Clustering Development Projects and Neighborhood Economics: Uncovering Patterns in Affordability for Renters

Ethan Shelburne

BrainStation: The Digital Learning Company

QUESTION: How do development projects shape affordability for renters in Vancouver?

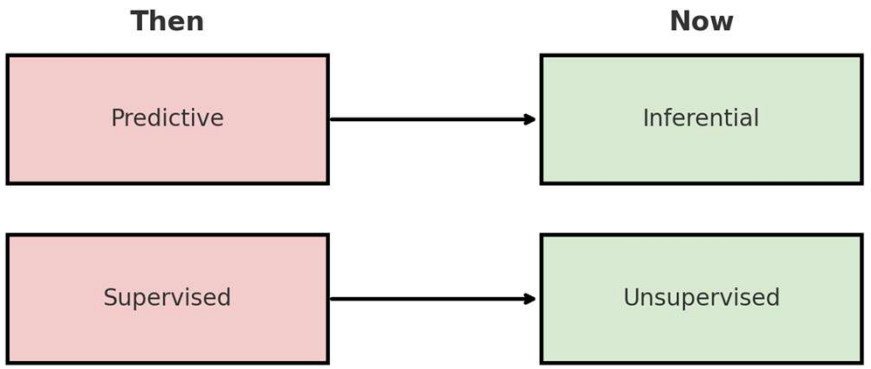
Goal: Understand how different types of building activity relate to changes in rent levels and vacancy rates.

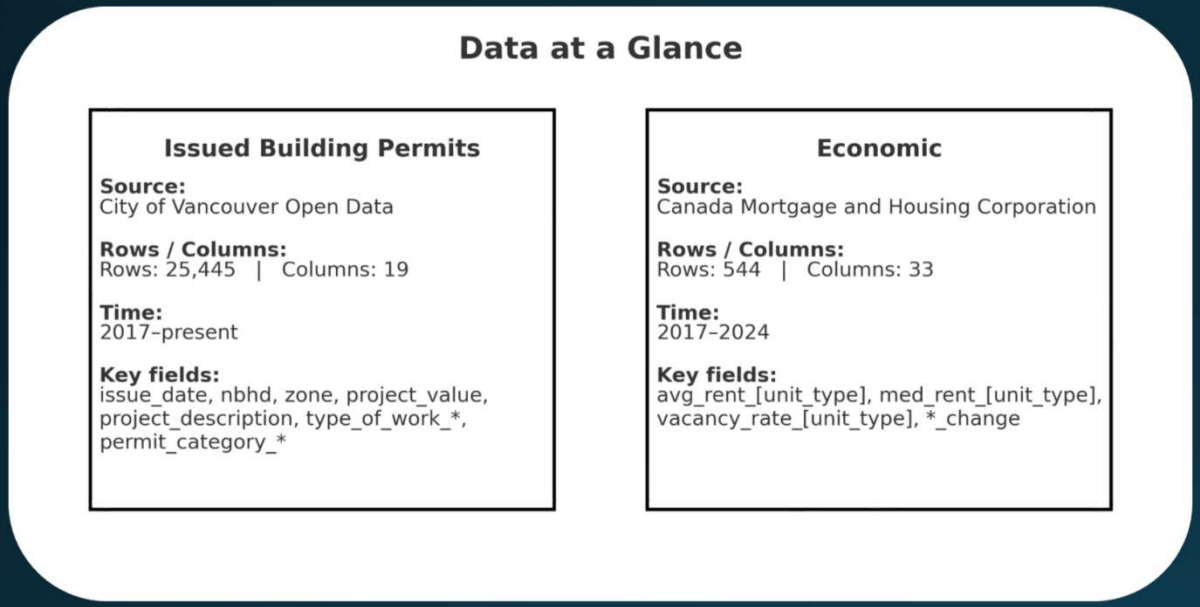
Context: Major policy shifts (such as the Broadway Plan) are reshaping neighbourhoods and could impact affordability.

Data: City-wide rental market trends and building permit records from 2017 to 2024.

How my approach has evolved:

- **Shift in aim:** Moved from *predicting* future rental trends to *inferring* patterns from past data.
- **Methodological change:** Transitioned from *supervised* models to *unsupervised* methods that discover natural groupings in the data.
- **Current focus:** Use *clustering* to identify links between types of development projects and changes in rental affordability.
- **New tools:** Apply *Natural Language Processing* to building permit descriptions to extract themes (e.g., high-rise, mixed-use) as additional features for analysis.



