

Housing Insights & Risk Dashboard

A Data-Driven Framework for Housing Forecasting and Risk
Assessment in Canada

Author: Yuri Spizhovy
Affiliation: MIT / MIT Emerging Talent
Course: Computer and Data Science
Date: Oct 2025

Co-Author: Max Spizhovy
Affiliation: Okanagan College
Course: Computer Information Systems

Executive Summary

The Canadian housing market faces unprecedented affordability challenges, driven by rapid immigration, limited housing supply, rising interest rates, and regional disparities. Policymakers, investors, and households require timely and reliable tools to anticipate price trends, assess risks, and support decision-making. The *Housing Insights & Risk Dashboard* was developed to address this need.

Built on a foundation of **economic theory, statistical modeling, and modern machine learning**, the dashboard integrates data from trusted sources such as the Canadian Real Estate Association (CREA), Canada Mortgage and Housing Corporation (CMHC), Statistics Canada, and the Bank of Canada. It combines **econometric models** for interpretability with **machine learning methods** for predictive accuracy, providing forecasts of housing prices, rental trends, and affordability indicators.

The system architecture follows a **transparent ETL pipeline** and robust database schema, ensuring reproducibility and traceability of all forecasts. The interactive dashboard allows users to explore market dynamics, compare scenarios, and monitor affordability stress across cities like Kelowna, Vancouver, and Toronto. Risk metrics, including affordability indices, volatility measures, and scenario stress tests, are presented in an intuitive, user-friendly interface.

By bridging academic research with practical implementation, the *Housing Insights & Risk Dashboard* contributes both as a **research tool** and as a **decision-support system**. Its design ensures scalability for future enhancements, including real-time monitoring, geospatial modeling, and climate risk integration. Ultimately, the project advances the transparency and reliability of housing market analysis in Canada, supporting evidence-based policy and informed decision-making.

Table of Contents

Executive Summary

Abstract

1. Introduction

- 1.1 Motivation and Purpose of the Project
- 1.2 Scope: Canadian Housing Market Focus
- 1.3 Research Questions

2. Background and Literature Review

- 2.1 Housing Market Dynamics: Economic Foundations
- 2.2 Key Drivers of House Prices and Rent Levels
- 2.3 Review of Forecasting Approaches in Real Estate Economics
- 2.4 Existing Housing Forecasting Models and Dashboards

3. Data Sources and Market Context

- 3.1 Canadian Housing Market Overview (CREA, CMHC, StatCan, BoC)
- 3.2 Rental Market Data (Rentals.ca, RentFaster, Craigslist, Castanet)
- 3.3 Economic Indicators and Risk Factors (GDP, Interest Rates, Inflation)
- 3.4 Data Challenges: Availability, Quality, and Bias

4. Statistical and Econometric Methods

- 4.1 Time Series Analysis (ARIMA, VAR, Cointegration)
- 4.2 Regression Approaches (Hedonic Pricing Models, OLS, Panel Data)
- 4.3 Risk Assessment Techniques (Volatility, Value-at-Risk for Housing)

5. Machine Learning and Predictive Modeling

- 5.1 Supervised Learning for Price Forecasting (Linear Models, Tree-Based Methods, Gradient Boosting, Neural Networks)
- 5.2 Unsupervised Learning for Market Segmentation and Risk Profiling
- 5.3 Hybrid Approaches (Econometrics + ML, e.g., Prophet, LightGBM)
- 5.4 Model Evaluation Metrics (MAPE, RMSE, R^2 , Backtesting)

6. Dashboard Design and Implementation

- 6.1 System Architecture (ETL → Database → ML Models → API → Dashboard)
- 6.2 Data Visualization Principles for Housing Markets
- 6.3 Risk Scoring and Forecast Representation (Confidence Intervals, Stress Scenarios)
- 6.4 User Interaction and Decision-Support Features

7. Case Study: Canadian Cities

- 7.1 Kelowna: Housing Affordability and Rent Trends
- 7.2 Vancouver: Price Volatility and Risk Factors
- 7.3 Toronto: Demand Pressures and Forecast Accuracy
- 7.4 Comparative Analysis Across Markets

8. Discussion

- 8.1 Insights from Forecasting Models
- 8.2 Implications for Policymakers, Investors, and Renters
- 8.3 Limitations of the Current Approach

9. Future Directions

- 9.1 Incorporating Geospatial and Demographic Data
- 9.2 Climate Change and Housing Risk Forecasting
- 9.3 Toward Real-Time Market Monitoring

10. Conclusion

- 10.1 Summary of Findings
- 10.2 Contribution of the Housing Insights & Risk Dashboard
- 10.3 Path Forward

11. References

- 11.1 Main References (all sources directly cited in the essay: scientific papers, policy reports, datasets, ML literature [1]–[69])
- 11.2 Additional References (datasets, geospatial/climate data, and ML tool documentation [70]–[78])

12. Appendices

- A. Database Schema and ETL Process
- B. Mathematical Derivations
- C. Model Hyperparameters and Configurations

13. Notes on Methodology and Tools

Abstract

The Canadian housing market has undergone unprecedented volatility in recent years, driven by structural and cyclical factors such as limited supply, rising immigration, changing monetary policy, and heightened investor participation. This has created affordability challenges for households while complicating policy responses at federal, provincial, and municipal levels [1][2][3]. Reliable forecasting of housing prices and rental costs is therefore crucial for supporting evidence-based policymaking, investment decisions, and risk management strategies.

This paper provides the conceptual and methodological foundation of the Housing Insights & Risk Dashboard, a decision-support platform designed to analyze and forecast housing trends in Canada. Drawing on both classical econometric approaches—such as time series analysis, hedonic regression, and volatility modeling—and contemporary machine learning techniques, including gradient boosting, neural networks, and hybrid models, the dashboard offers a robust and scalable framework for housing prediction. Data sources include the Canadian Real Estate Association (CREA), Canada Mortgage and Housing Corporation (CMHC), Statistics Canada, the Bank of Canada, and online rental platforms, ensuring a comprehensive perspective on both ownership and rental dynamics.

By integrating statistical rigor with modern predictive analytics, the dashboard contributes to both academic and applied discussions of housing market monitoring. It fills a gap in the Canadian context, where existing forecasting efforts remain fragmented and often lack transparency. The Housing Insights & Risk Dashboard seeks to enhance market transparency, improve affordability analysis, and provide stakeholders with a reliable tool for monitoring risk and forecasting future trends in Canada's housing sector.

1. Introduction

1.1 Motivation and Purpose of the Project

The housing market plays a critical role in the Canadian economy, both as a driver of household wealth and as a determinant of social well-being. However, affordability has deteriorated significantly across many urban centers, with housing prices and rents rising faster than income growth [1]. The COVID-19 pandemic, shifts in immigration, and interest rate volatility have further intensified uncertainty in the housing sector [2]. For policymakers, investors, and households, the ability to monitor risk and forecast housing outcomes is increasingly vital [3].

The *Housing Insights & Risk Dashboard* was developed in response to these challenges. Its purpose is to create a transparent, data-driven platform for analyzing housing affordability, forecasting price dynamics, and assessing risk across Canadian cities. By integrating economic theory, statistical modeling, and modern machine learning, the dashboard seeks to bridge the gap between academic research and real-world decision-making.

