

# Seattle Crime

By Alejandro Harrison

# Business Goal

- ▶ Seattle one of top 50 largest US cities
- ▶ One of the fastest growing in US
- ▶ Seattle recently ranked among least safest cities in US
- ▶ Make city and people feel safer
- ▶ Generate positive image
- ▶ Preventative action
- ▶ Classify/predict crimes as they occur

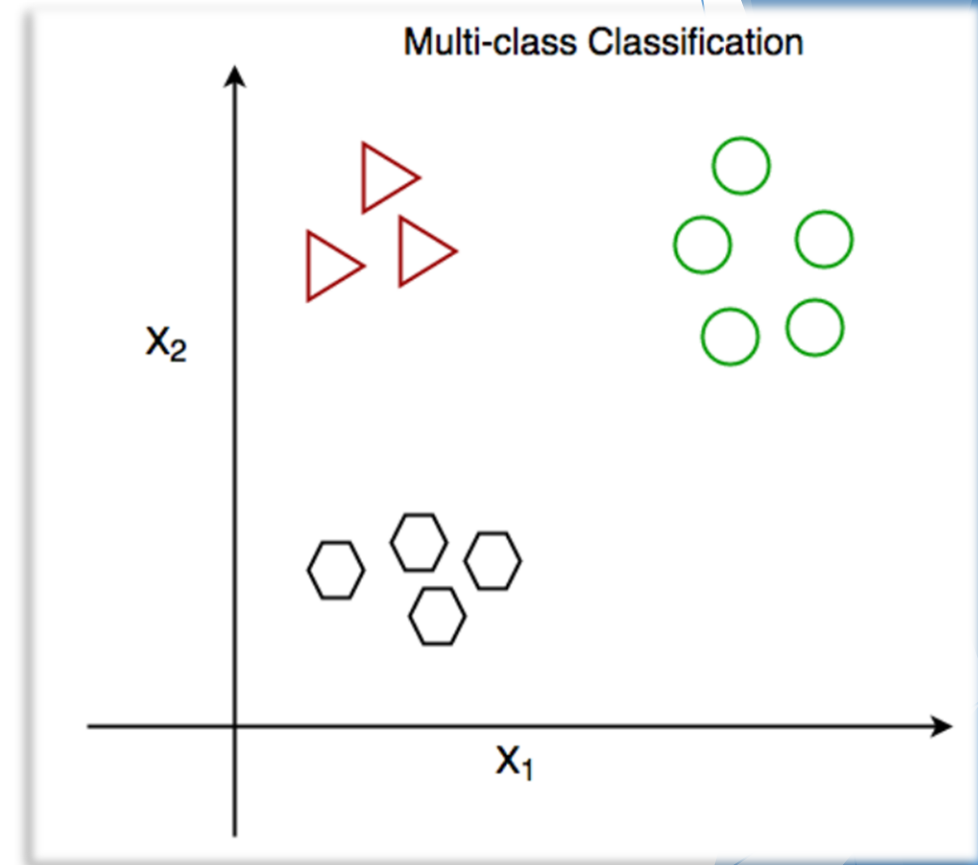


# Method

- ▶ Use of Multiclass Classification models
- ▶ Pick best one
- ▶ Able to classify types of new crimes
- ▶ New crime location
- ▶ What is Multiclass Classification?

