

Impact of Covid-19 on Domestic Violence in the Kansas City Metro

Korby A. Trautman

Northwest Missouri State University, Maryville MO 64468, USA
s543056@nwmissouri.edu and korby.trautman@hotmail.com

Abstract. The Covid-19 Pandemic has effected almost every citizen of the planet. Unfortunately, many of the negative impacts of the pandemic (outside of the viral infection itself) have yet to be fully discovered, as many nations are now returning to a somewhat normal existence. Through personal conversations with local police and social service workers, it was determined that one of the potential impacts could revolve around Domestic Violence crimes; as individuals were being forced to stay at home during the height of the pandemic. Utilizing Azure Machine Learning Pipelines, Covid-19 and Crime data from the Kansas City Metro region was analyzed for any correlation. It was determined that a very weak correlation occurred between Monthly New Covid-19 case counts and Domestic Violence crimes, with a Coefficient of Determination of 0.405, however, the large majority of the models ran showed a very low correlation between the two variables.

Keywords: Covid-19, · Crime · Domestic Violence · Kansas City

1 Introduction

On 20 January 2020, the first laboratory-confirmed Covid-19 case was discovered in the U.S [3]. Within 53 days, Covid-19 was declared a national emergency. Both federal and state governments began implementing various shutdowns and social distancing measures to help prevent the spread of the virus. Restaurants and bars would shut down, businesses would require their employees to work from home, and schools and universities quickly switched to virtual models; all with the intent to decrease the spread of the deadly virus by keeping individuals at home. With all the good intentions of keeping the virus at bay, however, another potential threat was presented when individuals were staying home. The safety of individuals that were now being required to stay at home was harder to monitor. For many victims of domestic violence, work or school would be considered a safe-haven. Teachers are trained to identify sign of neglect in children, and to keep an open eye for these signs when interacting with their students. Many employers also train their employees to look for potential signs of neglect when working with their colleagues. With the lockdowns brought on by the pandemic, many of these victims no longer had the protection of potential witnesses that they would interact with every day. In many countries, the rate of domestic violence

increased substantially in the first few weeks of the pandemic. For example, in the U.K, a specific domestic abuse hotline recorded a 25 percent increase in calls during the first few weeks of lockdowns and various social restrictions [4].

The goal of my research is to analyze Covid-19 and crime data for the Kansas City Metropolitan area, and identify if there is any significant correlation between Covid-19 and crime data, with a focus on Domestic Violence crimes.

2 Methodology

The analysis of Covid-19 and Crime data can be broken into 4 stages, shown in ??.

