**Housing Crisis in Canada - Fact or Fiction?**

“Unveiling Insights, Empowering Decisions - A Capstone Project Journey”

**Final Report**

**Capstone Project: Group 5**

Kaveriben Patel

Deep Sheta

Tirthak Bhingaradiya

Jatin Kansra

A green and white logo

Description automatically generated

**Mentor**

**Dr. Manjari Maheshwari**

Data Analytics for Business

St. Clair College, Windsor, Ontario, Canada

Table of Contents

[About Our Capstone Project: 3](#_Toc164969204)

[First, a Few Words About Who We Are: 3](#_Toc164969205)

[About the Housing Crisis: 3](#_Toc164969206)

[The Dataset: 3](#_Toc164969207)

[Project Outline: 4](#_Toc164969208)

[Exploratory Data Analysis: 4](#_Toc164969209)

[Model Selection: 8](#_Toc164969210)

[Model Building: 9](#_Toc164969211)

[1. LSTM: 9](#_Toc164969212)

[2. Autoregressive Integrated Moving Average (ARIMA): 9](#_Toc164969213)

[Model Implementation: 10](#_Toc164969214)

[LTSM: 10](#_Toc164969215)

[Observations: 11](#_Toc164969216)

[Predictions: 11](#_Toc164969217)

[Key findings: 14](#_Toc164969218)

[Recommendations: 16](#_Toc164969219)

[Conclusions and Possible Improvements in Project: 17](#_Toc164969220)

[Possible Improvements in Project: 17](#_Toc164969221)

# About Our Capstone Project:

In the backdrop of pervasive debates surrounding the housing crisis in Canada, our capstone project, titled "Housing Crisis - Fact or Fiction?", endeavors to dissect the realities behind the prevailing narratives. As part of our journey, we aim to leverage data-driven insights and advanced analytical techniques to unravel the complexities of this pressing issue, ultimately empowering stakeholders with actionable knowledge.

# First, a Few Words About Who We Are:

We are students at well-known institute St. Clair College, Windsor. We obtain Data Analytics for a business course. This is our capstone project, and we are a group of four people. Group #5 comprises dedicated individuals committed to the pursuit of truth and clarity.

Our team members are:

1. Kaveriben Patel
2. Tirthak Vipulbhai Bhingaradiya
3. Jatin Kansra
4. Deep Manshukhbhai Sheta

# About the Housing Crisis:

The housing crisis in Canada has emerged as a topic of widespread concern, driven by factors such as population growth, urbanization, and immigration. While media narratives often portray a dire situation characterized by soaring real estate prices and dwindling affordability, our project seeks to scrutinize these assertions, distinguishing between perception and reality.

## The Dataset:

Our analysis relies on a comprehensive dataset sourced from government repositories which is statistics of Canada. This dataset encompasses a wide range of variables, including population demographics, mortgage availability, interest rates, and housing market indicators. Here's an overview:

1. **Population Dataset:** This dataset offers insights into population trends over time, helping to understand the growth or decline of local populations. It includes REF\_DATE (time), GEO (geographical location), and population values.
2. **Constructions Dataset:** This dataset focuses on construction permits, housing starts, and building completions, offering insights into the supply side of the housing market. Key columns include REF\_DATE (time), GEO (geographical location), building type, construction type, and construction values.
3. **House & Land-selling Dataset:** This dataset contains information on housing and land sales, including property types, prices, locations, and transaction dates. Key columns include REF\_DATE (time), GEO (geographical location), new housing price indexes, and property values.

## Project Outline:

1. Exploratory Data Analysis
2. Model Selection
3. Model Building
4. LTSM
5. ARIMA
6. ARIMA with Grid Search
7. Model Comparisons
8. Predictions
9. Recommendations
10. Conclusions and Possible Improvements in Project

# Exploratory Data Analysis:

This phase involves delving into the dataset to identify trends, correlations, and outliers, providing valuable insights into the underlying dynamics of the housing market. Before diving into the analysis, it's essential to conduct exploratory data analysis to understand the dataset’s structure, identify patterns, and uncover potential insights. Here are some steps involved in EDA:

1. **Data Summary**: Begin by summarizing each dataset, including basic statistics such as mean, median, standard deviation, and quartiles for numerical features. For categorical features, provide frequency distributions.
2. **Data Visualization:** Visualize the distributions of key variables using histograms, box plots, and density plots. Scatter plots can help explore relationships between variables, while time series plots can reveal trends over time.

