



# Capstone Project – Sprint 2

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

*Allan Salamanca  
April 27, 2025*

# MLS Home Price Index (HPI)

---

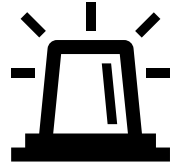


- Established by the Canadian Real Estate Association (CREA) to gauge home price levels and trends.
- Based on four property types, **detached, semi-detached, townhouse and apartment.**



# **TPS - Major Crime Indicators (MCI)**

---



- 5 crime categories that include:  
**Assault, Break and Enter, Auto Theft, Robbery and Theft.**

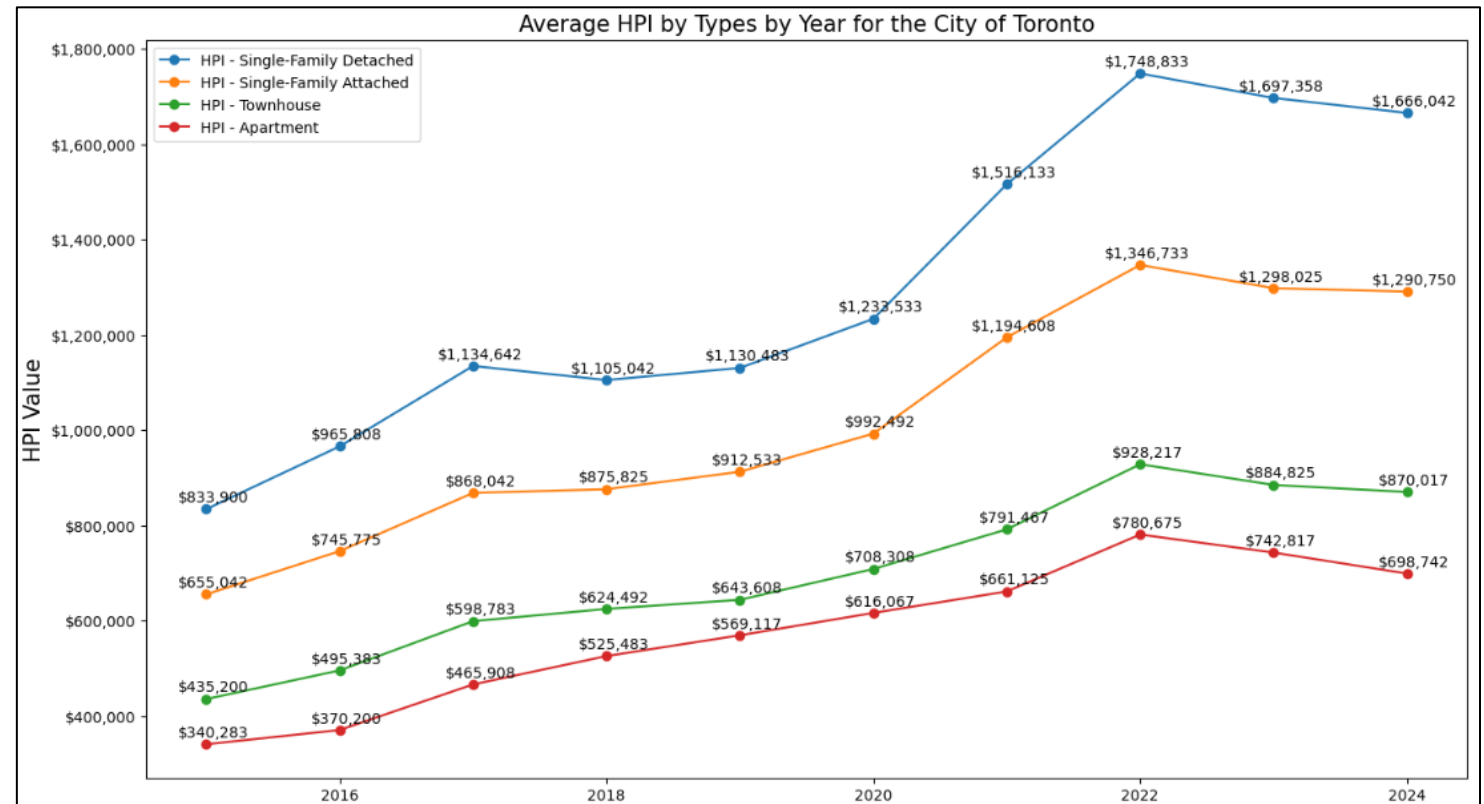
# Problem Area



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

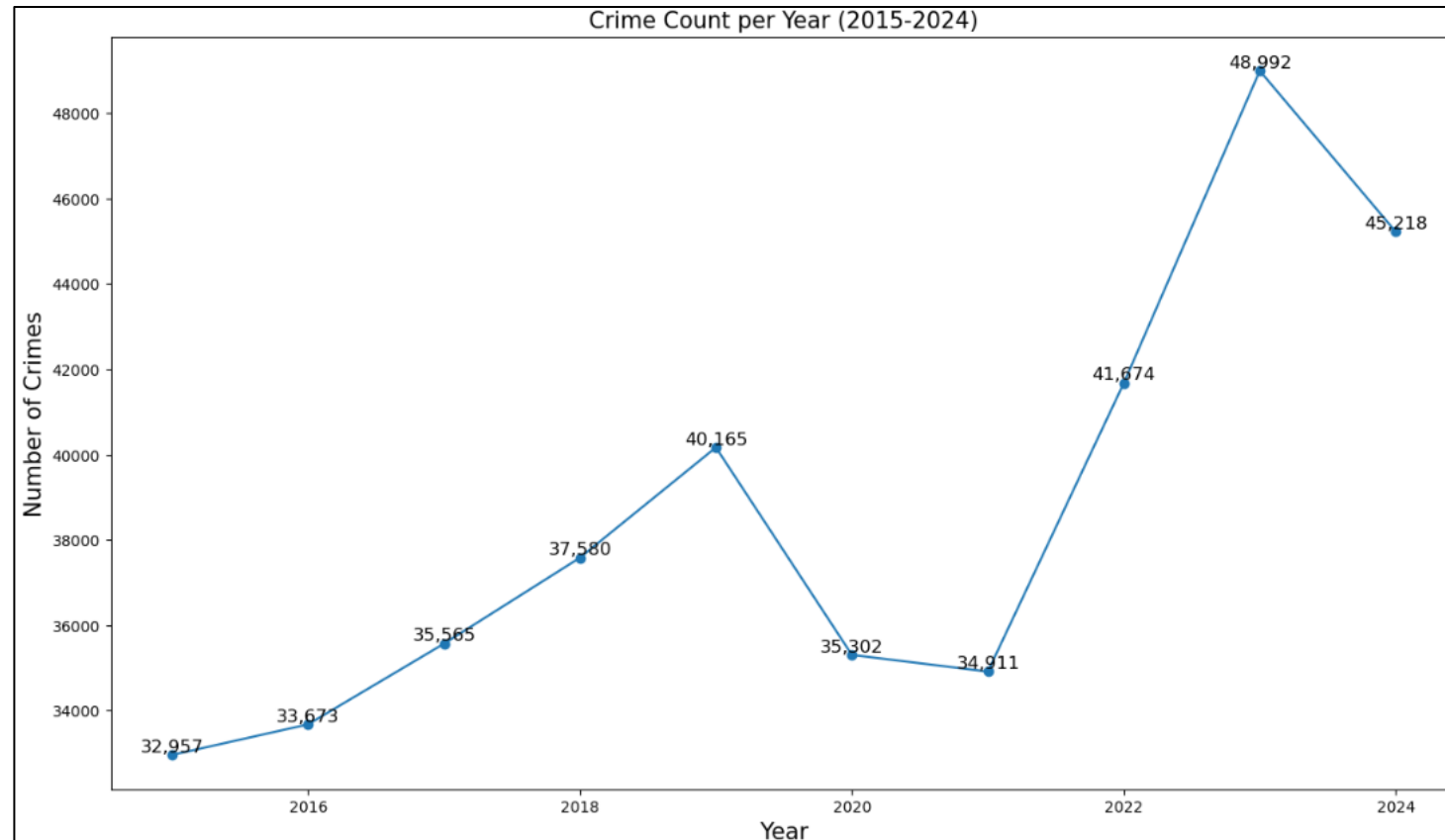


# Problem Area

---



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



# Problem Area: Data Solution

---



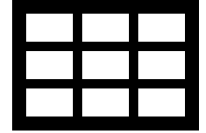
1. Understand what factors affect HPI values.
2. Develop a linear regression model that can predict HPI values by Time Period, Crime Incidents, Property Type and Neighborhoods (Municipality Code).

