



# **WQX7001 PRINCIPLE OF DATA SCIENCE**

## **ANALYSIS AND PREDICTION OF VANCOUVER'S CRIMINAL ACTIVITY USING MACHINE LEARNING**

# CONTENT

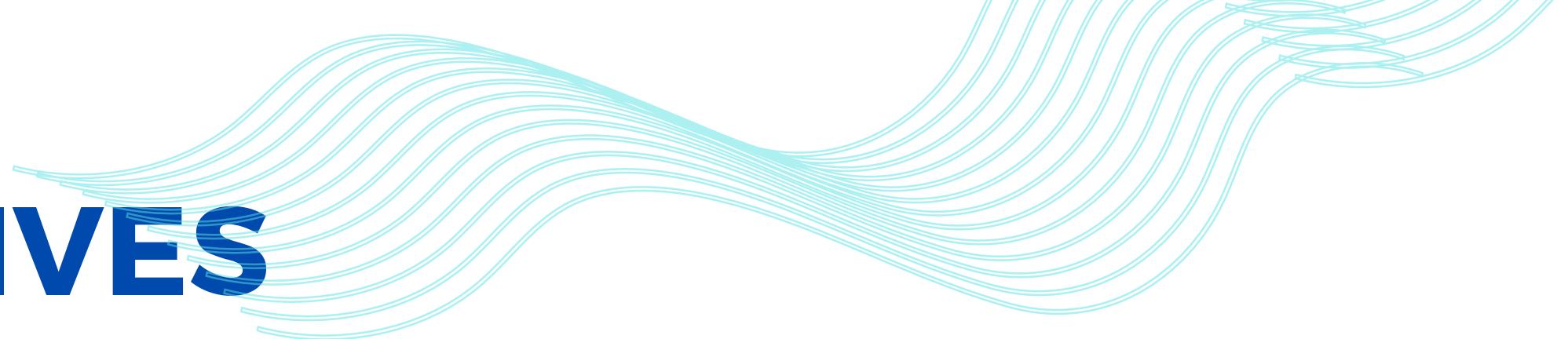
1. Project Background
2. Project Objectives
3. Methodology
4. Data Modelling
5. Data Interpretation
6. Deployment of Data Product
7. Insights & Conclusion
8. References

# PROJECT BACKGROUND

- The increasing need to decline crime activity is essential for promoting peace and justice, aligning with the Sustainable Development Goals (SDGs) aimed at fostering inclusive societies and effective institutions.
- This paper utilizes machine learning techniques to predict crime activities, which can significantly contribute to cost reduction and resource optimization in law enforcement and public safety initiatives.
- The dataset analyzed originates from the **Vancouver Open Data Catalogue**, comprising 881242 records spanning from 1 January 2003, to 24 November 2023.

The screenshot shows the official website of the Vancouver Police Department. At the top left is the department's crest. To its right is the text "VANCOUVER POLICE DEPARTMENT | Beyond the Call". Below this are social media icons for Facebook, Twitter, and YouTube, followed by a "MENU" button. A large banner image features a close-up of a police officer's uniform with the "POLICE VANCOUVER" badge. Below the banner, a navigation menu includes "Home", "News", and a link to the news article. The main headline reads "Vancouver Police Adopt New Technology to Predict Property Crime". Below the headline is a "VANCOUVER IS AWESOME" logo and a "JOIN OUR NEWSLETTER" button. The news article summary states "VPD data shows violent crime in Vancouver highest since 2013" and mentions "Robberies, attempted murders, assaults driving increase in violent crime".

# PROJECT OBJECTIVES



1

## Objective 1

To identify trend, patterns, and correlations in Vancouver's crime data from 2003 to 2021 through exploratory data analysis.

2

## Objective 2

To model crime prediction using machine learning techniques such as Naïve Bayes and Decision Trees, along with time-series forecasting to predict future crime trends.

3

## Objective 3

To evaluate the performance of the models by comparing their accuracy in predicting crime categories and trends, with a focus on improving prediction accuracy for law enforcement applications.

# METHODOLOGY - OBTAIN

- The data collected on crime in Vancouver includes offences against a person, such as robbery and sexual assault and property crimes like shoplifting and theft. This information is aggregated to protect personal identifiable information (PII) and is complemented by geospatial data.
- Data from **2003 - 2023**, with 881242 row, 10 col.
- Data Classification:**
  - Nominal:** TYPE, HUNDRED\_BLOCK, NEIGHBOURHOOD
  - Ordinal:** MONTH, DAY, X, Y
  - Ratio:** YEAR, HOUR, MINUTE

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
0	Break and Enter Commercial	2012	12	14	8	52		Nan	Oakridge	491285.0000 5.453433e+06
1	Break and Enter Commercial	2019	3	7	2	6	10XX SITKA SQ	Fairview	490612.9648 5.457110e+06	
2	Break and Enter Commercial	2019	8	27	4	12	10XX ALBERNI ST	West End	491004.8164 5.459177e+06	
3	Break and Enter Commercial	2021	4	26	4	44	10XX ALBERNI ST	West End	491007.7798 5.459174e+06	
4	Break and Enter Commercial	2014	8	8	5	13	10XX ALBERNI ST	West End	491015.9434 5.459166e+06	
...	...	...	...	...	...	...	...	...	...	...
881237	Vehicle Collision or Pedestrian Struck (with I...	2003	9	1	20	45	YUKON ST / W KING EDWARD AVE	Riley Park	491786.0000 5.455143e+06	
881238	Vehicle Collision or Pedestrian Struck (with I...	2005	7	20	18	57	YUKON ST / W KING EDWARD AVE	Riley Park	491786.0000 5.455143e+06	
881239	Vehicle Collision or Pedestrian Struck (with I...	2016	6	20	18	11	YUKON ST / W KING EDWARD AVE	Riley Park	491786.0000 5.455143e+06	
881240	Vehicle Collision or Pedestrian Struck (with I...	2013	12	3	9	49	YUKON ST / W KING EDWARD AVE	Riley Park	491835.0000 5.455126e+06	
881241	Vehicle Collision or Pedestrian Struck (with I...	2013	12	27	14	40	YUKON ST / W KING EDWARD AVE	Riley Park	491835.0000 5.455126e+06	

881242 rows × 10 columns

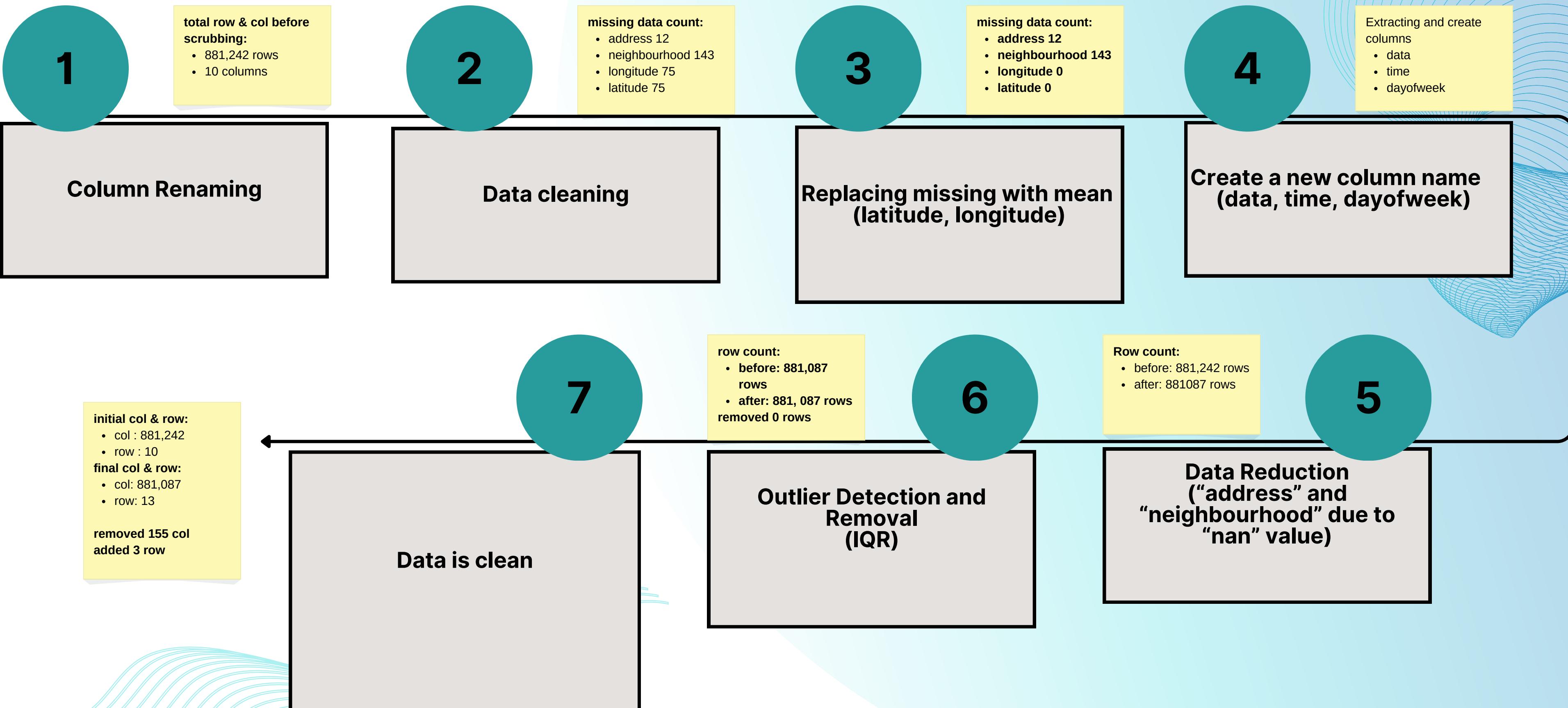
# METHODOLOGY - OBTAIN (CONT')

Features	Description	Data type
TYPE	The type of crime activity	object
YEAR	A four-digit field that indicates the year when the reported crime activity occurred	int64
MONTH	A numeric field that indicates the month when the reported crime activity occurred	int64
DAY	A two-digit field that indicates the day of the month when the reported crime activity occurred	int64
HOUR	A two-digit field that indicates the hour time (in 24 hours format) when the reported crime activity occurred.	int64
MINUTE	A two-digit field that indicates the minute when the reported crime activity occurred	int64
HUNDRED_BLOCK	Generalized location of the reported crime activity	OBJECT
NEIGHBOURHOOD	Neighbourhoods within the City of Vancouver are based on the census tract (CT) concept within census metropolita area (CMA)	OBJECT
X	Coordinate values are projected in UTM Zone 10	float64
Y	Coordinate values are projected in UTM Zone 10	float64