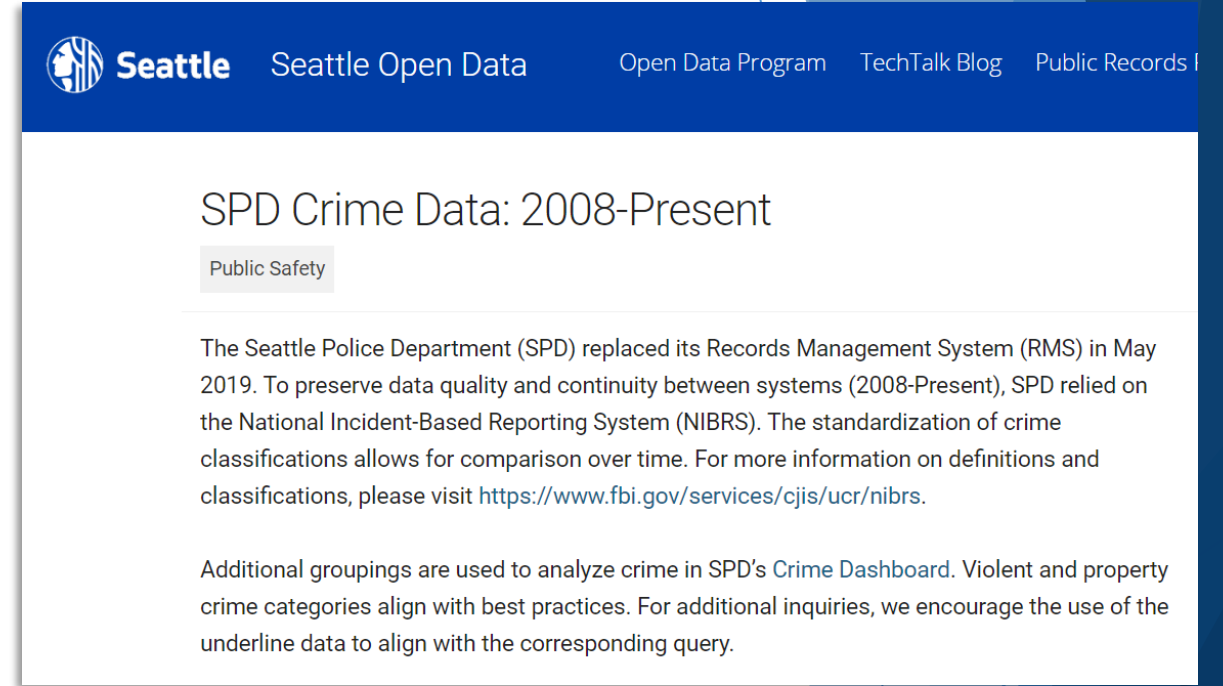
# Multiclass Classification

- ▶ Process of dividing data into classes/categories
- ▶ More than 2 categories
- ▶ Can be used to predict future values
- ▶ Crime type and location
- ▶ Crime data



# The Data

- ▶ City of Seattle's site (Seattle.gov)
- ▶ SPD crime dataset(2008-present)
- ▶ Publicly available
- ▶ Over 1 million rows
- ▶ Crime date, crime category, crime location, etc
- ▶ Data categorical



The screenshot shows the Seattle Open Data website. The header is dark blue with the Seattle logo and navigation links: "Seattle", "Seattle Open Data", "Open Data Program", "TechTalk Blog", and "Public Records". The main content area is white. The title "SPD Crime Data: 2008-Present" is displayed, followed by a "Public Safety" tag. The text describes the replacement of the Records Management System (RMS) in May 2019 by the National Incident-Based Reporting System (NIBRS) to preserve data quality and continuity. It mentions that the standardization of crime classifications allows for comparison over time and provides a link to the FBI's NIBRS page. A final paragraph states that additional groupings are used to analyze crime in the SPD's Crime Dashboard, noting that violent and property crime categories align with best practices and encouraging the use of the underlying data for queries.

Seattle Seattle Open Data Open Data Program TechTalk Blog Public Records

## SPD Crime Data: 2008-Present

Public Safety

The Seattle Police Department (SPD) replaced its Records Management System (RMS) in May 2019. To preserve data quality and continuity between systems (2008-Present), SPD relied on the National Incident-Based Reporting System (NIBRS). The standardization of crime classifications allows for comparison over time. For more information on definitions and classifications, please visit <https://www.fbi.gov/services/cjis/ucr/nibrs>.

Additional groupings are used to analyze crime in SPD's [Crime Dashboard](#). Violent and property crime categories align with best practices. For additional inquiries, we encourage the use of the underlying data to align with the corresponding query.

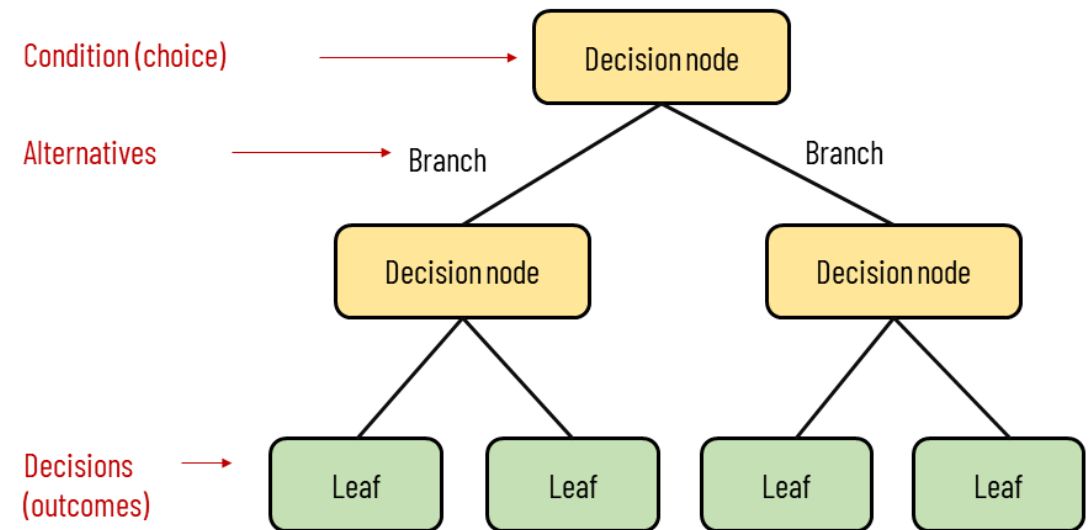
# The Models

- ▶ 2020-2023
- ▶ Built two separate models
- ▶ Used two different algorithms
- ▶ Built initial models then improved
- ▶ CatBoost (gradient boosted decision tree)
- ▶ Random Forest

# CatBoost

- ▶ Gradient boosted Decision Tree
- ▶ Used when data is categorical
- ▶ Decision tree splits data into series of decisions
- ▶ At end, final decision reached and data is categorized
- ▶ Decision Trees prone to overfitting