A graph with green and blue lines

Description automatically generated

A graph with a line going up

Description automatically generated

1. **Correlation Analysis:** Examine correlations between variables to identify potential relationships. Heatmaps and correlation matrices are useful for visualizing correlations.

A red and blue squares

Description automatically generated

1. **Temporal Analysis:** Since the datasets include temporal information, analyze trends over time. Group the data by quarters as specified and plot time series to observe seasonal patterns or long-term trends.
2. **Geographical Analysis:** Explore variations in housing metrics across different geographical locations (provinces). Maps and choropleth plots can visually represent regional differences.

A graph of a number of people

Description automatically generated with medium confidence

1. **Outlier Detection:** Identify outliers that may skew the analysis and determine whether they represent genuine anomalies or errors in the data.

A chart of different colored squares

Description automatically generated

1. **Missing Data Handling:** Assess the extent of missing data and decide on appropriate strategies for imputation or removal. Consider the impact of missing values on the analysis.

By conducting thorough exploratory data analysis, we can gain valuable insights into the housing market dynamics, population trends, mortgage affordability, interest rate impacts, and construction activities. These insights will inform subsequent analyses aimed at understanding the presence and severity of a housing crisis.

# Model Selection:

We evaluate various modeling techniques to determine the most suitable approach for forecasting and analysis, considering factors such as accuracy and interpretability. In our quest to accurately forecast and analyze housing trends while maintaining interpretability, we embark on evaluating various modeling techniques. Our goal is to select the most suitable approach that strikes a balance between predictive accuracy and ease of interpretation. Here's how we undertake this evaluation process:

1. **Model Selection:** We consider a range of modeling techniques that are commonly used in forecasting and analysis, including but not limited to:
2. **Time Series Models:** Such as LTSM, ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal ARIMA), which are well-suited for analyzing temporal data with seasonality and trends.
3. **Machine Learning Algorithms:** Including regression-based models like Linear Regression, tree-based models like Random Forest, and ensemble methods like Gradient Boosting, which can capture complex relationships in the data.
4. **Evaluation Metrics:** We define appropriate evaluation metrics to assess the performance of each model. These metrics may include:
5. Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) to measure predictive accuracy.
6. R-squared (R^2) or adjusted R-squared to quantify the proportion of variance explained by the model.
7. AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) for model selection and comparison, considering both goodness of fit and model complexity.
8. **Cross-Validation:** We employ cross-validation techniques such as k-fold cross-validation or time series cross-validation to ensure robustness and mitigate overfitting. This involves splitting the dataset into training and testing sets multiple times and evaluating the model's performance across different splits.
9. **Interpretability:** Alongside predictive accuracy, we prioritize the interpretability of the chosen model. Models that provide clear insights into the relationships between predictors and outcomes are preferred, especially in domains like real estate where stakeholders may require transparent explanations of forecasts.
10. **Ensemble Approaches:** We explore ensemble techniques that combine predictions from multiple models to leverage their respective strengths. Ensemble methods, such as model averaging or stacking, can often yield better performance than individual models while maintaining interpretability.
11. Model Complexity: We consider the trade-off between model complexity and performance. While more complex models may capture intricate patterns in the data, they also risk overfitting and may be harder to interpret. We aim to strike a balance by selecting models that are sufficiently complex to capture relevant information without sacrificing interpretability.
12. Sensitivity Analysis: We conduct sensitivity analyses to assess the robustness of our modeling approach to different assumptions or variations in the dataset. Sensitivity analysis helps us understand the stability of our forecasts and identify potential sources of uncertainty.

By meticulously evaluating various modeling techniques based on these criteria, we aim to identify the most suitable approach for forecasting and analyzing housing trends accurately and interpretable. This ensures that our insights are not only reliable but also actionable for stakeholders involved in addressing housing challenges.

# Model Building:

Implementing selected models, including linear regression and ARIMA, to develop predictive frameworks that capture the nuances of the housing crisis.

