# Capstone Project – Sprint 1

## Predict Home Price Index Based on Major Crime Incidents in Toronto

*Allan Salamanca*

*February 23, 2025*

# MLS Home Price Index (HPI)

- Established by the Canadian Real Estate Association (CREA) to gauge home price levels and trends.
- Statistical model that considers quantitative (# of rooms) and qualitative (finished basement) features.
- Based on four home categories, **detached, semi-detached, townhouse and apartment**.

# TPS - Major Crime Indicators (MCI)

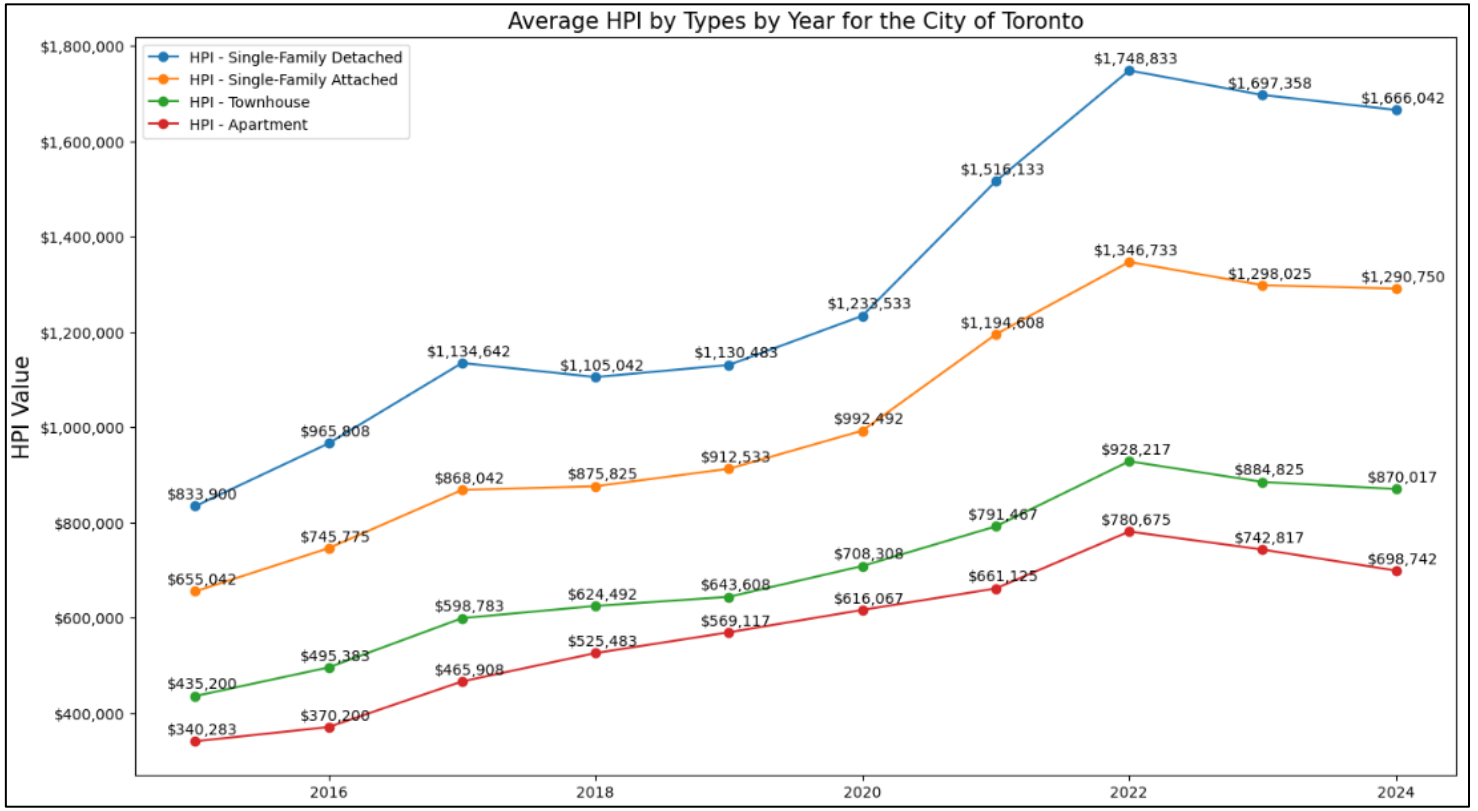

- 5 crime categories that include: **Assault, Break and Enter, Auto Theft, Robbery and Theft**.
- Data is either provided by the victim, or as the offence occurs or is reported.

