Topic: Weather and Crime Data

Big Data Project

Rudrani Bhadra
Master of Data Science and
Artificial Intelligence
University of Waterloo
r2bhadra@uwaterloo.ca

Lavanya Sharma
Master of Data Science and
Artificial Intelligence
University of Waterloo
l9sharma@uwaterloo.ca

INTRODUCTION

Recent research has correlated the rise in crime rates to an important environmental factor: the weather. Experts have attempted to understand the different factors that contribute to crime. Temperature and weather conditions seem to be one of the factors that contribute to increased criminal activity. Many law officers believe that certain atmospheric conditions might play a role in increased crime rates, but we want to find out whether the evidence actually shows a correlation between weather and crime [2].

Finding the effect of weather on crime is interesting for a few reasons. It can help law-enforcement officials to prepare for the rise in violent acts that comes with a higher temperature. It can also help researchers to understand the factors that contributes to the commission of a crime. The correlation between weather and crime is also important for the environmentalists as it helps them to understand how individuals are affected by environmental changes and how increase in global warming effects crime rates [1].

AIM

Here our main aim is to analyze the correlations between weather and crime in order to get a deeper insight about how crime changes with different weather conditions.

In this project, we analyzed the correlation between weather and crime. By processing the dataset, we analyzed the correlation between the important weather conditions such as temperature, humidity and pressure and the number of incidents that take place. We also trained a linear regression model in order to predict the number of crimes given the weather conditions.

METHODOLOGY AND DATA

1 Data

We have chosen the weather datasets from Kaggle [3]. The datasets contain 5 years of data of various weather attributes like temperature, humidity, air pressure, wind direction, and wind pressure from 2012 to 2017. The data is available for 30 US and Canadian cities as well as 6 Israeli cities. Each weather attribute has its own dataset. We have chosen to work with temperature, pressure, and humidity as the main weather conditions. The temperature is measured in Kelvin and pressure is in Pascal. We have chosen to analyze the relationship between weather and crime for three US cities- Los Angeles, Chicago, and Boston. So, we have used their respective crime datasets from Kaggle [4] [5] [6].

2 Data Aggregation

We have applied the following steps to the weather and crime datasets in order to make the dataset ready to be processed and analyzed:

i) Cleaned the temperature data by filling up the null values with the average based on date.

ii) Cleaned the humidity data by filling up the null values with the average based on date.

iii) Cleaned the pressure data by filling up the null values with the average based on date.

iv) For each of the datasets, the mean temperature, humidity and pressure is calculated for each date.

v) Joined the temperature, humidity and pressure datasets into one for each city. Temperature is converted from Kelvin to Celsius.

vi) Calculated the number of incidents that occurred on a particular date in the crime datasets by first grouping by date and summing up its frequency.

vii) Joined the weather data and crime datasets based on the date.

viii) The final dataset contains the attributes: date, temperature, humidity, pressure and number of crimes.

3 Analysis

Our analysis can be divided into four parts. In the first part, we are finding out whether certain weather conditions can influence the number of incidents taking place in a city. Second, we are analyzing the number of crimes that take place on Christmas and whether it differs from the normal pattern of crime that usually occurs. Third, we are finding out the average number of crimes that occur in the winter and summer months. Fourth, we use linear regression on training and testing data and then use the model in order to predict the number of crimes that can take place given certain weather conditions in a city.

EXPERIMENT RESULTS

1 Analysis of weather and crime data

First, we will take the data of Los Angeles, Chicago, and Boston and compare the temperature of each of these cities over the period of time of 2012-2017 to the number of crimes that take place over that time.
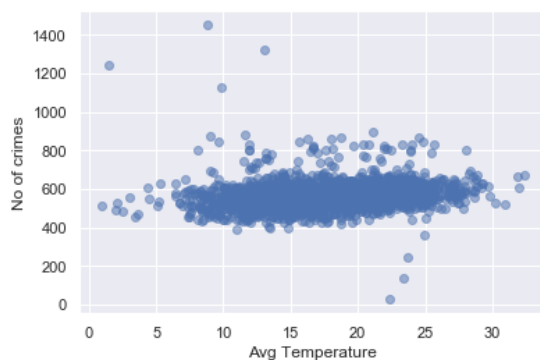
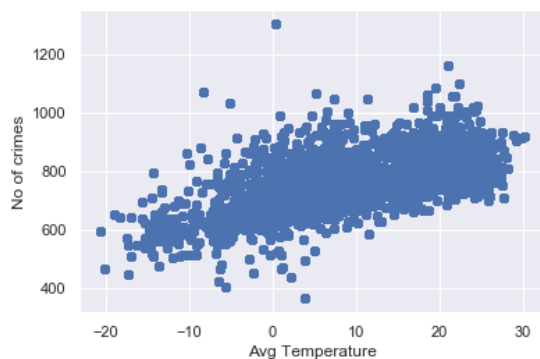**Figure 1: Relation between number of crimes and temperature in Los Angeles**

