Analyzing the Impact of COVID-19 Lockdown on Crime Numbers in Vancouver Using SARIMA and XGBoost Time Series Models

Hanze Wang 39408695

Yuchen Fan 54508932

Abstract

This study examines crime trends in Vancouver, Canada, using a monthly time series dataset spanning over two decades. The primary goal is to evaluate the impact of the COVID-19 lockdown on crime numbers through a counterfactual approach, comparing predicted crime numbers with actual observations. Using SARIMA and XGBoost models, the analysis also leverages several meaningful control variables to improve prediction accuracy.

Both models demonstrated strong predictive performance during validation, with Mean Absolute Percentage Errors (MAPE) of 9.95% for SARIMA and 8.94% for XGBoost. Following the implementation of the lockdown in March 2020, a counterfactual analysis revealed significant deviations between predicted and actual crime rates, with average percent differences of -25.31% (SARIMA) and -44.95% (XGBoost). These substantial differences suggest a strong association between the lockdown and decreased crime numbers.

Keywords

Crime number, Vancouver, COVID-19 Lockdown

Data

This study utilizes a variety of datasets spanning multiple socioeconomic and environmental factors to analyze crime trends in Vancouver. The selection of control variables was informed by suggestions from the FBI's guidelines on variables affecting crime, which include factors such as economic conditions, law enforcement presence, climate, and population dynamics (Federal Bureau of Investigation, 2011). The datasets used contained some missing values, which were addressed using forward-fill techniques to ensure continuity and consistency in the time series. The combined dataset spans the period from January 2003 to September 2024, with variables aggregated monthly. The key variables, their descriptions, and sources are summarized below:

Main Variable:

Crime Numbers: Daily recorded crime numbers in the City of Vancouver from 2003 to 2024, aggregated monthly for this study.

Source: Vancouver Police Department Open Data Portal (https://geodash.vpd.ca/opendata/#).

Control Variables:

Minimum Wage (2000-2024, annual updates): Provincial minimum wage in Canadian dollars.

Source: Canadian Government Minimum Wage Portal (https://minwage-salairemin.service.canada.ca/en/customsearch.html).

Consumer Price Index (CPI) (2000-2024, monthly): National index for inflation measurement.

Source: Statistics Canada (https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000601).

Unemployment Rate (2002-2020, monthly): Metro Vancouver unemployment rate, percentage.

Source: Statistics Canada (https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410029502).

Police Officers per 100,000 Population and Weighted Clearance Rate (2000-2023, annual): Indicators of law enforcement strength and crime resolution rates.

Source: Statistics Canada (https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510007701).

Temperature (1925-2024, daily): Daily temperature data for Vancouver from VANCOUVER HARBOUR climate station, aggregated monthly by using monthly average.

Source: Environment Canada (https://climate.weather.gc.ca/climate_data/daily_data_e.html).

Methodology

This study employs a counterfactual approach using SARIMA and XGBoost models to predict monthly crime numbers and assess the impact of the COVID-19 lockdown implemented in March 2020. The models compare predicted crime rates with actual observations to identify deviations attributable to the lockdown.

To ensure robust evaluation, the dataset was divided as follows:

1. Training Data (Pre-2019): Data from January 2003 to December 2018 were used to train the models.
2. Validation Data (2019): Data from January 2019 to December 2019 were used to validate the models and assess their performance prior to testing.
3. Testing Data (2020 and beyond): Data from January 2020 onwards were reserved for testing, including the lockdown period.

During validation, predictions were made using the validation data and compared to actual crime numbers to evaluate performance. Since we don't have many data points, after validation, both models were retrained using the full training dataset (January 2003 to December 2019) to generate predictions for the test period (January 2020 onwards). Finally, predictions for the test period were compared against actual crime numbers post-March 2020 using MAPE and average percent difference to quantify deviations. These deviations serve as indicators of the lockdown's impact on crime trends.

