

# Crime in Vancouver

Presented By:  
“Data Miners”

Aakash Patel [ak553155@dal.ca]

Bhumi Patel [bh842792@dal.ca]

Ganrong Tan [gn485145@dal.ca]

Sampark Pradhan [sm459977@dal.ca]

# Introduction

- find pattern and frequency of different type of crimes in specific neighbourhood.
- help to reduce the crime incidents.
- Guide to newcomers to find safest places to stay.

# Seminar Topics

1. Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data.
2. Narrowing the Gap: Random Forest in Theory and in Practice.
3. Time-varying Data Visual Analysis method based on Parallel Coordinate System

# 1. Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data.[2]

- Problem : Create helpful application from raw crime data using Data science and visualization.
- Relevance : Data analysis to identify crime patterns in particular area and correlation with time. Help to create project outline
- Solution : Data preprocessing, Different visualization tricks, different prediction models.
- Limitation : Low Scalability.
- Future work : Implement wide range applications with various dataset, visualization with multivariate graph.

## 2. Narrowing the Gap: Random Forest in Theory and in Practice[3].

→ Problem : Understand the workability and improve on tractability of Random Forest Classifier

→ Relevance : Working with different models and comparing their relative performance RF proved to be best among all.

→ Solution : Data preprocessing, Different visualization tricks, different prediction models.

→ Limitation : Random Forest proves to be biased to labels most occurring in the tree. It's prone to overfitting.

→ Future work : Work more on consistency and perform more theoretical analysis and empirical study to clarify current understanding of RF Classifier.

### 3. Time-varying Data Visual Analysis method based on Parallel Coordinate System[1]

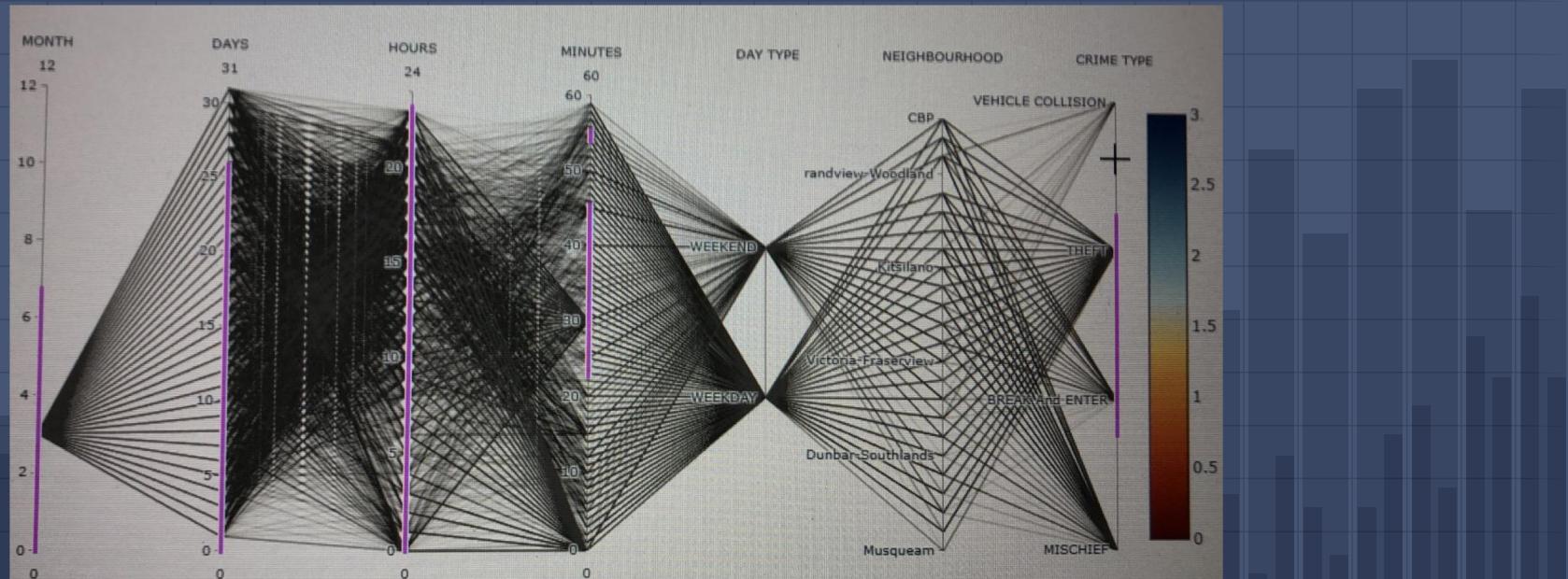
→ Problem : Trends identification for large scale, multidimensional time variance data.