Figure 2: Relation between number of crimes and temperature in Chicago
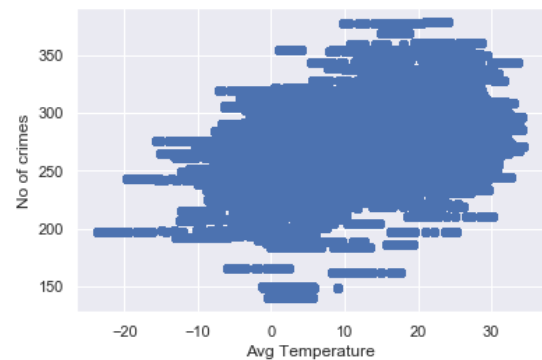

Figure 3: Relation between number of crimes and temperature in Boston

Now, we observe that in Los Angeles (Figure 1), there is hardly any correlation between temperature and number of crimes as the number of crimes remains pretty much the same as temperature increases. However, in both Chicago (Figure 2) and Boston (Figure 3), as temperature increases, the number of crimes seems to increase which indicates a positive correlation between the two. If the relationship between temperature and crimes is strongly correlated, we can observe that more extreme temperatures resulted in more number of incidents.

Next, we plotted the humidity of each city against the number of incidents.
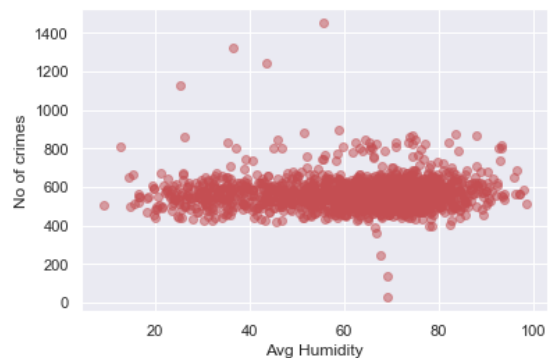

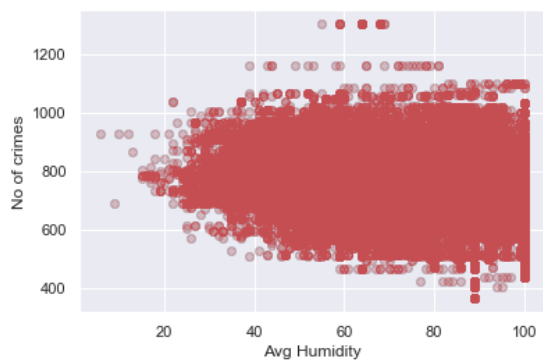**Figure 4: Relation between number of crimes and humidity for Los Angeles**

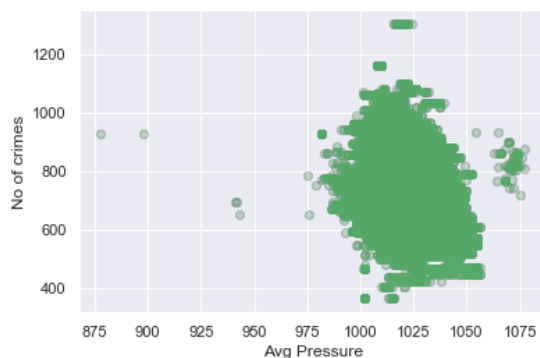**Figure 5: Relation between number of crimes and humidity in Chicago**