## Objective:

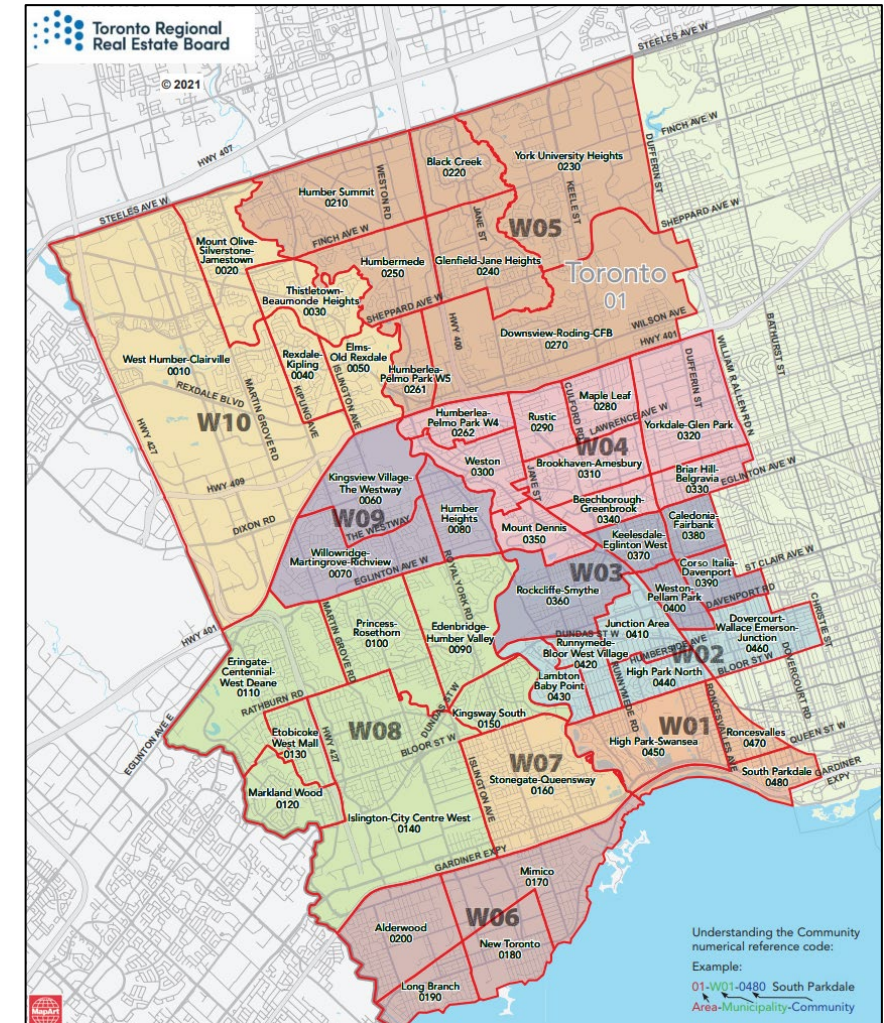
Select the neighbourhoods with the best return on investment



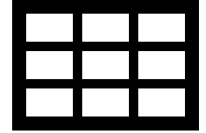
# Data Preprocessing



- Data Sets:
  - Toronto Home Price Index
  - Toronto Major Crime Incidents
  - Toronto Neighbourhood Index
- Joined 3 data sets to create a master data frame based on unique neighbourhoods to a municipality code.
- Dummies created for categories: property type and neighbourhood.
- Imputed missing data using KNN, for Townhouses in specific neighbourhoods.