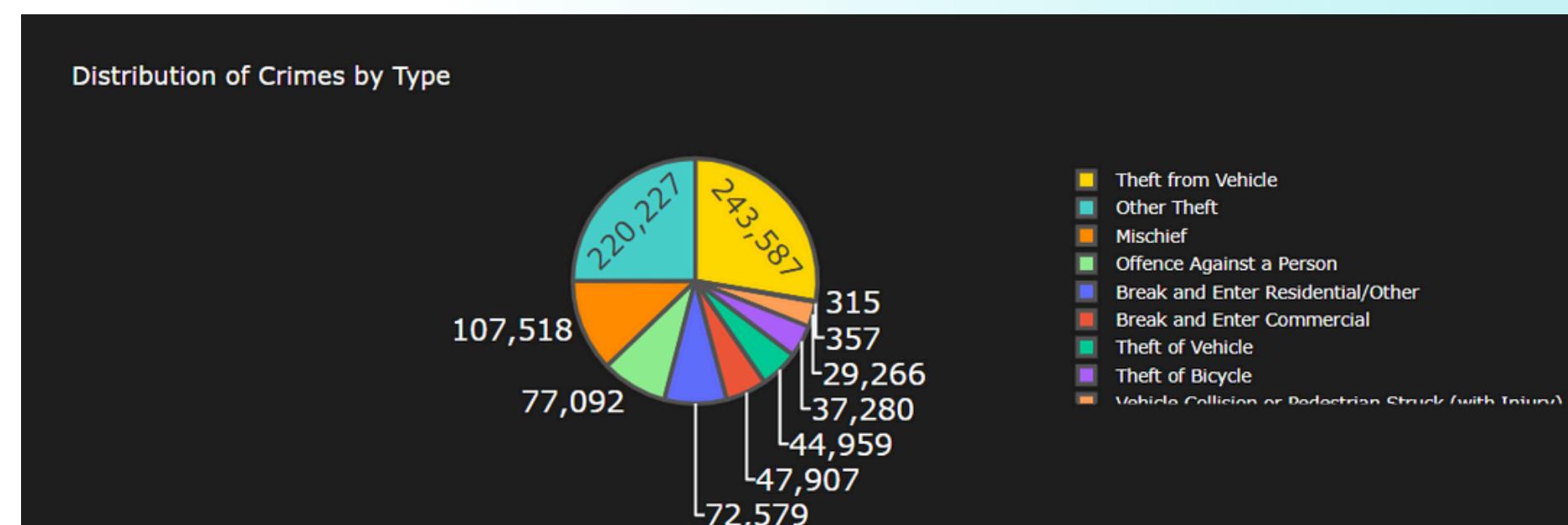
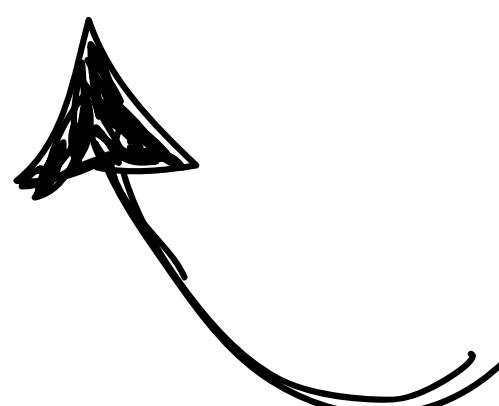
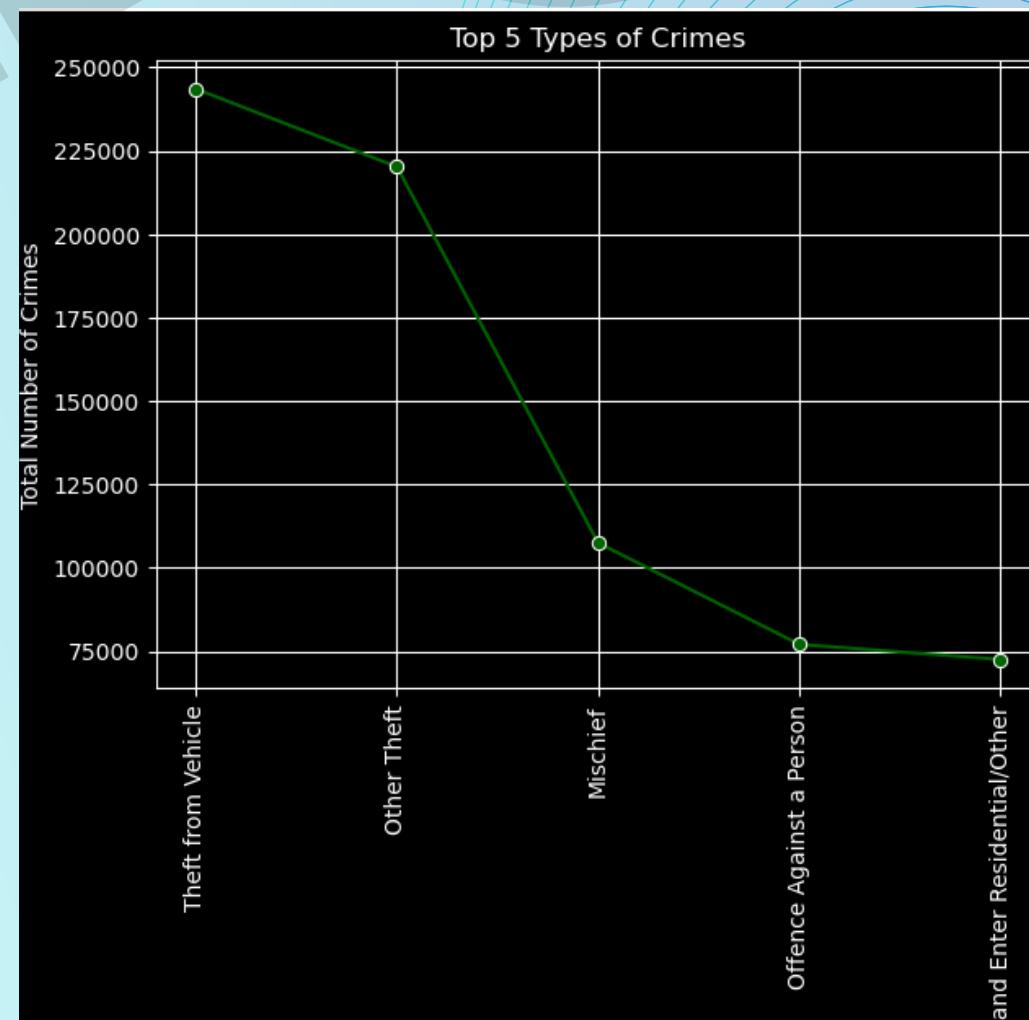
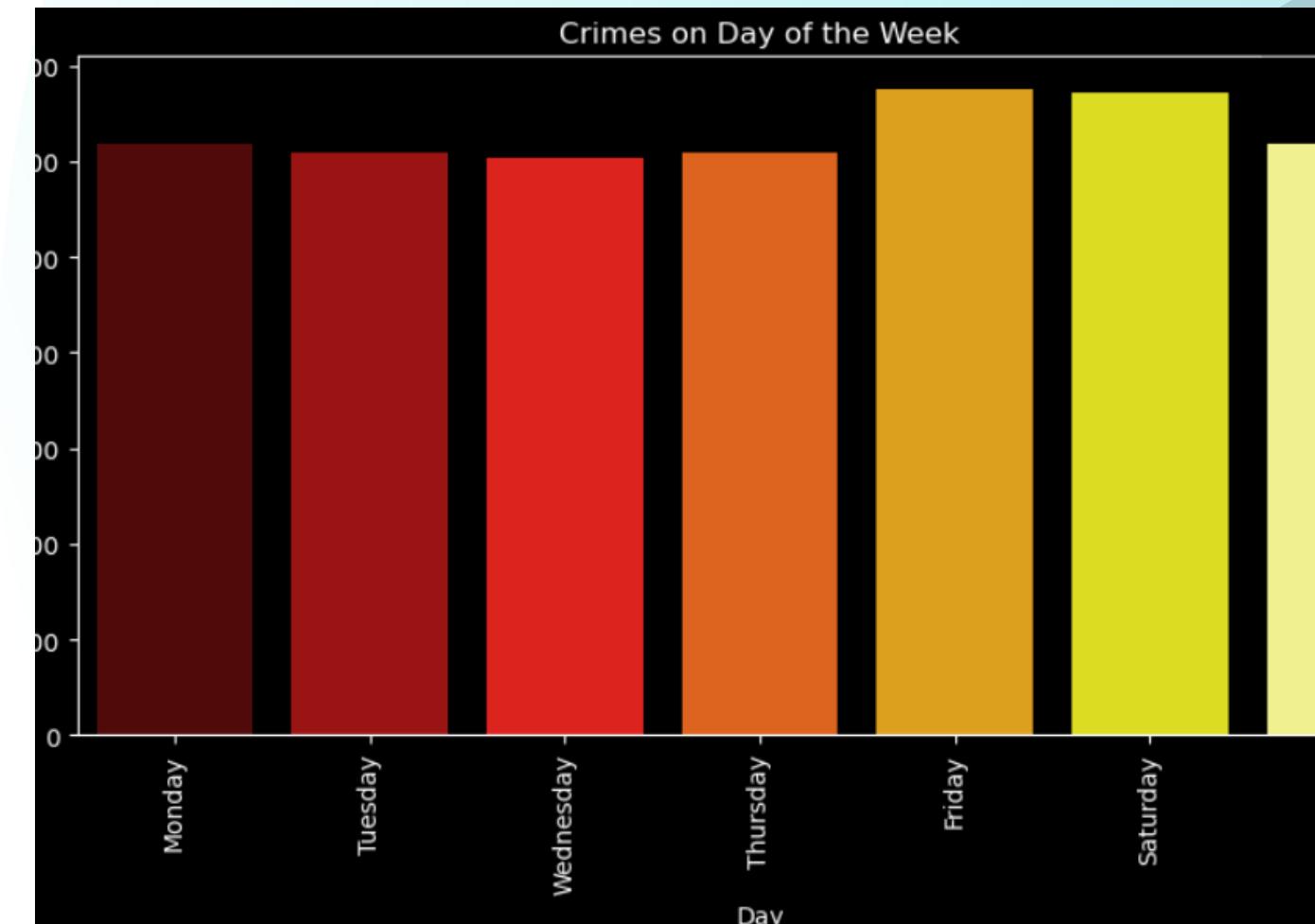
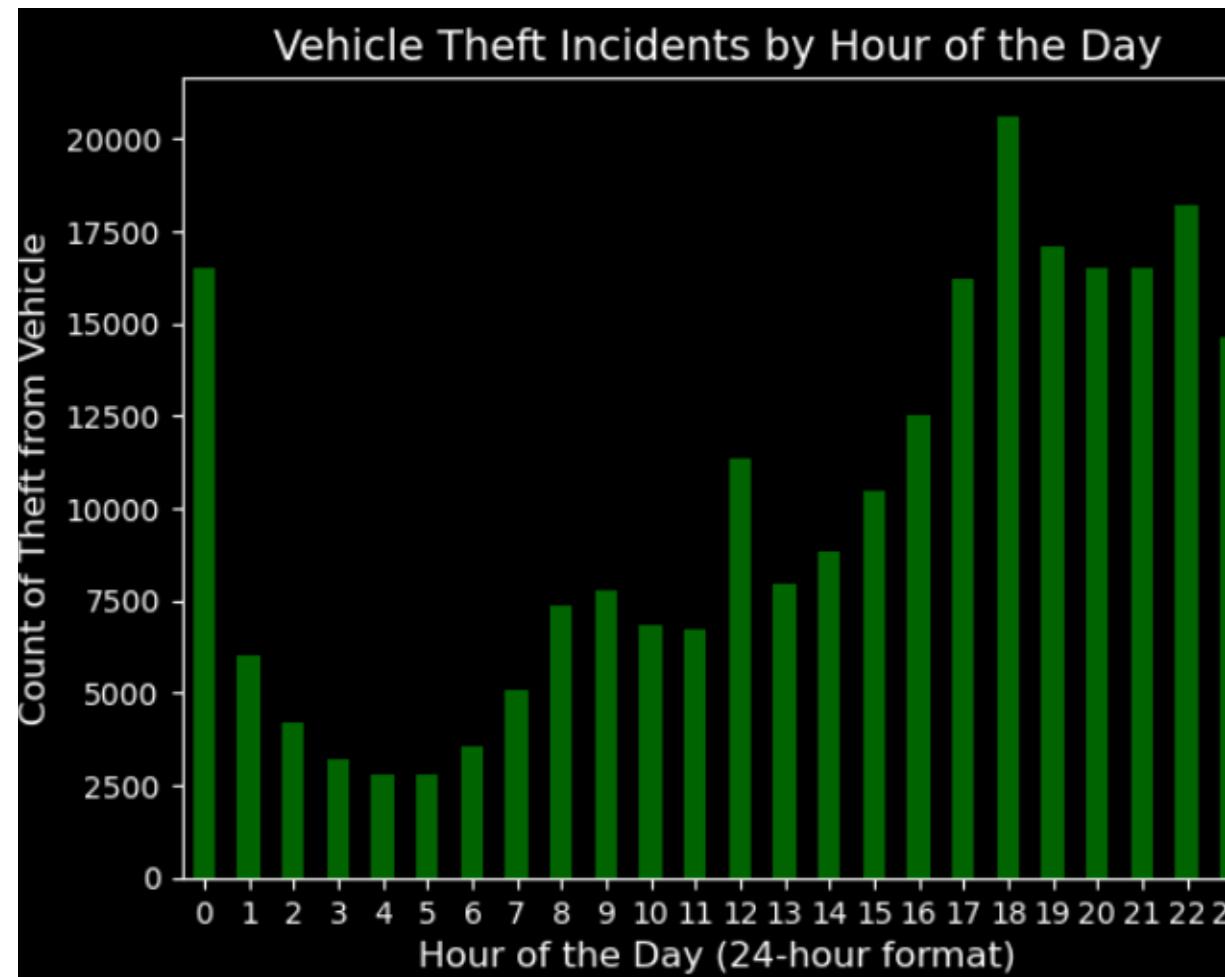
# METHODOLOGY - OBTAIN (CONT')