1.2 Scope: Canadian Housing Market Focus

While housing affordability is a global concern, Canada presents unique dynamics shaped by rapid immigration, regional disparities, and regulatory frameworks [3]. For example, Toronto and Vancouver have long experienced affordability crises due to high demand and limited land availability, while mid-sized cities such as Kelowna face mounting pressures from population growth and rental shortages [4].

The scope of this project therefore centers on Canadian housing markets, with case studies focused on Kelowna, Vancouver, and Toronto. Data is drawn from the Canadian Real Estate Association (CREA), Canada Mortgage and Housing Corporation (CMHC), Statistics Canada, the Bank of Canada, and rental listing platforms such as RentFaster and Rentals.ca. Together, these datasets provide a multi-dimensional view of housing dynamics, covering both ownership and rental markets.

1.3 Research Questions

This essay is guided by the following research questions:

1. **Price Forecasting:** How can statistical and machine learning models be used to predict housing prices and rental costs in Canadian cities?
2. **Risk Assessment:** Which economic and market indicators (e.g., interest rates, supply constraints, demographic trends) are most predictive of housing market volatility and affordability stress?

3. **Rental Dynamics:** To what extent can rental listing data serve as a timely leading indicator of broader market conditions?
4. **Decision Support:** How can predictive insights be effectively translated into a user-facing dashboard that supports policymakers, investors, and households?

By addressing these questions, the *Housing Insights & Risk Dashboard* positions itself as both a scientific contribution and a practical tool. It advances current research by combining econometric methods with machine learning, while also providing an interactive interface that makes complex forecasting accessible to non-specialists.

2. Background & Literature Review

2.1 Economic Foundations of the Housing Market

Housing markets combine features of consumption goods and long-lived capital assets. As a result, they are highly sensitive to macroeconomic conditions such as interest rates and credit availability, as well as structural constraints like land supply and regulation. In Canada, empirical research has shown that real house prices are cointegrated with macroeconomic variables such as income, population growth, and credit expansion, helping to explain long-run price trends [4].

On the supply side, studies of Canadian housing investment have demonstrated that the elasticity of investment with respect to interest rates has increased over time, indicating greater sensitivity of the sector to financing costs [5]. Because housing is illiquid and subject to construction delays, short-term deviations from fundamentals are common. This motivates the integration of traditional macroeconomic models with more flexible forecasting tools.

2.2 Key Drivers of House Prices and Rental Values

Empirical literature identifies several recurring predictors of housing outcomes:

- **Interest Rates and User Cost of Capital:** Changes in mortgage rates directly affect the cost of borrowing and demand for housing. Research modeling Canadian housing demand with user cost frameworks shows the critical role of financing costs in shaping affordability [6].
- **Income, Employment, and Demographics:** Household income growth, labor market stability, and immigration strongly influence demand for housing.
- **Supply Constraints and Construction Activity:** Housing starts, building permits, and zoning regulation directly impact available inventory.
- **Spatial Heterogeneity:** Location plays a major role in pricing. Spatial econometric studies of Canadian housing markets demonstrate that models accounting for spatial correlation outperform simple regression approaches [7].
- **Nonlinearities and Regime Changes:** Determinants of prices can shift dramatically under stress, such as during interest rate shocks. Flexible or hybrid models are better suited to capturing these dynamics.

Rental markets respond more quickly to demand changes, making rental listings a valuable leading indicator. Predictive models of Canadian housing need, such as those developed by CMHC, now incorporate rental market dynamics into their frameworks [8].

2.3 Econometric and Statistical Approaches in Housing Forecasting

Hedonic Pricing Models

Hedonic regression, which decomposes house prices into the value of their characteristics, has long been a standard tool. Surveys of hedonic methods highlight both traditional parametric models and newer nonparametric approaches [9]. Recent advances include hybrid methods that integrate hedonic pricing with machine learning, such as random forests, to improve predictive accuracy [10].

Spatial hedonic models allow coefficients to vary geographically, and empirical work on Canadian housing markets shows they outperform ordinary least squares (OLS) in capturing neighborhood-level variation [11]. Comparative studies confirm that spatial regression models adapt better to evolving parameter structures in Canadian cities [12].

Time Series and Multivariate Models

Time series models such as ARIMA and Vector Autoregression (VAR) have been widely applied to Canadian housing forecasting. Government reports show ARIMA performs well for short-term housing starts forecasting [13]. More recently, predictive modeling has been used to project flows into and out of core housing need [14].

Machine learning techniques are increasingly applied to Canadian housing data. A Bank of Canada study compared linear regression with machine learning approaches such as support vector regression and multilayer perceptrons, finding modest improvements in forecast accuracy under certain conditions [15].

2.4 Hybrid and Advanced Methods

Recent research emphasizes combining econometric interpretability with the flexibility of machine learning. Examples include:

- **Hybrid ARIMA–ML Models:** Using ML to model residuals from traditional time series models.
- **Spatial Functional Models:** Hierarchical spatial functional modeling decomposes housing value into latent surfaces and feature-based components, with strong empirical performance in Canadian datasets [16].
- **Image-Based Features:** Deep learning applied to housing façade images enhances predictive accuracy when integrated into hedonic models [17].
- **Joint Mean–Variance Models:** New semiparametric approaches explicitly model both mean and variance in housing price surfaces, enabling better risk assessment [18].

2.5 Gaps and Positioning of the Dashboard

From this literature, we observe that:

- Traditional econometric methods provide interpretability but struggle with nonlinearities and regime shifts.
- Pure ML models often lack transparency and policy relevance.
- No open-access, Canada-focused forecasting platform integrates both approaches with interactive visualization.

The *Housing Insights & Risk Dashboard* is designed to fill this gap by combining structured econometric models with ML enhancements, embedding spatial analysis, and producing risk metrics alongside forecasts.

3. Data Sources & Market Context

3.1 Canadian Housing Market Overview

The Canadian housing market is characterized by regional heterogeneity, with dynamics differing significantly between major metropolitan areas and mid-sized cities. Nationally, house price growth has been strongly correlated with macroeconomic conditions, immigration-driven demand, and supply-side constraints [20]. The Canadian Real Estate Association (CREA) publishes the **MLS® Home Price Index (HPI)**, which is widely used for monitoring long-term price trends and regional differences [21]. CREA data provides granular coverage across cities, making it central to any forecasting framework.

3.2 Rental Market Data

Rental housing plays an increasingly important role in affordability analysis. According to CMHC's **Rental Market Survey (RMS)**, vacancy rates in urban centers such as Toronto and Vancouver remain historically low, while demand pressures have spread to mid-sized cities like Kelowna [22].

In addition to official surveys, private rental listing platforms such as **Rentals.ca** and **RentFaster.ca** provide high-frequency, city-level rental price data. These sources are especially valuable for capturing near real-time changes in demand and affordability conditions [23]. While such datasets can be noisy due to inconsistencies in listing practices, they offer timeliness that complements lagged government reports.

3.3 Economic Indicators and Risk Factors

Housing affordability and price risk cannot be assessed in isolation. Broader economic indicators—including **interest rates, inflation, employment, and GDP growth**—are essential to understanding market fluctuations. The Bank of Canada provides detailed monetary policy reports and datasets on mortgage rates, household debt, and macroeconomic conditions that directly influence housing demand [24]. Statistics Canada publishes quarterly and annual reports on household income, labor market dynamics, and demographic growth, which are widely used in affordability and risk modeling [25].