# Problem Area

Home Prince Index in Toronto have increased 8% YoY from 2015 to 2024.

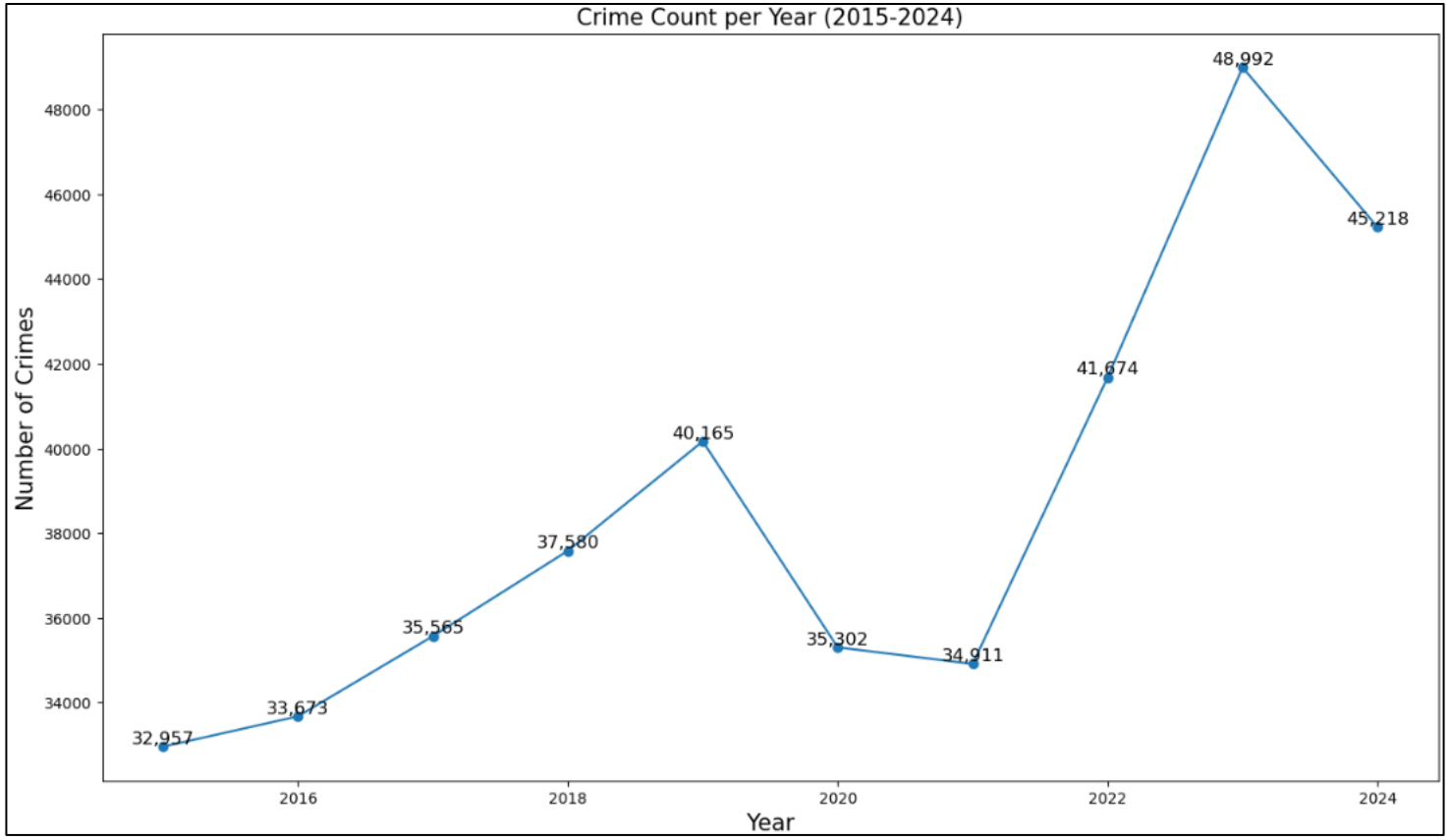
At the end of 2024, HPI for properties in Toronto:

- Detached - $1.6M
- Semi-detached - $1.2M
- Townhouse - $850K
- Apartments - $670K



# Problem Area

Major crime incidents have increased 4% YoY in Toronto from 2015 to 2024.