Figure 8: Relation between number of crimes and pressure in Chicago
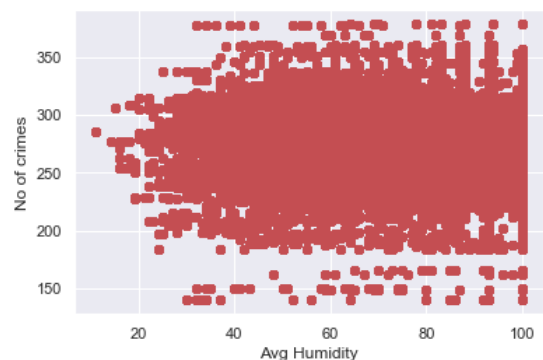

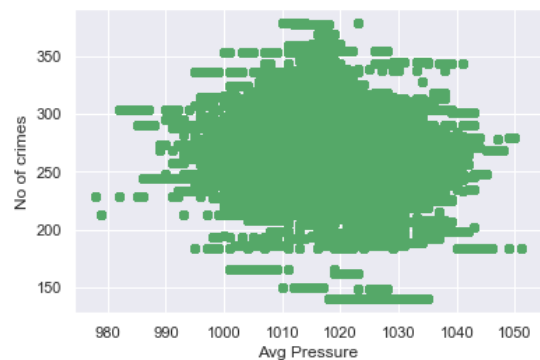**Figure 6: Relation between number of crimes and humidity in Boston**

Figure 9: Relation between number of crimes and pressure in Boston

In Los Angeles (Figure 4), there is no correlation between humidity and the number of crimes. Similarly, in Chicago (Figure 5) and Boston (Figure 6), there is not much correlation between crime and humidity although the range of crimes seems to increase slightly with an increase in humidity level.

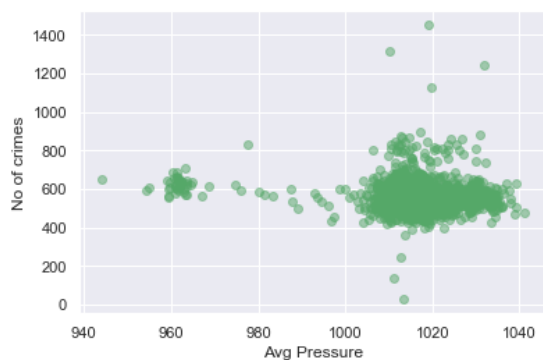Next, we plotted the average pressure of each city against the incidents.

Figure 7: Relation between number of crimes and pressure in Los Angeles

For Los Angeles (Figure 7), we notice that the range of crimes increases slightly with an increase in pressure. For Chicago (Figure 8), we see that there are hardly any crimes at the lower ranges of pressure. Crimes mostly occur when the pressure is between 1000Pa and 1050Pa. For Boston (Figure 9), we see that the range of crimes mostly remains the same throughout pressure between 960Pa and 1040Pa.

From the above analysis, we can infer that out of temperature, humidity and pressure, temperature has the most effect on the number of crime incidents while humidity has the least.

Finally, we have also plotted graphs showing the number of crimes that occur during different weather descriptions.
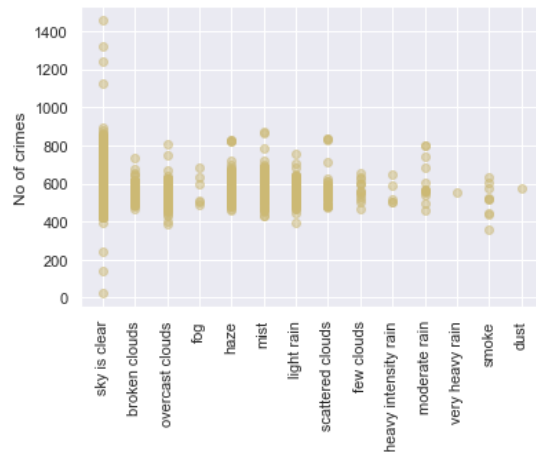
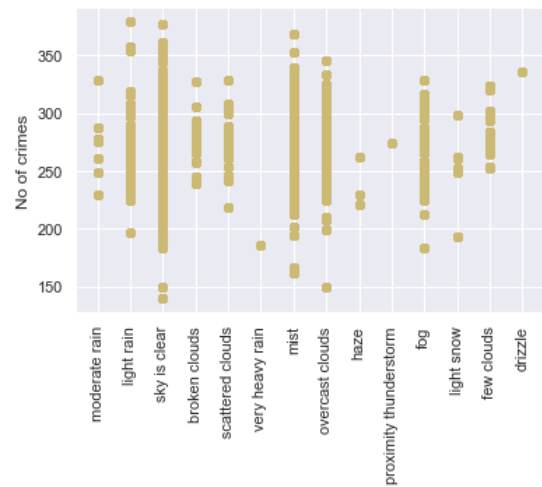**Figure 10: Relation between number of crimes and different weather descriptions in Los Angeles**

