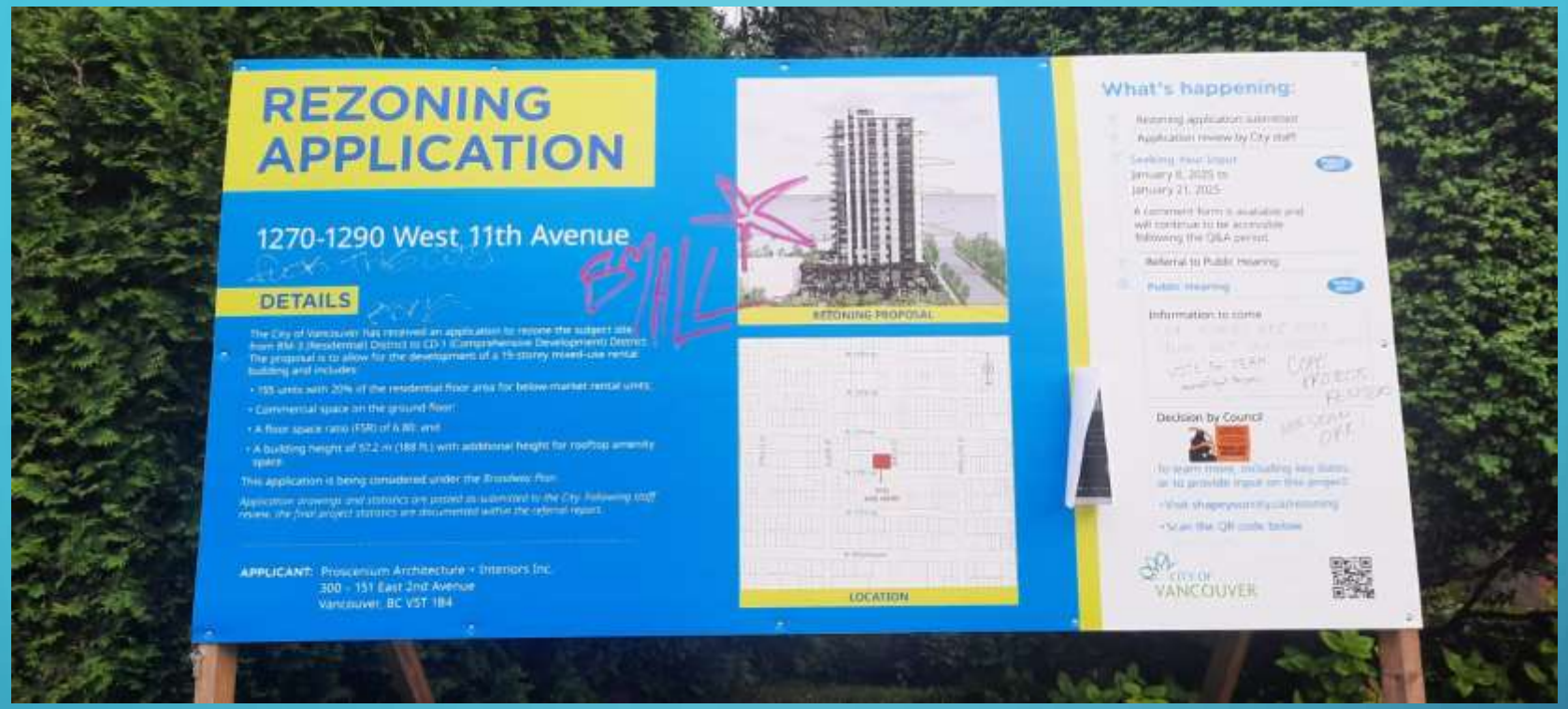
DATA-DRIVEN INSIGHTS INTO HOUSING TRENDS: FORECASTING AFFORDABILITY RISKS

ETHAN SHELBURNE

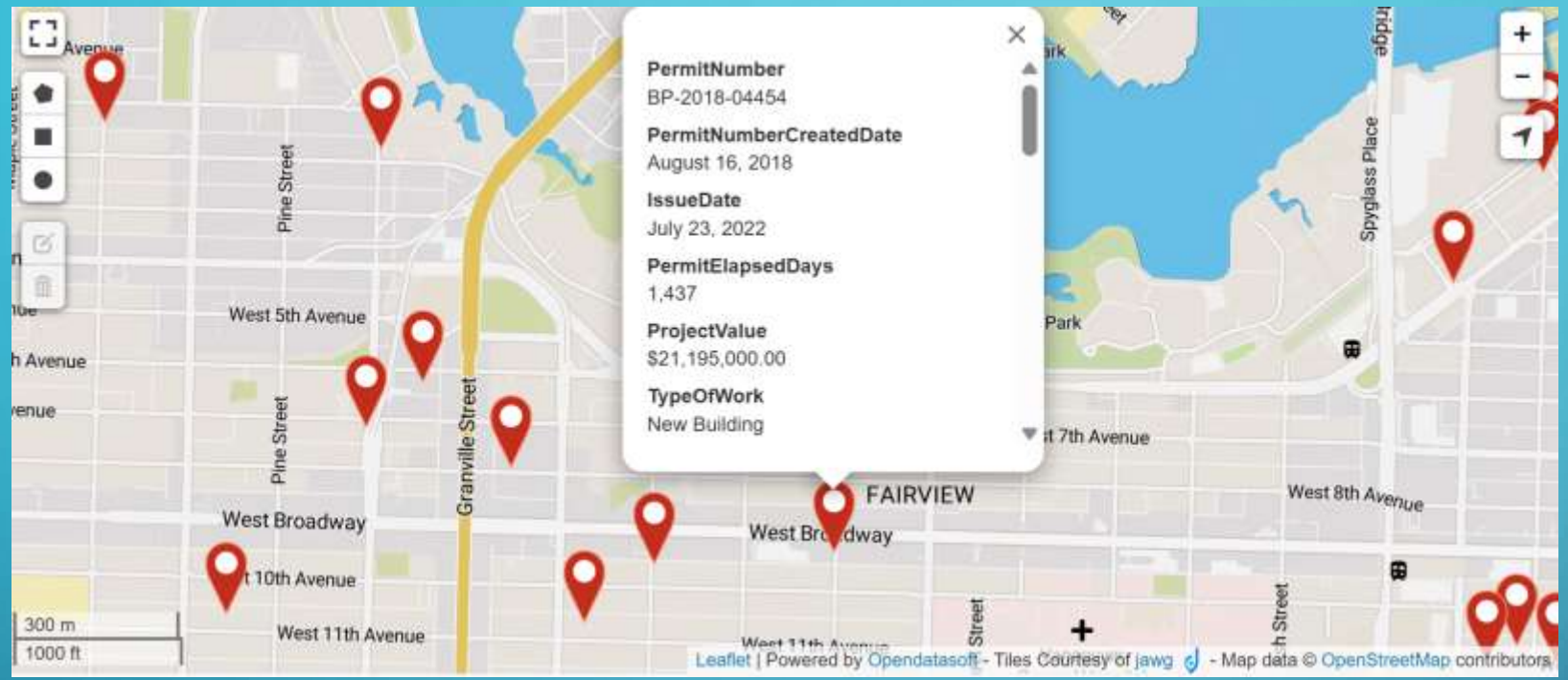
BRAINSTATION – THE DIGITAL LEARNING COMPANY



"OUR HOMES ARE MORE JUST THAN YOUR INVESTMENT"

- The Broadway Plan is a 30-year redevelopment strategy for Vancouver's Broadway Corridor, focused on building high-rise housing nearby new SkyTrain stations.
- Critics fear it will lead to the displacement of renters and worsening affordability.
- This project uses machine learning to model how new developments may impact local housing affordability.

PREDICTIVE MODELING WITH MACHINE LEARNING



Goal: Build a predictive tool to estimate the effects of developments like those in the Broadway Plan.

- Define affordability metrics at the neighborhood level.
- Utilize data set of building permits in Vancouver, each representing a housing development project.
- Use supervised machine learning regression models (such as **Linear Regression, Random Forest, and K-nearest neighbors**) to uncover correlations between development types and shifts in affordability.

PURSUING HOUSING EQUITY WITH DATA SCIENCE

- Activists have organized protests to challenge displacement linked to large-scale redevelopment plans.
- My goal is to provide a data-driven tool to help assess how specific developments may impact neighborhood affordability.

INTRODUCING THE DATA

Canada Mortgage and Housing Corporation (CMHC / SCHL)

City of Vancouver Open Data Portal

	Bachelor		1 Bedroom	
Zone	Oct-20	Oct-21	Oct-20	Oct-21
Zone 1 - West End/Stanley Park	1,229 a	1,280 a	1,537 a	1,539 a
Zone 2 - English Bay	1,340 a	1,337 a	1,637 a	1,621 a
Zone 3 - Downtown	1,450 a	1,470 a	1,741 a	1,727 a
West End/Downtown (Zones 1-3)	**1,389 a**	**1,406 a**	**1,673 a**	**1,660 a**