Risk factors also include financial stability indicators, such as household debt-to-income ratios and mortgage arrears rates. Bank of Canada reports emphasize the rising vulnerability of highly indebted households to interest rate increases, amplifying systemic housing risks [26].

3.4 Data Challenges: Availability, Quality, and Bias

Despite the richness of available sources, Canadian housing data faces challenges:

- **Fragmentation:** Data is spread across government agencies, industry associations, and private platforms.

- **Lag and Timeliness:** Official data (e.g., CMHC surveys, StatCan affordability reports) is often published quarterly or annually, while market shifts occur much faster.
- **Coverage Gaps:** Private rental data captures active listings but excludes informal rentals and may overrepresent specific market segments.
- **Bias and Consistency:** Methodological differences across sources can lead to inconsistent signals.

To address these issues, the *Housing Insights & Risk Dashboard* integrates multiple datasets, cross-validating official government sources with higher-frequency private listings. This blended approach provides a more complete, timely, and reliable representation of Canada's housing market dynamics.

4. Statistical and Econometric Methods

4.1 Time Series Analysis

Time series models remain foundational for housing price and rent forecasting. The **Autoregressive Integrated Moving Average (ARIMA)** model has been widely applied to Canadian housing starts and price indices, offering reliable short-term forecasts [27]. Extensions such as **Vector Autoregression (VAR)** allow modeling of interdependencies between housing prices, interest rates, and macroeconomic variables [28]. These approaches are valued for their interpretability and strong performance in stable economic environments, though they may underperform when structural breaks or nonlinearities emerge.

4.2 Regression Approaches

Regression-based models, especially **hedonic pricing models**, have long been used to estimate the impact of property characteristics (size, age, location) on housing values. While discussed in Section 2, here they are noted for their econometric rigor and policy relevance. **Panel data models**—leveraging variation across regions and time—are particularly useful in capturing regional heterogeneity in Canadian markets [29].

A common regression variant in housing analysis is the **error-correction model (ECM)**, which accounts for both short-run dynamics and long-run equilibrium relationships between housing prices and fundamentals such as income and interest rates [30]. These models provide policymakers with insights into how quickly housing markets adjust after shocks.

4.3 Risk Assessment Techniques

Beyond point forecasts, effective housing dashboards must evaluate **risk and volatility**. Several econometric frameworks are relevant:

- **ARCH/GARCH models** estimate volatility dynamics, allowing quantification of uncertainty in housing returns [31].
- **Value-at-Risk (VaR)** techniques, adapted from finance, have been applied to housing to estimate downside risks under stress conditions [32].
- **Scenario analysis** links macroeconomic shocks (e.g., rising interest rates, unemployment) to housing affordability, offering forward-looking risk metrics for policymakers [33].

4.4 Limitations of Purely Econometric Methods

While these statistical approaches offer interpretability and are grounded in economic theory, they face limitations:

- Difficulty capturing nonlinear interactions between variables.
- Sensitivity to structural breaks, such as the 2008 global financial crisis or the post-COVID housing surge.
- Dependence on assumptions (stationarity, homoskedasticity) that are often violated in real-world housing data.

These limitations motivate the integration of **machine learning methods** (Section 5), which can model complex nonlinearities and interactions, complementing the interpretability of econometrics.

5. Machine Learning and Predictive Modeling

5.1 Supervised Learning for Price Forecasting

Machine learning (ML) methods have gained traction in housing economics due to their ability to capture nonlinear relationships and complex interactions between variables. **Tree-based ensemble models** such as Random Forests, Gradient Boosting Machines (GBM), XGBoost, and LightGBM have been widely applied to housing price forecasting. Empirical studies find that these models often outperform traditional econometric approaches in predictive accuracy, particularly in short- to medium-term forecasts [34].

Neural networks, including feed-forward multilayer perceptrons and recurrent architectures, have also been employed in predicting housing prices. While these models can capture deep nonlinearities, they often require large training datasets and careful regularization to avoid overfitting [35].

5.2 Unsupervised Learning for Market Segmentation and Risk Profiling

Unsupervised methods provide complementary insights by uncovering patterns in housing data without predefined labels. **Clustering techniques** such as k-means and hierarchical clustering have been used to identify market segments with similar affordability pressures or risk exposure [36]. **Principal Component Analysis (PCA)** and other dimensionality reduction techniques help simplify high-dimensional datasets (e.g., rental listings enriched with geographic, demographic, and macroeconomic variables) into interpretable risk factors [37].

5.3 Hybrid Approaches

Recent research advocates for hybrid approaches that combine the **interpretability of econometric models** with the **flexibility of machine learning**. For example:

- **ARIMA-ML Ensembles:** Residuals from ARIMA forecasts can be modeled using gradient boosting, improving accuracy under structural breaks [38].
- **Spatial + ML Models:** Spatial econometric frameworks enriched with ML predictors capture both location dependence and nonlinear feature effects [39].
- **Deep Learning + Auxiliary Data:** Integration of nontraditional data sources (satellite imagery, street-view photos) into neural network architectures has been shown to enhance predictive accuracy in urban housing markets [40].

5.4 Model Evaluation Metrics

A rigorous evaluation framework is necessary to ensure predictive reliability. Common metrics include:

- **Mean Absolute Percentage Error (MAPE)** – measures relative error, useful for affordability discussions.
- **Root Mean Square Error (RMSE)** – penalizes large errors, standard in housing price forecasting [41].
- **R² (Coefficient of Determination)** – assesses explanatory power of features.
- **Backtesting and Cross-Validation** – evaluates out-of-sample performance under different market regimes [42].

These metrics, when applied consistently, allow transparent comparison between econometric, ML, and hybrid models within the *Housing Insights & Risk Dashboard*.

6. Dashboard Design and Implementation

6.1 System Architecture

The *Housing Insights & Risk Dashboard* is designed as a modular system integrating data collection, storage, modeling, and visualization layers. The architecture follows a classic **ETL (Extract–Transform–Load) pipeline**, feeding into a relational database and then into forecasting models before being exposed through an interactive API and dashboard interface [43].

- **Data Layer:** Sources include CREA (house prices), CMHC (rental/starts), Statistics Canada (demographics, income), Bank of Canada (interest rates), and private rental platforms.
- **Processing Layer:** Data cleaning, aggregation, and feature engineering are implemented in Python.
- **Modeling Layer:** Both econometric and machine learning models are deployed, enabling side-by-side comparison of forecasts.
- **Service Layer:** A RESTful API (FastAPI/Spring Boot) allows integration with front-end applications and external tools.
- **Visualization Layer:** Built with modern web frameworks (React + TypeScript, Recharts/D3.js) for interactive and responsive visualization of trends, forecasts, and risk metrics.

This layered design ensures scalability, transparency, and reproducibility [44].

6.2 Data Visualization Principles for Housing Markets

Dashboards must translate complex quantitative results into **intuitive insights** for policymakers, investors, and households. Research in visual analytics emphasizes the importance of:

- **Clear time-series visualization** for housing price and rent forecasts.
- **Uncertainty bands** (confidence intervals, fan charts) to convey risk rather than single-point predictions [45].
- **Geospatial visualization** to highlight regional disparities in affordability and risk.
- **Interactive filtering** (city, dwelling type, horizon) to allow users to tailor forecasts to their context.