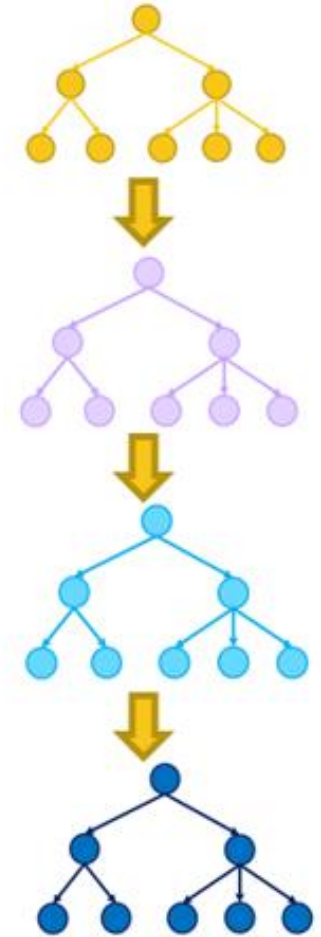
## Elements of a decision tree



# CatBoost Continued

- ▶ Reduce overfitting with gradient boosting
- ▶ Adding series of decision trees, each one improves upon last
- ▶ Great default models
- ▶ Train on computer part called GPU

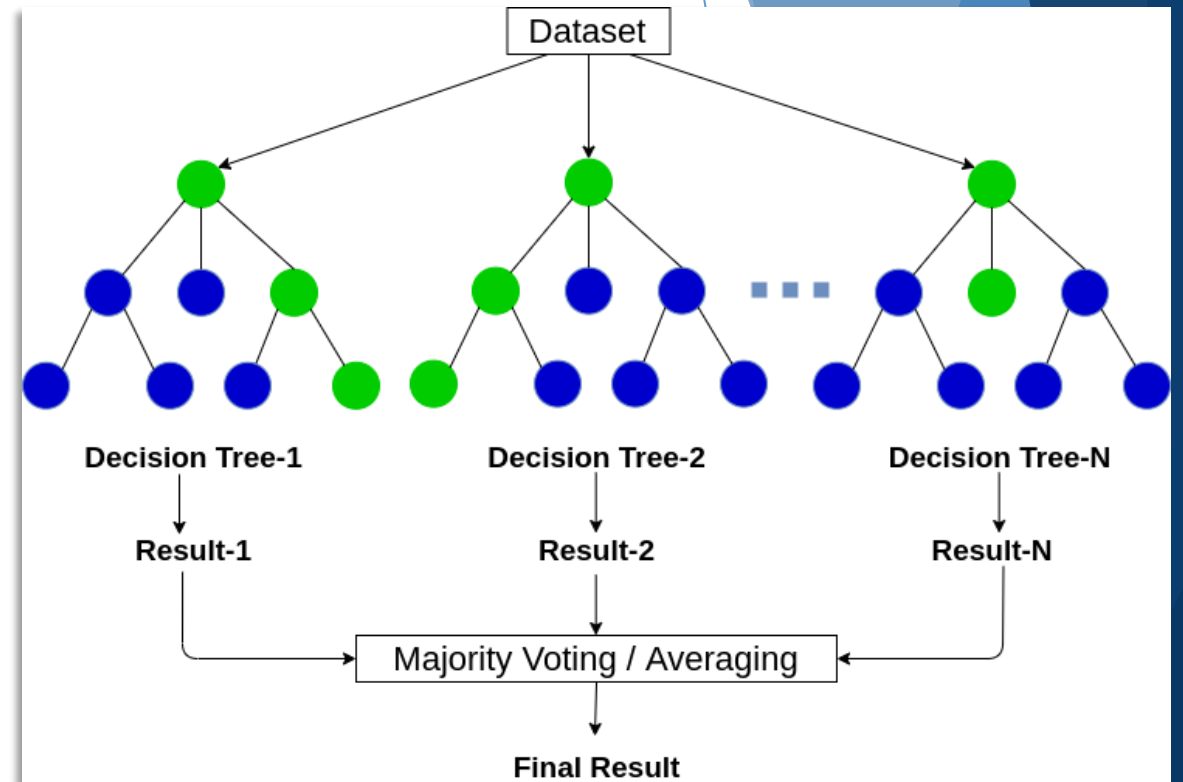
Gradient Boosted Trees





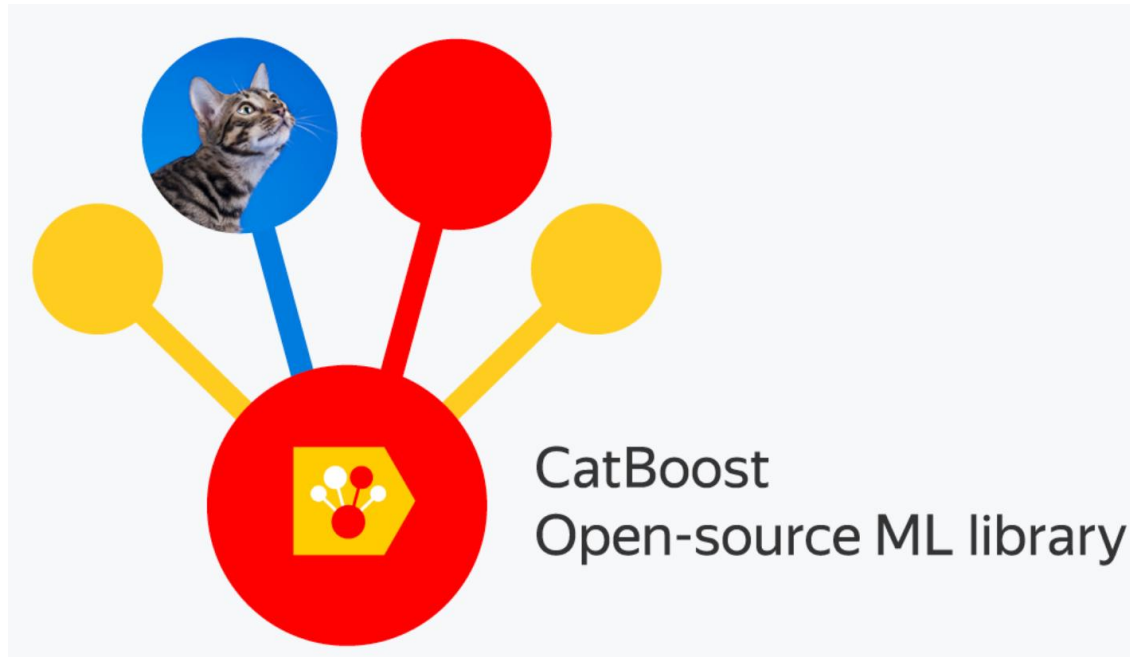
# Random Forest

- ▶ Random samples created from data
