Schulich School of Business
York University

Police Division Efficiency Improvement

Team 13: Pengbo Gao (220307831), Jinlun Zhang (220668810), Jingze Zhang (220240248), Yaosheng Liu (216173163), Yisong Lu(216379927)

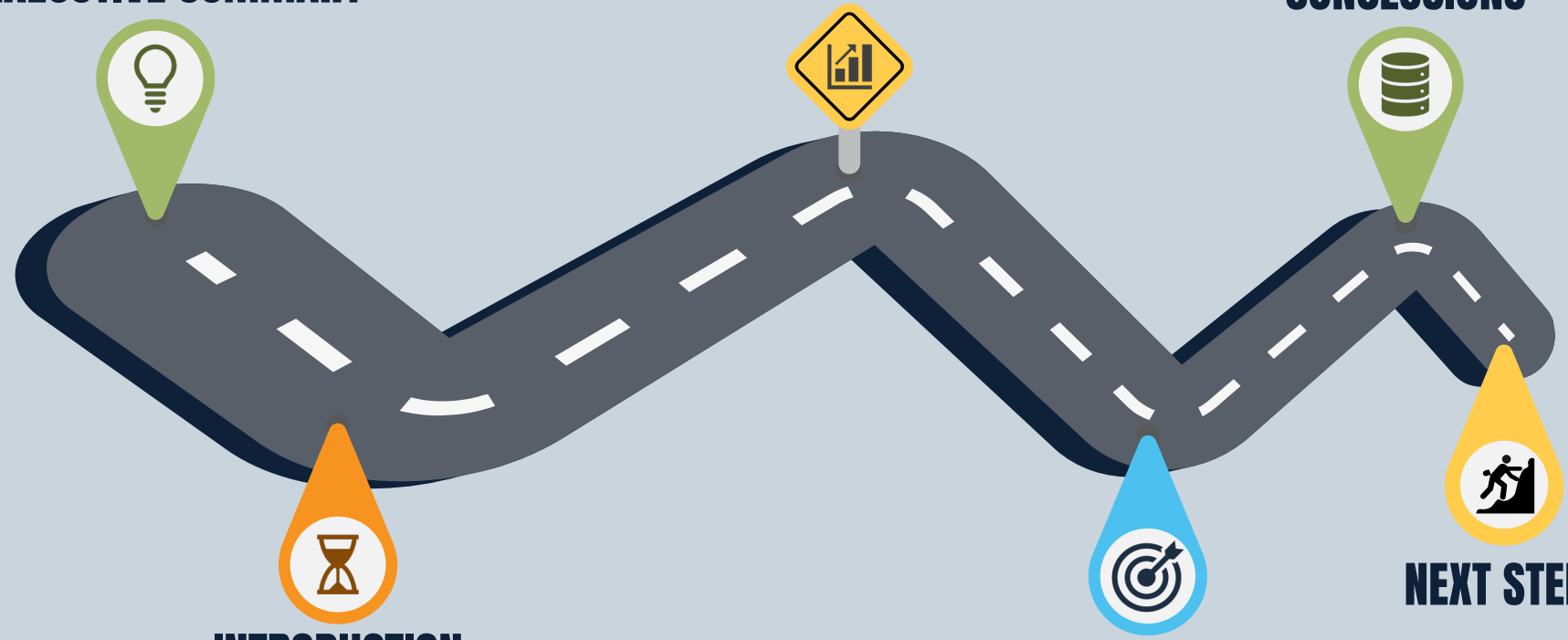
"The work we submitted as part of this assignment is original, and due credit is given to others where appropriate. We accept and acknowledge that that each member of our team is equally responsible for if the assignment found to be plagiarized in any way, and we will be subject to school's Academic Integrity policy"

AGENDA

EXECUTIVE SUMMARY

VISUAL ANALYSIS

CONCLUSIONS



INTRODUCTION

MODELING ANALYSIS

NEXT STEPS

EXECUTIVE SUMMARY

- The project analyzes incidents in various neighborhoods across Toronto using visualizations and clustering models, spotlighting crime "hot spots" and crime density.
- Nighttime sees more crime, indicating the need for targeted policing and optimized resource allocation.
- The elbow method determines optimal clusters, aiding in jurisdiction assignments and resource management.
- Clustering results identify high-incident divisions and assign jurisdictions, and possible multi-cluster divisions may suggest the need for additional inspections.
- Comparison of K-means model results reveals patterns in crime spatial distribution and the effectiveness of different policing strategies.
- Despite its simplicity, the K-means model outperforms Spectral Clustering for this specific analysis.