- Source
  - PRIME BC Police Records Management System (RMS)
    - Comprehensive source of structured police records.
    - Only "Founded" incidents included (verified through police investigation).
  - Vancouver Police Department (VPD)
    - Releases crime data to enhance community awareness of policing activities.
  - Users are cautioned against relying solely on this data for safety assessments.

# METHODOLOGY - SCRUB

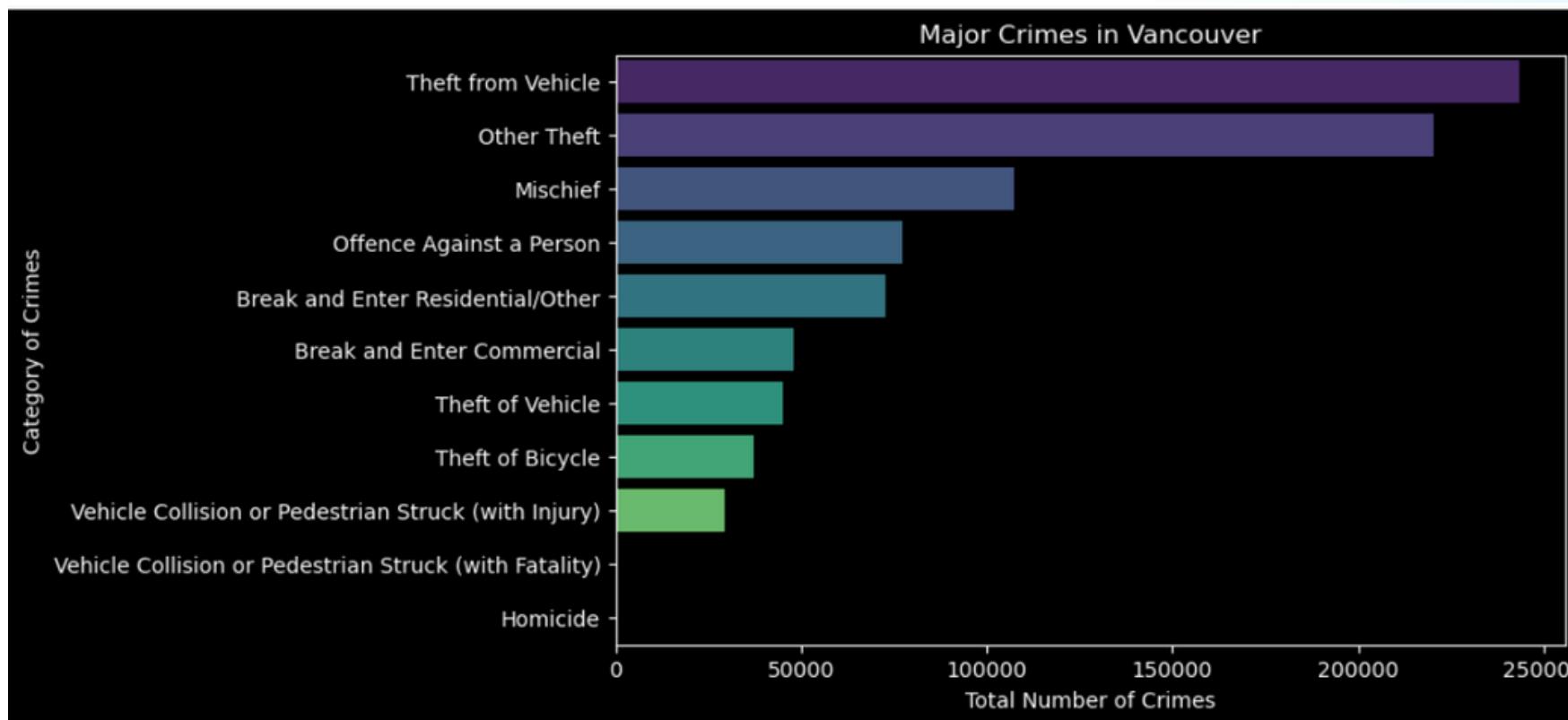


# METHODOLOGY - EXPLORE

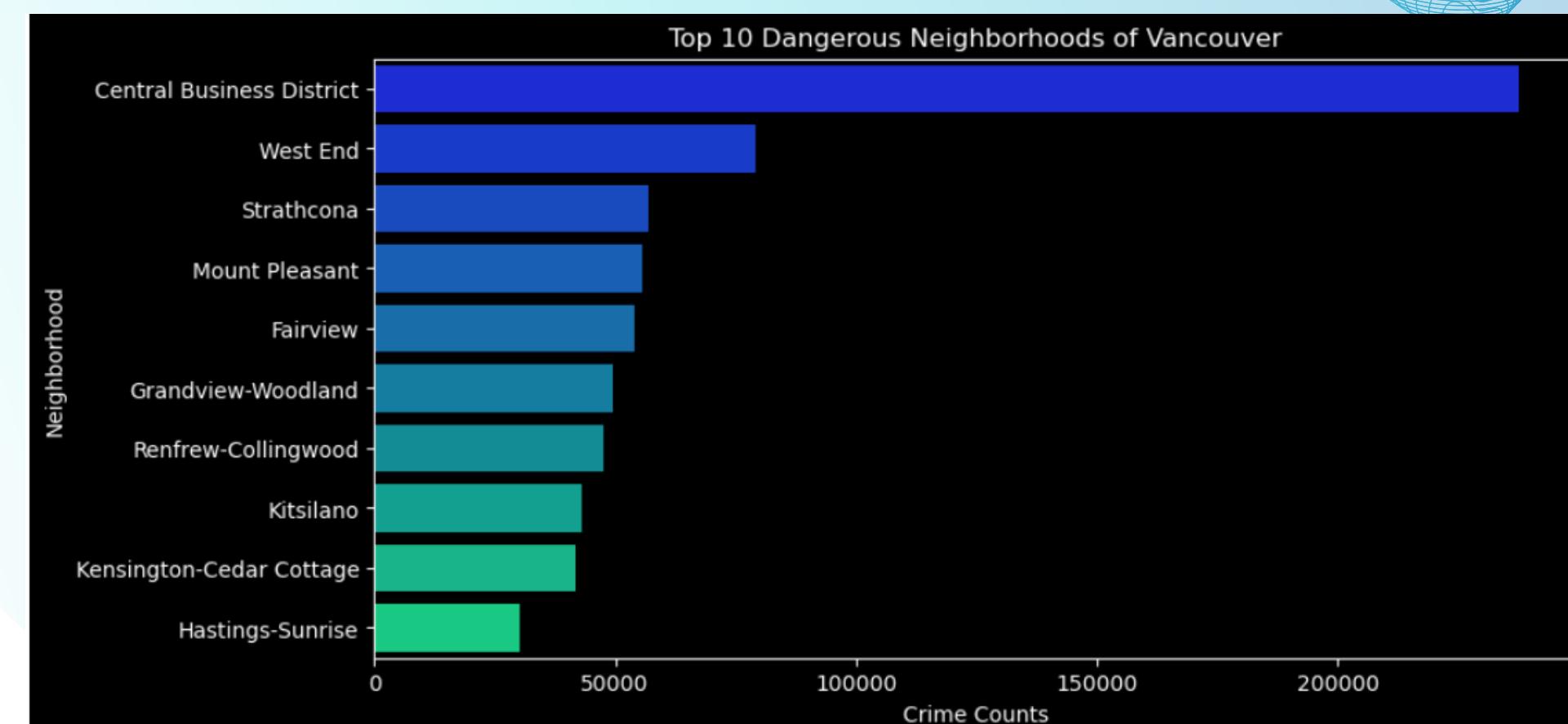
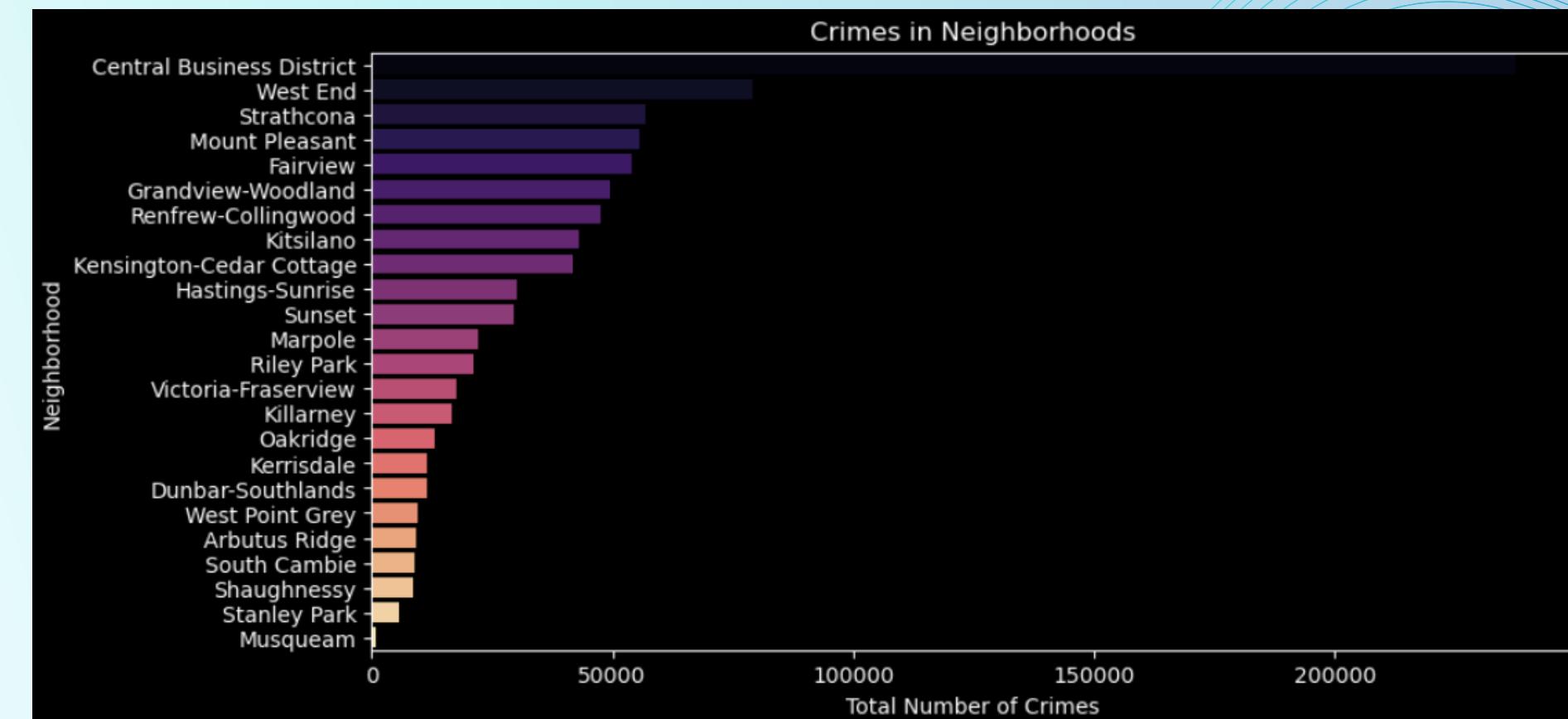