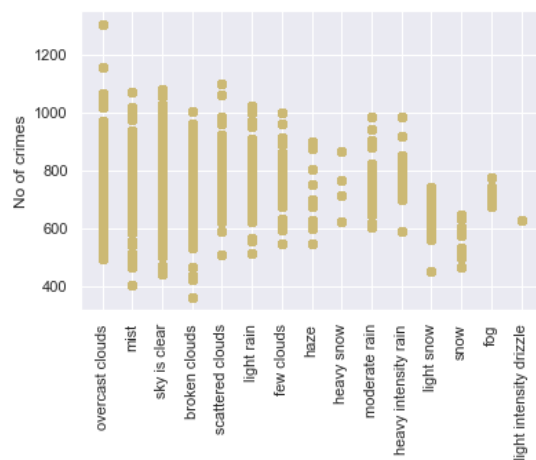**Figure 11: Relation between number of crimes and different weather descriptions in Chicago**

Figure 12: Relation between number of crimes and different weather descriptions in Boston

For Los Angeles (Figure 10), the range of crimes that occur is maximum when the sky is clear (0-1400). For Chicago (Figure 11), the range of crimes seems to be fairly similar given the different weather conditions (around 400-1000). For Boston (Figure 12), the range is high during clear skies, mist, and overcast clouds (around 150-350).

Overall, we see that the maximum number of crime incidents is around 1500 for Los Angeles, 1300 for Chicago and 400 for Boston. So, Boston seems to be the safest city out of all three cities.