INTRODUCTION

- Toronto has been dealing with varying incidents of crime, particularly Shooting and Firearm Discharges (SFD), dispersed throughout the city, leading to a necessity for a comprehensive analysis.
- The Toronto Police Service is seeking ways to optimize its resources, enhance operational efficiency, and minimize serious incidents, necessitating a data-driven approach.
- There's a need to understand the spatial distribution of these incidents and the effectiveness of law enforcement measures, focusing on different times of day and 'hot spots' of criminal activity.
- The problem at hand is to use clustering algorithms to find the optimal number of clusters of SFD incidents in Toronto neighborhoods, then assign the jurisdiction of the best-suited division to each cluster.
- The goal is to optimize police budget and manpower allocation, enhance operational efficiency, and minimize serious incidents.
- The challenge involves effectively implementing these algorithms for appropriate group assignments and comprehensive, effective policing across Toronto.

VISUAL ANALYSIS

Hot Spots

Areas with high concentrations of incidents and fatalities are indicated by strong warm colors.

High Risk Communities

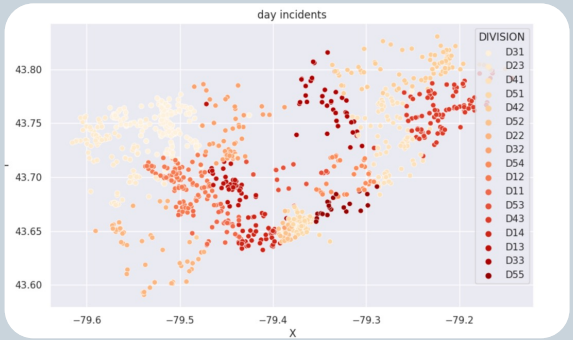
The heat map allows us to identify high-risk communities with high numbers of serious incidents resulting in fatalities.

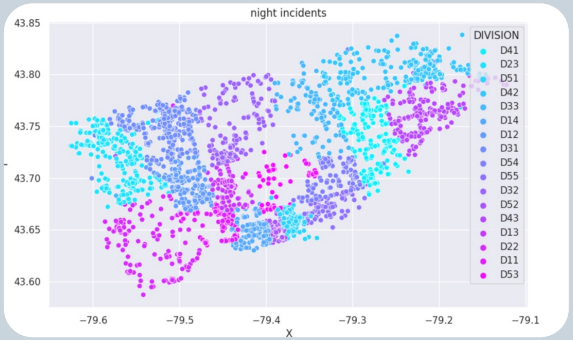
Spatial patterns

Observing incidents' spatial distribution and severity can identify patterns and trends.

Patterns and clusters of nighttime incidents may differ from those observed during the day, which may indicate differences in criminal activity and security issues between daytime and nighttime.

These two maps provide insight into the spatial distribution of nighttime incidents in different subsectors of Toronto.

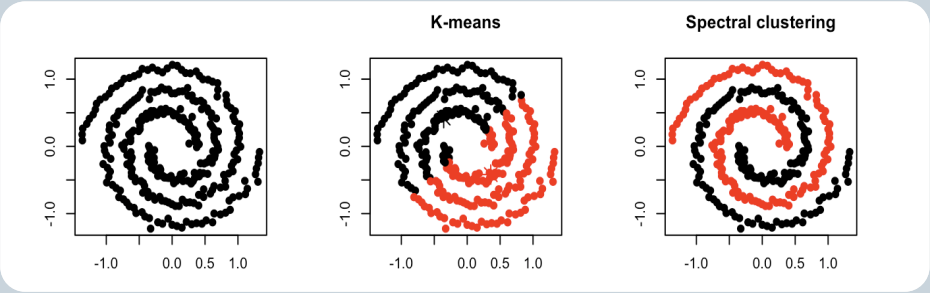




Modeling & Analysis | Feature Selection

Only select the longitude (X) and latitude (Y) as the training features

- ❑ Treat each SFD incident equally regardless of the number of injuries or deaths involved in each incident
- ❑ Clustering result purely based on the geographic distribution of the SFD incidents
- ❑ No data standardization/normalization