- ▶ Individual decision trees made for each sample
- ▶ All tree predictions compared together
- ▶ Majority vote for best prediction
- ▶ Which model best



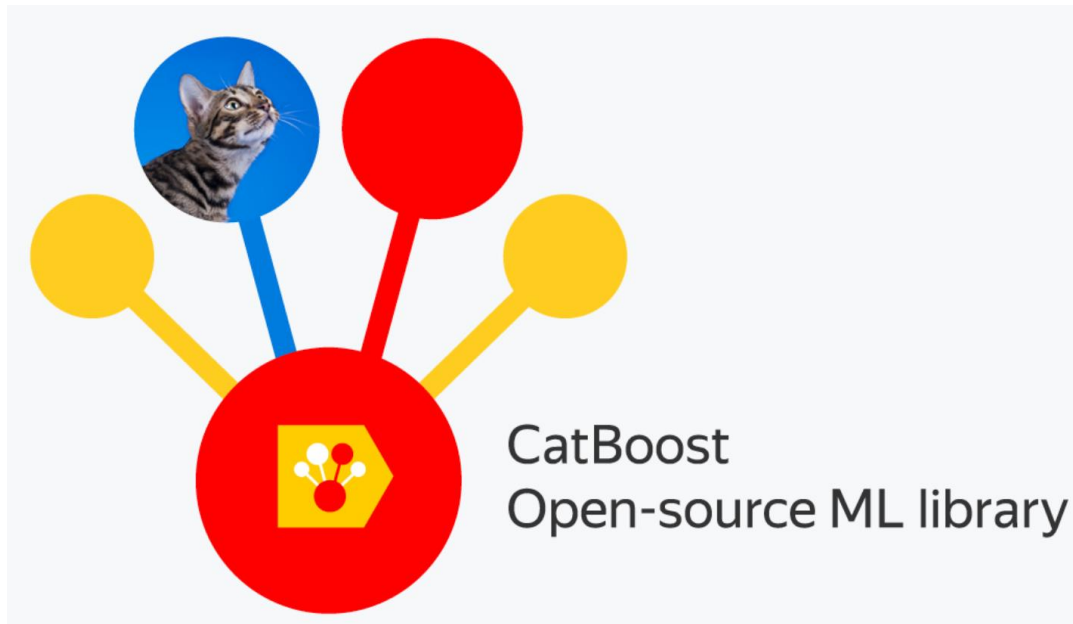
# Crime Type Best Model

- ▶ Compared Random Forest and CatBoost models
- ▶ Found our CatBoost model to be the best model



# Crime Location Best Model

- ▶ Compared a default CatBoost model and a tuned CatBoost model
- ▶ Found initial CatBoost model was our best one



