Idea #5: Predicting Housing Affordability Risk for Canadian Neighborhoods

Acceptance Likelihood: 4.0/5 — Solid Pick

The Problem

Canada is in the midst of a housing affordability crisis affecting renters and buyers across every major city. Housing costs have outpaced wage growth for years, pushing more Canadians into financial precarity.

Key Statistics

- **45%** of Canadians report being very concerned about housing affordability
- Owned accommodation costs have risen **25.1%** from early 2021 to October 2024
- Rent prices rose **24.0%** in the same period
- A household is "unaffordable" when it spends **>30%** of pre-tax income on shelter
- One-third of Canadian renters spend more than 30% of income on housing
- Housing challenges are worse for **lone-parent families, seniors, Indigenous peoples, and recent immigrants**
- Vacancy rates in major cities remain historically low

Why It Matters

Housing affordability affects everything: health, education, employment, family stability. Predicting which neighborhoods are at risk of becoming unaffordable enables **early intervention** — rent controls, affordable housing development, zoning changes, and targeted subsidies can be deployed before displacement occurs.

Why Machine Learning Can Help

Housing affordability follows economic and demographic patterns:

- **Local income trends** (median household income, employment rate)
- **New housing construction rates**
- **Population growth and migration patterns**
- **Interest rates and mortgage availability**
- **Transit proximity and walkability scores**
- **Neighborhood demographics** (age, immigration, family size)
- **Historical price trajectories**

ML can model the interaction of these factors to predict affordability trajectories at the neighborhood level.

ML Tasks

1. **Time-series forecasting:** Predict shelter-cost-to-income ratio for next 1-2 years
2. **Classification:** Flag neighborhoods transitioning from affordable to unaffordable
3. **Regression:** Predict rent/price changes from economic indicators

Kaggle Competition

A Kaggle competition "Housing Affordability in Canada" validates this as an ML-friendly problem with clean data.

Datasets

1. Canadian Housing Survey (CHS)

- **Source:** Statistics Canada
- **URL:** https://www150.statcan.gc.ca/n1/daily-quotidien/240910/dq240910b-eng.htm
- **Contains:** Shelter-cost-to-income ratios, housing satisfaction, affordability challenges by demographic group
- **Coverage:** National, annual since 2018
- **Format:** Public use microdata

2. New Housing Price Index (NHPI)

- **Source:** Statistics Canada via Open Canada Portal
- **URL:** https://open.canada.ca/data/en/dataset/324befd1-893b-42e6-bece-6d30af3dd9f1
- **Contains:** Monthly new housing price changes by CMA
- **History:** Monthly data since January 1981 — excellent for time-series ML
- **Format:** CSV, downloadable

3. Kaggle Housing Affordability Dataset

- **Source:** Kaggle competition
- **URL:** https://www.kaggle.com/competitions/housing-affordability-in-canada/data
- **Contains:** Curated features for predicting housing affordability across metropolitan regions
- **Format:** Clean CSV, competition-ready

Proposed ML Pipeline

```
1. Data collection
   - Download NHPI monthly data (1981-present)
   - Obtain CHS affordability metrics by CMA
   - Merge with Census demographic data (income, population, immigration)
   - Add economic indicators (employment rate, interest rates)

2. Feature engineering
   - Price-to-income ratios by CMA
   - Year-over-year price change rates
   - Population growth rates
   - Lagged economic indicators (interest rates, unemployment)
   - Construction starts per capita

3. Model training
   - XGBoost regressor (predict future affordability ratio)
   - Random Forest classifier (affordable vs. at-risk vs. unaffordable)
   - ARIMA/Prophet for time-series baseline
   - Compare ML to traditional econometric models

4. Evaluation
   - RMSE for price/ratio prediction
   - Classification accuracy for affordability tiers
   - Backtesting: use 2015-2020 data to predict 2021-2024 crisis
   - Geographic analysis of predictions vs. actuals
```

```
5. Deliverable
   - Affordability risk dashboard by CMA/neighborhood
   - Time-series forecasts with confidence intervals
   - Policy analysis: which interventions correlate with improved affordability?
```

Strengths

- **Universal relevance:** Housing is #1 concern for Canadians — every reviewer can relate
- **Excellent data depth:** NHPI monthly since 1981 provides rich time-series training data
- **Kaggle competition:** Pre-validated as ML-friendly problem with clean features
- **Novel for Let's SOLVE It:** Housing never explored in past cohorts
- **Clear ML task:** Regression/classification with structured tabular data

Risks & Mitigations

Risk	Mitigation
More economics than social good	Frame around vulnerable populations: seniors displaced, families forced to food insecurity
Weaker equity angle	Focus on differential impact: how affordability crisis hits marginalized groups hardest
Predictions are informational, not directly actionable	Partner output with policy recommendations
Kaggle competition means it's "known"	Use competition as starting point, extend with Canadian-specific features

References

- StatCan (2024): "Housing affordability in Canada, 2022" — https://www150.statcan.gc.ca/n1/daily-quotidien/240910/dq240910b-eng.htm
- StatCan (2024): "Housing challenges related to affordability" — https://www150.statcan.gc.ca/n1/daily-quotidien/241119/dq241119b-eng.htm
- Open Canada: New Housing Price Index — https://open.canada.ca/data/en/dataset/324befd1-893b-42e6-bece-6d30af3dd9f1
- Kaggle: Housing Affordability in Canada — https://www.kaggle.com/competitions/housing-affordability-in-canada/data
- StatCan (2025): "Research to Insights: Perspectives on Affordability and Inequality" — https://www150.statcan.gc.ca/n1/pub/11-631-x/11-631-x2025001-eng.htm