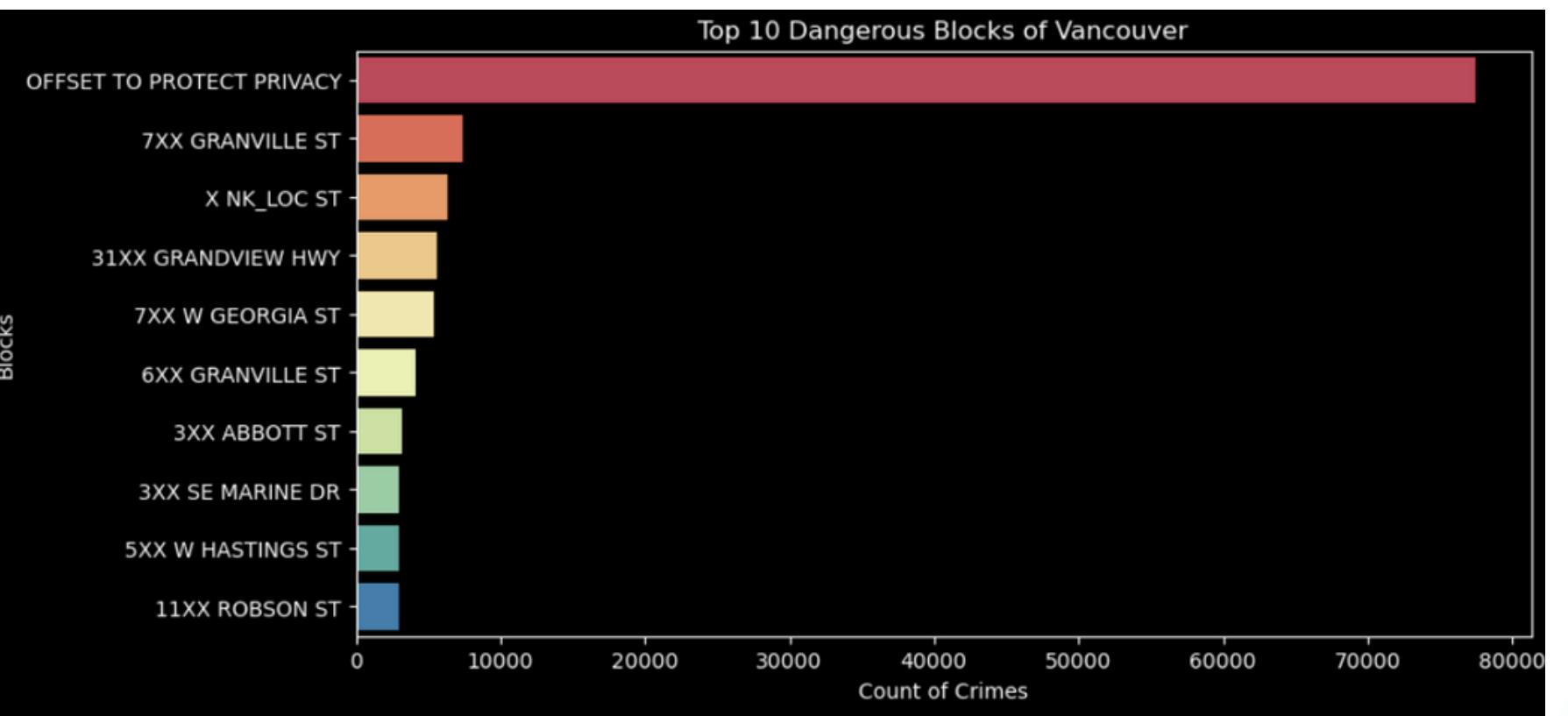
# EXPLORE

## EXPLORATORY DATA ANALYSIS (EDA)

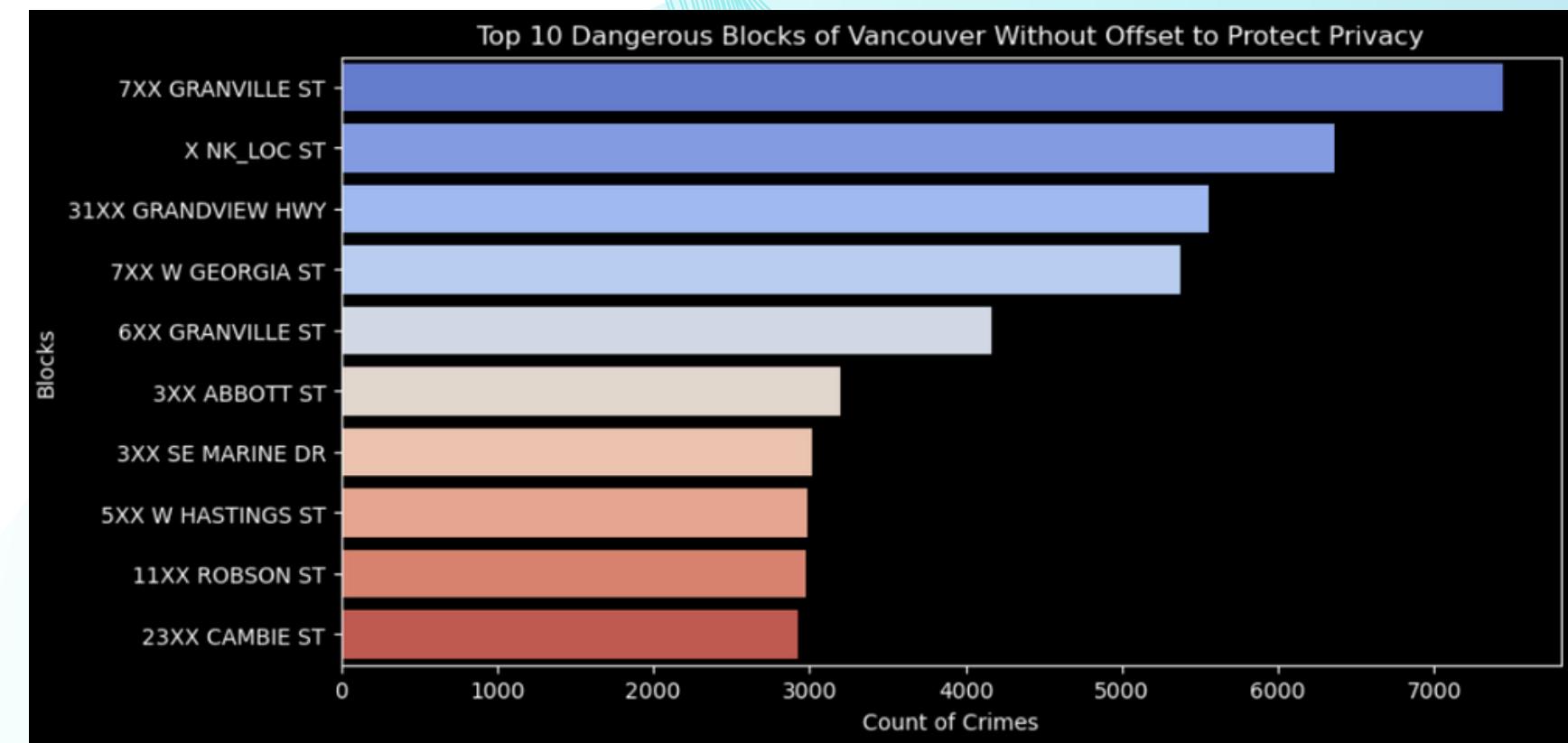


- MAJOR CRIMES IN VANCOUVER
- CRIMES IN NEIGHBORHOODS