Studies show that properly designed visualization tools improve user trust and decision-making, especially when accompanied by uncertainty representation [46].

6.3 Risk Scoring and Forecast Representation

To complement forecasts, the dashboard incorporates **risk indicators**, including:

- **Affordability Stress Index:** ratio of median price/rent to household income.
- **Volatility Index:** based on GARCH-modeled uncertainty in housing returns.
- **Scenario Stress Tests:** impact of hypothetical shocks (e.g., 2% interest rate rise) on affordability and prices.

These indicators are presented as **traffic-light style gauges** and **heat maps**, supporting quick interpretation by non-technical users [47].

6.4 User Interaction and Decision-Support Features

The dashboard is designed as a **decision-support system**, balancing technical rigor with accessibility. Users can:

- Select geographic focus (national, provincial, city-level).
- Compare econometric vs. ML forecasts.
- Explore rental vs. ownership dynamics.
- Download forecast datasets or reports for further analysis.

This aligns with best practices in **human-computer interaction (HCI)** for decision-support systems, where interactivity and transparency enhance user engagement and adoption [48].

7. Case Study: Canadian Cities

7.1 Kelowna: Housing Affordability and Rent Trends

Kelowna has emerged as one of the fastest-growing housing markets in Canada, driven by population growth, interprovincial migration, and lifestyle-oriented demand. CMHC data show that vacancy rates remain among the lowest in the country, while rental affordability has deteriorated sharply [49]. The affordability stress in Kelowna is amplified by a mismatch between income levels and rising median rents, a dynamic confirmed by recent CMHC Rental Market Reports [50].

7.2 Vancouver: Price Volatility and Risk Factors

Vancouver represents one of the most volatile and internationally influenced housing markets in Canada. Foreign investment, supply-side constraints due to geography, and persistent immigration-driven demand have contributed to sustained affordability challenges [51]. Studies highlight that Vancouver house prices are highly sensitive to changes in interest rates, with mortgage tightening policies and macroprudential regulations having measurable effects on price dynamics [52]. Vancouver's inclusion of vacant home and foreign buyer taxes has introduced unique policy variables into the forecasting landscape [53].

7.3 Toronto: Demand Pressures and Forecast Accuracy

Toronto remains Canada's largest housing market, with affordability challenges driven by strong immigration flows and persistent supply shortages. The Greater Toronto Area (GTA) continues to record some of the highest price-to-income ratios in the country [54]. Forecasting models applied to Toronto must account for both cyclical macroeconomic drivers and structural demographic pressures. Empirical studies suggest that incorporating rental market dynamics alongside ownership data improves the accuracy of Toronto housing forecasts [55].

7.4 Comparative Analysis Across Markets

Comparing Kelowna, Vancouver, and Toronto reveals important heterogeneity:

- **Kelowna:** affordability challenges stem from **income–rent mismatches** and rapid population growth.
- **Vancouver:** volatility and risk exposure reflect **international capital flows** and geographic supply constraints.
- **Toronto:** affordability crisis reflects **immigration-driven demand and persistent supply shortages**.

These cases demonstrate the importance of tailoring forecasting models to local market dynamics. The *Housing Insights & Risk Dashboard* provides a platform for such localized analysis by combining national economic indicators with city-level data sources.

8. Discussion

8.1 Insights from Forecasting Models

The integration of econometric and machine learning methods provides several key insights for understanding the Canadian housing market. First, while **traditional econometric models** (ARIMA, hedonic regression) offer interpretability and strong performance in stable conditions, they struggle to adapt to nonlinearities and structural breaks [56]. Second, **machine learning approaches** demonstrate improved predictive accuracy in volatile periods, particularly when incorporating diverse feature sets such as demographics, rental trends, and macroeconomic indicators [57]. However, they often trade interpretability for flexibility.

A hybrid approach—combining econometric foundations with machine learning enhancements—emerges as the most effective strategy, balancing explanatory power with predictive accuracy [58]. This aligns with recent international research on housing forecasting that shows blended models outperform either approach in isolation.

8.2 Implications for Policymakers, Investors, and Renters

For **policymakers**, the dashboard offers timely insights into housing affordability and systemic risks. Real-time rental data can serve as an early warning system for affordability stress, supporting more proactive policy interventions [59].

For **investors and developers**, integrating economic indicators with price forecasts provides a more robust framework for evaluating risk-return trade-offs. The dashboard's ability to run **scenario analyses**—for example, testing the effect of a 2% interest rate increase—enables better capital allocation under uncertainty [60].

For **renters and households**, the dashboard enhances market transparency. By presenting affordability indices and forecasts in an accessible format, households can better anticipate rental pressures or evaluate purchase timing, supporting informed decision-making [61].

8.3 Limitations of the Current Approach

Despite its contributions, the *Housing Insights & Risk Dashboard* faces limitations.

- **Data Constraints:** Private rental platforms may introduce bias, while official datasets often lag market conditions.
- **Model Uncertainty:** Forecasts are sensitive to model selection, parameter tuning, and structural assumptions.
- **External Shocks:** Housing forecasts are vulnerable to unpredictable shocks such as global financial crises, pandemics, or sudden regulatory changes.

- **User Interpretation:** Even with intuitive visualization, misinterpretation of probabilistic forecasts by non-experts remains a challenge [62].

Addressing these limitations requires ongoing refinement of data integration, model calibration, and user interface design.

9. Future Directions

9.1 Incorporating Geospatial and Demographic Data

One key future direction is the integration of **geospatial data** (e.g., satellite imagery, land use maps) and **fine-grained demographic data** into forecasting models. Geospatial analytics can improve prediction accuracy by capturing neighborhood-level heterogeneity [63]. For instance, combining census tract data with property-level characteristics allows for better identification of affordability hotspots. Demographic projections (immigration, aging population) will also be essential for long-term housing demand forecasting [64].

9.2 Climate Change and Housing Risk Forecasting

Climate change presents emerging risks to Canadian housing markets. Rising flood risks, wildfire exposure, and insurance availability are increasingly linked to property values and affordability [65]. Forecasting models must therefore account for **climate vulnerability indices** and **environmental risk data** alongside economic variables. Recent research shows that housing markets exposed to environmental risks may face long-term devaluation, particularly in regions where climate shocks interact with affordability challenges [66].

9.3 Toward Real-Time Market Monitoring

Another future direction is the development of **real-time housing market monitoring systems**, leveraging high-frequency data sources such as rental listings, mortgage approvals, and online search behavior. Advances in **natural language processing (NLP)** enable sentiment analysis of housing-related news and policy announcements, providing early signals of market sentiment shifts [67].

Integration of these data streams with machine learning pipelines can enable near-real-time updates of forecasts and risk indicators. This would enhance the dashboard's utility as an **early warning system**, aligning with global best practices in financial stability monitoring [68].

10. Conclusion

10.1 Summary of Findings

This essay has provided the scientific foundation for the *Housing Insights & Risk Dashboard*, a decision-support system designed to forecast housing prices, assess affordability, and evaluate systemic risks in the Canadian housing market. The review of the literature confirmed that housing dynamics are shaped by a combination of **economic fundamentals** (income, interest rates, demographics, supply constraints) and **market frictions** (delayed construction, geographic limits, investor behavior).