Modeling & Analysis | Model Training

1. Train 2 different clustering models: K-mean and Spectral Clustering

2. Visualize the clustering results produced by the 2 models

3. Compare their performances based on the evaluation metric score: Silhouette score

- ❑ The optimal number of clusters is determined based on the elbow method and the evaluation scores of the clustering results
- ❑ The number of clusters in this context is a hyper-parameter that has to be tuned.
- ❑ Validation / Testing set is not mandatory to be obtained and utilized

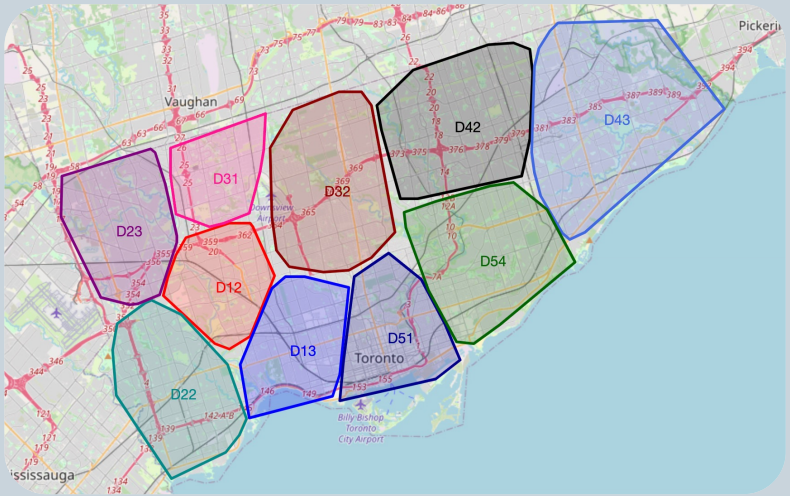
Modeling & Analysis | Model Training

K-mean on the nighttime incidents

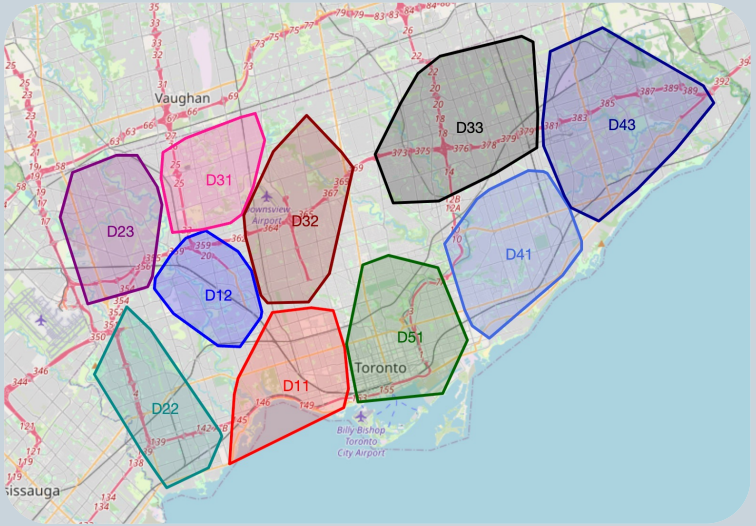
- ❑ Optimal Number of Cluster k = 10
- ❑ Identify the 10 critical divisions among the given 17 divisions.

K-mean on the daytime incidents

- ❑ Exactly the same conclusion with K-mean on nighttime incidents -> Optimal Number of Cluster k = 10
- ❑ Identify the 10 critical divisions among the given 17 divisions.