- TOP 10 DANGEROUS NEIGHBORHOODS OF VANCOUVER

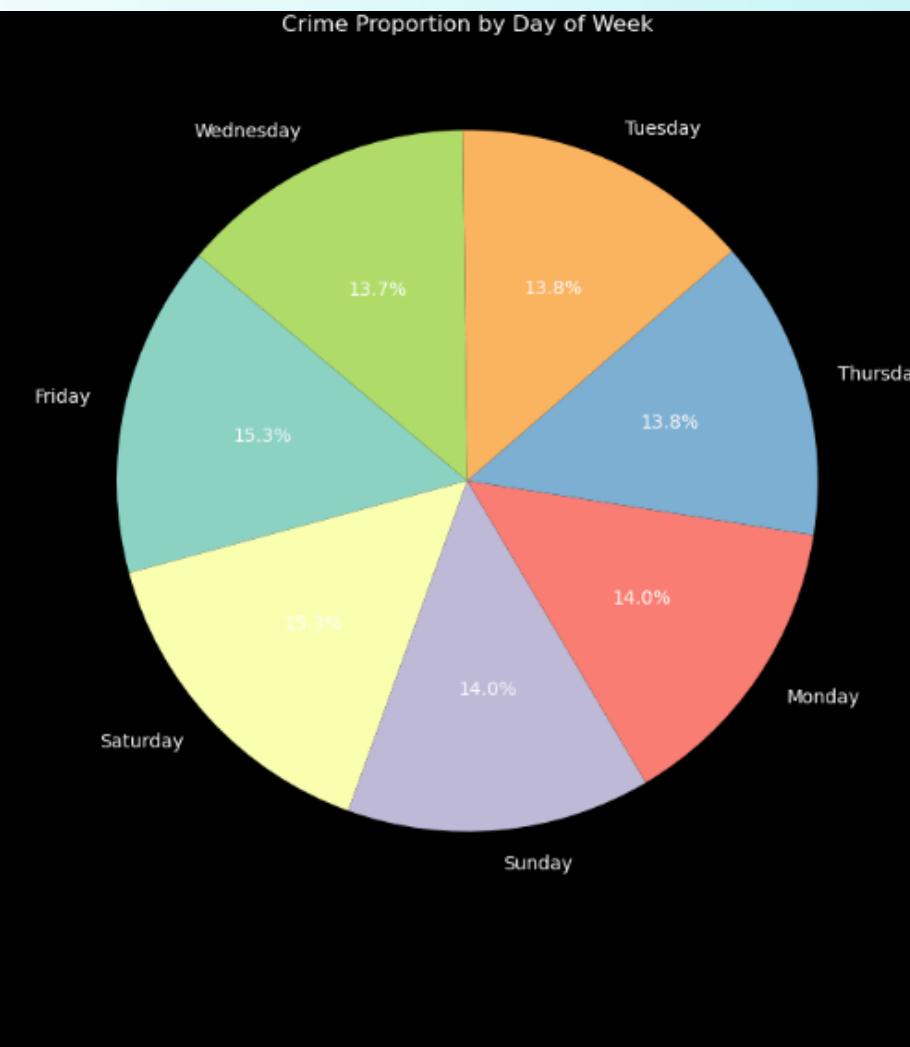
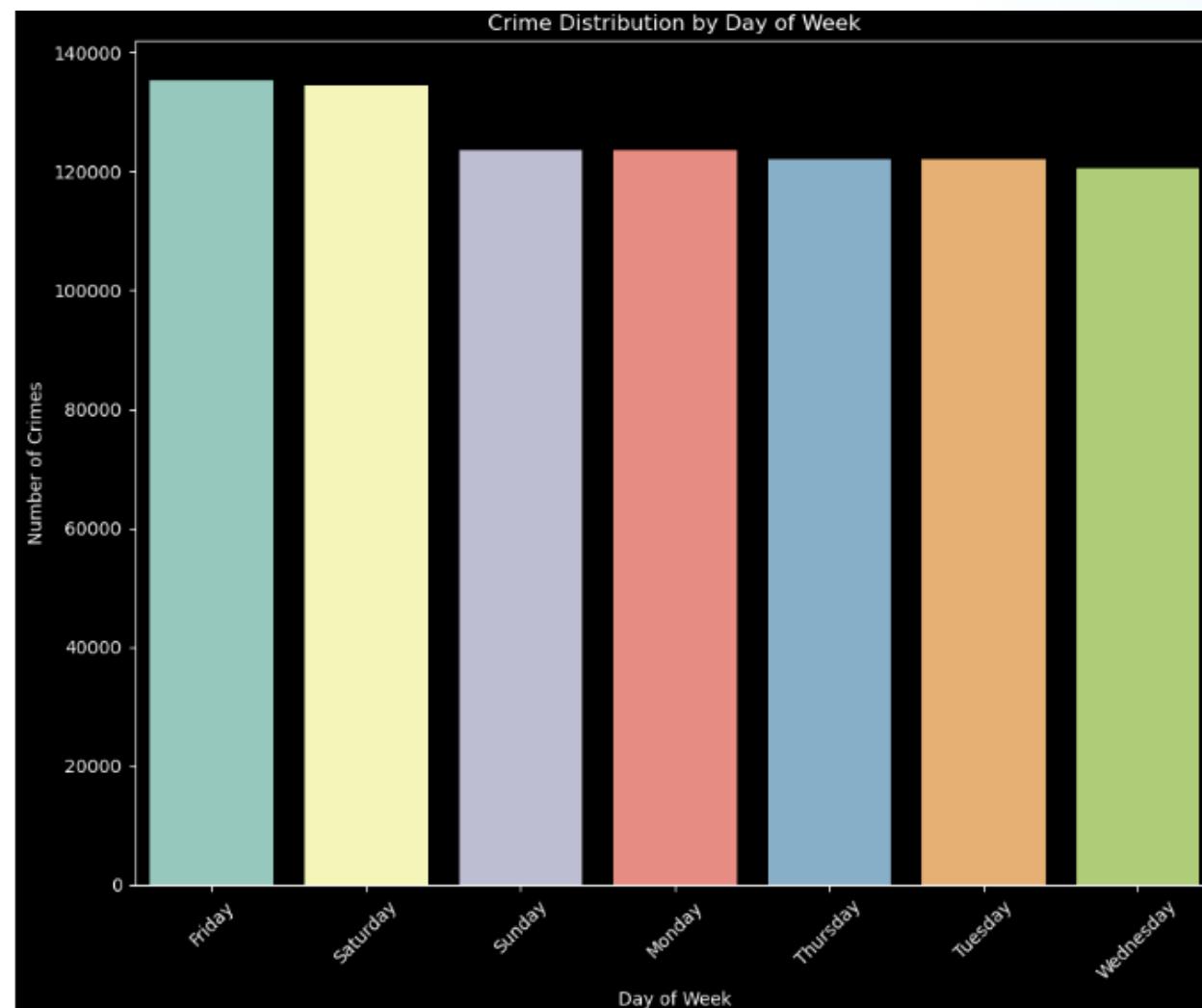




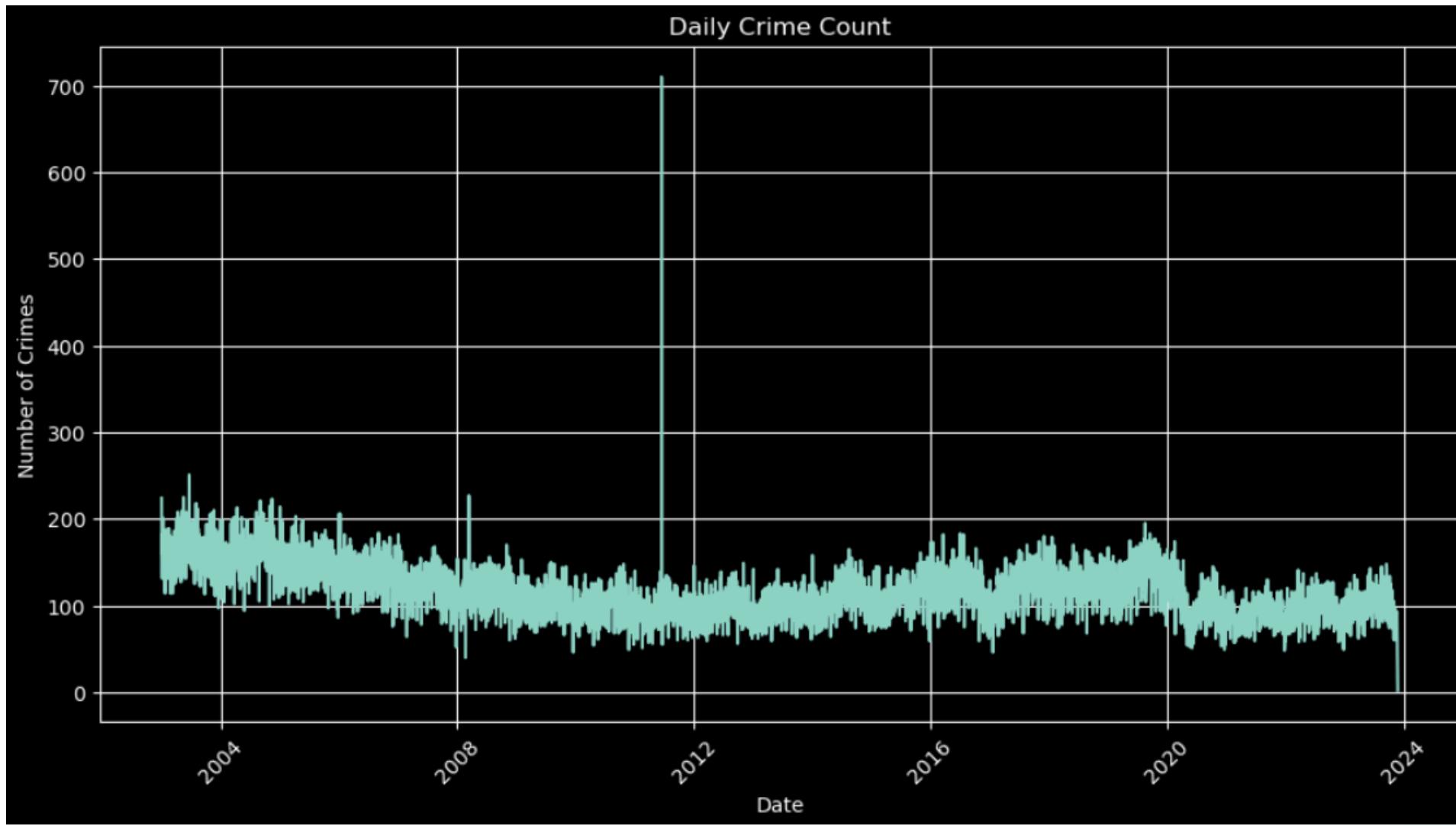
**TOP 10 DANGEROUS BLOCKS OF VANCOUVER**