Zone 4 - South Granville/Oak	1,243 a	1,296 a	1,517 a	1,566 a
Zone 5 - Kitsilano/Point Grey	1,251 a	1,300 a	1,563 a	1,579 a
Zone 6 - Westside/Kerrisdale	1,248 a	1,156 a	1,459 a	1,474 a
Zone 7 - Marpole	965 a	1,046 a	1,138 a	1,161 a
Zone 8 - Mount Pleasant/Renfrew Heights	1,281 a	1,333 a	1,361 a	1,388 a
Zone 9 - East Hastings	1,330 a	1,385 b	1,318 a	1,297 a
Zone 10 - Southeast Vancouver	1,165 a	1,249 b	1,413 a	1,403 a

```
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   PermitNumber           32744 non-null  object
 1   PermitNumberCreatedDate 32744 non-null  datetime64[ns]
 2   IssueDate              32744 non-null  datetime64[ns]
 3   PermitElapsedDays      32744 non-null  int64
 4   ProjectValue           32744 non-null  float64
 5   TypeOfWork             32744 non-null  object
 6   Address                32581 non-null  object
 7   ProjectDescription     32744 non-null  object
```

- The project combines **housing construction data** with **neighborhood-level economic indicators.**
- Preliminary EDA shows the datasets are rich and detailed.
- I anticipate challenges in merging datasets and handling inconsistent values.
- To clean the CMHC data, I need to impute missing census tract values (such as average rent) with values corresponding to a higher geographic level.
- Careful adjustments will also be needed to account for macroeconomic factors.

NEXT STEPS

- Continue cleaning and organizing datasets, building a cohesive database.
- Use Python to construct a geographic hierarchy (census tract < neighborhood < census subdivision < …) to impute missing values at different levels of granularity into CMHC data.
- Conduct feature engineering to align construction data with neighborhood-level economic indicators.
- Research machine learning models used in housing analysis to identify suitable algorithms.
- Begin baseline modeling!

QUESTIONS?