Traditional econometric models—such as ARIMA, VAR, and hedonic regression—remain valuable for their interpretability but often struggle with nonlinearities and sudden shocks. Machine learning methods, including tree-based ensembles and neural networks, provide enhanced predictive accuracy but at the cost of reduced transparency. The evidence suggests that **hybrid models**, combining econometric structure with machine learning flexibility, offer the most promising pathway for reliable forecasting [69].

10.2 Contribution of the Housing Insights & Risk Dashboard

The dashboard contributes both academically and practically by:

- Integrating **multiple Canadian data sources** (CREA, CMHC, StatCan, BoC, rental listings).
- Implementing **multi-model forecasting pipelines**, allowing users to compare econometric, ML, and hybrid forecasts.
- Providing **risk indicators** such as affordability stress indices, volatility measures, and scenario analysis tools.
- Delivering **accessible visualization** to policymakers, investors, and households through an interactive interface.

In doing so, it addresses a gap in Canada's housing landscape, where existing forecasting tools are fragmented, static, and often inaccessible to the general public.

10.3 Path Forward

Looking ahead, future enhancements should include the integration of **geospatial and climate risk data**, **real-time monitoring** through high-frequency rental and sentiment data, and improved **user interaction design** to ensure accessibility for non-experts. Such developments would position the dashboard as a scalable, adaptable platform capable of supporting **housing policy, investment planning, and household decision-making** under conditions of uncertainty.

By bridging economic theory, statistical methods, and modern machine learning, the *Housing Insights & Risk Dashboard* contributes to advancing both housing research and practical policy tools for one of Canada's most pressing socio-economic challenges.

11. References

Main References

- [1] Bank of Canada (2023). *Financial System Review 2023: Housing Market Risks*. Ottawa: Bank of Canada. [Link](#)
- [2] Canada Mortgage and Housing Corporation (CMHC) (2023). *Housing Market Outlook: Canada*. Ottawa: CMHC. [Link](#)
- [3] Statistics Canada (2022). *Housing Affordability in Canada: Demographic and Regional Perspectives*. Statistics Canada, Analytical Studies Branch Research Paper Series. [Link](#)
- [4] Canada Mortgage and Housing Corporation (CMHC) (2023). *Housing Market Outlook: Major Centres*. Ottawa: CMHC. [Link](#)
- [5] Yu Ma (2014). *Fundamentals Affecting Canadian Housing Market*. Simon Fraser University. [Link](#)
- [6] Demers, F. (2005). *Modeling and Forecasting Housing Investment: The Case of Canada*. Bank of Canada Working Paper 2005-41. [Link](#)
- [7] Forsyth, A. (2019). *A User Cost Approach to Housing Demand in Canada*. Queen's University Economics Department. [Link](#)
- [8] Farber, S., & Yeates, M. (2006). *A Comparison of Localized Regression Models in a Hedonic House Price Context*. *Canadian Geographer*, 50(2), 141–163. [Link](#)
- [9] Canada Mortgage and Housing Corporation (CMHC) (2018). *Estimating Annual and Long-Term Flows In and Out of Core Housing Need*. Ottawa: CMHC. [Link](#)
- [10] Owusu-Ansah, A. (2013). *A Review of Hedonic Pricing Models in Housing Research*. *Journal of Property Research*, 30(2), 123–137. [Link](#)
- [11] Rey-Blanco, C., et al. (2024). *Hybrid Hedonic and Machine Learning Models for Housing Price Prediction*. *Expert Systems with Applications*, 243, 122795. [Link](#)
- [12] Chan, W. M. (2014). *A Comparison of Spatial Hedonic Models: Evidence from Vancouver*. Simon Fraser University. [Link](#)
- [13] Fotheringham, A. S., et al. (2021). *An Over Time Comparison of Linear and Spatial Regression Models for Housing Price Prediction*. *Canadian Association of Geographers Journal*. [Link](#)
- [14] Canada Mortgage and Housing Corporation (CMHC) (1991). *Forecasting Canadian Housing Starts Using ARIMA Models*. Government of Canada. [Link](#)
- [15] Canada Mortgage and Housing Corporation (CMHC) (2021). *Forecasting Trends in Canada's Core Housing Need*. Government of Canada. [Link](#)

- [16] Brännlund, E., et al. (2023). *Predicting Changes in Canadian Housing Markets with Machine Learning*. Bank of Canada Staff Discussion Paper 2023-21. [Link](#)
- [17] Liu, X., et al. (2018). *Hierarchical Spatial Functional Models for Housing Price Prediction*. *Annals of Applied Statistics*, 12(3), 1613–1637. [Link](#)
- [18] Law, S., et al. (2024). *Image-Informed Hedonic Models for Housing Prices Using Deep Learning*. *arXiv preprint*. [Link](#)
- [19] Toloza-Delgado, A., et al. (2024). *Joint Modeling of Housing Price Mean and Variance with Spatial Semi-Parametric Regression*. *arXiv preprint*. [Link](#)
- [20] Canada Mortgage and Housing Corporation (CMHC) (2023). *Canadian Housing Market at a Glance*. Ottawa: CMHC.
<https://www.cmhc-schl.gc.ca/professionals/housing-markets-data-and-research>
- [21] Canadian Real Estate Association (CREA) (2023). *MLS® Home Price Index Methodology*. Ottawa: CREA. <https://www.crea.ca/housing-market-stats/mls-home-price-index>
- [22] Canada Mortgage and Housing Corporation (CMHC) (2023). *Rental Market Survey*. Ottawa: CMHC.
<https://www.cmhc-schl.gc.ca/professionals/housing-markets-data-and-research/rental-market-reports-canada-and-provincial-highlights>
- [23] Rentals.ca (2023). *National Rent Report*. Toronto: Rentals.ca.
<https://rentals.ca/national-rent-report>
- [24] Bank of Canada (2023). *Monetary Policy Report – October 2023*. Ottawa: Bank of Canada.
<https://www.bankofcanada.ca/publications/mpir>
- [25] Statistics Canada (2023). *Household Economic Well-being Dashboard*. Ottawa: Statistics Canada.
https://www150.statcan.gc.ca/n1/en/subjects/economy/household_economic_well-being
- [26] Bank of Canada (2022). *Household Indebtedness and Housing Vulnerabilities*. *Financial System Review*. Ottawa: Bank of Canada.
<https://www.bankofcanada.ca/2022/06/financial-system-review>
- [27] Canada Mortgage and Housing Corporation (CMHC) (1991). *Forecasting Canadian Housing Starts Using ARIMA Models*. Government of Canada.
https://publications.gc.ca/collections/collection_2018/schl-cmhc/nh15/NH15-743-1991-eng.pdf
- [28] Cushman, D. O., & Zha, T. (1997). *Identifying Monetary Policy in a Small Open Economy under Flexible Exchange Rates*. *Journal of Monetary Economics*, 39(3), 433–448.
[https://doi.org/10.1016/S0304-3932\(97\)00029-9](https://doi.org/10.1016/S0304-3932(97)00029-9)
- [29] Holly, S., Pesaran, M. H., & Yamagata, T. (2010). *A Spatio-Temporal Model of House Prices in the USA*. *Journal of Econometrics*, 158(1), 160–173.
<https://doi.org/10.1016/j.jeconom.2010.03.040>