→ Relevance : Proposed better visual approach of processing Multidimensional time varying data. Can relate to effective crime occurrence exploration across years to the respective month, days. Potential for trends identification based on timeliness of crime events and their types.

→ Solution : A three step approach based on Parallel Coordinate System.

- Reducing Dimensions of MD TV data
- Clustering after dimension reduction
- Visualization of Time-varying parallel Coordinate System  
Based on Edge Binding

- Limitation : Doesn't help much after size increases after certain amount.
- Future Work: Complex spatiotemporal correlation model of multidimensional data for further understanding and analysis about the mining of data implicit patterns.



# Dataset Recall

Source: Crime data extracted from the Vancouver Open Data Catalog [1].

- Size: 530,652 records from 2003-01-01 to 2017-07-13.
- Information: Crime occurrence depicted across exact locations of city over the duration of 14 years along with timestamp.

# Dataset Attributes

TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_B	NEIGHBOUR_X	Y	Latitude	Longitude	
Other Theft	2003	5	12	16	15	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Other Theft	2003	5	7	15	20	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Other Theft	2003	4	23	16	40	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Other Theft	2003	4	20	11	15	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Other Theft	2003	4	12	17	45	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Other Theft	2003	3	26	20	45	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Break and Er	2003	3	10	12	0	63XX WILTSI	Kerrisdale	489325.58	5452817.95	49.2280508	-123.14661
Mischief	2003	6	28	4	13	40XX W 19T	Dunbar-Sout	485903.09	5455883.77	49.2555592	-123.19373
Other Theft	2003	2	16	9	2	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Break and Er	2003	7	9	18	15	18XX E 3RD /	Grandview-V	495078.19	5457221.38	49.2677339	-123.06765
Other Theft	2003	1	31	19	45	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376
Mischief	2003	9	27	1	0	40XX W 21S	Dunbar-Sout	485852.96	5455684.11	49.253762	-123.19441
Break and Er	2003	4	19	18	0	18XX E 3RD /	Grandview-V	495093.69	5457230.31	49.2678143	-123.06744
Break and Er	2003	9	24	18	30	18XX E 3RD /	Grandview-V	495103.82	5457221.02	49.2677308	-123.0673
Break and Er	2003	11	5	8	12	63XX WINDS	Sunset	493790.48	5452630.9	49.2264298	-123.08528
Break and Er	2003	9	26	2	30	10XX ALBER	West End	491067.65	5459114.22	49.2847148	-123.12282
Break and Er	2003	10	21	10	0	18XX E 3RD /	Grandview-V	495119.32	5457229.95	49.2678113	-123.06709
Other Theft	2003	1	25	12	30	9XX TERMIN	Strathcona	493906.5	5457452.47	49.269802	-123.08376