After conducting exploratory data analysis and understanding the characteristics of the datasets, the next step is to build predictive frameworks to capture the nuances of the housing crisis. We will evaluate various modeling techniques to determine the most suitable approach for forecasting and analysis, considering factors such as accuracy and interpretability.

1. LSTM: Sequential models are useful for predicting numerical outcomes, making them suitable for forecasting housing prices, population growth, mortgage values, and other related metrics. We will assess the performance of linear regression models by training them on historical data and evaluating their predictive accuracy using appropriate metrics such as mean squared error (MSE) or R-squared.
2. Autoregressive Integrated Moving Average (ARIMA): ARIMA models are well-suited for time series forecasting, making them valuable for analyzing trends and patterns in the housing market over time. By capturing the temporal dependencies present in the data, ARIMA models can provide insights into future housing market dynamics. We will evaluate ARIMA models by fitting them to the time series data and assessing their ability to capture seasonal variations, trends, and fluctuations.

# Model Implementation:

Once we have selected the most appropriate modeling techniques, we will proceed with implementing the chosen models. This will involve the following steps:

1. **Data Preparation:** Prepare the datasets for model training, validation, and testing. This may include feature engineering, normalization, and splitting the data into training and testing sets.
2. **Model Training:** Train the selected models using historical data. For linear regression, this involves fitting a linear equation to the training data, while for ARIMA, it involves estimating the model parameters based on the time series data.
3. **Model Evaluation:** Evaluate the performance of the trained models using appropriate evaluation metrics. This helps assess how well the models generalize to unseen data and provides insights into their predictive accuracy.
4. **Model Tuning:** Fine-tuning the model parameters and hyperparameters to optimize performance further. This may involve conducting grid search or cross-validation to find the optimal settings for the models.
5. **Prediction and Analysis:** Use the trained models to make predictions on future housing market trends and analyze the results. This involves interpreting the model outputs and identifying key insights that can inform decision-making regarding the housing crisis.

By implementing selected models, including linear regression and ARIMA, we aim to develop predictive frameworks that accurately capture the dynamics of the housing market. These frameworks will enable us to forecast key metrics such as housing prices, population growth, mortgage values, and interest rates, facilitating a comprehensive analysis of the housing crisis and its contributing factors.

## LTSM:

Forecasting future population trends is essential for anticipating housing demand and devising effective policy interventions. In this report, we delve into the application of LSTM, a sophisticated deep learning technique, to forecast population trends in Canada.

Methodology:

The methodology adopted in this study involves the following steps:

1. **Data Collection:** Historical population data spanning several decades is collected from reliable sources such as Statistics Canada.
2. **Data Preprocessing:** The collected data is preprocessed to handle missing values, outliers, and ensure consistency in format.
3. **LSTM Model Development:** An LSTM model is constructed using TensorFlow or a similar deep learning framework. The model architecture is tailored to accommodate the temporal dependencies inherent in population data.
4. **Training and Validation:** The LSTM model is trained on historical population data and validated using appropriate evaluation metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).
5. **Forecasting:** The trained LSTM model is utilized to forecast future population trends for Canada.

### Observations:

The results obtained from the LSTM-based population forecasting model reveal valuable insights into the potential trajectory of population growth in Canada. By analyzing historical data and extrapolating future trends, the model generates forecasts that can aid policymakers, urban planners, and stakeholders in making informed decisions regarding housing development, infrastructure investment, and resource allocation.

The application of LSTM for population forecasting offers several advantages, including its ability to capture nonlinear dependencies and long-term temporal patterns. However, it is crucial to acknowledge the limitations and uncertainties inherent in any forecasting exercise. Factors such as economic fluctuations, immigration policies, and demographic shifts may influence population dynamics in ways that are difficult to predict accurately.

