November 2020 | Carolina Santos

Forecasting Vancouver Crime

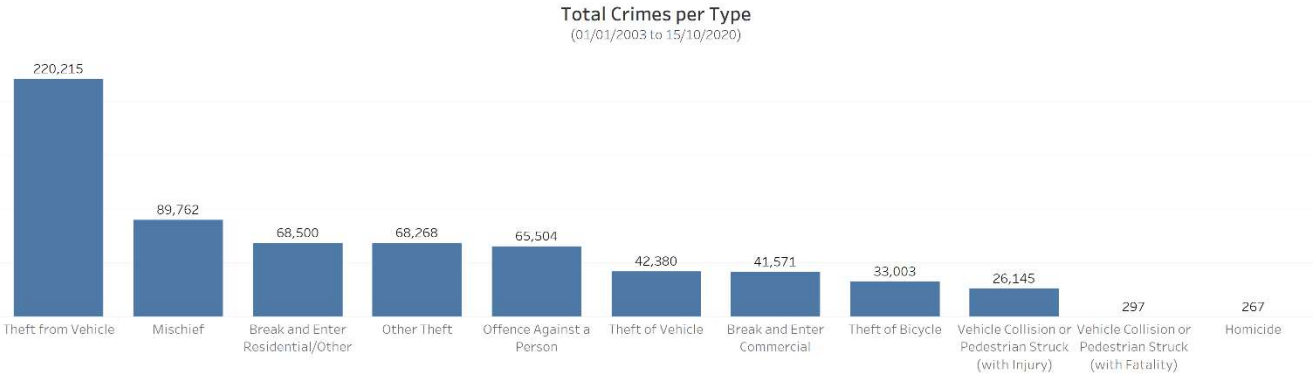
Introduction

The city of Vancouver, Canada, has a population of 675,218 (2017)[1] people and is considered the third best city in the world for quality of life[2]. The Vancouver Police Department (VPD) provides daily crime data intended to enhance community awareness of policing activity across several types of crimes. In this study we will use time series analysis to determine: ***is it possible to accurately forecast crime***?

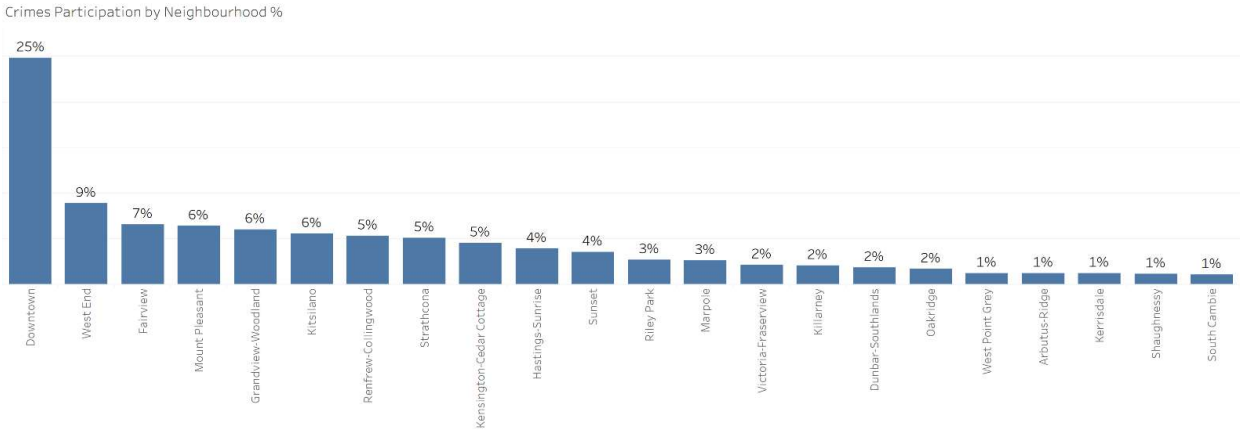
The current data available for visualization from the VPD GeoDash[3] is limited to 5,000 cases and it is static. By using the underlying data, we were able to provide insights about the trend on the number of crimes, seasonality, and forecast it until 31/12/2022.

Data Overview

The data used for this report carries information from January 1st, 2003 until October 15th, 2020, with details on the Type of Crime, Date and Time of occurrence and location. There was a total of 655,912 crimes recorded on the VPD Geodash Open Data with 9 features for this date range. On the chart below, we can see that the 3 most common crimes in the city are: Theft from Vehicle (34%), Mischief (14%), Break and Enter Residential (11%).