The underlying distribution patterns between the incidents in the 2-time ranges may resemble.

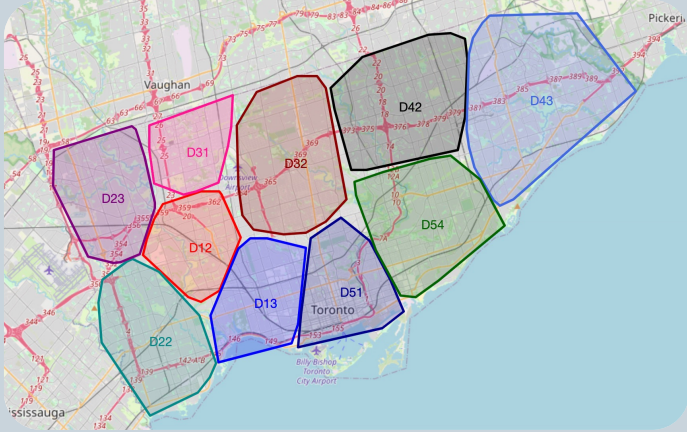


Modeling & Analysis | Model Training

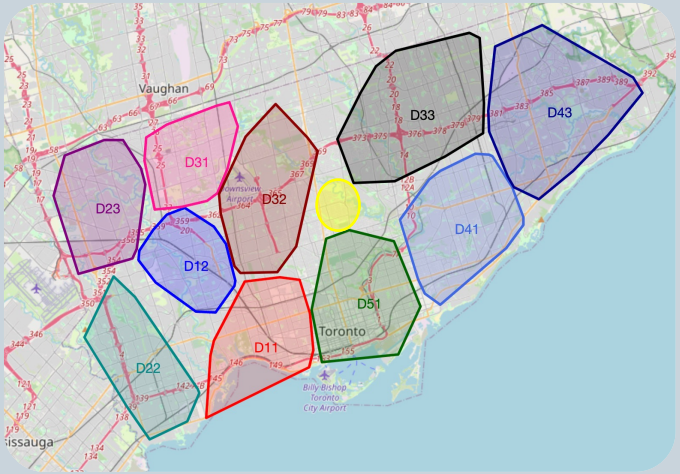
Daytime clustering result is less-condensed than nighttime clustering

- ❑ The regions that are not covered by any division (i.e. the one highlighted in the yellow circle on the Day Time ring) can be less emphasized/patrolled on as those regions have historically been ‘SFD-free’.
- ❑ No repeated division for multiple clusters in either the Nighttime Clustering or the Daytime Clustering,
- ❑ Allocate sufficient funds and manpower to the critical divisions that are assigned to both nighttime and daytime clusters as those divisions are essential for the regional security and residents’ safety.
- ❑ The budgets and manpower of the divisions that are not assigned to either the daytime or nighttime clusters can be reduced to increase the cost-efficiency of police budgets.

Divisions in Both Night & Day Time Clusters (need more manpower and budgets)	Divisions Not in Either Night / Day Time Clusters (need to be 'optimized' or restructured to boost overall police division operation efficiency)
D31, D51, D23, D12, D32, D43, D22	D14, D52, D55, D53