### Predictions:

1. **Construction Activity:** Construction activity is anticipated to respond to the evolving demand for housing infrastructure. In regions experiencing population growth and rising housing prices, construction activity is likely to increase to meet the demand for new housing units. Conversely, areas with stagnant population growth or oversupply conditions may see a slowdown in construction activity. The model incorporates variables such as population forecasts, housing market indicators, and regulatory frameworks to project construction trends.
2. **Crisis Assessment:** By integrating predictions for population growth, housing prices, and construction activity, stakeholders can assess the likelihood of a housing crisis in Canada. Key indicators such as rapid population growth coupled with stagnant construction activity and soaring housing prices in certain regions may signal the emergence of a crisis. Conversely, balanced population growth, stable housing prices, and proactive construction activity suggest a more resilient housing market. The model enables policymakers and stakeholders to monitor these indicators and implement timely interventions to mitigate the risk of a housing crisis.
3. **Population Growth:** The LSTM-based population forecasting model predicts a steady increase in the population of Canada over the next decade. Factors such as immigration, birth rates, and economic conditions are likely to contribute to this growth trend. By analyzing historical data and projecting future trends, the model estimates the magnitude of population growth in different regions of the country.

A graph with a line going up

Description automatically generated

1. **Housing Prices:** The forecast for housing prices is expected to exhibit variation across different regions of Canada. While urban centers may experience continued upward pressure on prices due to population growth and limited housing supply, rural areas may witness more moderate price dynamics. The model considers factors such as demand-supply dynamics, economic indicators, and government policies to generate forecasts for housing prices.

A graph of a graph showing a graph of a graph

Description automatically generated with medium confidence

Overall, the integrated predictions provided by the LSTM-based forecasting model offer valuable insights into the interplay between population dynamics, housing market trends, and the potential for a housing crisis in Canada. By leveraging these predictions, stakeholders can make informed decisions and take proactive measures to ensure the long-term sustainability and affordability of housing for Canadians.

### Key findings:

A diagram of different colored circles

Description automatically generated

**1**A table with numbers and text

Description automatically generated

1. **Regional Disparity in the Housing Crisis:**

The housing crisis in Canada presents itself with stark regional disparities, exhibiting varying degrees of impact across provinces and within cities. This multifaceted issue requires an in-depth examination to understand its nuances and implications for different regions.

1. **British Columbia:**

In British Columbia, the housing crisis is characterized by a modest 3.5% crisis rate, which is accompanied by a declining trend. Despite this seemingly favorable statistic, the province faces significant challenges related to housing affordability. High house and land prices persist, exacerbating the affordability crisis for residents. While the crisis rate may be lower compared to other regions, the affordability challenges remain a pressing concern for policymakers and residents alike.

1. **Manitoba:**

Manitoba emerges as one of the provinces grappling with the most acute housing crisis, although it often receives less media attention compared to Ontario. The severity of the crisis in Manitoba underscores the need for focused interventions to address housing affordability and accessibility issues. Despite the relative lack of media coverage, the challenges faced by Manitobans in securing affordable housing demand urgent attention and proactive measures from policymakers.

1. **Ontario:**

Contrary to widespread media and public perception, Ontario experiences a surprisingly high crisis rate of around 20%. This disparity between perception and reality highlights the complexity of the housing crisis and the importance of data-driven analysis in understanding its true extent. The prevalence of the crisis in Ontario underscores the need for comprehensive policy measures to tackle housing affordability issues and ensure access to affordable housing for all residents.

1. **Price Dynamics:**

Across most provinces, the housing market exhibits a dynamic interplay between house prices and land prices. While house prices surge, land prices generally remain relatively stable, indicating the intricate dynamics at play within the housing market. This divergence in price dynamics underscores the need for a holistic approach to addressing housing affordability, which considers both the cost of housing and the underlying land values.

In conclusion, the housing crisis in Canada manifests differently across provinces and cities, with varying degrees of impact and underlying dynamics. Addressing this multifaceted issue requires tailored interventions that consider regional disparities, price dynamics, and the complex interplay between housing affordability and land values.

# Recommendations:

Based on our analysis, providing actionable recommendations to address the housing crisis and mitigate its impact on Canadians.