Descriptive Statistics:

Figure 1: Variable Definitions and Summary Statistics

Variables	N	Minimum	Maximum	Mean	Std. Deviation
Main Variable					
crime_number	261	2316	5297	3434.01	672.74
Control Variable					
MinWage	261	8	17.4	10.7	2.96
CPI	261	102.1	161.3	125.58	15.71
temperature	261	2.3	21	11.42	5.35
Police officers per 100,000 population	261	183.8	234.4	204.14	15.4
Weighted clearance rate	261	22.74	30.85	26.33	2.26
Vancouver Unemployment Rate	261	3.2	14.2	6.16	1.57

Figure 2: Crime Number Over Time

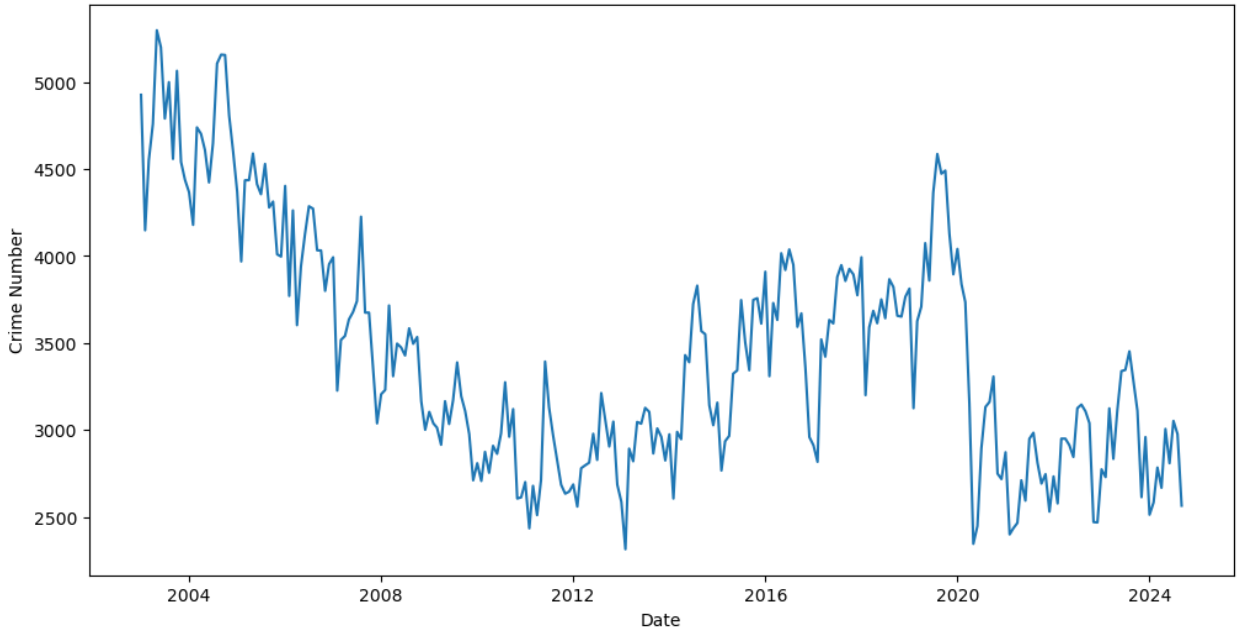
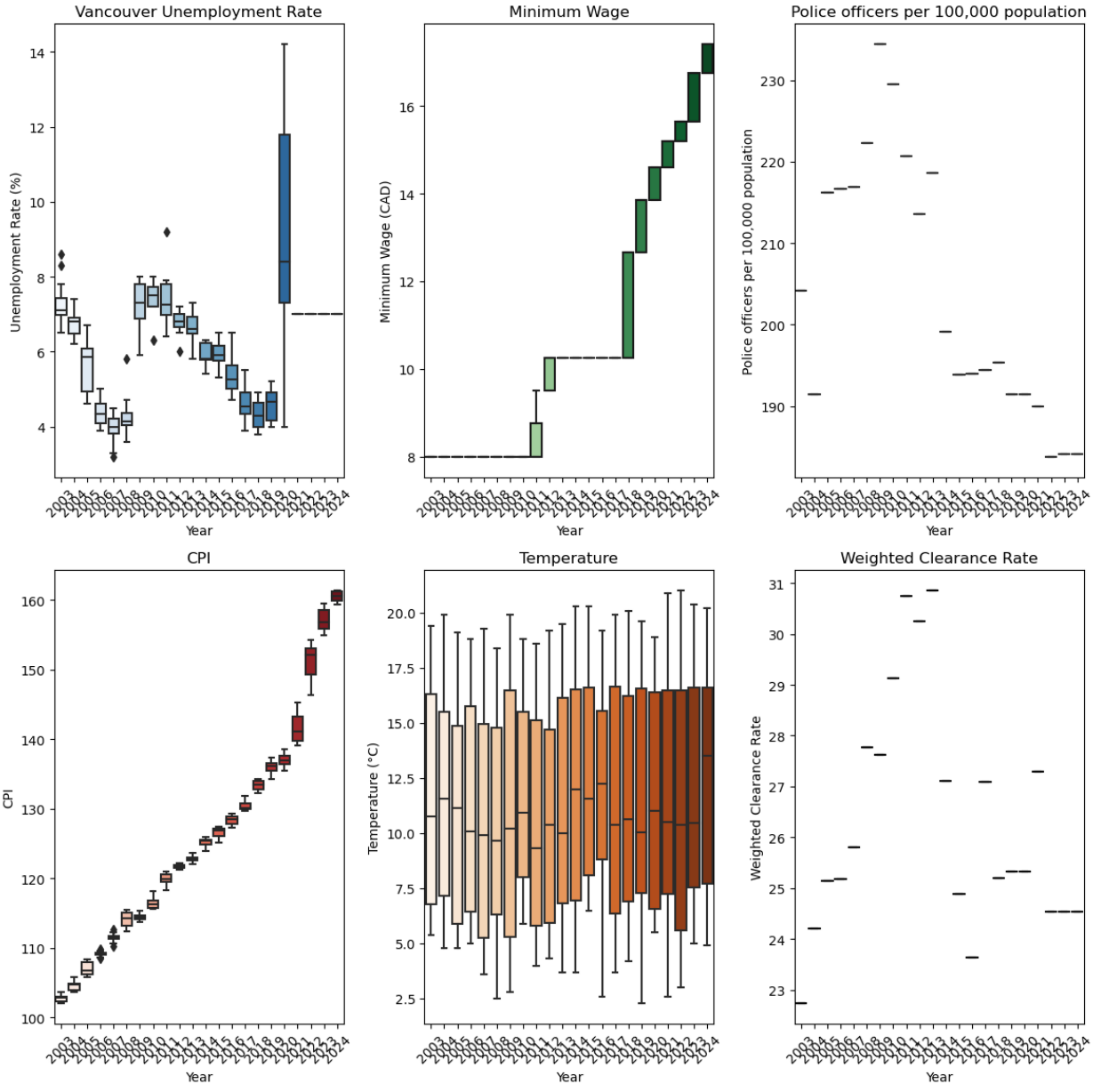
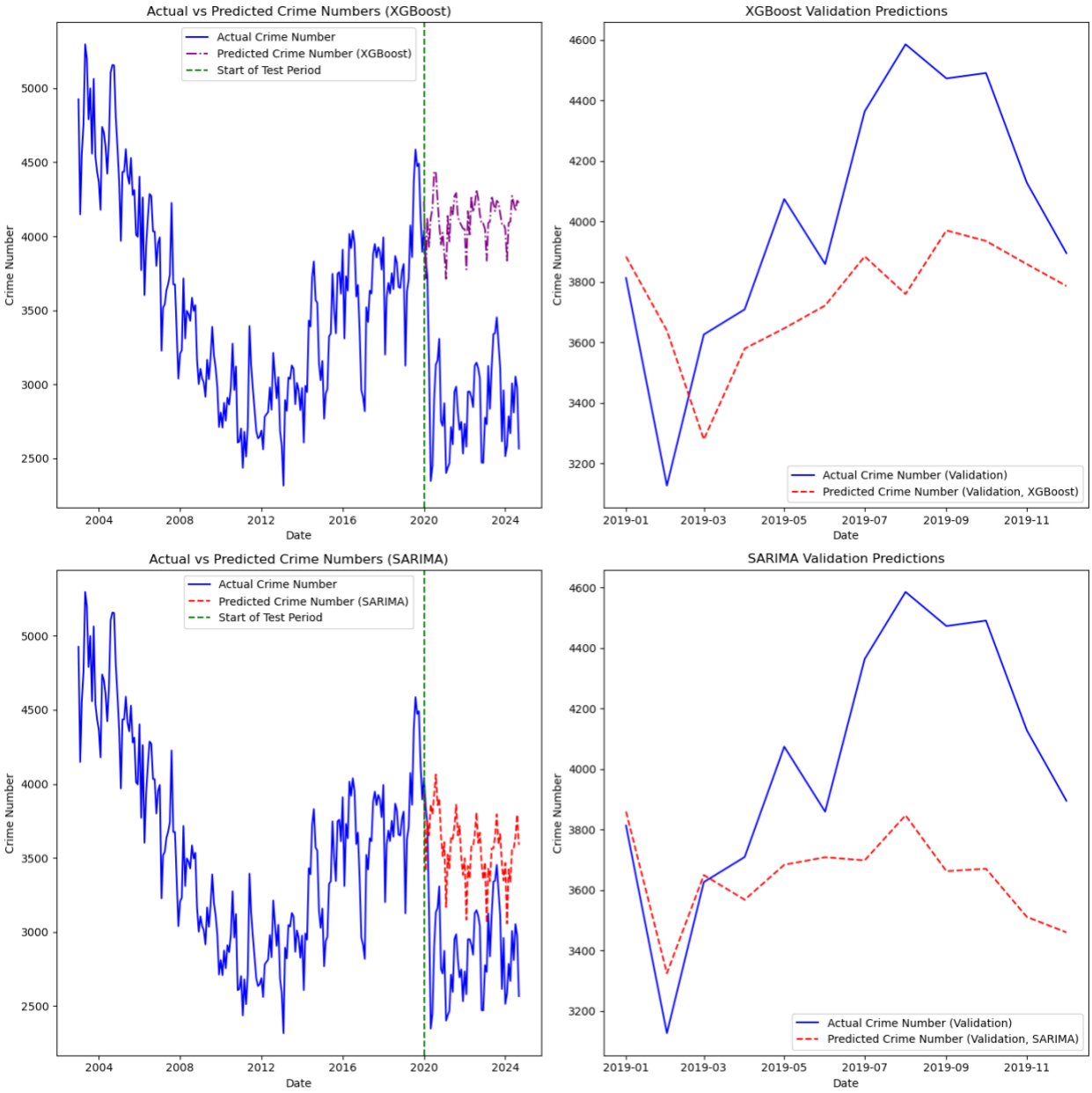


