries and generate plots, we engineered some new columns:
  - **Month**: We extracted the month out of the Date column and mapped it to their character value.
  - **Year**: We also extracted the year out of the Date column.
  - **Mean_Temperature**: We built a new attribute by taking the mean of high and low temperature attributes.
  - Our **time_to_report** column had date in the format of #days, HH, MM, SS, so we converted it strictly into days. If crime was reported the same day, it took a value of 0.

V. Database Design

A. ER Diagram

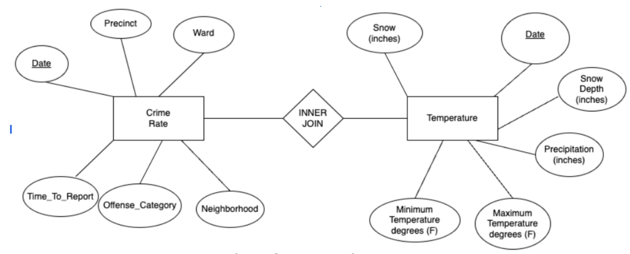

Fig. 6: ER Diagram

VI. Analysis & Key Questions

1) **List the total number of reported crimes between 2019 and 2022.**

   SELECT COUNT(*) AS $TotalReportedCrimes$
   FROM $final\_data$
   WHERE $YEAR(Date)$ BETWEEN
   2019 AND 2020;

2) **What are the 3 most common crimes reported and what percentage amount are they of the total amount of reported crimes?**

   SELECT $offense\_category$,COUNT(date) AS
   $crime\_count$, COUNT(date) * 100.0 /
   (SELECT COUNT(*) FROM $final\_data$)
   AS $crime\_rate\_percentage$
   FROM $final\_data$
   GROUP BY $offense\_category$
   ORDER BY $crime\_count$ desc
   LIMIT 3;

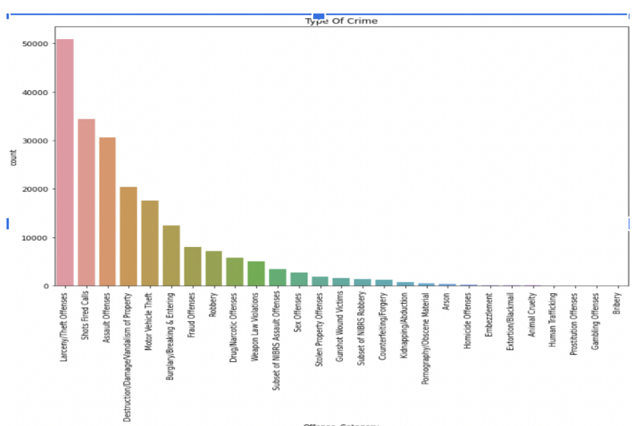

Fig. 7: Crime type countplot

3) **What are the top 5 neighborhoods that had the MOST and the LEAST number of reported crimes?**

SELECT $neighborhood$, COUNT(*) AS
$crime\_count$ FROM $final\_data$ AS fd
GROUP BY $neighborhood$
ORDER BY $crime\_count$ asc/desc
LIMIT 5;

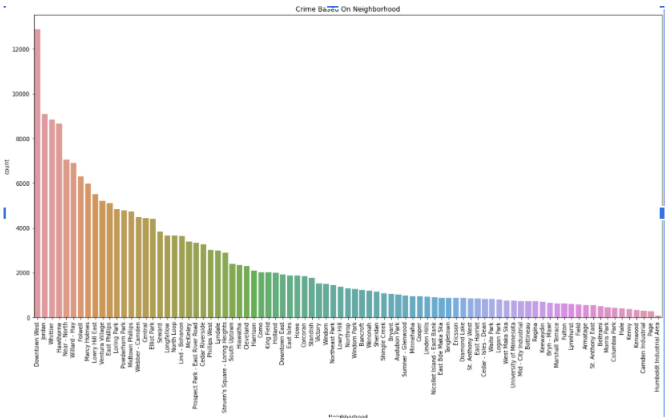

Fig. 8: Crime by neighborhood

4) **What month had the most crimes reported and what was the average temperature per month in the last four years?**

SELECT $month$, COUNT(date) AS $crime\_count$,
ROUND(AVG($mean\_temperature\_f$), 2)
as $average\_temperature_F$
FROM $final\_data$ AS fd
GROUP BY $month$
ORDER BY $month$;