# Data Preprocessing



Data Summary - 16,800 rows & 45 columns

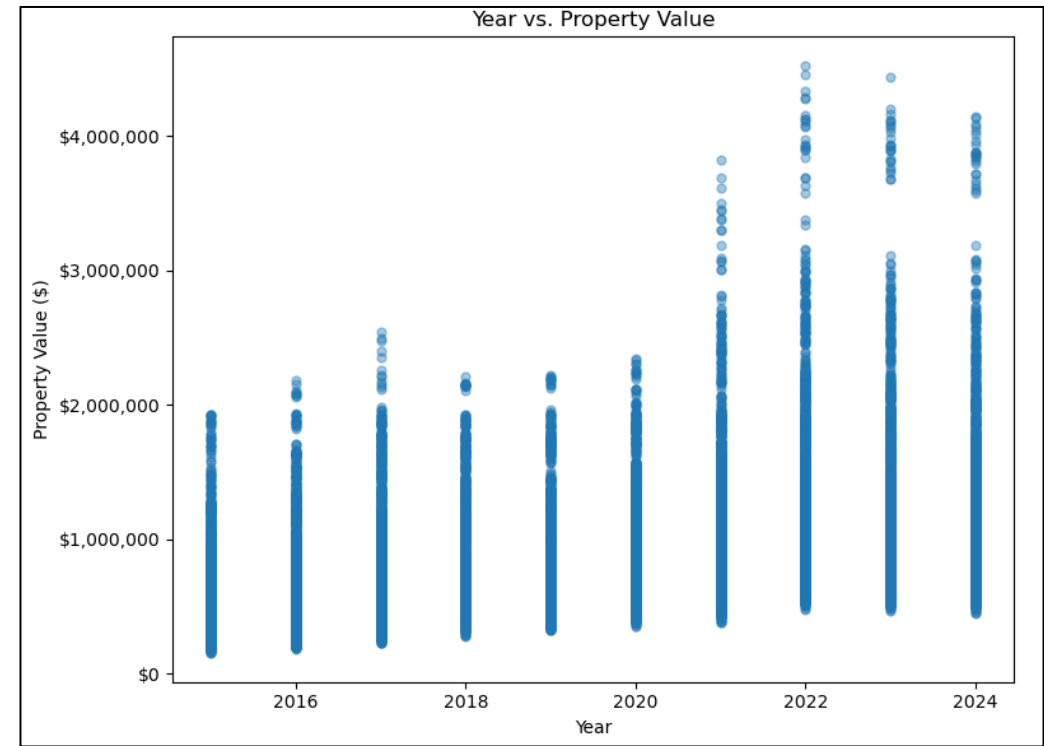
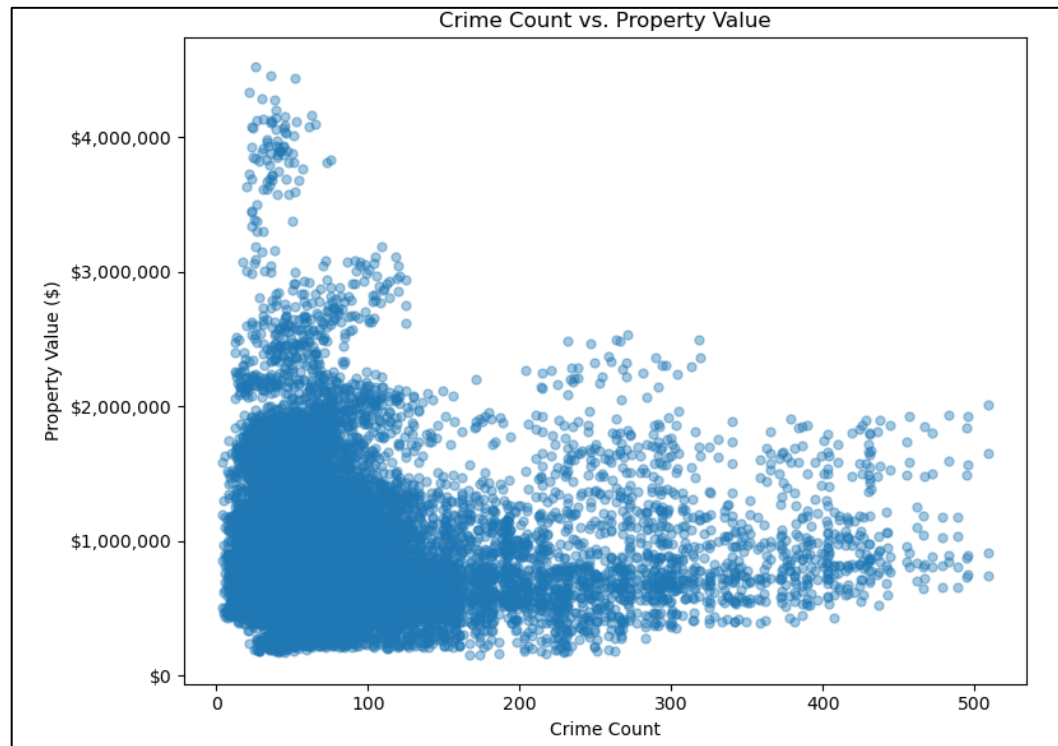
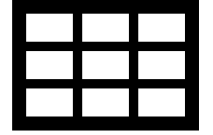
**Dependent Variable** - Value (HPI)