1. **Investment in Affordable Housing Initiatives:** Given the projected population growth and potential strain on housing affordability, governments at various levels should prioritize investment in affordable housing initiatives. This includes subsidizing affordable housing developments, incentivizing the construction of rental units, and implementing policies to protect vulnerable populations from displacement due to rising housing costs.
2. **Promotion of Sustainable Urban Development:** Encouraging sustainable urban development practices can help alleviate pressure on housing markets in densely populated areas. This involves promoting mixed-use developments, improving public transportation infrastructure, and implementing zoning policies that support density and infill development.
3. **Streamlining Regulatory Processes:** Simplifying regulatory processes and reducing bureaucratic barriers to new construction can expedite housing development and increase housing supply. Governments should explore measures such as expedited permitting, zoning reforms, and tax incentives to encourage developers to build more housing units.
4. **Incentives for Affordable Housing Construction:** Providing financial incentives and regulatory relief to developers who commit to building affordable housing units can stimulate the production of much-needed housing stock for low- and moderate-income households.
5. **Monitoring and Evaluation:** Continuous monitoring and evaluation of housing market dynamics, population trends, and policy interventions are essential for informed decision-making. Governments and relevant stakeholders should establish robust data collection mechanisms and regularly assess the effectiveness of housing policies in addressing affordability challenges.
6. **Addressing Demand-Side Factors:** In addition to supply-side interventions, addressing demand-side factors contributing to housing affordability challenges is crucial. This may involve measures such as enhancing access to homeownership assistance programs, promoting sustainable mortgage lending practices, and combating speculative real estate investment.
7. **Long-Term Planning:** Planning for housing affordability and sustainability should be approached with a long-term perspective. Governments and stakeholders should develop comprehensive housing strategies that anticipate future population growth, demographic shifts, and economic trends, ensuring that housing policies remain adaptive and responsive to evolving needs.

By implementing these recommendations in a coordinated manner, policymakers and stakeholders can work towards ensuring access to affordable, safe, and adequate housing for all Canadians, thereby fostering inclusive and resilient communities.

# Conclusions and Possible Improvements in Project:

Reflecting on our findings, identifying areas for further exploration, and proposing enhancements to our methodology for future research endeavors. Through our collective efforts, we endeavor to shed light on the housing crisis, demystifying misconceptions and paving the way for informed decision-making and policy interventions.

1. **Insights into Population Dynamics:** The utilization of LSTM for population forecasting has provided valuable insights into the future trajectory of population growth in Canada. By analyzing historical data and projecting future trends, the model has facilitated a deeper understanding of the demographic factors driving housing demand in different regions of the country.
2. **Potential Housing Market Trends:** The integrated predictions for housing prices, construction activity, and the likelihood of a housing crisis offer stakeholders a comprehensive view of the housing market landscape. These predictions can inform policy decisions, investment strategies, and proactive measures to address housing affordability challenges.
3. **Foundation for Informed Decision-Making:** The project lays a foundation for informed decision-making by policymakers, urban planners, and other stakeholders involved in housing policy and development. By incorporating advanced forecasting techniques, the project equips decision-makers with the tools and insights necessary to anticipate future housing market dynamics and formulate effective interventions.

### Possible Improvements in Project:

1. **Enhanced Data Quality and Granularity:** Improving the quality and granularity of data used for population forecasting could enhance the accuracy of predictions. Access to more detailed demographic data, including information on migration patterns, age cohorts, and household composition, could provide a more nuanced understanding of population dynamics and their implications for housing demand.
2. **Integration of External Factors:** Incorporating external factors such as economic indicators, interest rates, and policy changes into the forecasting model could enhance its predictive capabilities. By considering a broader range of variables that influence housing market dynamics, the model could generate more robust and actionable predictions for policymakers and stakeholders.
3. **Validation and Sensitivity Analysis:** Conducting rigorous validation and sensitivity analysis of the forecasting model can help assess its reliability and robustness. By testing the model's performance under different scenarios and assumptions, researchers can identify potential sources of uncertainty and refine the forecasting methodology accordingly.
4. **Long-Term Monitoring and Evaluation:** Establishing mechanisms for long-term monitoring and evaluation of housing market trends and policy interventions is essential for assessing the impact of the project over time. By continuously monitoring key indicators and evaluating the effectiveness of policy measures, stakeholders can iteratively refine their strategies and adapt to changing market conditions.

By addressing these potential areas for improvement, the project can enhance its utility as a tool for informing evidence-based decision-making and addressing the complex challenges of housing affordability and accessibility in Canada.