# Results

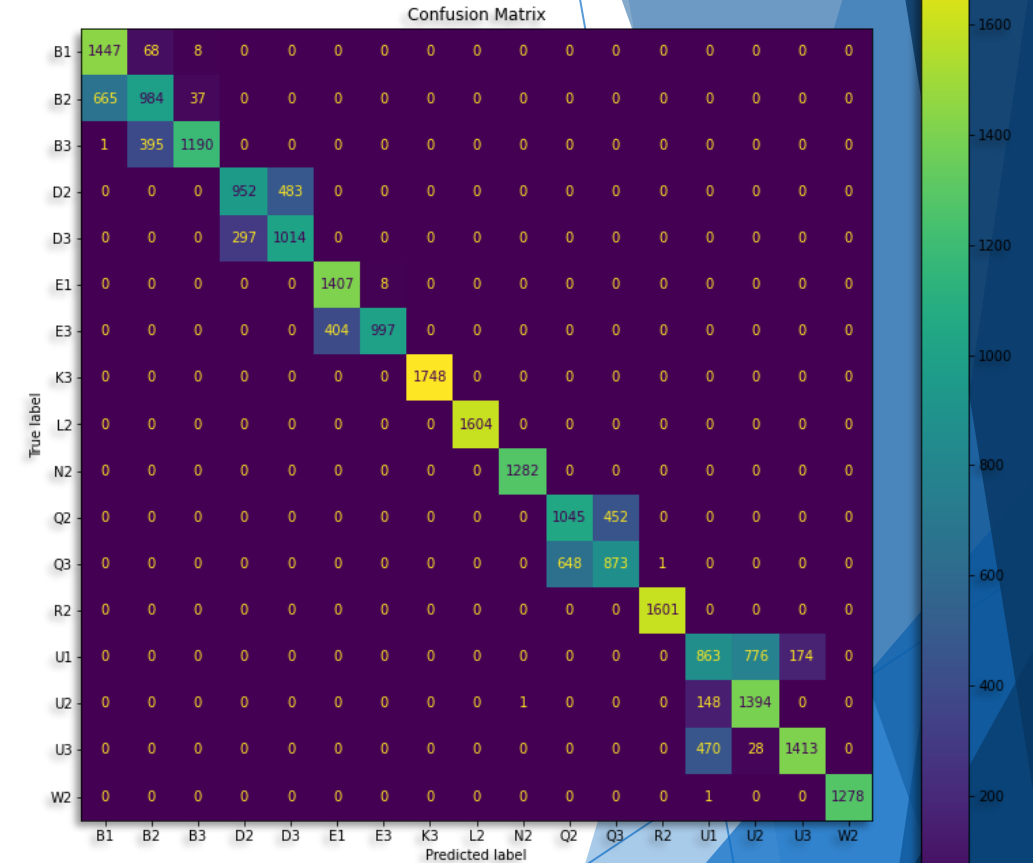
- ▶ Crime type, crime locations
- ▶ Cover major ones
- ▶ Recall, True Positives, False Negatives

# Results Continued

- ▶ Recall
- ▶ Total number of certain type crime/crime location, what percent did model correctly predict

# Results Continued

- ▶ TP, FN
- ▶ How many crimes/crime locations were correctly predicted
- ▶ How many crimes/crime locations wrongly predicted as not belonging to a certain class