Figure 3: Boxplots showing annual averages over the period.


Vancouver Unemployment Rate
Unemployment Rate (%)
Year
Minimum Wage
Minimum Wage (CAD)
Year
Police officers per 100,000 population
Police officers per 100,000 population
Year
CPI
CPI
Year
Temperature
Temperature (°C)
Year
Weighted Clearance Rate
Weighted Clearance Rate
Year

Results:

Figure 4: Comparison of SARIMA and XGBoost Predictions for Crime Numbers and Validations



Based on the validation results for SARIMA during the period from January 2019 to December 2019, the Mean Absolute Percentage Error (MAPE) was 9.95%. As shown in Figure 4 (bottom right), the SARIMA model successfully captured the overall trend of crime number fluctuations during the validation period. After fitting the model on the full training dataset and forecasting crime numbers for the post-2020-03 period, the average percent difference between

predicted and actual values (actual - prediction) was -25.31%, highlighting a substantial deviation from the validation period's accuracy (Figure 4 bottom left).

Similarly, the XGBoost model achieved a MAPE of 8.94% during validation, indicating strong predictive performance (Figure 4 top right). However, for the test period after March 2020, the average percent difference between predicted and actual values was -44.95%, showing a much larger deviation compared to the validation period (Figure 4 top left).

Conclusion

The findings of this study reveal a significant decrease in crime numbers in Vancouver during the COVID-19 lockdown, as both SARIMA and XGBoost models consistently overestimated crime rates when compared to actual observations post-lockdown. These results contrast with the conclusions of Martin A. Andresen and Tarah Hodgkinson (2022), who suggested that overall crime levels in Vancouver remained relatively stable at the city level during the pandemic, with most changes occurring in the spatial distribution and patterns of crime.

Literature review

Martin A. Andresen and Tarah Hodgkinson (2022) employed differential local Moran's I and ANOVA to analyze the spatial changes in crime patterns in Vancouver pre- and post-COVID-19. Their methodology focused on identifying clusters of neighborhoods based on crime level changes. They further analyzed these clusters using social disorganization theory to investigate the relationship between neighborhood characteristics and changes in crime due to the COVID-19 pandemic. Their work emphasizes spatial variation and neighborhood-level changes rather than overall citywide changes in crime levels.

In contrast, our research adopts a time series analysis framework using SARIMA and XGBoost models to evaluate citywide crime trends. Instead of focusing on spatial clusters, our study evaluates temporal trends and employs a counterfactual approach to compare predicted crime numbers with actual observations. This broader temporal scope provides insights into citywide crime trends over time rather than neighborhood-level dynamics.

Boman and Gallupe (2020) and others also investigated change of crime rates in the U.S. during the pandemic, primarily using police calls for service (911 calls) as a metric. It does not account for broader socioeconomic and environmental factors influencing crime trends. Our study, on the other hand, incorporates control variables to improve prediction accuracy and

provide a more comprehensive analysis of factors affecting crime trends during the COVID-19 lockdown.

References

Andresen, M. A., & Hodgkinson, T. (2022). In a world called catastrophe: The impact of COVID-19 on neighbourhood level crime in Vancouver, Canada. Journal of Experimental Criminology. Advance online publication. https://doi.org/10.1007/s11292-021-09495

Boman, J. H., & Gallupe, O. (2020). Has COVID-19 changed crime? Crime rates in the United States during the pandemic. American Journal of Criminal Justice, 45(4), 537–545. https://doi.org/10.1007/s12103-020-09551-3

Environment and Climate Change Canada. (n.d.). Historical climate data. Retrieved November 26, 2024, from https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

Federal Bureau of Investigation. (2011). Variables affecting crime. Retrieved November 26, 2024, from https://ucr.fbi.gov/hate-crime/2011/resources/variables-affecting-crime

Government of Canada. (n.d.). Minimum wage database: Custom search. Retrieved November 26, 2024, from https://minwage-salairemin.service.canada.ca/en/customsearch.html

Statistics Canada. (n.d.). Table 18-10-0006-01: Consumer Price Index, monthly, not seasonally adjusted. Retrieved November 26, 2024, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000601

Statistics Canada. (n.d.). Table 14-10-0295-02: Labour force characteristics by province, seasonally adjusted. Retrieved November 26, 2024, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410029502

Statistics Canada. (n.d.). Table 35-10-0077-01: Crime severity index and weighted clearance rates by police service. Retrieved November 26, 2024, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510007701

Vancouver Police Department (VPD). (n.d.). Open data portal. Retrieved November 26, 2024, from https://geodash.vpd.ca/opendata/