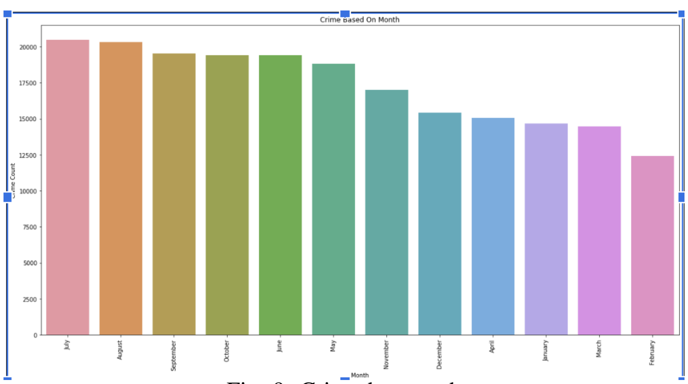

Fig. 9: Crime by month

5) **What is the yearly percentage change in the crime rate over the past 4 years?**

SELECT YEAR(Date) AS $Year$,
COUNT(*) AS $Crime\_Count$,
LAG(COUNT(*)) OVER (ORDER BY
YEAR(Date)) AS $Previous\_Year\_Crime\_Count$,
(COUNT(*) - LAG(COUNT(*)) OVER
(ORDER BY YEAR(Date))) /
LAG(COUNT(*)) OVER (ORDER BY
YEAR(Date)) * 100 AS
$Yearly\_Growth\_Percentage$
FROM $final\_data$
GROUP BY $Year$
ORDER BY $Year$;

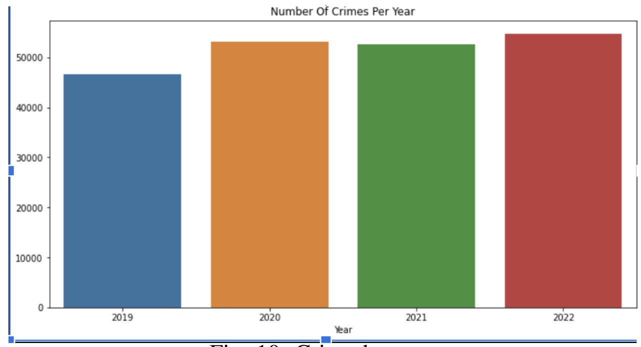

Fig. 10: Crime by year

6) **List the month and the average snowfall with the corresponding number of crimes.**

SELECT $Month$,
COUNT(*) AS $Crime\_Count$,
ROUND(AVG(COALESCE(NULLIF(‘$Snow(inches)$‘, ”), 0)),3)
AS $Average\_Snowfall$
FROM $final\_data$
GROUP BY $Month$;

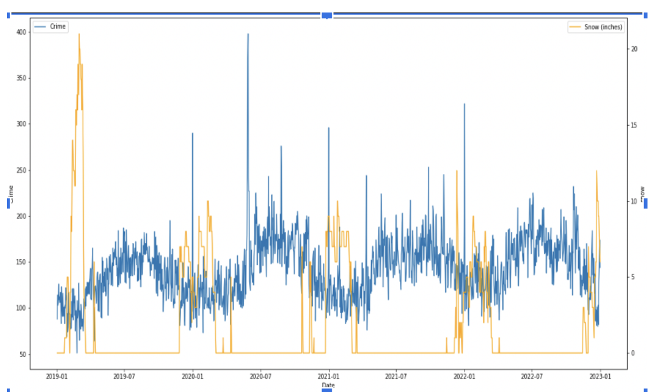

Fig. 11: Crime and snowfall lineplot

7) **Is there a relationship between crime rate and temperature? Support with a graphic.**

8) **For the year 2021, list the number of crimes reported monthly, and what was the average temperature for each month during the year?**

SELECT $Month$,
COUNT(*) AS $Crime\_Count$,
Round(AVG($Mean\_Temperature\_F$),2)

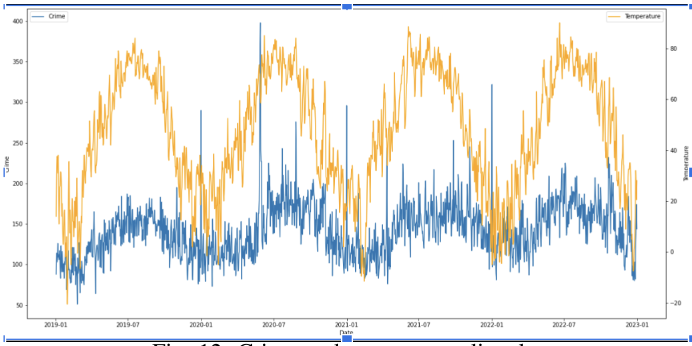

Fig. 12: Crime and temperature lineplot

AS *Average_Temperature*
FROM *final_data*
WHERE YEAR(Date) = 2021
GROUP BY *Month*
ORDER BY *Crime_Count* desc;