2   Analysis of crime during Christmas

Here, we wanted to observe whether the mean rate of crime incidents that occur during the Christmas week is similar to the mean rate of incidents that occur throughout the given dates of other parts of the year. In this part, we generate the linear regression plot between each of the weather conditions (temperature, humidity, and pressure) and the number of crimes along with the mean incident value. In the same figure, we also plot between each of the weather conditions (temperature, humidity, and pressure) and the number of crimes along with the mean incident value only during the Christmas week (highlighted in red).
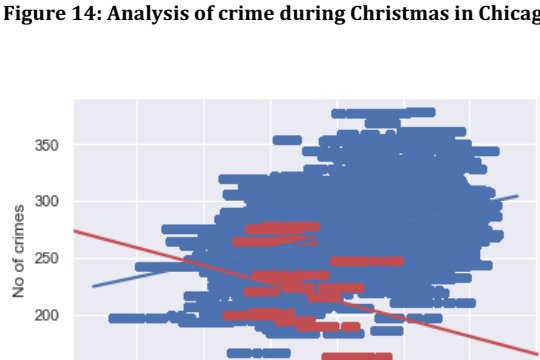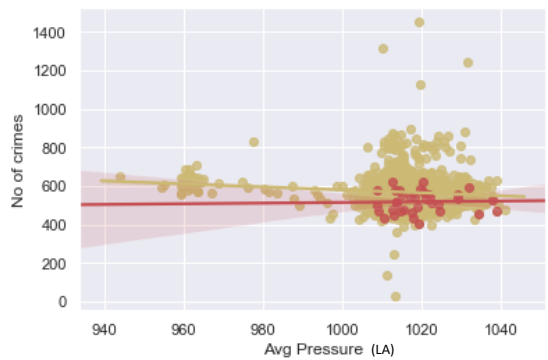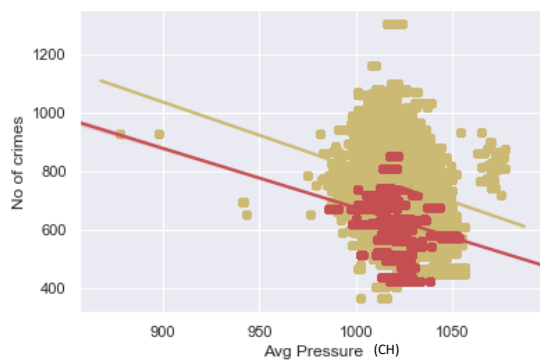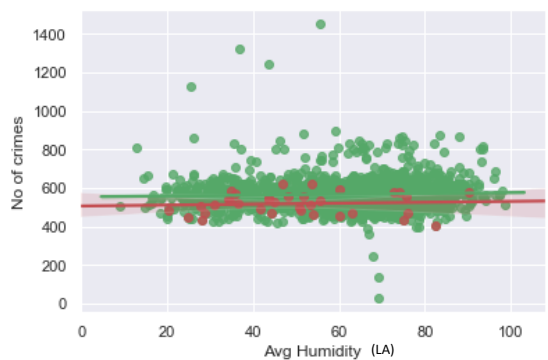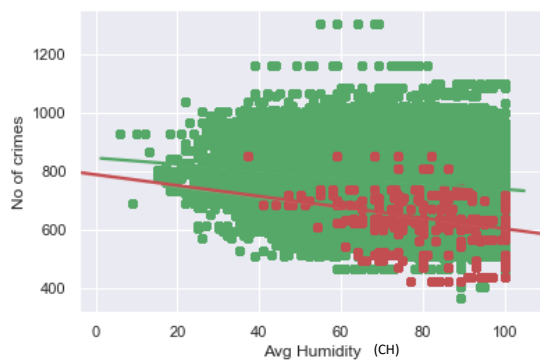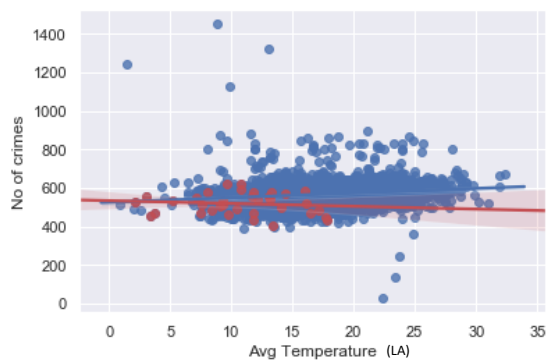
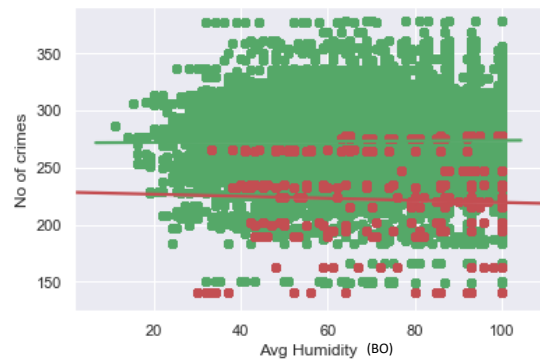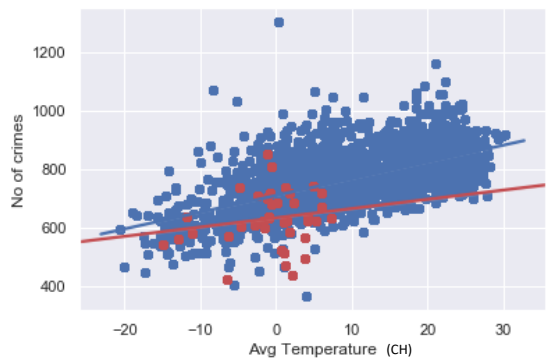Figure 13: Analysis of crime during Christmas in Los Angeles

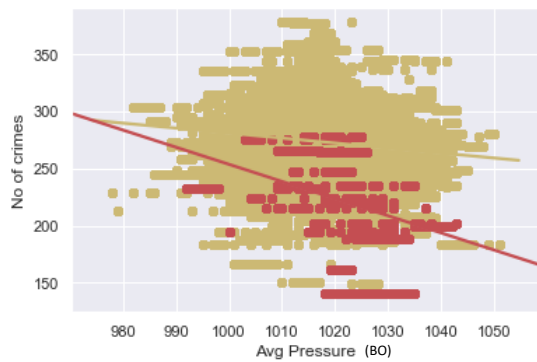Figure 14: Analysis of crime during Christmas in Chicago

Figure 15: Analysis of crime during Christmas in Boston

The mean for the entire period is shown in

i) blue line when plotted against temperature. The blue dots represent the incidents during the entire year.

ii) green line when plotting against humidity. The green dots represent the incidents during the entire year.

iii) yellow line when plotting against pressure. The yellow dots represent the incidents during the entire year.

while the mean for the Christmas week is shown in the red line and the red dots indicate incidents during the Christmas week.