For the neighbourhood analysis, the (x, y) coordinates were transformed into geometries and then joined with the GeoJSON file that carries the neighbourhood polygons from the City of Vancouver Local Boundary data. By doing so we could correctly identify the location of crimes that did not have a neighbourhood identified.



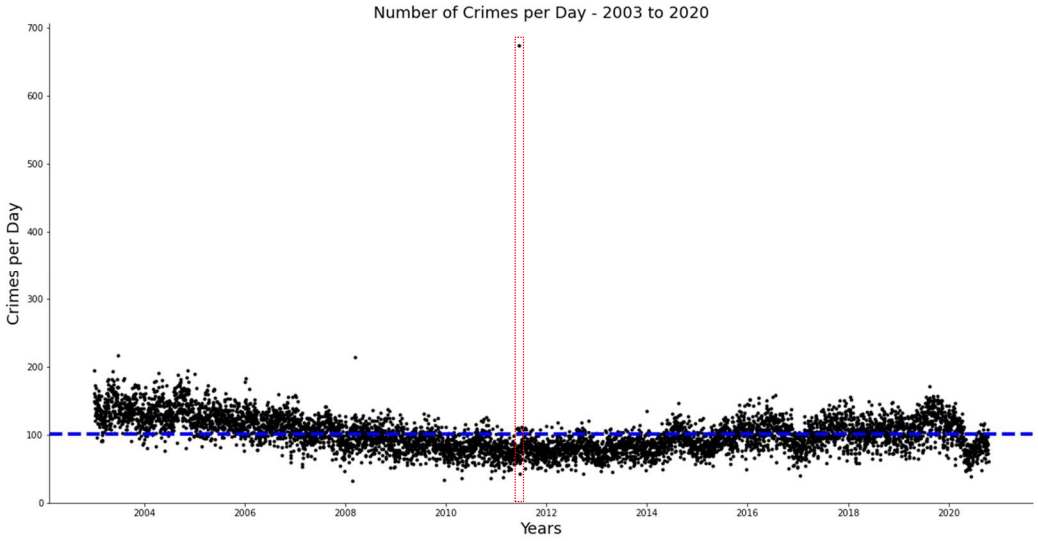
[1] Data Commons – Place Explorer <https://datacommons.org/place/wikidataId/Q24639>

[2] Vancouver ranks third best in the world for quality of life <https://globalnews.ca/news/5055257/vancouver-third-quality-of-life/>

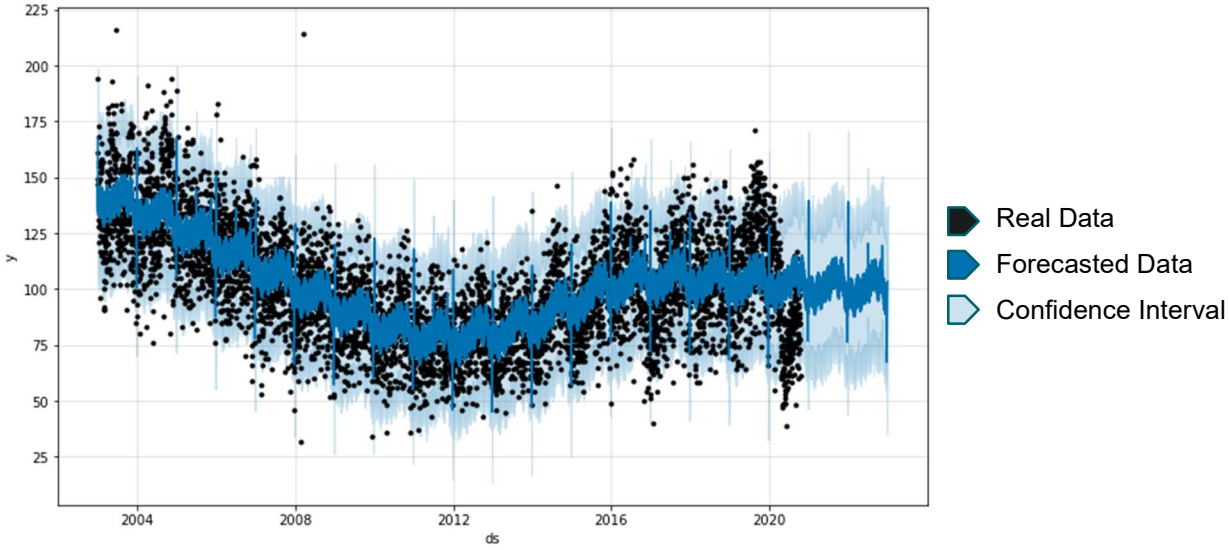
[3] Geodash Map <https://geodash.vpd.ca/Html5Viewer/?disclaimer=on&viewer=VPDPublicRefresh_gvh&x=144&y=51>