# Development cycle

1. Data Analysis and Data preprocessing

2. Exploratory Data Analysis

3. Data Modeling.

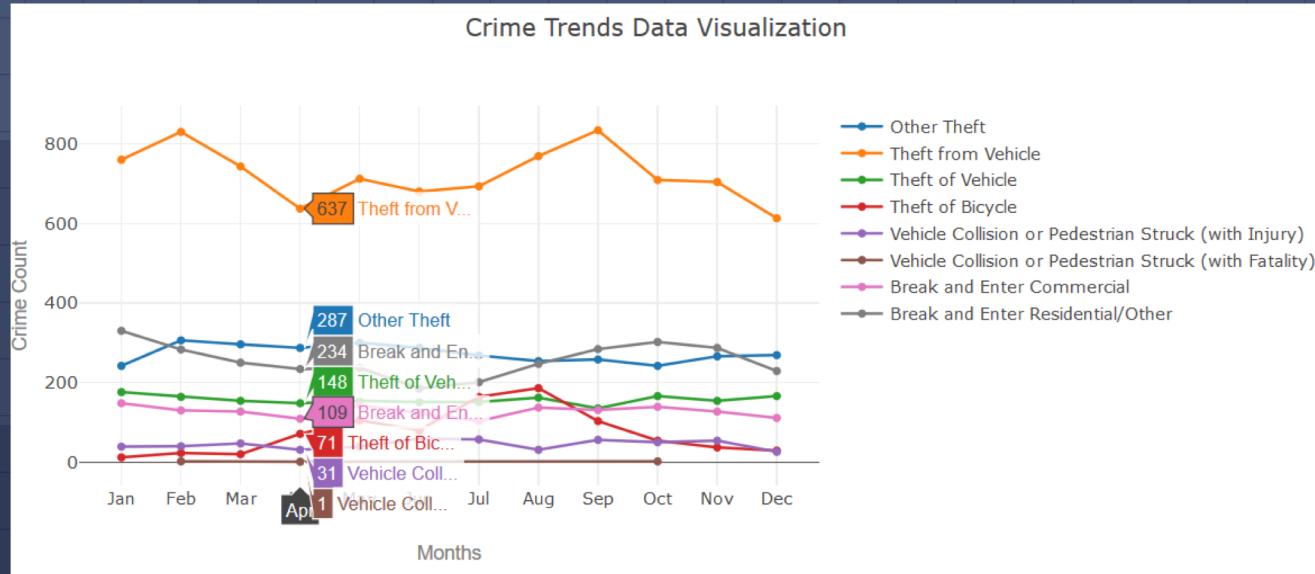
4. Model Visualization.

# Preprocessing and Analysis

- Cleaning :
  - Bad Data
  - NAN
  - Outlier Identification
- Attribute Analysis
  - Relevance comparison Ex: Min, 'X', 'Y'
  - Preprocessing with Importance to the context
    - Categorizing sub 'vehicle collision' categories to single one
    - Preprocessing for categorizing Hundred\_Block attribute