Overall from all the plots above, we can infer that the mean rate of crime incidents is lower during Christmas week than other parts of the year which means that people are busy celebrating with their families in their homes or they might just be busy in the holiday season or spending a lot of time indoors which reduces the overall crime rate.

3 Analysis of crime during summer and winter months

In this section, we have analyzed the general trend of crimes in the winter and summer months and calculate the average temperature.

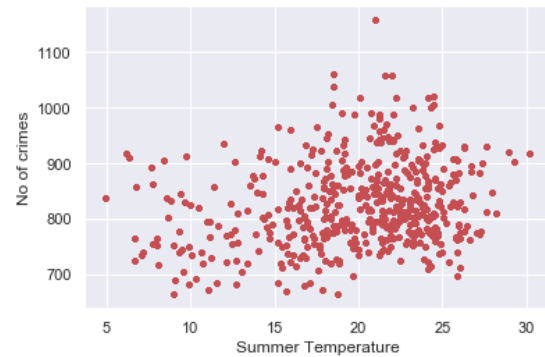**Figure 16: Analysis of crime during summer months in Los Angeles**

Figure 17: Analysis of crime during summer months in Chicago

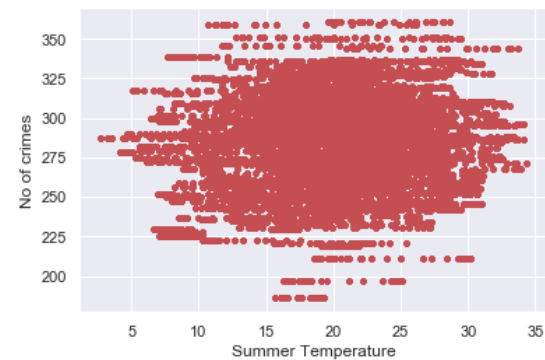**Figure 18: Analysis of crime during summer months in Boston**

For Los Angeles (Figure 16) and Chicago (Figure 17), the number of crimes seem to increase slightly with the increase in temperature while for Boston (Figure 18), there does not seem to be much correlation between the temperatures during the summer months and the number of crimes.
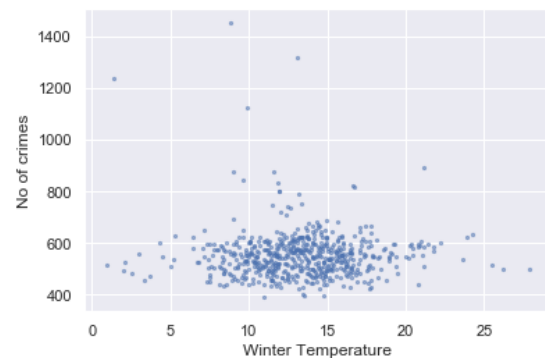

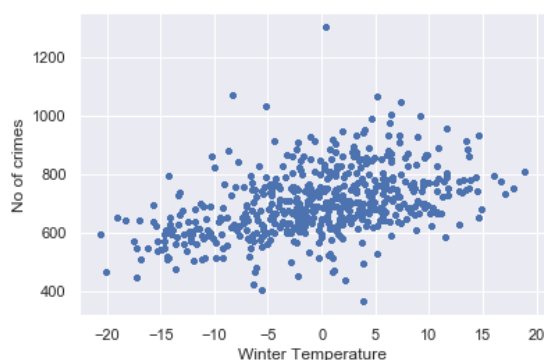**Figure 19: Analysis of crime during winter months in Los Angeles**