**TOP 10 DANGEROUS BLOCKS OF VANCOUVER  
WITHOUT OFFSET TO PROTECT PRIVACY**

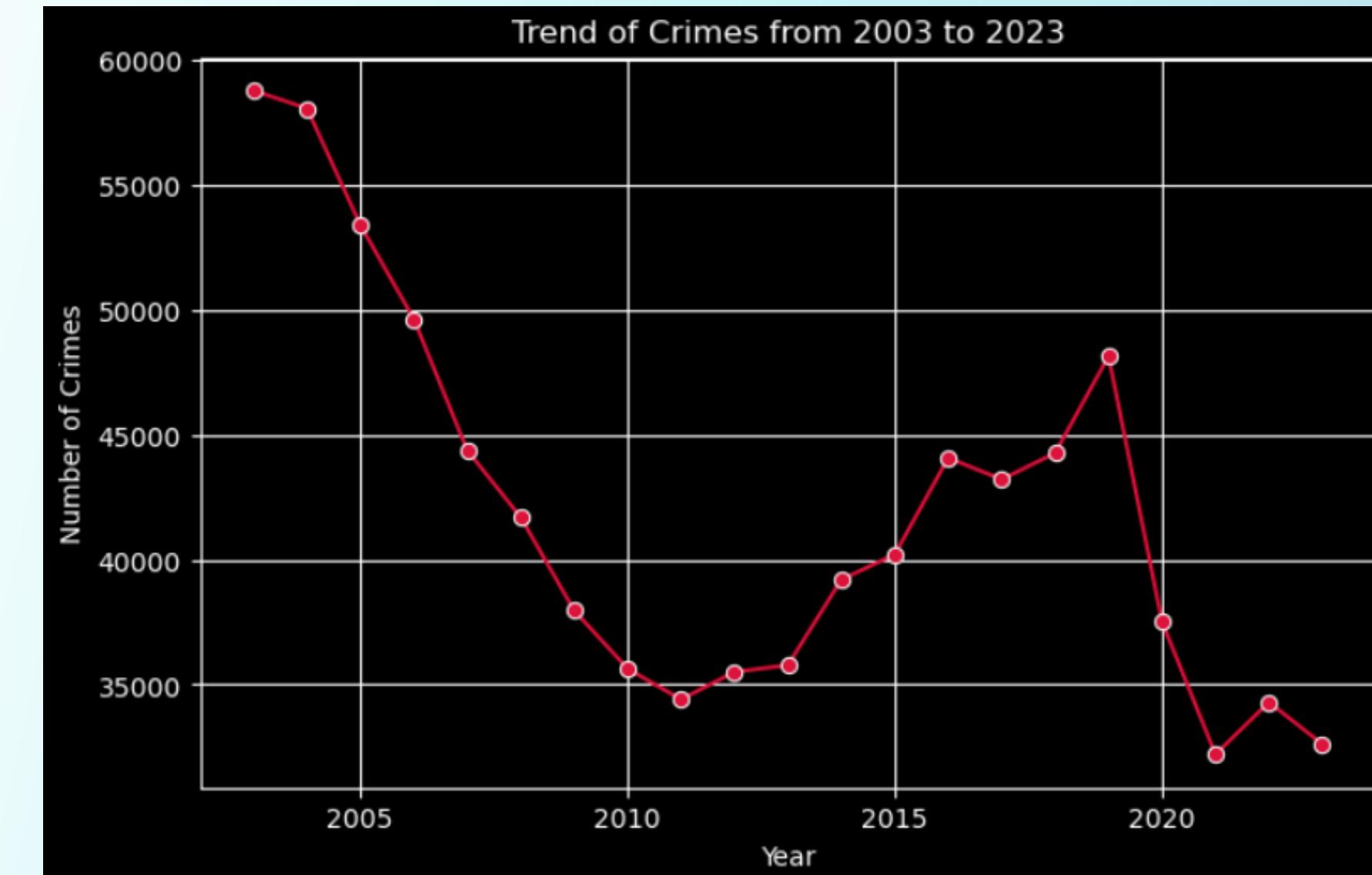


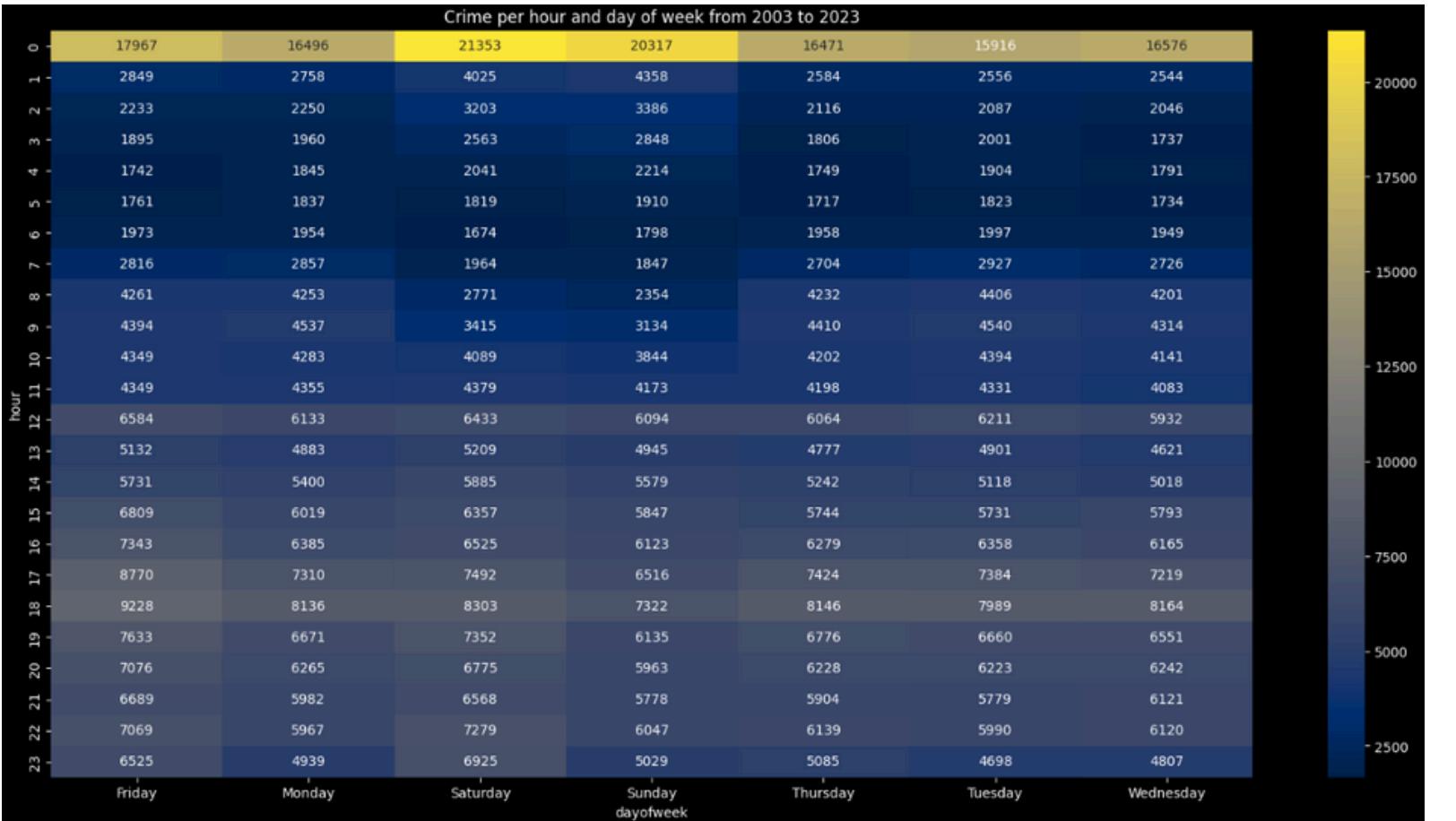
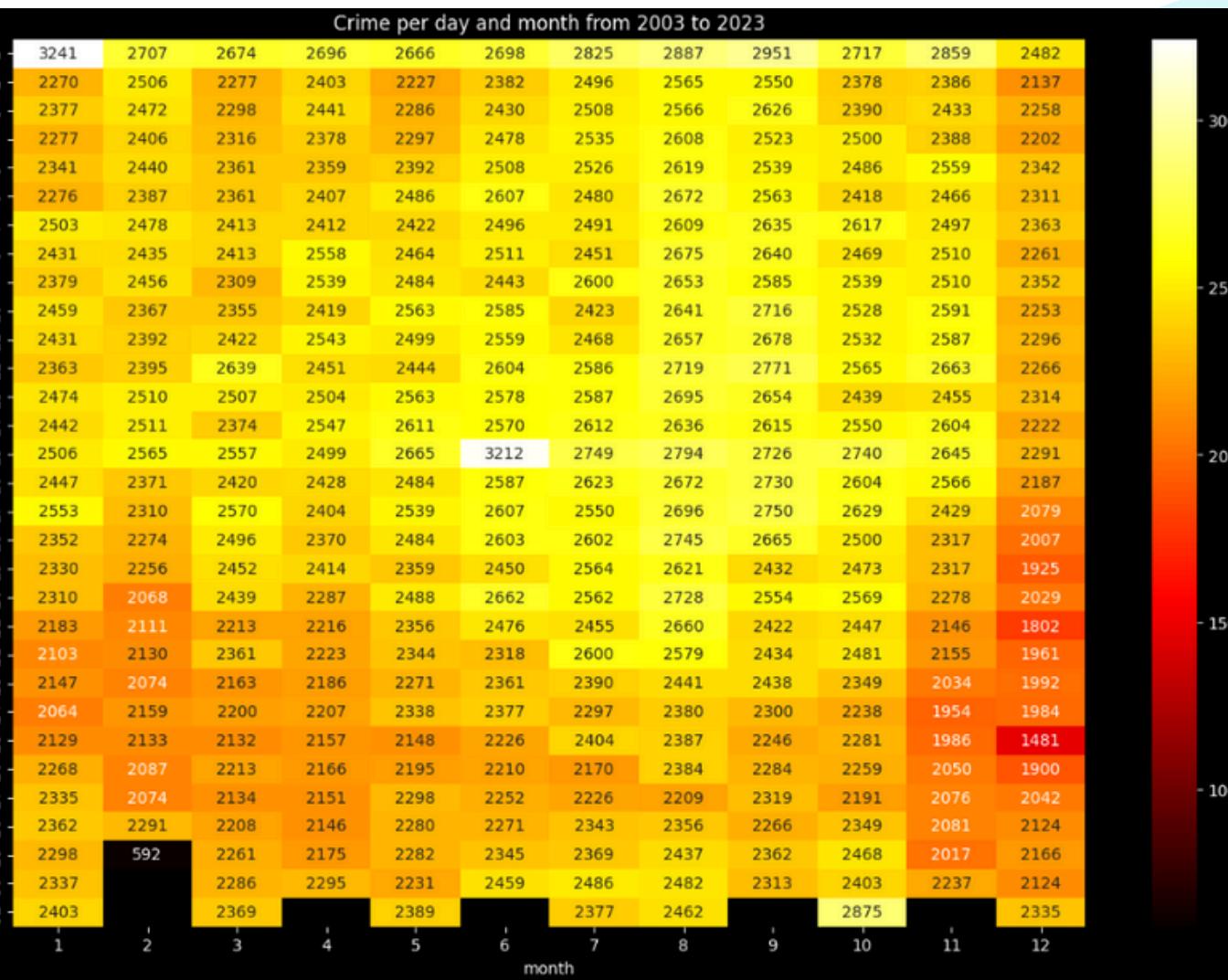
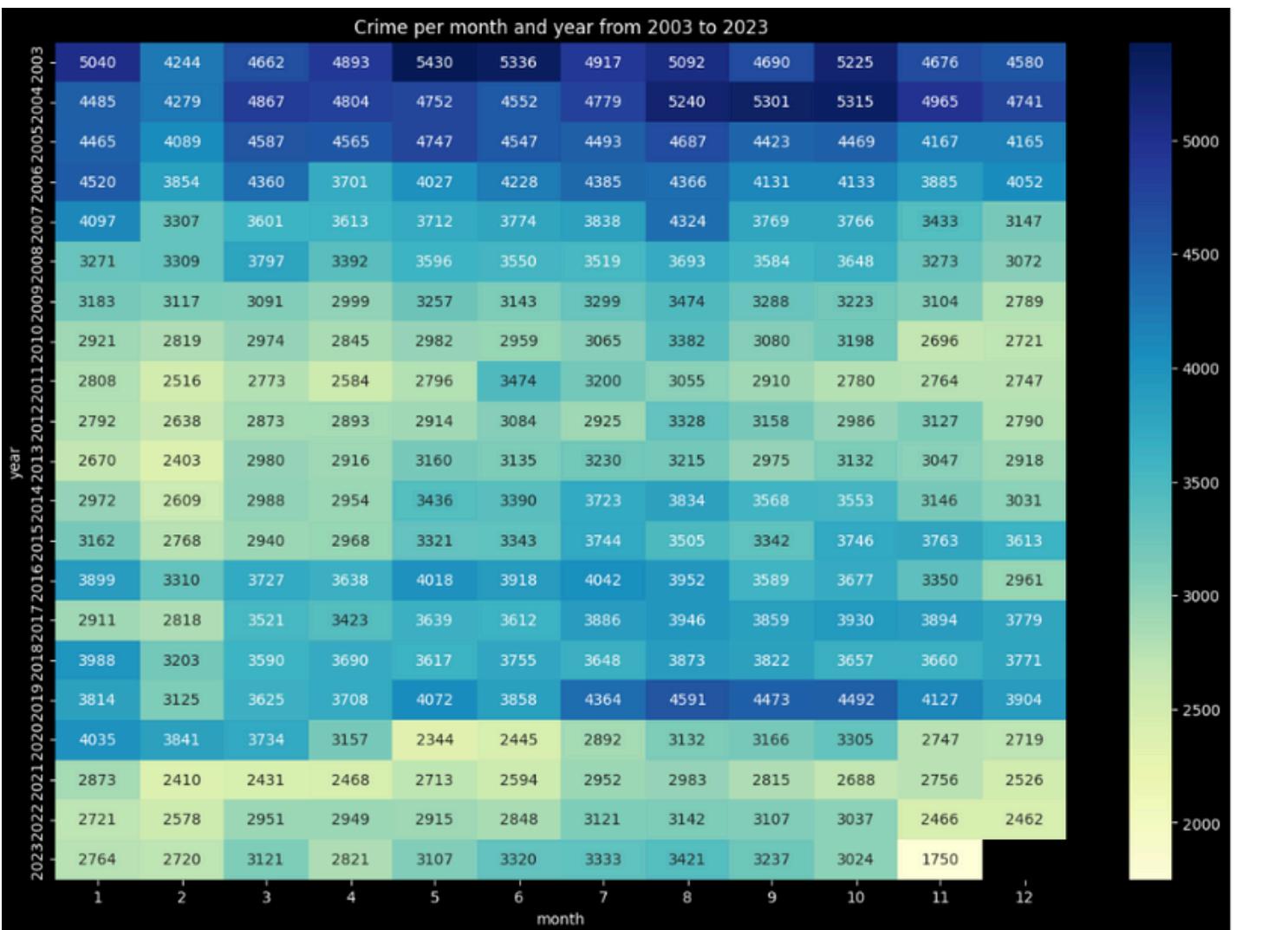
**CRIMES ON DAY OF THE WEEK'**