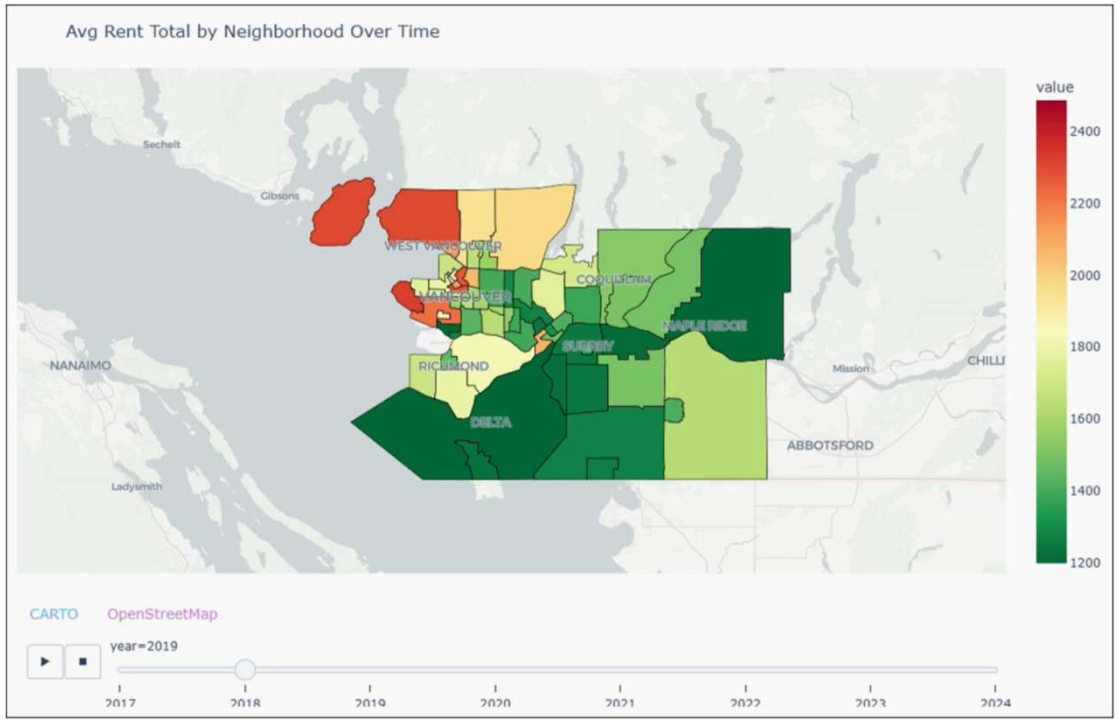


Data preprocessing overview

- Combined ~100 CSVs into a single dataframe storing economic metrics by neighborhood and year.
- Dropped irrelevant columns/rows and standardized column names and formats across sources.
- Imputed missing economic metrics hierarchically using higher level geography (zone → neighbourhood).
- Inflation-adjusted all monetary values to 2024 CAD using Consumer Price Index (CPI).