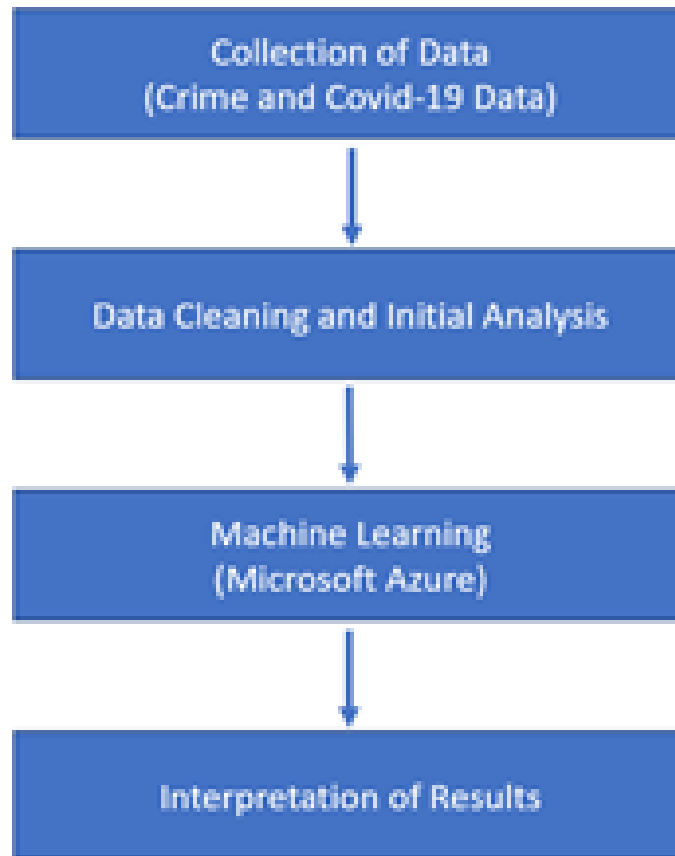


Fig. 1. Stages of analysis

Stage 1 involves the Collection of Data. For this specific study, the data will be limited to the Kansas City Metropolitan area for all of 2019-2021. Stage 2 involves data cleansing, and initial investigations with basic data visuals and calculations within Microsoft Excel. Stage 3 involves running various machine learning modules using Microsoft Azure. Finally, Stage 4 involves the final interpretation of the results.

3 Collection of Data

For this study, four data sets were used from the OpenDataKC project [1]. The OpenDataKC project is a repository of various databases open to the public, providing raw data as well as visuals and data analysis on various topics related to the city (i.e. Budget data, surveys, Covid-19 Data, Crime Data) The first three data sets consisted of 2019 [7], 2020 [8], and 2021 [9] Crime data. The final data set consisted of Covid-19 Data from 2020-present [6]. All data was accessible via a .csv file download.

4 Data Cleansing and Initial Analysis

The first stage of the data cleansing process involved the identification of the important features. Table 1 identifies the features used from the Covid-19 Data set, while Table 2 identifies the features used from the 2019, 2020, and 2021 crime data sets.

Table 1. Covid-19 Data Features

Feature	Description
Date	Specific Date
Total Cases	Total active cases on the given date
New Cases	New cases reported on the given date

Table 2. Crime Data Features

Feature	Description
Report _{NO}	Unique crime report number
Reported _{date}	Specific date of crime
Offense	Offense category (i.e Domestic Assault)
Description	Further description of crime (i.e Domestic assault with a deadly weapon)
Address	Address of Crime/Arrest/Witness
City	City of Crime/Arrest/Witness
Zip Code	Zip code of Crime/Arrest/Witness
DVFlag	"Domestic Violence" flag
Involvement	Specific line is related to either victim, suspect, or witness

The data was then cleansed to remove duplicate lines and to correct various data entry errors. The Covid-19 Data did not require any cleansing, however, the three crime data sets required multiple steps of cleansing. The first step was to remove duplicate lines created by Involvement feature. If a crime consisted of one victim, one suspect, and one witness, the data was repeated three times in order to capture the address of each individual. For this study, it was decided that only the victims (Involvement feature) would be kept in the final data set. In addition, if multiple victims were reported, then each victim line would be considered unique and included. Another major cleansing that occurred involved the City feature. Many of the cities were spelled incorrectly. For example, the city of Lee's Summit was repeated multiple times with various spellings, including Lee Summit, Lees' Summit, and Lees Summit. Additionally, the Covid-19 data only included data for the KC Metro area. The crime data had multiple lines that occurred outside of the KC Metro area. For example, a crime may have had two victims; one with an address of Maryville, while the other had an address of Overland Park. Based off of the cities within the KC Metro area [2], the Maryville case would be excluded from the final data set. The last data cleanse that occurred involved the DVFlag feature. The 2019 and 2021 data sets had either a Yes, No, or Unknown values, while the 2020 data set had it listed as either True, False, or Unknown. The data was aligned so that the values were only listed as True, False, or Unknown. Once the data was cleaned, initial analysis began. Using simple Pivot tables within Microsoft Excel, the data was broken out to view the general crime rates compared to the Covid-19 rates (both total and new cases and deaths). This not only allowed for an initial look at potential features to be loaded into the Machine Learning process, but allowed for a quick overview to see if any major discrepancies occurred with the cleaning of data. The data remained consistent across the entire time period for both Covid-19 cases as well as crime data.

The final step before loading the data into a Machine Learning pipeline involved filtering out the unwanted data. Only victim, Domestic Violence Flag = True, cases occurring in the KC Metro area were included in the final dataset.