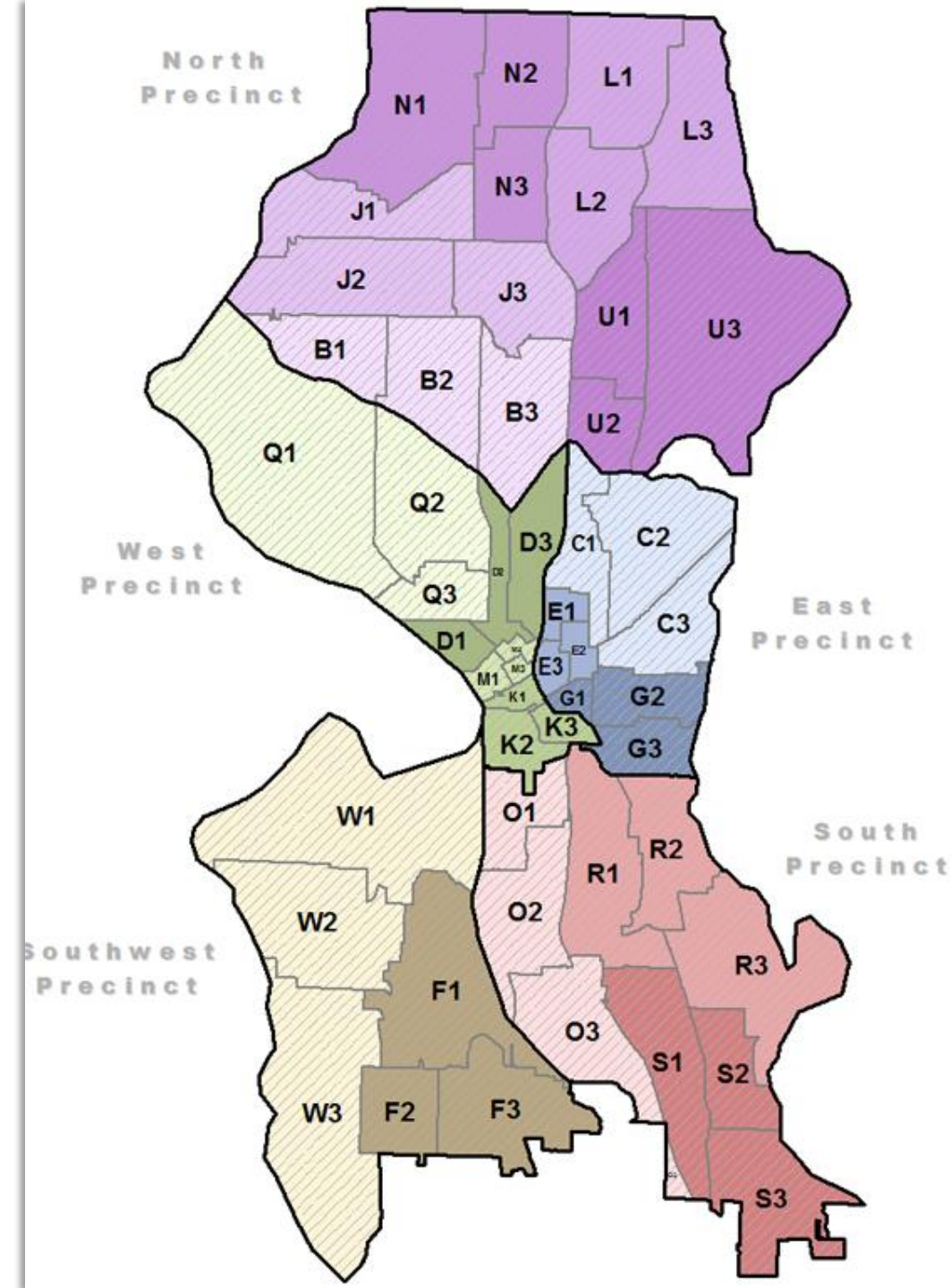


# Crime Type Results

- ▶ Best predicted crimes were Shoplifting and Simple Assault
- ▶ Shoplifting recall was .62
- ▶ 1,859 true positive instances, 1,116 false negative instances
- ▶ Simple Assault recall was .56
- ▶ 2,139 true positive instances, 1,651 false negative instances
- ▶ Low recall
- ▶ High FN instances

# Crime Location Results

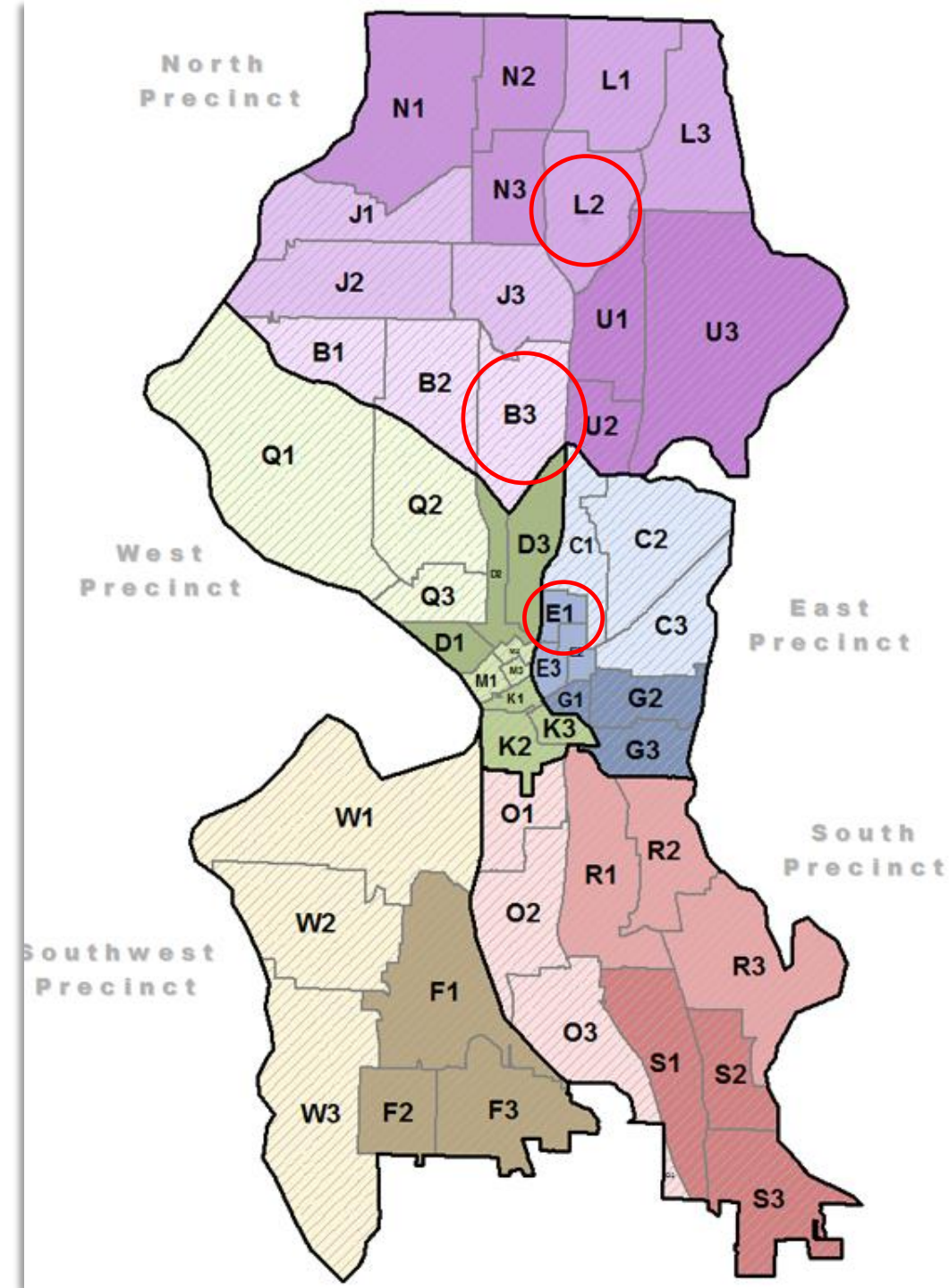
- ▶ Important to know how Seattle separated
- ▶ Seattle separated into precincts, sectors, and beats
- ▶ Precincts are police station general locations
- ▶ Sectors smaller areas within precincts
- ▶ Smallest area within sectors called Beats
- ▶ Beats are what we use