Forecasting

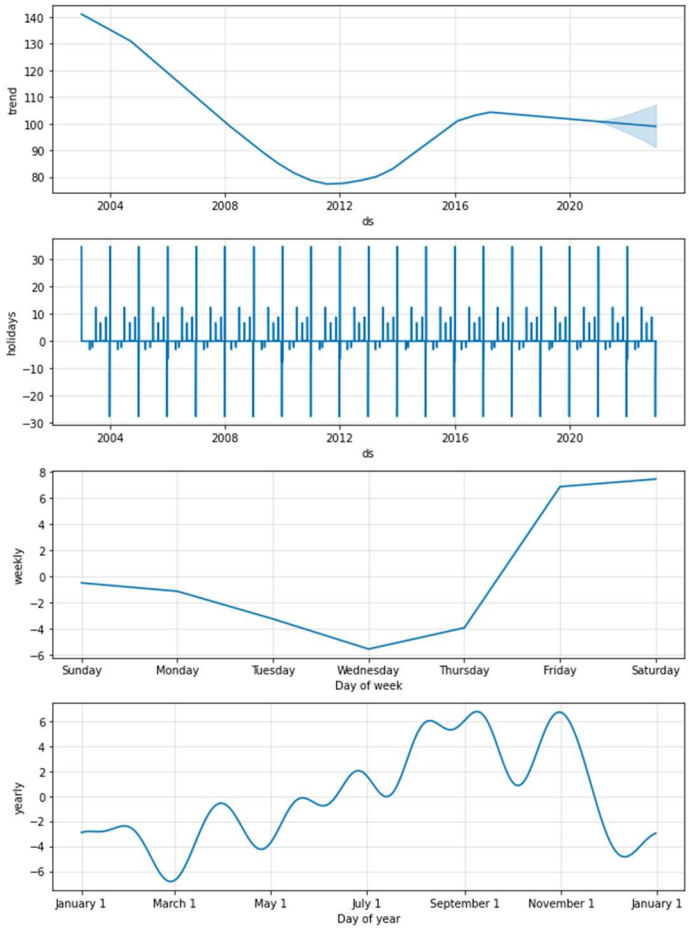
For the forecasting, we grouped the number of daily crimes and the date when they occurred and plotted to identify outliers. In the chart below, we highlighted June 15th, 2011, which was the day of the Stanley Cup riot that broke out almost immediately after the conclusion of the Boston Bruins' win over the Vancouver Canucks in game seven of the 2011 Stanley Cup Finals. At least 140 people were injured during the incident, including 1 critically.



The historic daily average crimes in Vancouver is around 100, so having close to 700 in one day is a clear outlier which was removed from the dataset to not affect the forecast. The forecast spans from 16/10/2020 to 31/12/2022 and was modeled using a python package called Prophet, an open source software released by Facebook's Core Data Science team. Prophet works best with time series that have strong seasonal effects and several seasons of historical data. The historical and the forecasted data is as follows:



The components of the forecasting chart are:



Trend

There is a downward trend since 2003, with an increase between 2013 to 2016 and stabilizing around 100 cases per day.

Holidays

Canadian holidays are included in the forecasting calculation showing that lower number of crimes occur around December 24th and 25th (Christmas) and higher numbers on December 31st (New Year's Eve)

Weekly Seasonality

Considering the weekly seasonality, Friday and Saturday are the most popular days for crimes, and the opposite from Tuesday to Thursday.

Yearly Seasonality

Considering the yearly seasonality, the second half of the year shows a higher number of crimes than the first half.

Results

We used time series cross validation to measure forecast error using historical data, which is provided by the Prophet package. This is done by selecting cutoff points in the history and for each of them fitting the model using data only up to the cutoff point. We then compare forecasted values and actual values.

Measuring the mean absolute percent error (MAPE), the result was a 12.95% error for forecasts of 37 days, and a 14.89% error for forecasts of 365 days. We can conclude that the forecast has a low error and that the occurrence of crimes in Vancouver can be accurately forecasted, which was a surprise given the random nature of the data.

horizon	mape	coverage
37 days	12,95%	95,38%
365 days	14,89%	92,69%

Next Steps

The results of this study can be used for VPD resource management and for community awareness. For future steps we are looking to fit a Holt-Winters Exponential Smoothing as an alternative forecasting method.