- [30] Meen, G. (2002). *The Time-Series Behaviour of House Prices: A Transatlantic Divide?* *Journal of Housing Economics*, 11(1), 1–23. <https://doi.org/10.1006/jhec.2002.0307>
- [31] Gupta, R., & Miller, S. M. (2012). *The Time Series Properties of House Prices: A Case Study of the Southern California Market*. *Journal of Real Estate Finance and Economics*, 45(2), 452–473. <https://doi.org/10.1007/s11146-010-9267-0>
- [32] Hoesli, M., & Reka, K. (2013). *Volatility Estimation for Direct Real Estate Investments*. *Journal of Real Estate Finance and Economics*, 47(2), 347–369. <https://doi.org/10.1007/s11146-011-9353-4>
- [33] Bank of Canada (2023). *Financial System Review: Stress Testing and Scenario Analysis*. Ottawa: Bank of Canada. <https://www.bankofcanada.ca/2023/05/financial-system-review-2023>
- [34] Brännlund, E., et al. (2023). *Predicting Changes in Canadian Housing Markets with Machine Learning*. Bank of Canada Staff Discussion Paper 2023-21. <https://www.bankofcanada.ca/wp-content/uploads/2023/09/sdp2023-21.pdf>
- [35] Selim, H. (2009). *Determinants of House Prices in Turkey: Hedonic Regression vs. Artificial Neural Network*. *Expert Systems with Applications*, 36(2), 2843–2852. <https://doi.org/10.1016/j.eswa.2008.01.044>
- [36] Balemi, N., Füss, R., & Weigand, A. (2021). *Clustering in Real Estate Markets: Detecting Housing Market Segments*. *Journal of Real Estate Finance and Economics*, 62(3), 422–449. <https://doi.org/10.1007/s11146-020-09755-2>
- [37] Glaeser, E. L., & Nathanson, C. G. (2017). *An Extrapolative Model of House Price Dynamics*. *Journal of Financial Economics*, 126(1), 147–170. <https://doi.org/10.1016/j.jfineco.2017.07.008>
- [38] Zhang, G. P. (2003). *Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model*. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- [39] Pace, R. K., Barry, R., & Sirmans, C. F. (1998). *Spatial Statistics and Real Estate*. *Journal of Real Estate Finance and Economics*, 17(1), 5–13. <https://doi.org/10.1023/A:1007703229506>
- [40] Law, S., Seresinhe, C. I., Yao, J., & Shen, Y. (2019). *Street-Frontage and Housing Prices: Integrating Deep Learning with Hedonic Modeling*. *Journal of the Royal Statistical Society: Series A*, 182(3), 1411–1437. <https://doi.org/10.1111/rssa.12436>
- [41] Chatzis, S. P., & Siakoulis, V. (2016). *Forecasting Real Estate Prices with Machine Learning and Time Series Models: An Empirical Comparison*. *Expert Systems with Applications*, 55, 492–502. <https://doi.org/10.1016/j.eswa.2016.02.020>
- [42] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts: Melbourne, Australia. <https://otexts.com/fpp2>
- [43] Kimball, R., & Caserta, J. (2011). *The Data Warehouse ETL Toolkit*. Wiley Publishing.
- [44] Inmon, W. H. (2005). *Building the Data Warehouse*. John Wiley & Sons.

- [45] Bank of England (2015). *Visualising Uncertainty in Macroeconomic Forecasts*. London: Bank of England.
<https://www.bankofengland.co.uk/quarterly-bulletin/2015/q2/visualising-uncertainty-in-macroeconomic-forecasts>
- [46] Hullman, J., Resnick, P., & Adar, E. (2015). *Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering*. *PLoS ONE*, 10(11), e0142444. <https://doi.org/10.1371/journal.pone.0142444>
- [47] Acolin, A., Bricker, J., Calem, P., & Wachter, S. (2019). *Borrowing Constraints and Homeownership*. *Journal of Housing Economics*, 44, 1–13.
<https://doi.org/10.1016/j.jhe.2019.03.001>
- [48] Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2017). *Designing the User Interface: Strategies for Effective Human–Computer Interaction* (6th ed.). Pearson.
- [49] Canada Mortgage and Housing Corporation (CMHC) (2023). *Rental Market Report: Kelowna*. Ottawa: CMHC.
<https://www.cmhc-schl.gc.ca/professionals/housing-markets-data-and-research/rental-market-reports-major-centres>
- [50] Canada Mortgage and Housing Corporation (CMHC) (2022). *Housing Market Information Portal – Kelowna*. Ottawa: CMHC. <https://www03.cmhc-schl.gc.ca/hmip-pimh/en>
- [51] Canada Mortgage and Housing Corporation (CMHC) (2023). *Housing Market Outlook: Vancouver*. Ottawa: CMHC.
<https://www.cmhc-schl.gc.ca/professionals/housing-markets-data-and-research/housing-market-outlook>
- [52] Bank of Canada (2021). *Impact of Monetary Policy on Regional Housing Markets*. Ottawa: Bank of Canada. <https://www.bankofcanada.ca/research>
- [53] Davidoff, T., & Wong, B. (2020). *Do Foreign Buyer Taxes Work? Evidence from Vancouver*. *Canadian Journal of Economics*, 53(4), 1675–1701. <https://doi.org/10.1111/caje.12474>
- [54] Statistics Canada (2023). *Housing Affordability Indicators, Toronto CMA*. Ottawa: Statistics Canada. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3410016201>
- [55] Allen, J., & Ka Kit, W. (2019). *Rental Market Signals and Housing Market Forecasting in Toronto*. *Urban Studies*, 56(12), 2547–2566. <https://doi.org/10.1177/0042098018798574>
- [56] Meen, G. (2018). *Evaluating Housing Market Models: Econometrics vs. Machine Learning*. *Housing Studies*, 33(8), 1231–1257. <https://doi.org/10.1080/02673037.2018.1424805>
- [57] Rafiei, M., & Adeli, H. (2016). *A Novel Machine Learning Model for Estimation of Housing Price Index*. *Journal of Construction Engineering and Management*, 142(2), 04015066.
[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001045](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001045)