Raw data and LaTeX files for this paper can be found in the appropriate GitHub repository ¹.

5 Machine Learning

The final dataset was then uploaded to Microsoft Azure to be ran through a Machine Learning Pipeline (see Figure 2). The data was split into an 80:20 Training and Test set. The training set was then trained using either a Linear Regression or a Neural Network Regression Machine Learning model. The data was then scored and evaluated.

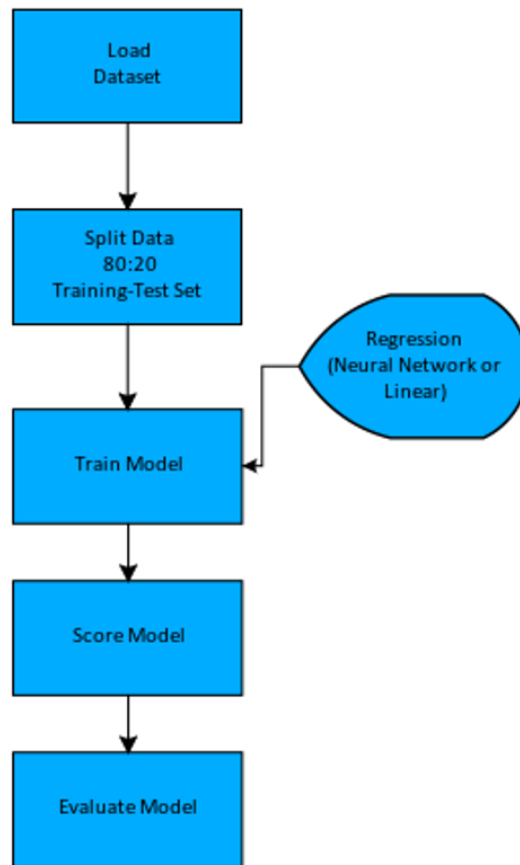


Fig. 2. Machine Learning Pipeline

¹ <https://github.com/ktrautman/CAPSTONE>

Three interpretations of data were loaded into the system. Interpretation A (Table 3) contained a breakout of the data on a daily level, and was analyzed using a Linear Regression. This same data was then ran through a Neural Network Regression (Table 4)

Table 3. Interpretation A - Daily Cases - Linear Regression

Feature	Coefficient of Determination	Mean Absolute Error	Root Mean Squared Error
New Count	0.164	6.82	8.256
Total Count	0.336	5.879	7.355

Table 4. Interpretation A - Daily Cases - Neural Network Regression

Feature	Coefficient of Determination	Mean Absolute Error	Root Mean Squared Error
New Count	0.223	6.16	8.064
Total Count	0.372	5.71	7.249

Interpretation B (Table 5) contained a breakout of all of the data on a monthly level.

Table 5. Interpretation B - Monthly Cases - Linear Regression

Feature	Coefficient of Determination	Mean Absolute Error	Root Mean Squared Error
New Count	0.405	115.573	135.1
Total Count	0.257	121.034	150.877

Interpretation C (Table 6) contained a breakout of all of the daily cases, however, only from 14March2020 forward, since this was the date the first Covid-19 case was recorded in the KC Metro area.

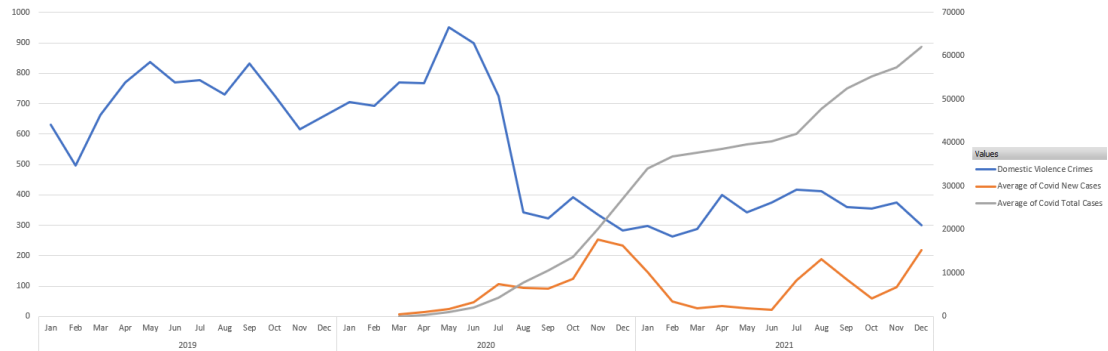
Table 6. Interpretation C - Daily Cases Post 14March2020 - Linear Regression