Fig. 13: Crime and temperature by month for year 2021

VII. Hypothesis Testing

- By now, it is pretty clear that temperature and crime rate in the city of Minneapolis are positively correlated. We will now conduct a basic hypothesis test to formalize this insight.
- As our data is cyclic in nature, if we take the whole data it would not be ideal to fit a SLR model to it. Hence, we are just conducting the test for the year 2021.
- We have transformed our data a bit for this experiment. As we want to keep it simple, we have taken monthly data for the year of 2021. Each month has a total crime count and a mean temperature for that month.
- Final data was as following:

	month	crime_count	Mean_Temperature_F
0	January	3851	22.034796
1	February	2973	12.581063
2	March	3509	40.576803

Fig. 14: Data for hypothesis testing

- We tried a SLR model with crime_count as the response and mean_temperature as the predictor. For this test:
  - **Null hypothesis**: Coefficient of mean_temperature is not statistically significant and thus = 0.
  - **Alternative hypothesis**: Coefficient of mean_temperature is statistically significant and hence != 0.

```
> model1=lm(crime_count~Mean_Temperature_F,data=df)
> summary(model1)

Call:
lm(formula = crime_count ~ Mean_Temperature_F, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-658.22 -409.99   29.06  348.59  650.21 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        3226.043    343.290   9.397  2.8e-06 ***
Mean_Temperature_F   23.195      6.351   3.652  0.00445 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 469.4 on 10 degrees of freedom
Multiple R-squared:  0.5715,	Adjusted R-squared:  0.5286 
F-statistic: 13.34 on 1 and 10 DF,  p-value: 0.004448
```