**Independent Variables:** Year, Crime Count, Property Type (Attribute) and Neighbourhood (Toronto Municipality)

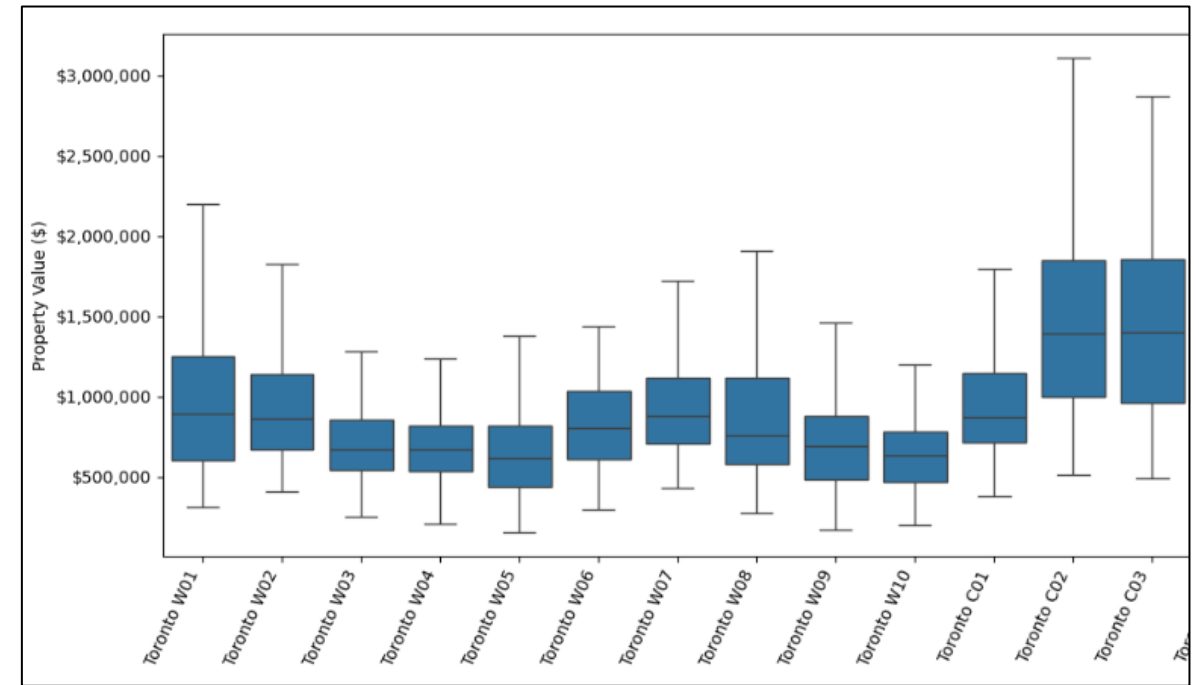
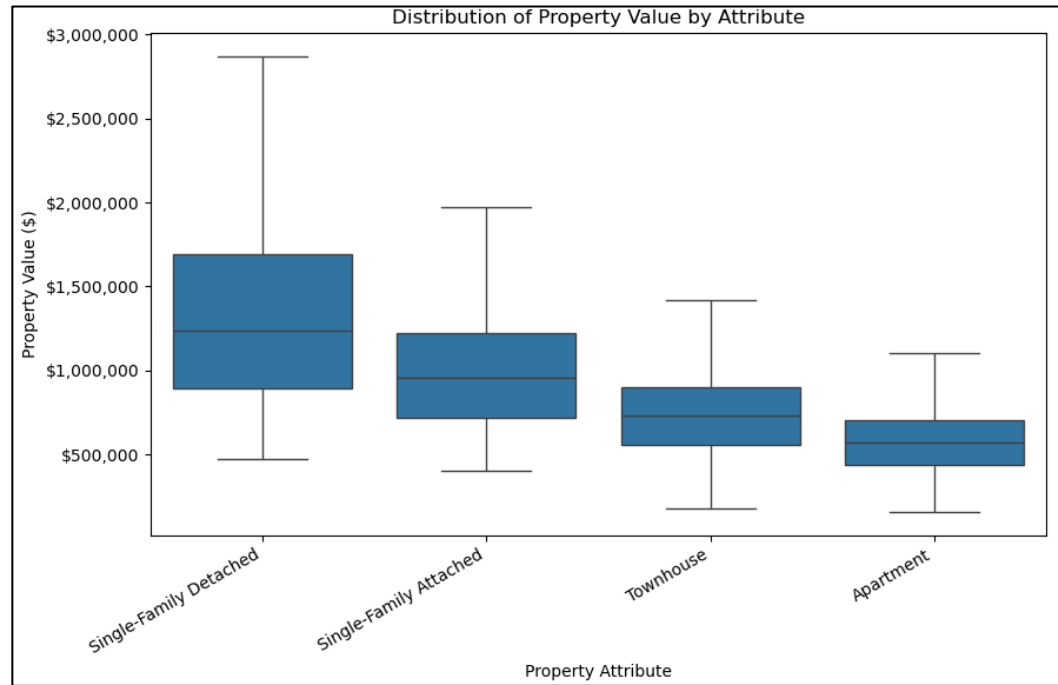
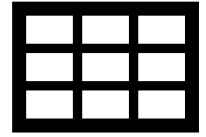
Attribute_HPI - Single- Family Detached	Attribute_HPI - Townhouse	Crime Count	Month	Toronto Municipality_Toronto C01	Toronto Municipality_Toronto C02	Toronto Municipality_Toronto C03	...	Toronto Municipality_Toronto W04	Toronto Municipality_Toronto W05	Toronto Municipality_Toronto W06	Toronto Municipality_Toronto W07	Toronto Municipality_Toronto W08	Toronto Municipality_Toronto W09	Toronto Municipality_Toronto W10	Value	Year
1.0	0.0	44.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	849400.0	2015.0
0.0	0.0	44.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	677500.0	2015.0
0.0	1.0	44.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	415100.0	2015.0
0.0	0.0	44.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	326300.0	2015.0