# Exploratory Data Analysis

- Crime Trends Data Visualization



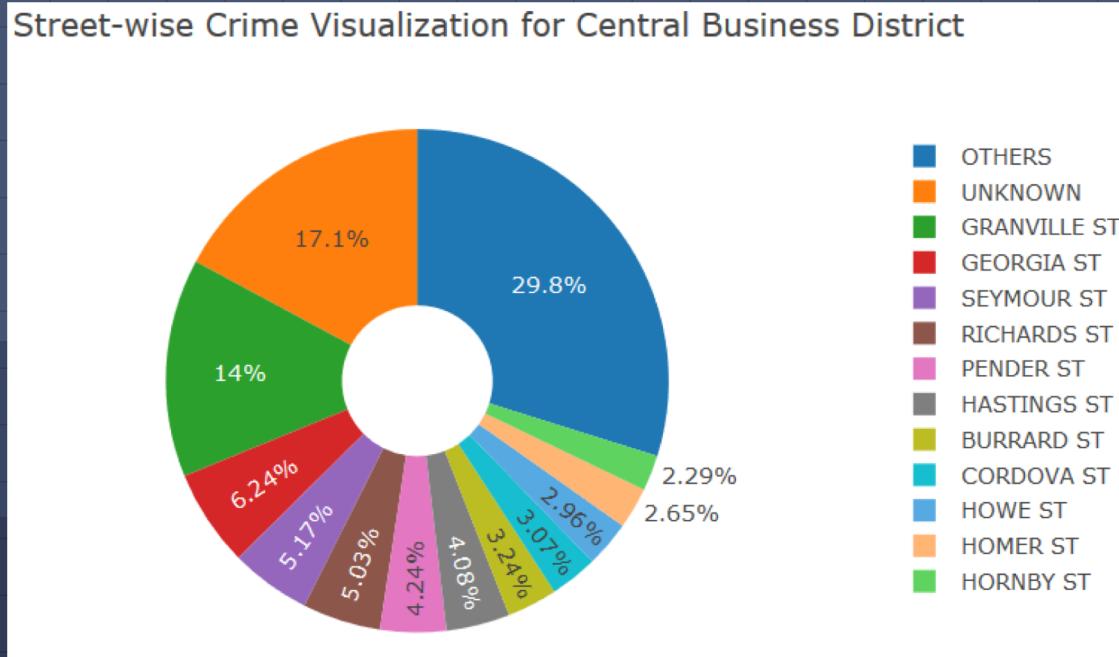
# Exploratory Data Analysis

- Neighbourhood Crime Data Visualization



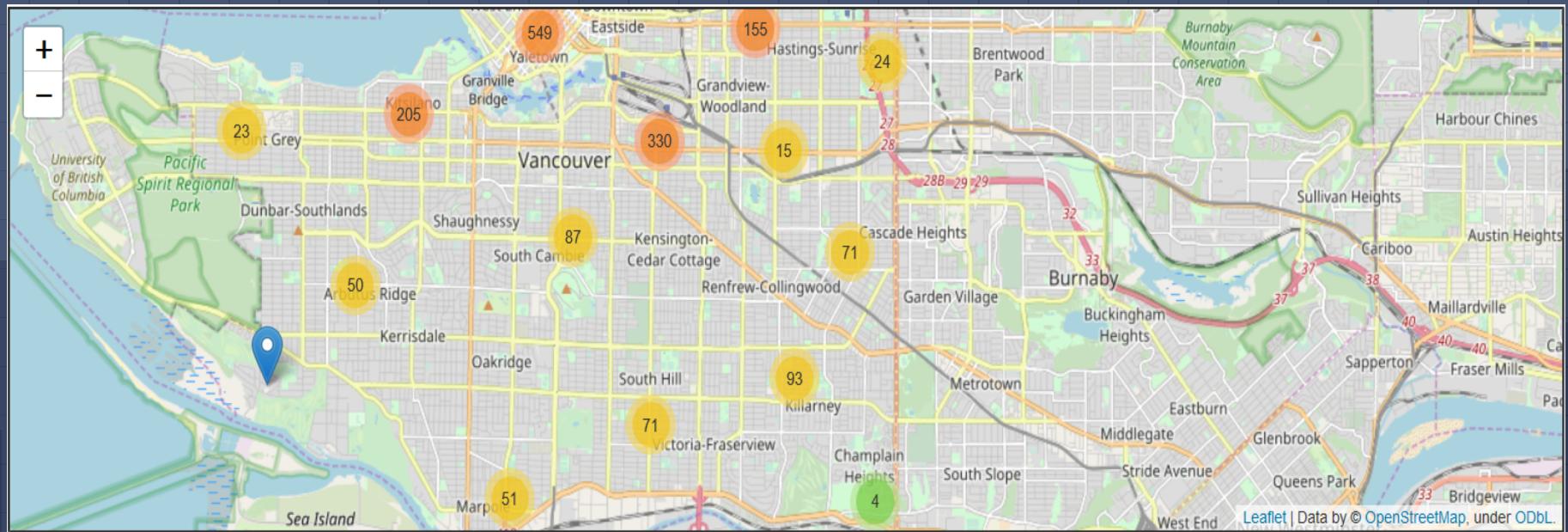
# Exploratory Data Analysis

- Streetwise crime analysis of Neighbourhood

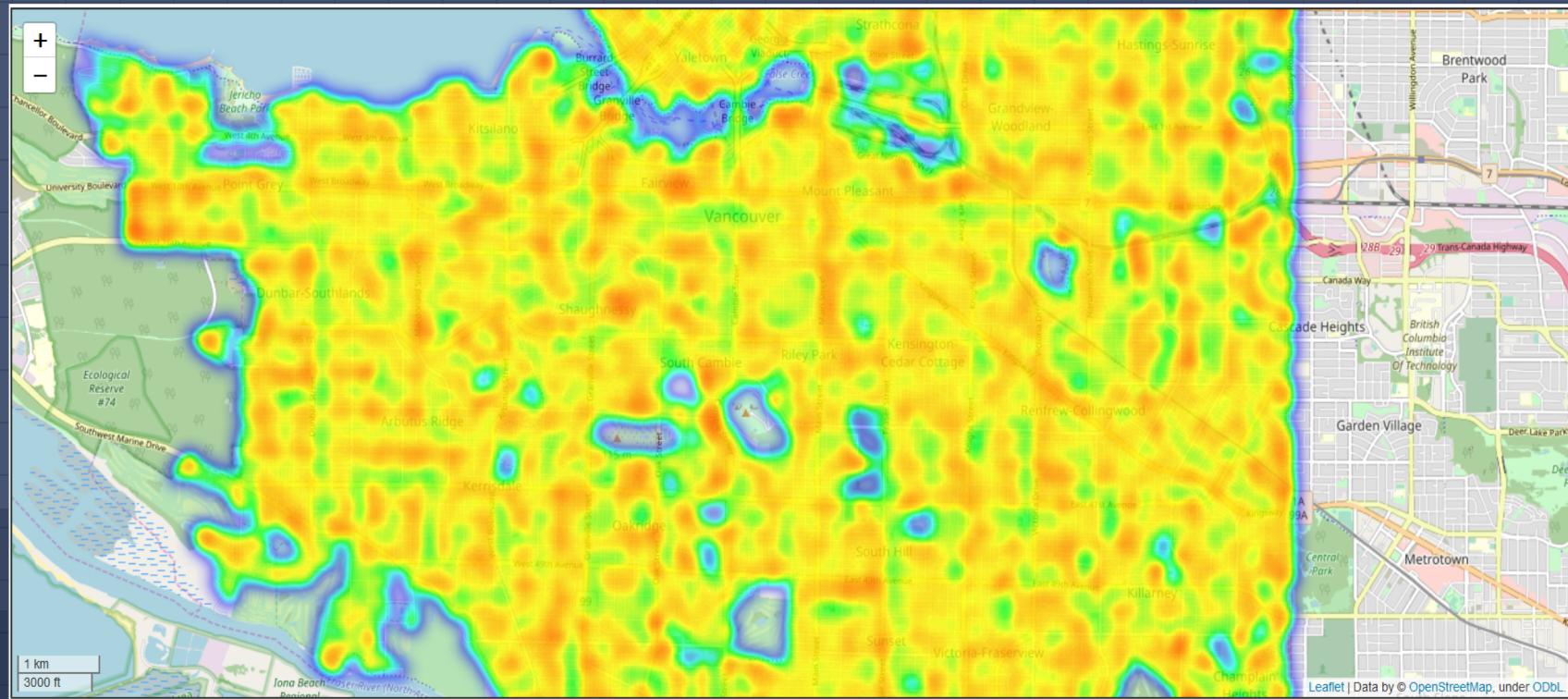


# Crime Zone Visualizations

- We aimed to visualize importance of crime zones across the city according to the crime count.



# Crime Zone Severity Analysis



# Project Demo

# Challenges and Bottlenecks

- Working with Dash to render web app.
- Charting every graph into single webapp.
- Integrating global intractability across all the charts.
- Overcoming Limitations of Dash for loading large amounts of data.
- Deployment across cloud options.
- Porting everything in Flask and Bringing everything together.
- Depicting Modelling Visualization into appropriate format.
- Optimizing on Modelling time efficiency.
- Transforming Vbox widget to integrate in Flask Frontend.

# Reference

1. Zihui Wang, E Haihong, Meina Song, and Zhijun Ren. 2019. Time-varying Data Visual Analysis method based on Parallel Coordinate System. In 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC).  
<https://doi.org/10.1109/itnec.2019.8728990>.
2. Mingchen Feng, Jiangbin Zheng, Jinchang Ren, Amir Hussain, Xiuxiu Li, Yue Xi, and Qiaoyuan Liu. 2019. Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data. *IEEE Access* 7: 106111–106123.
3. Denil, M., Matheson, D. & De Freitas, N.. (2014). Narrowing the Gap: Random Forests In Theory and In Practice. *Proceedings of the 31st International Conference on Machine Learning*, in PMLR 32(1):665-673

THANKS!