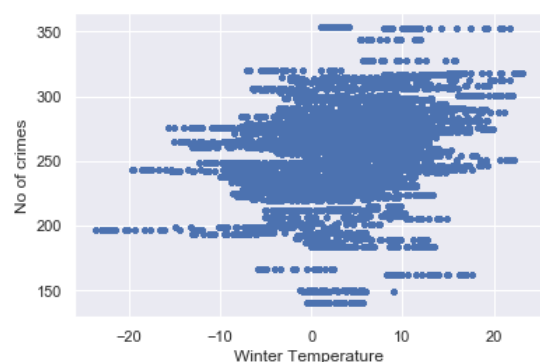**Figure 20: Analysis of crime during winter months in Chicago**

Figure 21: Analysis of crime during winter months in Boston

For Chicago (Figure 20) and Boston (Figure 21), the number of crimes tend to increase with an increase in temperature during winter.

Surprisingly, for Los Angeles (Figure 19) and Chicago (Figure 20), the maximum number of crime incidents is approximately 1500 and 1300 respectively, each which is higher than the maximum of number of crimes during summer (900 for Los Angeles from Figure 16 and 1200 for Chicago from Figure 17), although they are rare occurrences.

City	Average number of crimes in summer	Average number of crimes in winter
Los Angeles	582.089	553.678
Chicago	826.456	700.132
Boston	282.583	258.904

Table 1: Average number of crime incidents during summer and winter months in Los Angeles, Chicago and Boston

From Table 1, we can observe that the average number of crimes is more in the summer months than in the winter months in all three cities. There can be two reasons for this [2]:

i) Warm weather may create more opportunities for crime since in warmer weather, people tend to spend time outdoors and like to indulge themselves in outdoor activities. When people spend more time outside, they will usually interact with other people. More interactions provide a higher chance of criminal activity or encounter. Warm weather brings the victims and criminals all in the same place.

ii) Hot weather also makes people more irritable, impulsive and aggressive which can make them lose their temper easily and in turn can cause them to commit crimes.

4 Prediction by Linear Regression

To develop a model to do prediction, we used a linear regression model to predict the number of crimes given certain weather details.

The model uses three parameters: temperature, humidity and pressure, all of numeric type. The target variable is the number of crime incidents.

The model is trained as follows:

For Los Angeles:

$$Incidents = 1001.826 + 2.2 * temperature + 0.2389 * humidity - 0.48 * pressure$$

For Chicago:

$$Incidents = 1099.413 + 5.49 * temperature - 0.605 * humidity - 0.336 * pressure$$

For Boston:

$$Incidents = 413.721 + 1.22 * temperature - 0.032 * humidity - 0.15 * pressure$$

Therefore, using this model, we can predict the approximate number of crimes which might occur given the temperature, humidity and pressure level in one of the following cities – Los Angeles, Chicago or Boston.

CONCLUSION

Based on the analysis we did above, we can conclude the following interesting phenomena:

• In Los Angeles, there isn't much correlation between crime and weather.

• In Chicago and Boston, an increase in temperature seems to increase the number of crime incidents.

• Humidity does not seem to have much correlation with the number of crimes taking place for all three cities.

• In Chicago, most crimes take place where pressure is in the higher ranges (1000Pa-1050Pa).

• During Christmas week, the mean rate of incidents is lower than the overall mean rate of incidents that happen over the year so it is a bit safer to go outdoors during that period.

• Summer months have higher amount of crime counts than winter months. Probably because people get irritable and aggressive in hot weather and act impulsively.

• Overall, Boston seems to be the safest city out of Chicago and Los Angeles.

REFERENCES

[1] Weather Group Television, LLC (2018), Do Certain Weather Conditions Increase Crime Rate?, Retrieved on 25th November, 2019 from https://weloveweather.tv/weather-crime/

[2] ZME SCIENCE (2019), Weather and Crime: is there a connection?, Retrieved on 25th November, 2019 from https://www.zmescience.com/science/weather-crime-connection-04234/#WHY_A_CONNECTION_BETWEEN_WEATHER_AND_CRIME_MATT ERS

[3] https://www.kaggle.com/selfishgene/historical-hourly-weather-data

[4] https://www.kaggle.com/cityofLA/crime-in-los-angeles

[5] https://www.kaggle.com/currie32/crimes-in-chicago

[6] https://www.kaggle.com/ankkur13/boston-crime-data