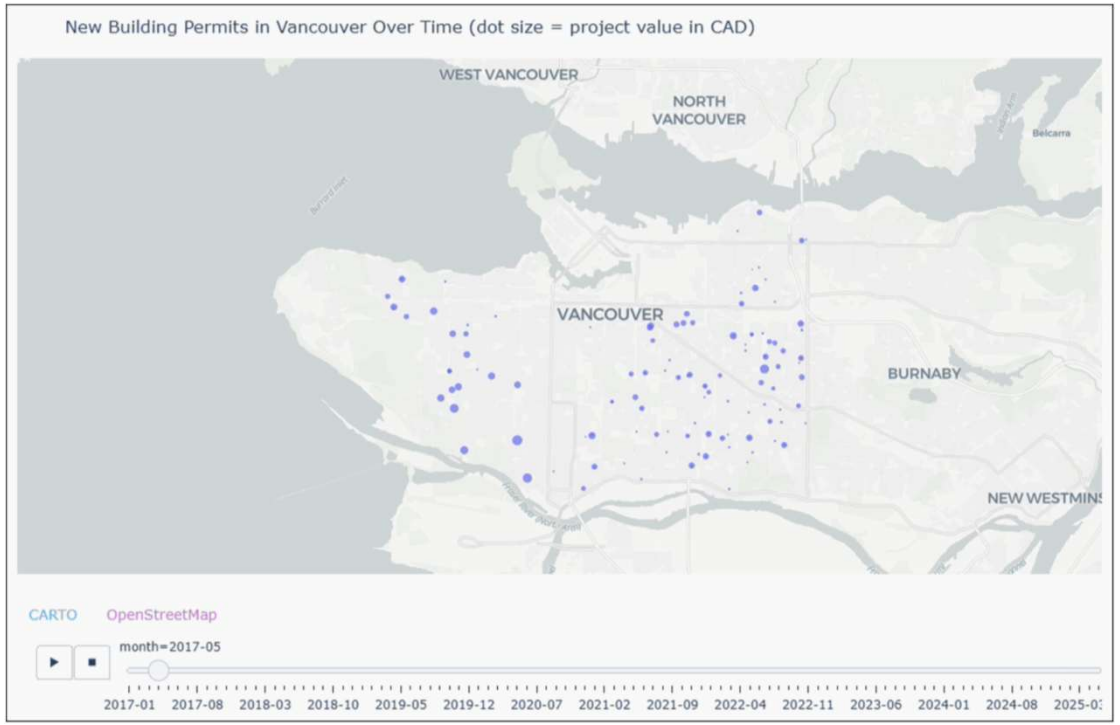
Economic EDA: Rents & Vacancy

- Our dataset reveals large disparities across neighbourhoods in average/median rents and vacancy rates.
- Vacancy is very low: across all neighbourhoods/years the mean is 1.23%, well below the ~3% "healthy" benchmark.
- After CPI-adjusting to 2024 CAD, rents still rise substantially in many neighbourhoods.
- Patterns vary by unit type, year, and neighborhood but the overall trend is tightening markets and rising rents.

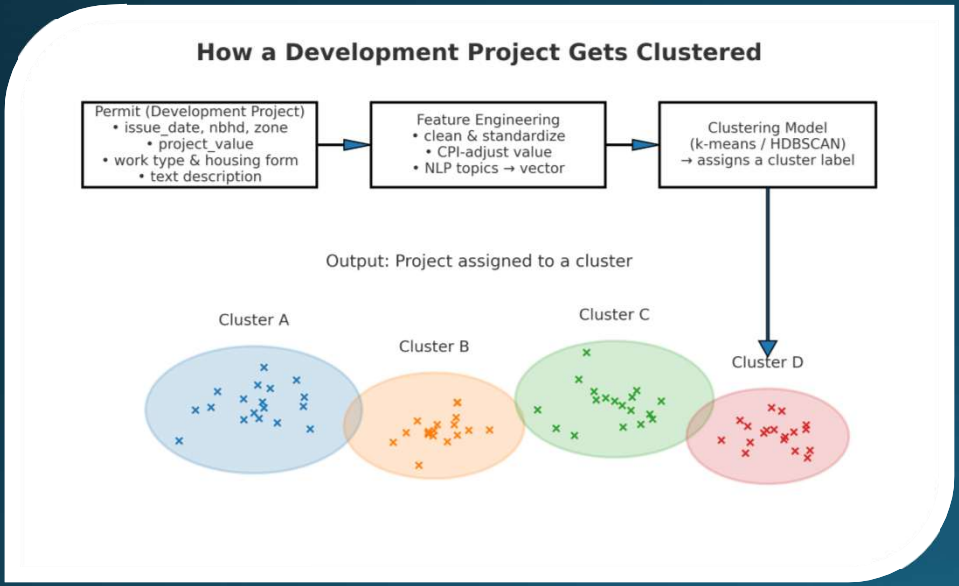


Permit EDA: Development Activity & Project Values

- What's included: new buildings, demolitions, and renovations (laneway house, duplex, secondary suite, single-detached suite, multiple dwelling, multiple conversion dwelling, ...).
- Project value (in 2024 CAD) is highly skewed:
  - median ≈ $101K
  - mean ≈ $0.88M
  - max ≈ $840.8M
- Trend: Even after CPI adjustment, project values rise over time.
- Early NLP signals: storey_building, demolish_existing, high_density, multiple_dwelling, strata, zoning_development, parking_spaces,...



Modeling approach: Clustering Development Projects



- **Main objective:** apply unsupervised clustering models to our building permit projects to classify development types.
- **Relate to economics:** Aggregate project-type intensity by neighbourhood and year and examine lagged correlations with rent and vacancy.
- **Economic typologies:** Separately cluster the economic dataset (rents, vacancy, etc.) to identify neighbourhood "status" classes.
- **Paired clustering:** cluster "development decisions" (pairings of building permits with neighbourhoods' concurrent economic metrics) to infer more patterns.

Next Steps

- Refine the NLP features by building TF-IDF vectors (uni/bi-grams), lemmatizing, and curating a focused keyword list.
- Apply PCA to mitigate the "curse of dimensionality" before clustering.
- Experiment with major clustering algorithms (k-means, GMM, HDBSCAN, etc.) and tune hyperparameters based on interpretability and key metrics (silhouette, Davies–Bouldin, Calinski–Harabasz, etc.).
- Connect development clusters to economic patterns by aggregating cluster mix by neighbourhood and year and testing lagged associations with rents and vacancy.
- Compile detailed visualizations into a public dashboard summarizing key patterns and statistical findings.

Questions?