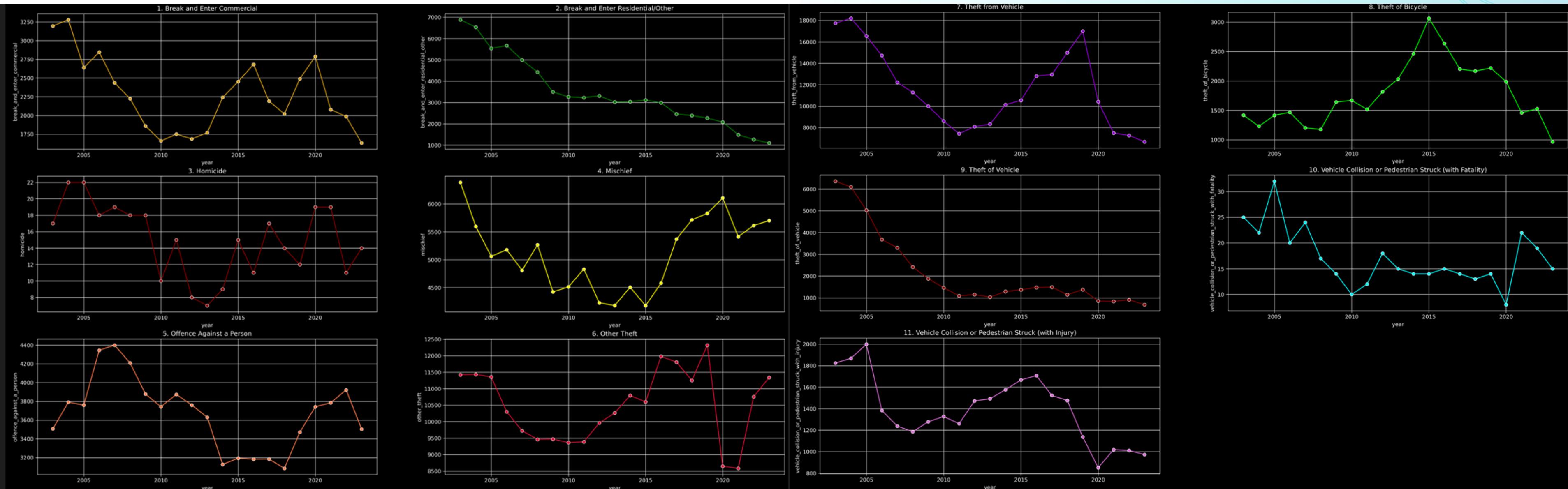
TREND IN VANCOUVER'S CRIME RATE

## TIME SERIES GRAPH WITH CRIMES PER DAY





# DISTRIBUTION OF EACH CRIME BY TYPE FROM 2003 TO 2023



# DATA INTERPRETATION

## MODEL RESULTS

Classification Report:				
	precision	recall	f1-score	support
0	0.38	0.28	0.32	9514
1	0.37	0.34	0.35	14548
2	0.00	0.00	0.00	50
3	0.38	0.27	0.32	21654
4	1.00	1.00	1.00	15392
5	0.63	0.68	0.66	44006
6	0.50	0.65	0.56	48594
7	0.21	0.08	0.11	7507
8	0.22	0.12	0.15	9050
9	0.19	0.05	0.07	64
10	0.88	0.80	0.84	5839
accuracy			0.55	176218
macro avg	0.43	0.39	0.40	176218
weighted avg	0.53	0.55	0.53	176218
Accuracy: 55.08%				

## RANDOM FOREST

Classification Report:				
	precision	recall	f1-score	support
0	0.35	0.32	0.34	9514
1	0.36	0.39	0.37	14548
2	0.00	0.00	0.00	50
3	0.37	0.30	0.33	21654
4	1.00	1.00	1.00	15392
5	0.63	0.68	0.66	44006
6	0.51	0.62	0.56	48594
7	0.23	0.07	0.10	7507
8	0.22	0.09	0.13	9050
9	0.10	0.05	0.06	64
10	0.86	0.80	0.83	5839
accuracy			0.55	176218
macro avg	0.42	0.39	0.40	176218
weighted avg	0.53	0.55	0.53	176218
Accuracy: 54.96%				

## DECISION TREE

Classification Report:				
	precision	recall	f1-score	support
0	0.35	0.02	0.04	9514
1	0.37	0.08	0.13	14548
2	0.00	0.00	0.00	50
3	0.42	0.03	0.06	21654
4	1.00	1.00	1.00	15392
5	0.56	0.67	0.61	44006
6	0.39	0.81	0.53	48594
7	0.42	0.02	0.03	7507
8	0.31	0.00	0.00	9050
9	0.00	0.00	0.00	64
10	0.49	0.15	0.23	5839
accuracy			0.50	176218
macro avg	0.39	0.25	0.24	176218
weighted avg	0.49	0.50	0.41	176218
Accuracy: 49.62%				

## XGBOOST

Classification Report:				
	precision	recall	f1-score	support
0	0.29	0.33	0.31	9514
1	0.33	0.41	0.36	14548
2	0.00	0.00	0.00	50
3	0.34	0.33	0.33	21654
4	1.00	1.00	1.00	15392
5	0.62	0.64	0.63	44006
6	0.50	0.56	0.52	48594
7	0.22	0.06	0.10	7507
8	0.21	0.08	0.12	9050
9	0.00	0.00	0.00	64
10	0.85	0.73	0.79	5839
accuracy			0.52	176218
macro avg	0.39	0.38	0.38	176218
weighted avg	0.51	0.52	0.51	176218
Accuracy: 52.32%				

## KNEIGHBORS

# DATA INTERPRETATION

## BEST MODEL

Classification Report:				
	precision	recall	f1-score	support
0	0.38	0.28	0.32	9514
1	0.37	0.34	0.35	14548
2	0.00	0.00	0.00	50
3	0.38	0.27	0.32	21654
4	1.00	1.00	1.00	15392
5	0.63	0.68	0.66	44006
6	0.50	0.65	0.56	48594
7	0.21	0.08	0.11	7507
8	0.22	0.12	0.15	9050
9	0.19	0.05	0.07	64
10	0.88	0.80	0.84	5839
accuracy			0.55	176218
macro avg	0.43	0.39	0.40	176218
weighted avg	0.53	0.55	0.53	176218

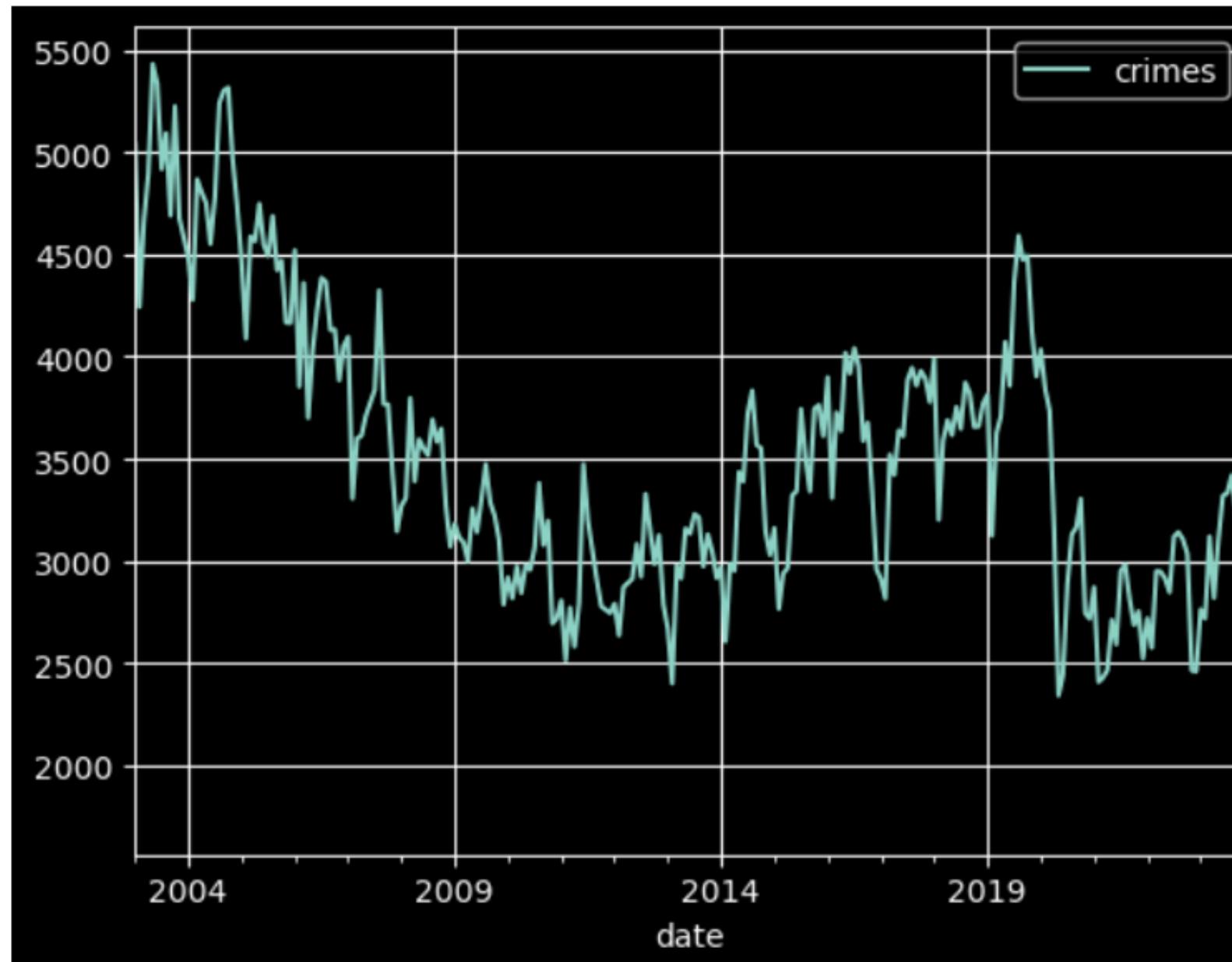
Accuracy: 55.08%

Classifier	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)
RandomForestClassifier	55.08%	0.43	0.39	0.40
DecisionTreeClassifier	54.96%	0.42	0.39	0.40
XGBClassifier	49.52%	0.38	0.25	0.24
KNeighborsClassifier	52.30%	0.40	0.38	0.38