- [58] Zhang, D., & Yi, Y. (2022). *Hybrid Econometric and Machine Learning Models for Real Estate Price Forecasting*. *Expert Systems with Applications*, 195, 116532. <https://doi.org/10.1016/j.eswa.2022.116532>
- [59] Canada Mortgage and Housing Corporation (CMHC) (2022). *Rental Market Dynamics and Housing Affordability*. Ottawa: CMHC. <https://www.cmhc-schl.gc.ca>
- [60] Bank of Canada (2023). *Scenario Analysis in Housing Finance*. Ottawa: Bank of Canada. <https://www.bankofcanada.ca>
- [61] Glaeser, E. L., & Gyourko, J. (2018). *The Economic Implications of Housing Affordability*. *Journal of Economic Perspectives*, 32(1), 3–30. <https://doi.org/10.1257/jep.32.1.3>
- [62] Spiegelhalter, D. (2017). *Risk and Uncertainty Communication in Policy and Practice*. *Annual Review of Statistics and Its Application*, 4, 31–60. <https://doi.org/10.1146/annurev-statistics-010814-020148>
- [63] Kok, N., & Jennen, M. (2012). *The Value of Buildings in a Changing Climate: Evidence from the Commercial Property Sector*. *Journal of Sustainable Real Estate*, 4(1), 1–28. <https://doi.org/10.1080/10835547.2012.12091828>
- [64] Statistics Canada (2022). *Demographic Projections for Canada, Provinces and Territories*. Ottawa: Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/91-520-x/91-520-x2022001-eng.htm>
- [65] Bank of Canada & Office of the Superintendent of Financial Institutions (OSFI) (2022). *Using Scenario Analysis to Assess Climate Transition Risk*. Ottawa: Bank of Canada. <https://www.bankofcanada.ca/2022/01/staff-discussion-paper-2022-1>
- [66] Bernstein, A., Gustafson, M., & Lewis, R. (2019). *Disaster on the Horizon: The Price Effect of Sea Level Rise*. *Journal of Financial Economics*, 134(2), 253–272. <https://doi.org/10.1016/j.jfineco.2019.03.013>
- [67] Rekabsaz, N., et al. (2021). *Housing Market Forecasting with News Sentiment Analysis: Evidence from Natural Language Processing*. *Decision Support Systems*, 144, 113516. <https://doi.org/10.1016/j.dss.2021.113516>
- [68] International Monetary Fund (IMF) (2020). *Real-Time Monitoring of Housing Markets: Lessons for Financial Stability*. Washington, DC: IMF. <https://www.imf.org/en/Publications/WP/Issues/2020/12/18/Real-Time-Monitoring-of-Housing-Markets-Lessons-for-Financial-Stability-49964>
- [69] Ng, W. L., & De Silva, D. (2023). *Hybrid Approaches to Housing Market Forecasting: Combining Econometrics and Machine Learning*. *Journal of Housing Economics*, 62, 101012. <https://doi.org/10.1016/j.jhe.2023.101012>

Additional References

Note on Additional References

The **Additional References** listed above serve a practical purpose distinct from the academic and policy sources cited in earlier sections. They document the **datasets** (CREA, CMHC, Rentals.ca, StatCan Census, NRCan, IBC) and **machine learning tools** (scikit-learn, LightGBM, Prophet) that underpin the technical implementation of the *Housing Insights & Risk Dashboard*.

Including these ensures that the essay acknowledges both the **scientific foundations** (Sections 1–10) and the **operational resources** required for reproducibility, data acquisition, and model deployment.

[70] Canadian Real Estate Association (CREA) (2023). *MLS® Home Price Index (HPI) Data*. Ottawa: CREA. <https://www.crea.ca/housing-market-stats/mls-home-price-index>

[71] RentFaster.ca (2023). *Rental Listings Database*. Calgary: RentFaster.ca. <https://www.rentfaster.ca>

[72] Rentals.ca (2023). *National Rent Report Data Portal*. Toronto: Rentals.ca. <https://rentals.ca/national-rent-report>

[73] Statistics Canada (2021). *Census of Population Program: Housing, Demographic, and Socioeconomic Data*. Ottawa: Statistics Canada. <https://www12.statcan.gc.ca/census-recensement/index-eng.cfm>

[74] Natural Resources Canada (2022). *Geospatial Data for Canada*. Ottawa: Government of Canada. <https://www.nrcan.gc.ca/maps-tools-publications>

[75] Insurance Bureau of Canada (IBC) (2020). *Climate Risk and Flood Mapping Data Portal*. Toronto: IBC. <https://www.ibc.ca/on/natural-disasters/climate>

[76] Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research*, 12, 2825–2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>

[77] Ke, G., et al. (2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. *Advances in Neural Information Processing Systems (NeurIPS)*. <https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>

[78] Taylor, S. J., & Letham, B. (2018). *Forecasting at Scale with Prophet*. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>

Appendices

Appendix A. Database Schema and ETL Process

The *Housing Insights & Risk Dashboard* uses a **PostgreSQL relational database** structured to support both **short-term predictions** (e.g., monthly rental prices) and **long-term forecasts** (e.g., housing affordability trends over 5–10 years). The schema is organized into four main groups:

1. Raw Data Tables

- **listings_raw** – contains raw scraped data (URLs, dates, city, postal code, property type, listing type, price, bedrooms, bathrooms, area, year built, description).
 - Indexed by **city + postal code** to support neighborhood-level joins.

2. Processed & Feature-Engineered Tables

- **listings_features** – cleaned and feature-engineered dataset, including price per square foot, property age, one-hot encoded property types, etc.
- **model_predictions** – stores pre-computed forecasts served to the API. Includes model name, horizon, city, forecast values (yhat, yhat_lower, yhat_upper), and metadata such as feature version and artifact URI.

3. Aggregated Time-Series Tables

- **house_price_index** – CREA HPI and aggregated ownership market metrics (median prices by type, active listings, average listing days).
- **rent_index** – CMHC and private rental aggregates (median rents by bedroom type, vacancy proxies).
- **demographics** – Statistics Canada demographic and income data (population, net migration, income, age cohorts).

4. Macro-Economic & Contextual Tables

- **macro_economic_data** – macro fundamentals (unemployment, GDP growth, prime lending rates, housing starts).

- **news_sentiment** – NLP-derived sentiment scores for housing-related headlines, capturing demand-side market mood.
- **construction_permits** – supply pipeline indicators (units approved, postal code, property type, approval date).

The ETL pipeline is central to the *Housing Insights & Risk Dashboard*. It not only ingests heterogeneous housing data but also ensures reproducibility, timeliness, and model readiness.

ETL Pipeline Workflow

1. Ingestion

- **CREA MLS® HPI** (ownership prices).
- **CMHC Rental Market Survey** and **Rentals.ca/RentFaster** (rents).
- **Statistics Canada & Bank of Canada** (macroeconomic indicators).
- **News RSS feeds** (CBC, Globe & Mail, local sources) for sentiment.
- Orchestrated using **Prefect / Airflow** or **GitHub Actions** for automation.

2. Transformation

- Cleaning raw rental listings, handling missing values, outlier detection.
- Feature engineering (price per sq.ft., affordability ratios, property age).
- Aggregation: monthly medians, city-level indices.
- NLP preprocessing on news headlines → embeddings → sentiment/topic classification.

3. Loading

- **PostgreSQL (+PostGIS)** for structured data: metrics, features, forecasts, risks, sentiment.
- **S3/MinIO object store** for unstructured data: raw HTML snapshots, PDFs, ML artifacts.

4. Model Integration

- Processed data flows into **ML pipelines** (Prophet, LightGBM, crisis_RF).

- Predictions written to `model_predictions`.
- Results served to downstream APIs.

This pipeline ensures a continuous flow from **raw inputs** → **clean features** → **aggregated datasets** → **model-ready tables** → **API outputs**.

Design Rationale

This schema ensures:

- **Transparency** (raw data stored for reproducibility).
- **Scalability** (supports multiple cities, long-term growth).
- **Auditability** (linking back to raw sources via URLs and snapshot storage).
- **Model compatibility** (tables structured for both econometrics and ML).

Appendix B. Mathematical Derivations

Key mathematical foundations underlying the dashboard:

1. Hedonic Pricing Model

$$P_i = \beta_0 + \sum_{k=1}^K \beta_k X_{ik} + \epsilon_i$$

Where P_i is the price of property i , X_{ik} are attributes (e.g., size, location), and β_k are estimated coefficients.

2. Error-Correction Model (ECM)

$$\Delta P_t = \alpha(P_{t-1} - \beta Y_{t-1}) + \gamma \Delta Y_t + \epsilon_t$$

Where P_t is housing price, Y_t represents fundamentals (income, rates), and α measures speed of adjustment to equilibrium.