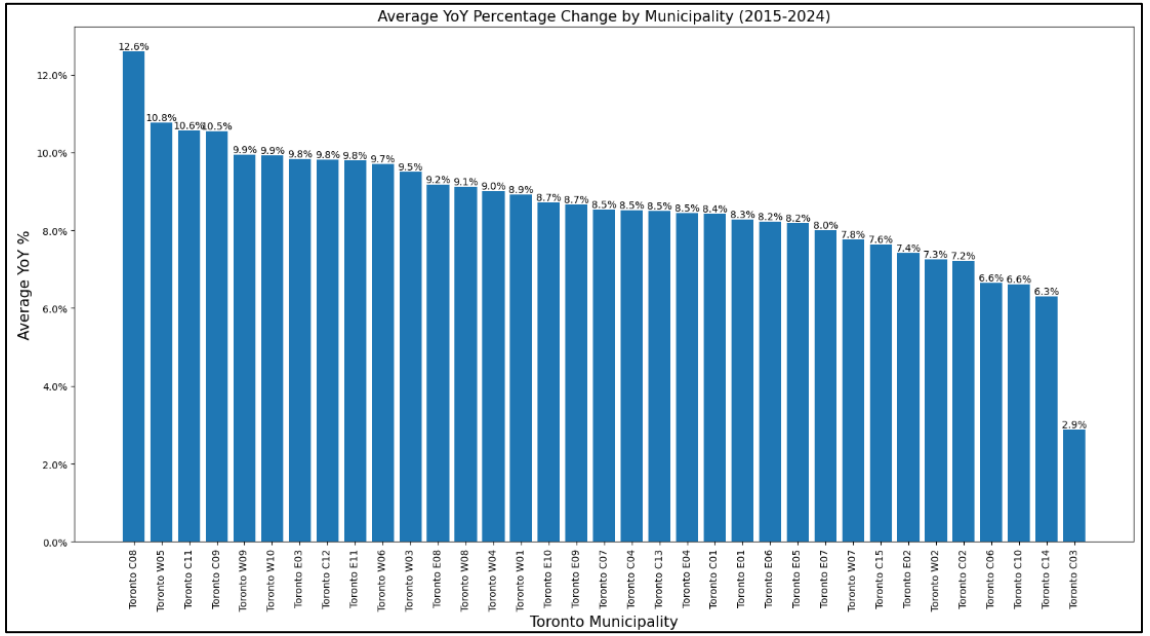
# Problem Area: Solution

**It’s then crucial to understand what factors can affect housing prices, to maximize equity and return on investment.  With this tool, we aim to predict the housing price impact by neighborhood, based on the estimated crime rate incidents.**

# Data Set: Toronto Home Price Index

- Toronto Municipality Code (Neighbourhood)
- HPI - Single-Family Detached
- HPI - Single-Family Attached
- HPI – Townhouse
- HPI - Apartment
- Date (2015-2024)