## RANDOM FOREST

# DATA INTERPRETATION

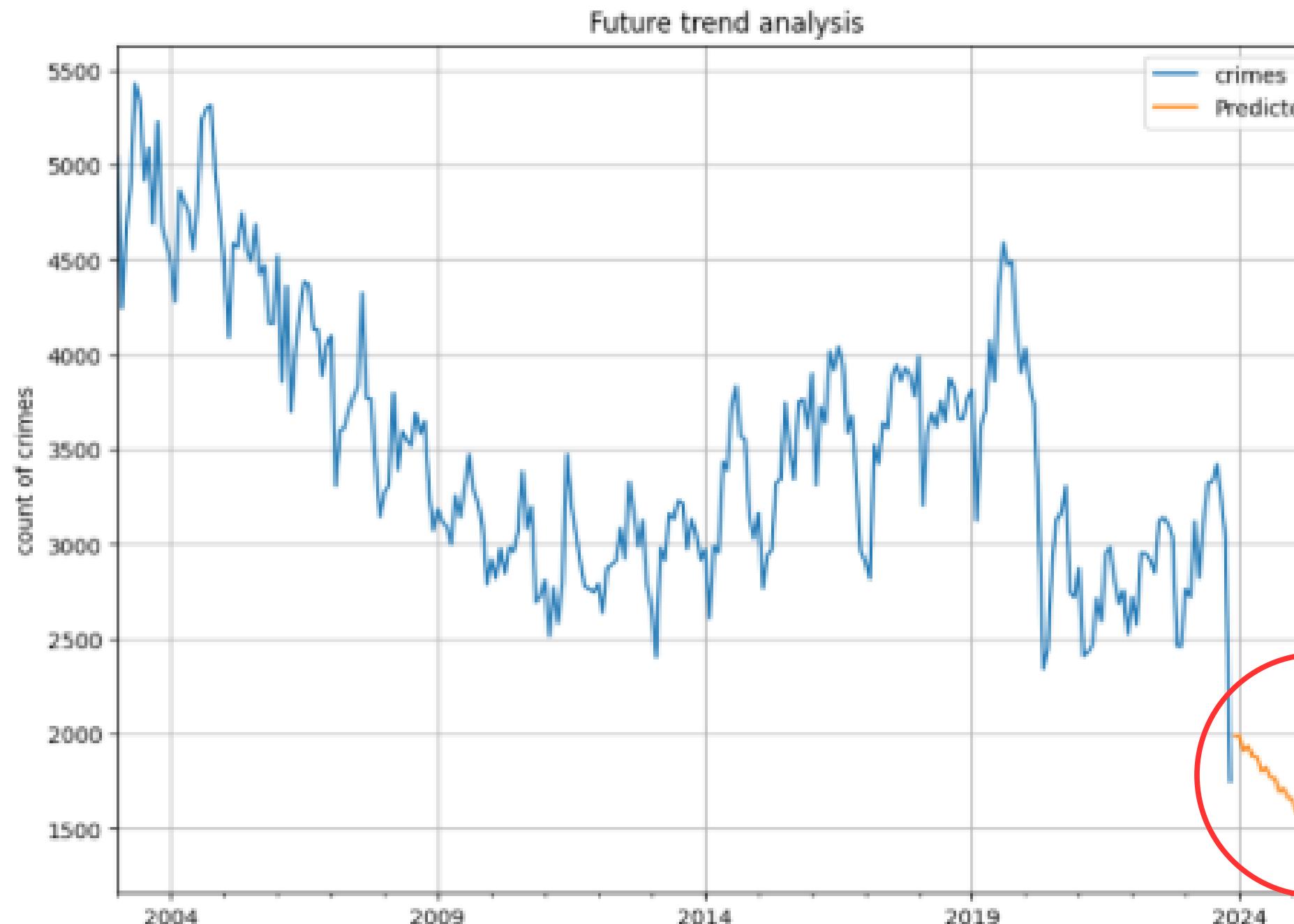
## WHAT CAN WE EXPECT IN TERMS OF CRIME IN VANCOUVER IN THE FUTURE?



- 2003–2010: Steady decline in crime rates.
- 2010–2019: Fluctuating trend with notable spikes, especially around 2019.
- Post-2019: Stabilization at lower levels compared to 2004.
- Peaks & Dips: Sharp spikes and dips indicate external factors/events affecting crime rates.
- Overall: Long-term decline but periods of fluctuation in recent years.

# DATA INTERPRETATION

## TIME-SERIES FORECASTING



- The data was analysed using a time-series approach, specifically the Holt-Winters forecasting method, to predict future crime trends in Vancouver
- the future crime trend is forecasted to decline significantly, indicating a reduction in crime rates in the coming years
- Seasonal trends revealed that crime rates were generally higher in summer months compared to winter

# DEPLOYMENT OF DATA PRODUCT

GROUP 8

Prediction of Vancouver's Criminal Activity

Home

Prediction  
Model

About Us

Home

The growing need to reduce crime activity is crucial for promoting peace and justice, in line with the Sustainable Development Goals (SDGs) that aim to foster inclusive societies and effective institutions. This page employs machine learning techniques, specifically using a Random Forest model, which is trained on historical crime data, to predict criminal activities. Random Forest is known for providing high accuracy in classification tasks, making it an effective tool for crime prediction. These predictions can significantly aid in cost reduction and resource optimization for law enforcement and public safety initiatives.

The dataset used for this modeling originates from the Vancouver Open Data Catalogue, consisting of 530,652 records spanning from January 1, 2003, to April 26, 2021.

Data sourced from the Vancouver Open Data Catalogue.

**The application integrates a trained Random Forest model with an accuracy of 55.08% and is built using the Flask web framework.**

# DEPLOYMENT OF DATA PRODUCT

GROUP 8

Prediction of Vancouver's Criminal Activity

[Home](#)

[Prediction Model](#)

[About Us](#)

## Prediction Model

This tool allows you to predict crime activities based on the geographical coordinates (easting and northing in UTM format, Zone 10). The trained Random Forest model, which has an accuracy of 55.08%, analyzes historical crime data to predict the likelihood of crime activity in the given area. By providing these inputs, the model helps optimize law enforcement resources and enhance public safety.

Please enter the easting and northing (in UTM format, Zone 10) to make a prediction.

UTM Easting

491117.2295

UTM Northing

5458138.764

[Reset](#)

[Submit](#)

Prediction: Vehicle Collision or Pedestrian Struck (with Injury)

**Intuitive web application designed to forecast potential criminal activities based on user-provided geographic inputs in the UTM coordinate system**

# INSIGHTS & CONCLUSION

- Theft from vehicles was the most common crime in Vancouver, followed by other theft and mischief
- Peak crime activity was noted during weekends, especially on Fridays and Saturdays
- The most crime-prone times were during the evening hours (6:00 PM to midnight), particularly for theft-related incidents.
- The crime rate in Vancouver is in a decreasing state from 2003 until 2010 but observed a sharp rise in 2019
- This study underscores the potential of machine learning in understanding and combating crime, providing a foundation for more targeted interventions and informed decision-making

# REFERENCE

- **References List**

- Catlett, C., Cesario, E., Talia, D., & Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*, 53, 62-74. <https://doi.org/https://doi.org/10.1016/j.pmcj.2019.01.003>
- Cesario, E., Lindia, P., & Vinci, A. (2024). Multi-density crime predictor: an approach to forecast criminal activities in multi-density crime hotspots. *Journal of Big Data*, 11(1), 75. <https://doi.org/10.1186/s40537-024-00935-4>
- Hajela, G., Chawla, M., & Rasool, A. (2020). A Clustering Based Hotspot Identification Approach For Crime Prediction. *Procedia Computer Science*, 167, 1462-1470. <https://doi.org/10.1016/j.procs.2020.03.357>
- Hossain, S., Abtahee, A., Kashem, I., Hoque, M. M., & Sarker, I. H. (2020, 2020//). Crime Prediction Using Spatio-Temporal Data. Computing Science, Communication and Security, Singapore.
- Ingilevich, V., & Ivanov, S. (2018). Crime rate prediction in the urban environment using social factors. *Procedia Computer Science*, 136, 472-478. <https://doi.org/https://doi.org/10.1016/j.procs.2018.08.261>
- Khan, M., Ali, A., & Alharbi, Y. (2022). Predicting and Preventing Crime: A Crime Prediction Model Using San Francisco Crime Data by Classification Techniques. *Complexity*, 2022, 1-13. <https://doi.org/10.1155/2022/4830411>
- Patterson, E. (2018). Poverty, income inequality, and community crime rates. *Criminology*, 29, 755-776. <https://doi.org/10.1111/j.1745-9125.1991.tb01087.x>
- Schleimer, J. P., Pear, V. A., McCort, C. D., Shev, A. B., De Biasi, A., Tomsich, E., Buggs, S., Laqueur, H. S., & Wintemute, G. J. (2022). Unemployment and Crime in US Cities During the Coronavirus Pandemic. *Journal of Urban Health*, 99(1), 82-91. <https://doi.org/10.1007/s11524-021-00605-3>
- Sundaram, S., Perumal, S., Parthiban, D., & Revathy, D. (2024). Crime Hotspot Analysis Using Six Layered Deep Recurrent Neural Networks 1. 2022. <https://doi.org/10.48047/intjccse/V14I8.309>
- Venter, Z. S., Shackleton, C., Faull, A., Lancaster, L., Breetzke, G., & Edelstein, I. (2022). Is green space associated with reduced crime? A national-scale study from the Global South. *Science of The Total Environment*, 825, 154005. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.154005>
- Yu, C. H., Ward, M. W., Morabito, M., & Ding, W. (2019, 11-11 Dec. 2011). Crime Forecasting Using Data Mining Techniques. 2011 IEEE 11th International Conference on Data Mining Workshops,

# **THANK YOU**

# DATA INTERPRETATION

## CLASSIFICATION REPORT

→ Classification Report:

	precision	recall	f1-score	support
0	0.38	0.28	0.32	9514
1	0.37	0.34	0.35	14548
2	0.00	0.00	0.00	50
3	0.38	0.27	0.32	21654
4	1.00	1.00	1.00	15392
5	0.63	0.68	0.66	44006
6	0.50	0.65	0.56	48594
7	0.21	0.08	0.11	7507
8	0.22	0.12	0.15	9050
9	0.19	0.05	0.07	64
10	0.88	0.80	0.84	5839
accuracy			0.55	176218
macro avg	0.43	0.39	0.40	176218
weighted avg	0.53	0.55	0.53	176218

Accuracy: 55.08%

RANDOM FOREST

→ Classification Report:

	precision	recall	f1-score	support
0	0.35	0.32	0.34	9514
1	0.36	0.39	0.37	14548
2	0.00	0.00	0.00	50
3	0.37	0.30	0.33	21654
4	1.00	1.00	1.00	15392
5	0.63	0.68	0.66	44006
6	0.51	0.62	0.56	48594
7	0.23	0.07	0.10	7507
8	0.22	0.09	0.13	9050
9	0.10	0.05	0.06	64
10	0.86	0.80	0.83	5839
accuracy			0.55	176218
macro avg	0.42	0.39	0.40	176218
weighted avg	0.53	0.55	0.53	176218

Accuracy: 54.96%

DECISION TREE