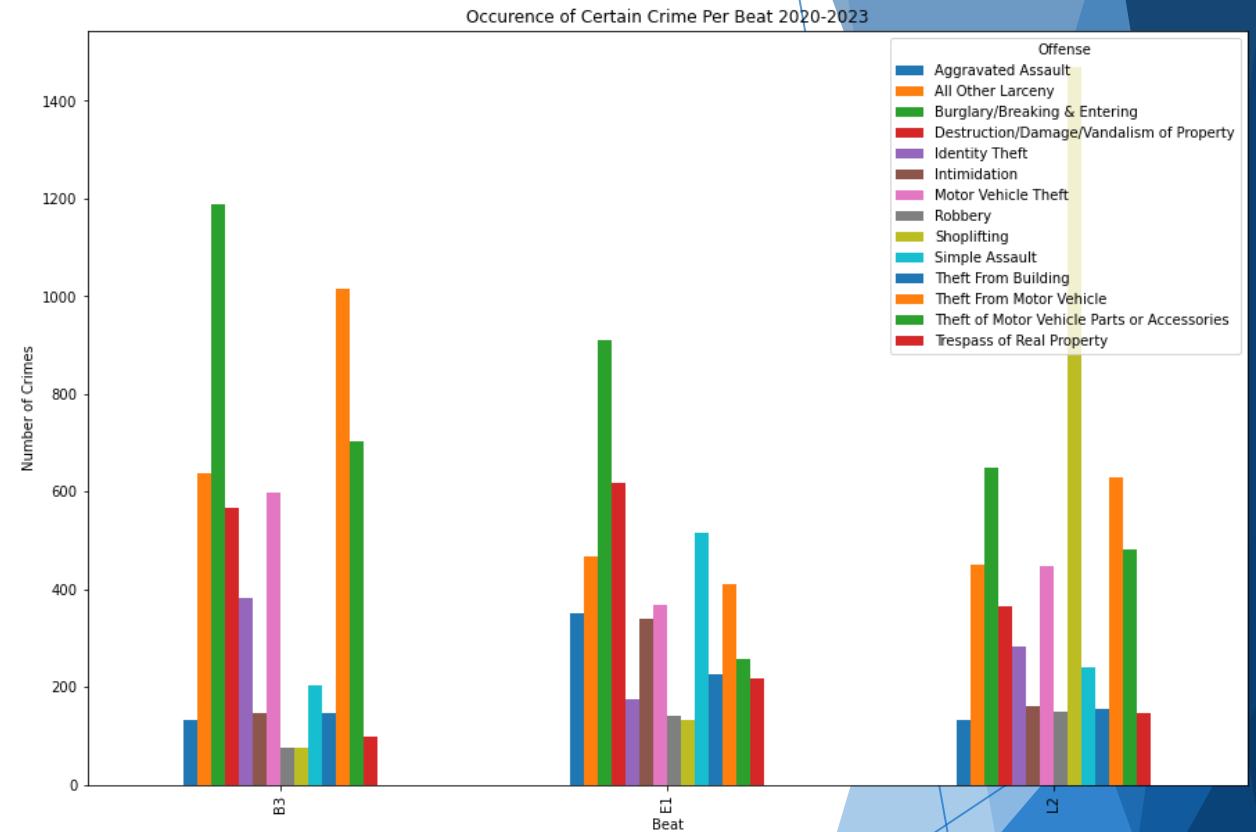
# Crime Location Results

- ▶ Best predicted crime locations were B3, E1, L2 Beats
- ▶ For recall B3(.75), E1(.99), L2 (1.0)
- ▶ For TP instance B3(1,187), E1(1,406), L2(1,604)
- ▶ For FN instances B3(399), E1(9), L2(0)
- ▶ High recall, High TP
- ▶ Low FN



# Additional Analysis

- ▶ Looked at data without the use of ML models
- ▶ Looked at which crimes occurred the most in our best predicted Beats
- ▶ We can see top crimes between the 3 Beats are Burglarly/Breaking & Entering and Theft From Motor Vehicle crimes