Fig. 15: Model summary

- From the figure in the last slide, we can see that both the coefficients are statistically significant.
- Let us conduct an ANOVA test to formalize this even further:

```
> anova(model1)
Analysis of Variance Table

Response: crime_count
                   Df  Sum Sq Mean Sq F value   Pr(>F)   
Mean_Temperature_F  1 2938559 2938559  13.337 0.004448 **
Residuals          10 2203376  220338                    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig. 16: ANOVA test

VIII. Assumptions for Linear Regression Analysis

Here are the four assumptions for our SLR model:

- **Linearity**: We assume that our data is linear to a high extent.

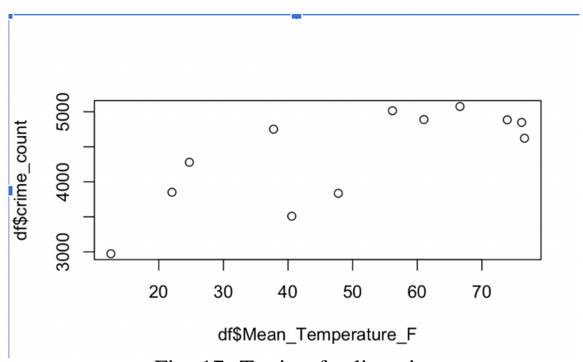

Fig. 17: Testing for linearity

- **Independence**: We assume that our residuals are independent of each other.

- **Homoscedasticity**: We assume that the residuals of our data have constant variance. We can check for constant variance by plotting residual Vs fitted values for the model. The assumption of constant variance holds if the points are distributed randomly, showing no pattern. From the plot below, we can see this is more or less the case for our data.

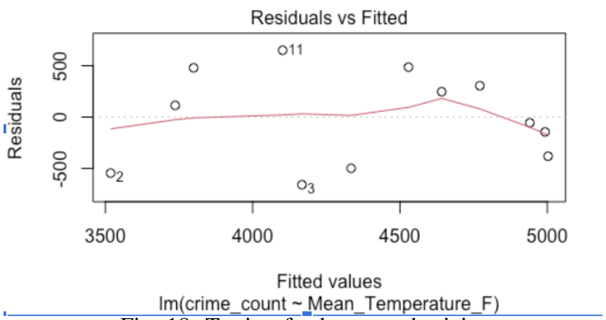

Fig. 18: Testing for homoscedasticity

- **Normality**: The residuals are assumed to be normally distributed. We can check for this by plotting a Q-Q plot. It's a plot between standardized residuals Vs the theoretical quantiles. The points plotted should fall on the Q-Q line. From the plot below, that is almost the case.

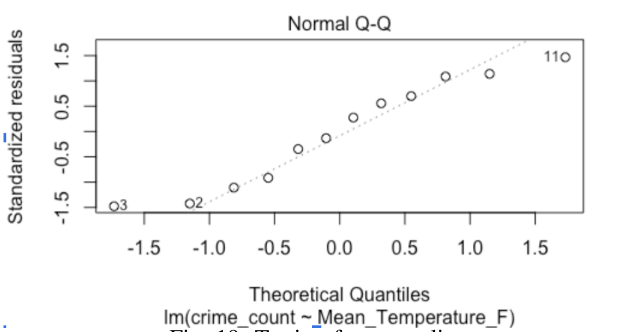

Fig. 19: Testing for normality

IX. Conclusion

The scrutiny of graphical data has unveiled a conspicuous and intriguing correlation between temperature and crime rates in Minneapolis, painting a nuanced picture of the environmental influences on criminal activity. Specifically, our analysis illuminated a discernible inverse relationship, wherein crime rates experience a noteworthy decline during periods characterized by heavy snowfall and lower temperatures. This intriguing phenomenon becomes particularly pronounced during the winter months, suggesting a compelling association between climatic conditions and the occurrence of criminal offenses within the city. The temporal alignment of reduced criminal activity with colder weather and snowfall prompts a deeper exploration into the interplay of external factors and human behavior. To substantiate the observed correlation, we employed a linear model to fit the crime rates against temperature. The outcome yielded a statistically significant predictor coefficient, affirming the robustness of the relationship between temperature variations and crime occurrences. Notably, the application of ANOVA (Analysis of Variance) provided further support by allowing us to reject the null hypothesis, which posited that temperature was not a significant factor. Instead, we embraced the alternative hypothesis, underscoring the meaningful impact of temperature on criminal behavior in Minneapolis. This empirical evidence not only deepens our understanding of the intricate dynamics influencing crime rates but also highlights the importance of considering environmental factors in comprehensive crime prevention strategies. The interconnection between climate and criminality in Minneapolis calls for a nuanced approach to public safety that recognizes the role of weather conditions as a potential influencer of societal behavior.

X. Authors & contribution

The team had organized a comprehensive plan for the successful completion of the project, with each member assigned specific tasks. Prakrit had taken on the responsibility of finding suitable datasets for study, ensuring a solid foundation for the project. Following this, Khushi was entrusted with the critical task of pre-processing the data, while Jithendra focused on treating null and missing values, as well as aggregating the data. Akshara was responsible for cleaning the data and removing duplicates, ensuring the dataset's integrity. Further down the line, Prakrit worked on removing irrelevant attributes, contributing to the refinement of the dataset. Khushi then created the ER diagram, providing a visual representation of the data structure. Moving into the analysis phase, Akshara conducted summary statistics, and Jithendra delved into data visualization and aggregation, enhancing the team's understanding of the dataset. In the later stages, Akshara identified potential patterns and relationships within the data, paving the way for hypothesis testing, a task assigned to Prakrit. Subsequently, Akshara conducted statistical analysis, adding a quantitative dimension to the team's findings. Finally, Khushi was responsible for reporting the observations, consolidating the team's efforts into a coherent and insightful project conclusion. This well-organized division of tasks ensured a systematic and collaborative approach to the project, with each member contributing their expertise to achieve a comprehensive and well-documented outcome.

Acknowledgment

We would like to extend our heartfelt gratitude to Professor **Jaideep Srivastava** for their invaluable guidance and unwavering support throughout the completion of our project. Professor Srivastava's expertise in the field of Database Systems has been instrumental in shaping our understanding and approach. Their insightful feedback and constructive suggestions have not only enriched our project but have also contributed significantly to our overall learning experience. Additionally, we would like to express our appreciation to the Teaching Assistant of CSCI 5707: Principles of Database Systems, **Jiacheng Eric Liu**. Mr. Liu's dedicated efforts and willingness to provide the necessary facilities and correct direction have been crucial to the success of our project. His responsiveness and commitment to fostering a conducive learning environment have greatly facilitated our understanding of the course material. We are truly grateful for the mentorship provided by Professor Jaideep Srivastava and the support extended by Teaching Assistant Jiacheng Eric Liu. Their contributions have played a pivotal role in our academic journey, and we feel fortunate to have had such experienced and dedicated individuals guiding us. Once again, thank you for your exceptional mentorship and support.

REFERENCES

[1] https://www.dnr.state.mn.us/climate/historical/daily-data.html?sid=mspthr&sname=Minneapolis/St%20Paul%20Threaded%20Record&sdate=2010-01-01&edate=pormain_page_content

[2] https://opendata.minneapolismn.gov/datasets/cityoflakes::crime-data/about

[3] https://medium.com/@minhta16/temperature-and-crime-whats-the-story-ffa23c1851b3

[4] https://medium.com/@WeatherDecTech/how-does-the-temperature-affect-crime-rates-fb91ff8c3167: :text=Theirtunity