1.0	0.0	62.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	805300.0	2015.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.0	0.0	56.0	12.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	486100.0	2024.0
1.0	0.0	58.0	12.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1062900.0	2024.0
0.0	0.0	58.0	12.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	885800.0	2024.0
0.0	1.0	58.0	12.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	709300.0	2024.0
0.0	0.0	58.0	12.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	500300.0	2024.0

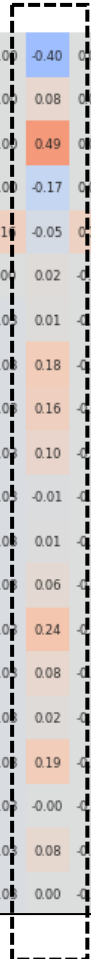


# Exploratory Data Analysis

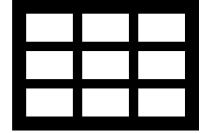


# Exploratory Data Analysis





# Baseline Modelling



## Standard Scaler & Linear Regression

- Train score: 77%
- Test score: 77%
- Test RMSE: \$247K
- Test MAE: \$167K

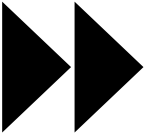
## OLS Results

- R-Squared – 0.771
- No p-value > 0.05



# Next Steps

---



- Poor RMSE and MAE scores.
- Review wide data set when it comes to HPI values – log/scale.
- Compare and asses more advanced modelling, such as RandomForestRegressor, DecisionTreeRegressor or XGBoost.
- Hyperparameter Tuning.