Night Time Clustering



Day Time Clustering

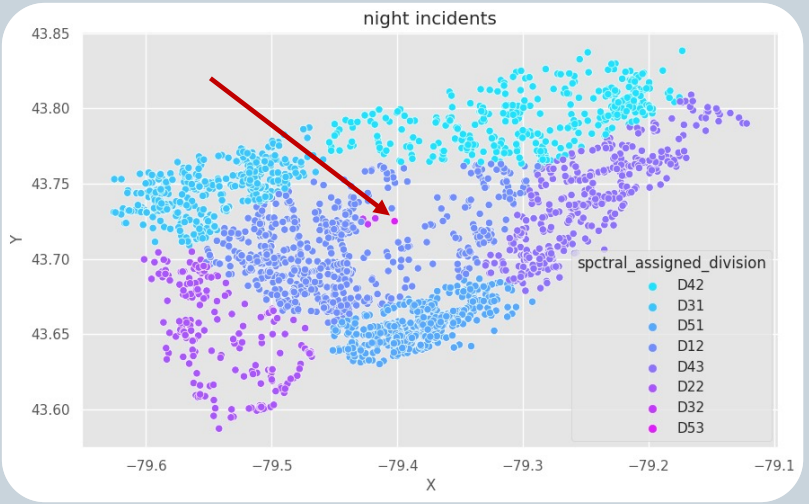
Modeling & Analysis | Model Training

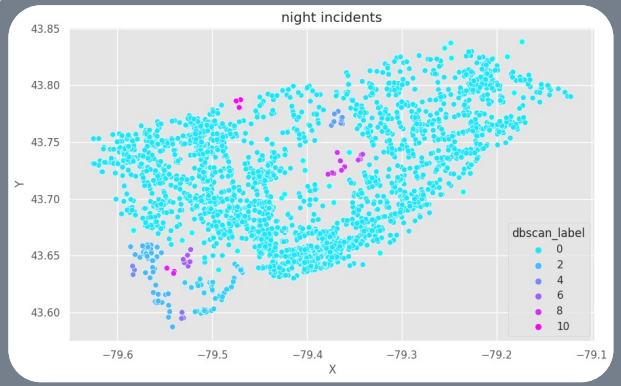
Schulich School of Business, York University

Spectral Clustering model (on nighttime incidents):

- Unlikely to give a useful result like the K-mean can yield, because the spectral clustering performance is worse than K-mean once the number of clusters is larger than 3.
- Exist clusters with only one instance in it (i.e. the cluster assigned to division D53 as highlighted by the red arrow), further demonstrating the deficiency of the clustering result produced by Spectral Clustering.
- A simple and computationally cheap algorithm (K-mean) can yield better results than sophisticated and computationally expensive models depending on our business context and objectives.

```
For n_clusters = 2, silhouette score is 0.5331476026460312
For n_clusters = 3, silhouette score is 0.444528033469475
For n_clusters = 4, silhouette score is 0.33921824580586746
For n_clusters = 5, silhouette score is 0.3071028605278935
For n_clusters = 6, silhouette score is 0.2902155174018957
For n_clusters = 7, silhouette score is 0.20480388869298072
For n_clusters = 8, silhouette score is 0.1918115314360118
For n_clusters = 9, silhouette score is 0.17899858818058415
For n_clusters = 10, silhouette score is 0.17533593433584937
For n_clusters = 11, silhouette score is 0.17212187505898308
For n_clusters = 12, silhouette score is 0.14642462918819518
For n_clusters = 13, silhouette score is 0.15828166922492387
For n_clusters = 14, silhouette score is 0.15579156854007997
```





Extra: Training DB Scan

- The model performs very poorly for this task without lengthy hyper-parameter tuning
- Most data points cannot be assigned to a cluster (as indicated by the blue colour, label 0)

CONCLUSIONS

- In conclusion, this report provides a comprehensive analysis and visualization of Toronto incidents, highlighting high-incident "hot spots" and emphasizing the need for targeted law enforcement.
- Using clustering models and silhouette scores, optimal clusters for daytime and nighttime incidents are determined, allowing for effective jurisdiction assignments and resource allocation.
- Comparison of K-means clustering results for different times of day offers valuable insights into spatial distribution patterns and patrol intensity requirements.
- These findings enable Toronto Police to optimize resource and personnel allocation, adopt data-driven safety measures, and demonstrate the effectiveness of simpler models like K-means.
- Overall, the analysis enhances police efficiency, facilitates effective resource allocation, and aids proactive crime addressing, improving public safety.

NEXT STEPS

❑ **Continuous monitoring**:

Implement a real-time monitoring system to continuously observe the effectiveness of newly adopted strategies. This involves tracking the incidence rates by area and making required modifications.

❑ **Further model development**:

Exploring other machine learning models or further adapting existing models. Merge additional features or data sources that can enhance the clustering capabilities.

❑ **Regular reporting**:

Establish a regular reporting system to inform stakeholders of the progress and effectiveness of implemented strategies.

Thank You!