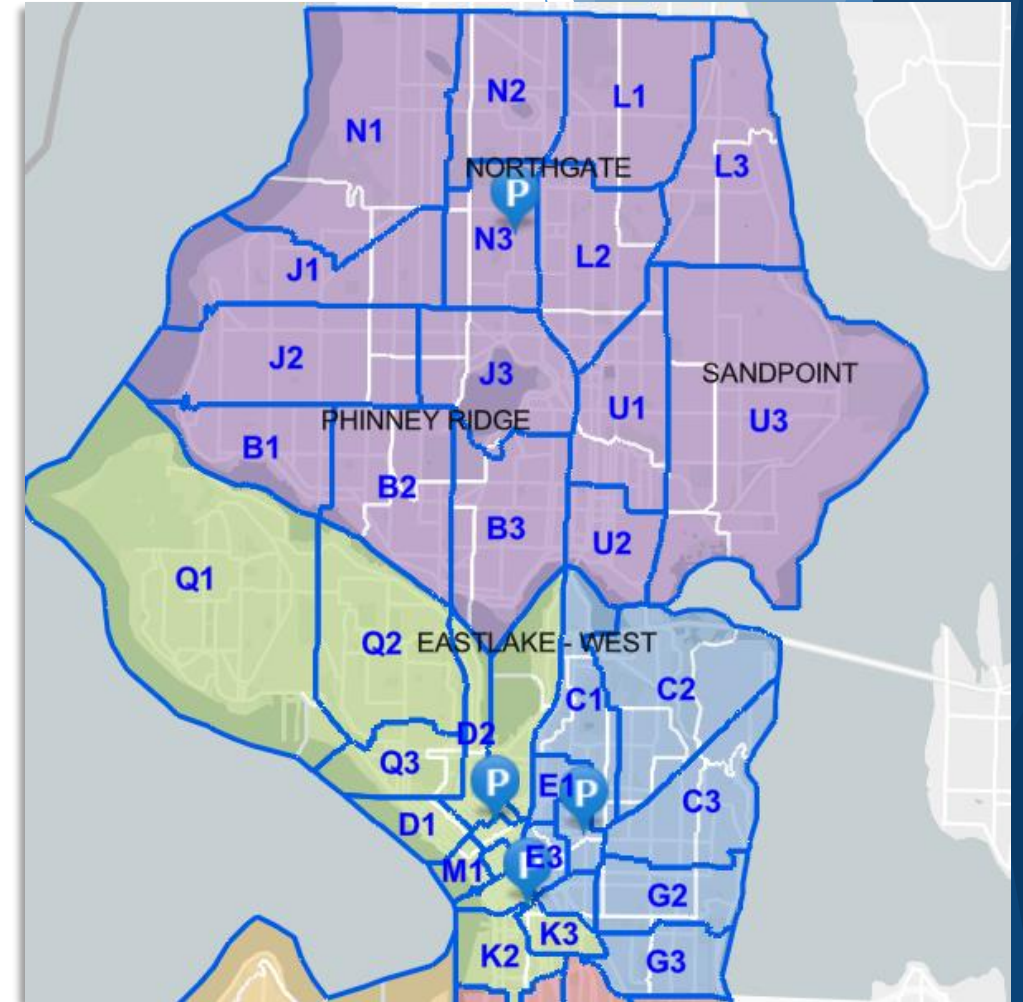


# Final Results

- ▶ Our model best classifies new crimes as occurring in the B3, E1 and L2 Beats
- ▶ Top crimes are Shoplifting, Simple Assault, Burglary/Breaking & Entering and Theft From Motor Vehicle
- ▶ Can focus on preventative measures

# Recommendations

- ▶ B3, E1, L2 and police stations on map
- ▶ Need more station(s) between B3 and L2
- ▶ Alternatively divert some southern police forces north
- ▶ Lead to faster response times, deterrence of future crimes (all types)
- ▶ Shortage of officers, due to staffing crisis and quitting
- ▶ Another option to form neighborhood watch for each Beat
- ▶ Build/improve relations with the public





# Recommendations Continued

- ▶ Install more security cameras in strategic places, with 24-hour surveillance
- ▶ Dimly lit areas, areas of lower foot traffic
- ▶ Easier to track all crimes, may also deter crime
- ▶ Lastly, build more monitored, affordable parking
- ▶ Reduce theft and create positive feedback from public



## Recommendations Continued

- ▶ Implementing model
- ▶ Implement model in Dispatch center
- ▶ Police laptops
- ▶ Will allow for instant notification and response
- ▶ Integrate with social media
- ▶ Any social media posts will alert model



# Next Steps

- ▶ Results are great starting point
- ▶ Reduce overfitting in location model
- ▶ Obtain higher recall scores in crime type model
- ▶ Don't filter data
- ▶ Model takes while to train
- ▶ Run faster on GPU
- ▶ May be expensive to buy GPUS, but can rent GPUs (less than 3\$ per hour)

# Next Steps

- ▶ Once model successfully tuned and implemented, will have numerous positive effects
- ▶ Decrease crime, improve police-public relations, increase business/tourism
- ▶ Once proven successful, model may even be implemented in other parts WA, and then even other states



# Sources

- ▶ <https://www.rexmontrealestate.com/blog/seattle-economy/#:~:text=Seattle%2C%20Washington%2C%20has%20a%20long,is%20home%20to%20many%20industries.>
- ▶ <https://wallethub.com/edu/safest-cities-in-america/41926>
- ▶ <https://www.infoplease.com/us/cities/top-50-cities-us-population-and-rank>
- ▶ <https://www.king5.com/article/news/local/seattle/mayor-bruce-harrell-new-seattle-police-recruitment-plan-staffing-reaches-30-year-low/281-e98429ee-fc88-47f6-8890-b961393e2047>

# Questions and Contact Information

- ▶ Questions?
- ▶ Any additional inquiries can be directed to linkedin:
  - ▶ <https://www.linkedin.com/in/alejandro-harrison/>