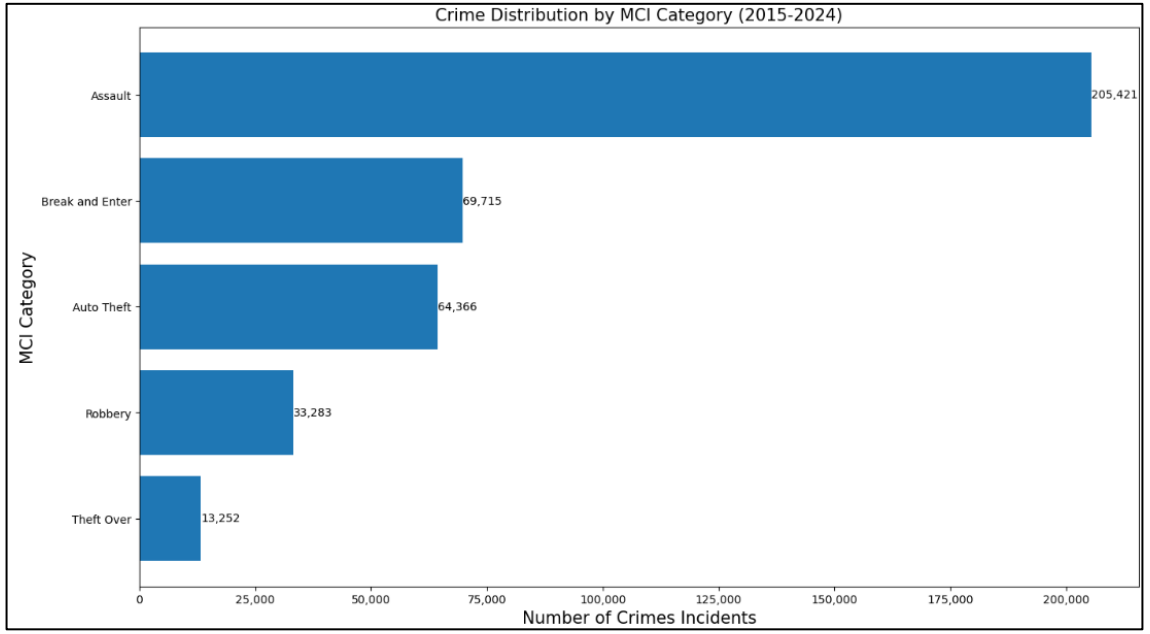
| | Toronto Municipality | HPI - Single-Family Detached | HPI - Single-Family Attached | HPI - Townhouse | HPI - Apartment | Effective Date |
|---|---|---|---|---|---|---|
| 0 | City of Toronto | 787300 | 625000 | 413100.0 | 330200 | 2015-01-01 |
| 1 | Toronto W01 | 849400 | 677500 | 415100.0 | 326300 | 2015-01-01 |
| 2 | Toronto W02 | 805300 | 661100 | 410300.0 | 563000 | 2015-01-01 |
| 3 | Toronto W03 | 509900 | 493700 | NaN | 257400 | 2015-01-01 |
| 4 | Toronto W04 | 538000 | 501900 | 337000.0 | 206400 | 2015-01-01 |
| ... | ... | ... | ... | ... | ... | ... |
| 4315 | Toronto E07 | 1219200 | 1013700 | 824900.0 | 593600 | 2024-12-01 |
| 4316 | Toronto E08 | 1219900 | 872600 | 752500.0 | 530000 | 2024-12-01 |
| 4317 | Toronto E09 | 1026700 | 884600 | 656600.0 | 575000 | 2024-12-01 |
| 4318 | Toronto E10 | 1171300 | 887000 | 620300.0 | 486100 | 2024-12-01 |
| 4319 | Toronto E11 | 1062900 | 885800 | 709300.0 | 500300 | 2024-12-01 |



# Data Set: Toronto Major Crime Indicators

- Occurrence Date
- Division
- MCI Category
- Offence (Subcategory)
- Neighbourhood
- Date (2015-2024)