Feature	Coefficient of Determination	Mean Absolute Error	Root Mean Squared Error
New Count	0.063	6.693	8.736
Total Count	0.087	6.614	8.626

6 Interpretation of Results

Based on the models ran, no significant correlation existed between the number of Covid-19 cases and the number of Domestic Violence Crimes that occurred in the Kansas City Metro area. The highest Coefficient of Determination occurred when analyzing Monthly New cases (Interpretation B), with a value of 0.405. This is a fairly weak value, and does not constitute a strong correlation between the two values. When viewing just dates with Covid-19 cases (Interpretation B), there is a much weaker correlation, with a Coefficient of Determination of 0.063 and 0.087 for New and Total counts, respectfully.

When graphing the data, it does appear that there is a slight negative correlation between the Covid-19 cases crime data (Figure 3).

**Fig. 3.** Monthly Cases vs. New and Total Covid-19 Case Counts

However, the models do not show a significant correlation between the two. When viewing the data on a daily level instead of monthly, the negative correlation mentioned above is much less noticeable (Figure 4 and 5).

7 Limitations

One of the most significant limitations experienced during this project was the lack of available data in regards to Covid-19 data for the Kansas City Metro Area. Case and Death counts were readily available, however, one of the initial goals was to determine if lockdown restriction protocols had a significant impact