3. ARIMA Forecasting

$$\phi(L)(1 - L)^d P_t = \theta(L)\epsilon_t$$

Where L is the lag operator, d differencing order, $\phi(L)$ autoregressive polynomial, and $\theta(L)$ moving average polynomial.

4. Risk Metrics

- **Affordability Index:**

$$AI = \frac{\text{Median Price or Rent}}{\text{Median Household Income}}.$$

- **Volatility (GARCH(1,1)):**

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

Where σ_t^2 is conditional variance.

These derivations provide theoretical grounding for the dashboard's metrics.

Appendix C. Model Hyperparameters and Configurations

To ensure reproducibility, the following ML model configurations were used:

- **ARIMA:** Orders (p,d,q) tuned by AIC/BIC, typical ranges p=0–3, d=1, q=0–2.
- **LightGBM:**
 - learning_rate = 0.05
 - num_leaves = 31
 - max_depth = -1 (unlimited)
 - feature_fraction = 0.8
 - bagging_fraction = 0.8
- **XGBoost:**
 - eta = 0.05
 - max_depth = 6
 - subsample = 0.7
 - colsample_bytree = 0.8
- **Neural Networks (MLP):**
 - Hidden layers: [128, 64]
 - Activation: ReLU
 - Optimizer: Adam, learning_rate=0.001
 - Epochs: 100, Batch size: 32
- **Validation:** 5-fold cross-validation with time-series split.
- **Evaluation Metrics:** RMSE, MAPE, R^2 across all models.

D. Project Resources

The following resources provide access to the prototype dashboard, source code, and API documentation developed as part of this project:

- Application (Dashboard): <https://hird.netlify.app/>
- Source Code Repository:
<https://github.com/yuri-spizhovi-mit/housing-insights-risk-dashboard>
- API Documentation: <https://housing-insights-risk-dashboard.vercel.app/docs#>

This configuration ensures balance between accuracy and generalization, while maintaining comparability across modeling approaches.

Notes on Methodology and Tools

The *Housing Insights & Risk Dashboard* combines the rigor of academic research with the pragmatism of modern data engineering. The methodological foundation rests on **econometric models** (hedonic regression, ARIMA, error-correction frameworks) that provide interpretability and policy relevance, complemented by **machine learning techniques** (LightGBM, XGBoost, neural networks) that capture nonlinearities and complex feature interactions.

From a technical perspective, the dashboard is built on a robust **ETL and database architecture**, ensuring that raw housing and economic data are collected, cleaned, and stored transparently for reproducibility. Forecasting models are implemented using open-source libraries such as **scikit-learn, LightGBM, and Prophet**, enabling scalable and reproducible pipelines. Visualization and user interaction rely on **React, TypeScript, and Recharts/D3.js**, translating quantitative results into actionable insights.

Together, these methodological and technical choices reflect a deliberate balance between **scientific rigor, practical implementation, and accessibility**. The dashboard is not only a research tool but also a decision-support system, bridging the gap between data science, economics, and housing policy. Its design ensures adaptability, allowing for the integration of new datasets (e.g., climate risk, geospatial indicators) and evolving modeling techniques.

In this way, the project demonstrates how **data science and applied research can be combined to address one of Canada's most pressing socio-economic challenges: housing affordability and market risk**.

Methodological Foundation

- **Econometric models:** hedonic regression, ARIMA, and error-correction frameworks, which provide interpretability and policy relevance.
- **Machine learning methods:** LightGBM, XGBoost, and neural networks, which capture nonlinearities and complex interactions.
- **Hybrid approaches:** econometric baselines enhanced with ML residual modeling, producing more accurate forecasts without losing explanatory power.

Technical Architecture

The system is organized in modular layers, aligned with the project's repository structure:

1. **Infrastructure Layer (*infra/*)**
 - **Database:** PostgreSQL with PostGIS for structured data (prices, rents, demographics, macro indicators).

- **Object Storage:** MinIO/S3 for raw snapshots, PDFs, and model artifacts.
- **Migrations/Backups:** SQL initialization scripts, migration history, and backup files.

2. Data and ETL Layer (**ml/src/etl/**)

- Adapters for CREA, CMHC, Rentals.ca, RentFaster, Castanet, Statistics Canada, and Bank of Canada.
- ETL orchestrated with Prefect/Airflow or GitHub Actions.
- Raw outputs stored in **_raw** tables and snapshots; transformed outputs aggregated into monthly indices and feature tables.

3. Feature Engineering (**ml/src/features/**)

- Generates model-ready variables: price per square foot, property age, rental aggregates by bedroom type, affordability ratios.
- Encodes categorical data and handles outlier detection.

4. Modeling Layer (**ml/src/models/**)

- **Forecasting:** Prophet, LightGBM, ARIMA variants.
- **Risk:** affordability stress indices, volatility modeling (e.g., GARCH), scenario analysis.
- **Anomaly detection:** (planned) methods to identify sudden spikes in prices or rents.
- Predictions written to **model_predictions** with metadata for traceability.

5. NLP & Sentiment (**ml/src/nlp/**)

- News ingestion pipeline with DistilBERT for sentiment and topic classification.
- Outputs monthly sentiment indices per city, stored alongside quantitative indicators.

6. Reporting (**ml/src/reporting/**)

- Templates for PDF generation with Jinja2 → HTML → PDF conversion.

- Deployed via FastAPI microservice; artifacts stored in S3/MinIO.

7. Pipelines (**ml/pipelines/**)

- Scripts for daily ingestion, smoke tests (e.g., CREA to DB), and retraining schedules.
- Orchestrates the full data flow: extract → transform → model → store → serve.

8. API Layer (**services/api/**)

- Built with **Spring Boot (Java)**, exposing public endpoints:
 - `/v1/cities`, `/v1/forecast/{city}`, `/v1/risks/{city}`,
`/v1/sentiment/{city}`, `/v1/report/{city}.pdf`.
- Provides authentication (API key/JWT), request validation, Swagger/OpenAPI docs.
- Acts as a secure gateway, routing to model-serving and report services.

9. Model Serving (**FastAPI, Python**)

- Forecast, risk, and anomaly services containerized for scalability.
- Integrated with MLflow (planned) for model registry and lifecycle management.

10. Frontend Layer (**services/ui/**)

- React + TypeScript (Vite build system).
- Features:
 - City and scenario selection.
 - Forecast bands, risk gauges, sentiment timelines.
 - Downloadable PDF reports.
- Libraries: Recharts/D3.js, Leaflet/Mapbox for geospatial visualization.

Model Store and Lifecycle (Planned)

- **Artifacts:** Models, feature sets, and metadata stored in S3/MinIO with semantic versioning.
- **Registry:** MLflow to manage training runs, metrics, and deployment status.
- **Automation:** Retraining pipelines with time-series CV; promotion of models to “production” if evaluation metrics improve.
- **Monitoring:** Data drift, concept drift, and backtesting errors tracked; planned integration with Great Expectations for data validation.
- **Explainability:** SHAP values for feature importance, to ensure forecasts remain transparent for policymakers.

Current vs Planned

- **Implemented:** ETL adapters, Prophet + LightGBM forecasts, affordability indices, sentiment pipeline, REST APIs, React dashboard with visualizations and PDF export.
- **Planned:** Automated retraining, MLflow registry, anomaly service, geospatial enrichment, explainable AI outputs, scenario engine for macro shocks.