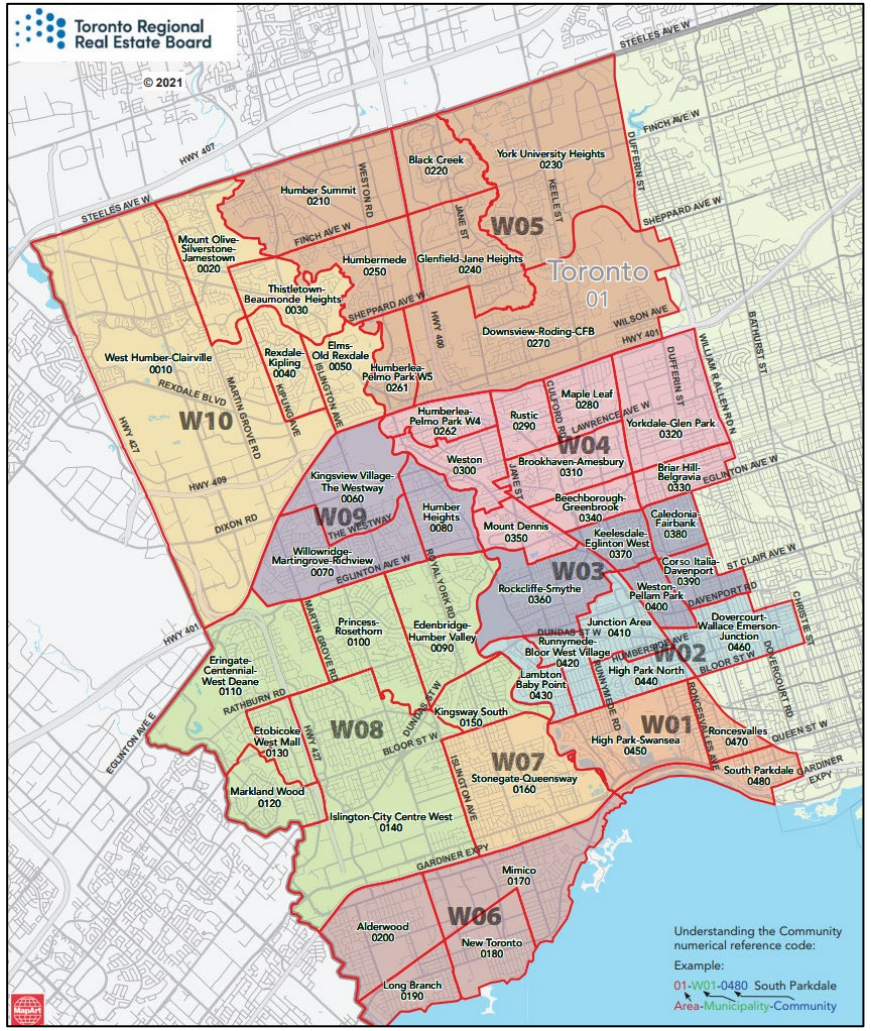
| | EVENT_UNIQUE_ID | OCC_DATE | DIVISION | PREMISES_TYPE | OFFENCE | MCI_CATEGORY | NEIGHBOURHOOD_140 |
|---|---|---|---|---|---|---|---|
| **26** | GO-20151956 | 2015-01-01 05:00:00 | D32 | House | Assault With Weapon | Assault | Newtonbrook West (36) |
| **27** | GO-20151923 | 2015-01-01 05:00:00 | D14 | Outside | Assault | Assault | Waterfront Communities-The Island (77) |
| **28** | GO-20151955 | 2015-01-01 05:00:00 | D42 | Outside | Robbery - Mugging | Robbery | Malvern (132) |
| **29** | GO-20151945 | 2015-01-01 05:00:00 | D41 | Apartment | Assault | Assault | Wexford/Maryvale (119) |
| **30** | GO-20152177 | 2015-01-01 05:00:00 | D14 | Apartment | Assault | Assault | Dufferin Grove (83) |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **386176** | GO-20242816627 | 2024-12-24 05:00:00 | D32 | Outside | Theft Of Motor Vehicle | Auto Theft | Clanton Park (33) |
| **386177** | GO-20242815828 | 2024-12-30 05:00:00 | D31 | Outside | Theft Of Motor Vehicle | Auto Theft | Glenfield-Jane Heights (25) |
| **386178** | GO-20242814680 | 2024-12-31 05:00:00 | D51 | Outside | Robbery - Mugging | Robbery | Church-Yonge Corridor (75) |
| **386179** | GO-20242815237 | 2024-12-31 05:00:00 | D32 | Commercial | B&E | Break and Enter | Englemount-Lawrence (32) |
| **386180** | GO-20242815267 | 2024-12-31 05:00:00 | D14 | Commercial | B&E | Break and Enter | Trinity-Bellwoods (81) |



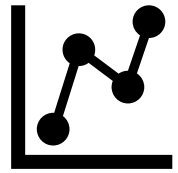
# Data Set: Toronto Neighbourhood Index

- Manually created table.
- Helper table to join TPS and MLS neighbourhood to MLS municipality codes.

| | Toronto Police Neighbourhood Categories | MLS Neighbourhood Categories | MLS Municipality Code |
|---|---|---|---|
| 0 | Agincourt North (129) | Agincourt North | Toronto E07 |
| 1 | Agincourt South-Malvern West (128) | Agincourt South-Malvern West | Toronto E07 |
| 2 | Alderwood (20) | Alderwood | Toronto W06 |
| 3 | Annex (95) | Annex | Toronto C02 |
| 4 | Banbury-Don Mills (42) | Banbury-Don Mills | Toronto C13 |
| ... | ... | ... | ... |
| 136 | Wychwood (94) | Wychwood | Toronto C02 |
| 137 | Yonge-Eglinton (100) | Yonge-Eglinton | Toronto C03 |
| 138 | Yonge-St.Clair (97) | Yonge-St.Clair | Toronto C02 |
| 139 | York University Heights (27) | York University Heights | Toronto W05 |
| 140 | Yorkdale-Glen Park (31) | Yorkdale-Glen Park | Toronto W04 |



# Data Science Approach



- **Target:** Predict home price index (HPI) by house type (Detached, Semi-Detached, Town House, or Apartment), and neighbourhood, based on major crime incident occurrences.
- **Features:** Home Price Index, House Type, Neighbourhood, MCI Category.
- Utilize **Logistic Regression** to predict HPI.

# Next Steps

- Address the null values in the HPI - Townhouse columns, in the MLS data. Utilize the K Nearest Neighbours approach to get approximate values.
- Address the null values (NSA) in the NEIGHBOURHOOD_140 column, in the Toronto Police data. Since these areas are unknown, we can omit this from the total data set.
- Create a joined data frame, to link all the data sets together.
- Conduct logistics regression, to determine relationship and statistical significance.
- Feature engineering, to determine which levers we can use for our prediction.