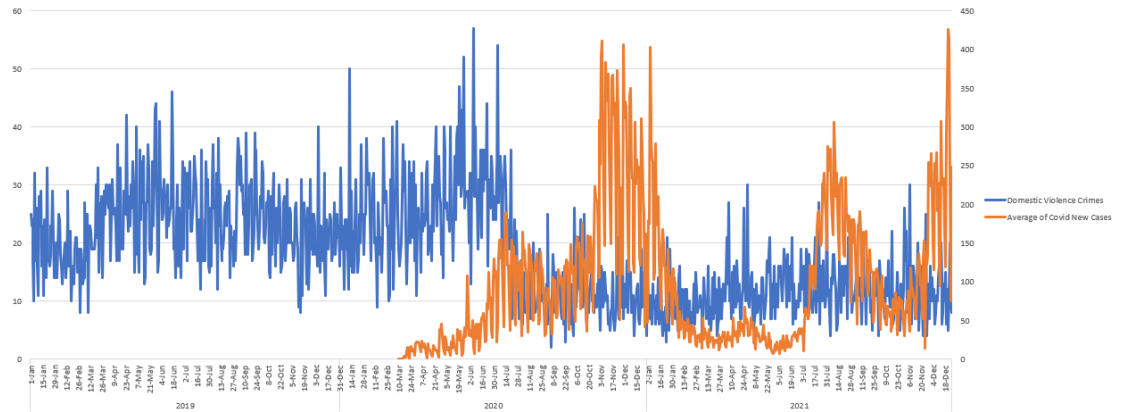


Fig. 4. Daily Cases vs. New Covid-19 Case Counts

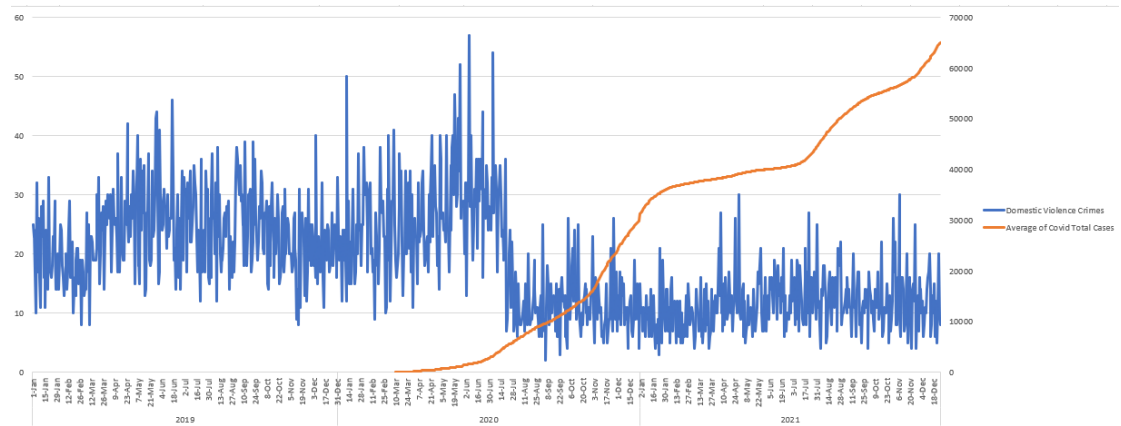


Fig. 5. Daily Cases vs. Total Covid-19 Case Counts

on Domestic Violence cases. The only lockdown protocols that could be found were for a four week period, and then the information was no longer recorded. In addition, the overall data quality for the Kansas City COvid-19 data was fairly poor. Large amounts of data had to be cleaned in order to align the data so it could be used. Typos and inconsistency in features made it initially difficult to identify what the data was actually saying. It was identified that the largest portion of data that needed to be cleaned involved the 2020 data. The 2019 and 2021 data required little work, however, the 2020 data set was in a much poorer state of condition. It would be interesting to investigate if Covid-19 had a large impact on the data quality that was entered into the system, since the police force was heavily impacted by Covid-19 [5].

Additional investigation was going to be done on other types of crime outside of Domestic Violence, however, the description of the crimes were not aligned between the various year data sets. If the researcher for this project had a legal education or background, they would have been inclined to group the data based on the crimes performed.

8 Conclusion

Overall, there is not a significant statistical correlation between Covid-19 cases and Domestic Violence Crimes in the Kansas City Metro area. All models created show a Coefficient of Determination below 0.5, constituting weak to little correlation. Future investigations would look at more features, including lockdown protocols and a deeper investigation into different crimes that occurred. Additional jurisdictions could be included as well. Even though ample information has been made public regarding Covid-19 data, the pandemic is still in its infancy, and much more detailed accounts may be available at a later time.

References

1. Open data kc, <https://data.kcmo.org/>
2. Overview of the kansas city area (metro area), <https://statisticalatlas.com/metro-area/Missouri/Kansas-City/Overview>
3. Cdc museum covid-19 timeline (Jan 2022), <https://www.cdc.gov/museum/timeline/covid19.html>
4. Bradbury-Jones, C., Isham, L.: The pandemic paradox: The consequences of covid-19 on domestic violence. *Journal of Clinical Nursing* **29**(13-14), 2047–2049 (2020). <https://doi.org/10.1111/jocn.15296>
5. Jennings, W.G., Perez, N.M.: The immediate impact of covid-19 on law enforcement in the united states. *American Journal of Criminal Justice* **45**(4), 690–701 (2020). <https://doi.org/10.1007/s12103-020-09536-2>
6. Kansas City, M.H.D.: Covid-19 case amp; death trends by date: Open data kc (Apr 2022), <https://data.kcmo.org/Health/COVID-19-Case-Death-Trends-by-Date/nfta-sjx6>
7. Technology, K.I.: Kcpd crime data 2019: Open data kc (Mar 2021), <https://data.kcmo.org/Crime/KCPD-Crime-Data-2019/pxaa-ahcm>

8. Technology, K.I.: Kcpd crime data 2020: Open data kc (Sep 2021), <https://data.kcmo.org/Crime/KCPD-Crime-Data-2020/vsgj-uufz>
9. Technology, K.I.: Kcpd crime data 2021: Open data kc (Jan 2022), <https://data.kcmo.org/Crime/KCPD-Crime-Data-2021/w795-ffu6>