Analyzing the Impact of Temporary Residents on Canada's Housing Market

CSCI 6612 - Visual Analytics

Group TRHA (Temporary Resident Housing Analysis)

Ovaiz Ali
ov778752@dal.ca

Bhavya Mukesh Dave
bh392017@dal.ca

December 2024

1 Abstract

This project focuses on exploring how temporary residents, such as international students and foreign workers, impact the Canadian housing market. By analyzing data on housing prices, study permits, worker numbers, and economic factors like CPI and interest rates, the project examines trends in housing affordability and availability across provinces. Advanced visualizations, including time-series and scatter plots, help reveal these trends. A Large Language Model (LLM) was used to generate insights and summaries, assisting decision-makers such as policymakers, urban planners, and real estate investors. The goal was to provide a clearer understanding of the link between immigration and housing market changes in Canada.

2 Project Explanation: Problem, Importance, and Solution

2.1 Problem Overview

The Canadian housing market is under significant pressure due to rapid population growth, economic changes, and an increasing number of temporary residents. International students and foreign workers have contributed substantially to the housing demand, particularly in urban centers such as Toronto and Vancouver [1] - [3]. Traditional analyses often overlook this component, making the understanding of the dynamics of the housing market incomplete.

Our project addresses the following key questions:

- How do factors such as province, study permits issued, total workers, and common occupational titles influence housing prices in Canadian regions?
- Can past values of the New Housing Price Index and related economic indicators predict housing prices for future periods in Canadian regions?
- Which Canadian regions exhibit similar trends in housing prices and economic metrics (e.g., housing price index, worker counts, consumer price index), suggesting common economic characteristics?
- How do Consumer Price Index metrics and Canadian bank interest rates correlate with housing price variations across different provinces?
- Are certain occupations (e.g., legislators and other common occupational titles) linked with higher or lower housing price indices in various regions?

Addressing these questions would help policymakers, urban planners, and investors make informed decisions about housing affordability and market dynamics.

2.1.1 Importance of the Project

The issue of housing affordability is a growing challenge in Canadian cities, worsened by the increase in temporary residents. Our analysis explores how these residents, along with economic factors like CPI and interest rates, influence housing prices. This adds value by helping in the following areas:

- **Inform Policy**: Provide government policymakers with insights to help balance immigration and housing policies.
- **Guide Urban Planning**: Identify regions facing significant housing pressures to focus intervention efforts.
- **Assist Investment Decisions**: Offer data-driven insights into housing market trends for real estate stakeholders.

2.1.2 Solution Approach

The project integrates machine learning models, advanced interactive visualizations, and a fine-tuned Large Language Model (LLM) to deliver actionable insights. These components allow stakeholders to explore data relationships, identify trends, and receive detailed analysis. The following is an explanation of each aspect of our implemented solution.

Data Sources and Pre-processing: The unified dataset spans 2015–2024 and combines several data sources to capture housing market trends and their relationship with temporary residents and economic factors. The following is an explanation of each data source:

- **New Housing Price Index (1981-2024)**: Includes 62,880 records which shows monthly price indices for houses and land across Canadian regions. Here, the key attributes include geographical regions, housing categories (total, house-only, and land-only), and monthly index values. This dataset provides the backbone for understanding long-term housing price trends [7].

Table 1: Housing Price Index Data Sample [7]

REF_DATE	GEO	Housing Category	Value
Jan-1981	Canada	Total(house and land)	38.2
Jan-1981	Canada	House only	36.1

- **Study Permit Holders Dataset (2015-2024)**: Contains 230 records of quarterly and annual study permit issuance, segmented by country of citizenship. Here, the attributes include total annual permit holders and quarterly breakdowns, enabling the analysis of international student influx and its impact on housing demand [5].

Table 2: Study Permit Holders Data Sample [5]

Countryof Citizenship	2015	Q1	Q2	Q3	Q4	2016
Afghanistan	15	20	35	25	25	30
Albania	15	25	45	30	25	35

- **TFWP & IMP Work Permit Holders Dataset (2015-2024)**: This dataset comprises 4,810 records with 166 columns covering work permits issued under the Temporary Foreign Worker Program (TFWP) and the International Mobility Program (IMP). Here, the attributes include province/territory of destination, National Occupational Classification (NOC) codes, and quarterly permit data, offering insights into workforce dynamics across regions [6].

Table 3: Work Permit Holders Data Sample [6]

Province / Occupation	2015	Q1	Q2	Q3	Q4	2016
Senior Managers (NOC 0012)	0	0	0	0	5	5
Financial Managers(NOC 0013)	5	10	20	10	10	20

- **Consumer Price Index (CPI) Data (1995-2024)**: This dataset includes the comprehensive CPI data reflecting inflationary trends. Here, the key metrics include Total CPI, CPI Trim, CPI Median, and static CPI changes. The dataset includes 358 records and 6 columns, spanning from 1995 to 2024 [3].

Table 4: CPI Data Sample [3]

Date	Total CPI	Total CPI(Seasonally Adjusted)	Static CPIChange	CPI Trim	CPI Median
1995-01-01	86.6	86.6	0.6	1.8	1.7
1995-02-01	87	87	1.9	1.8	1.8
1995-03-01	87.2	87.2	2.1	1.9	1.8

- **Canadian Interest Rates and Monetary Policy Variables (2015-2023)**: Provides data on bank interest rates, variable mortgage rates, and effective interest rates for households and businesses. The dataset contains 470 records and 5 columns, ranging from 2015 to 2024 [9].

Table 5: Interest Rates Data Sample [9]

Date	**Bank Rate (Weekly)**	**Estimated Variable Mortgage Rate**	**Weekly Effective Household Interest Rate**	**Weekly Effective Business Interest Rate**
1/3/2024	5.25	6.26	6.69	6.41
12/27/2023	5.25	6.26	6.73	6.44
12/20/2023	5.25	6.25	6.80	6.47

Final Combined Dataset: After pre-processing and integrating all the datasets, which involved handling missing values (dropping null values) and ensuring data consistency, we finalized a single dataset that includes house prices, study permits, worker information, and economic indicators for analysis. This pre-processing step ensured the data was clean, complete, and ready for use in machine learning models and visualizations. Below is a part-by-part snapshot of the combined dataset, along with an explanation of the columns we merged and feature-engineered.

Table 6: Final Combined Dataset Snapshot (Part 1)

Year	**Province**	**Location**	**New Housing Price Index**	**Avg. Housing Value**	**Number of Study Permits Issued**	**Study Permit to Value Impact Metric**
2016	Ontario	Windsor, Ontario	Land only	97.42	103720	1064.67
2017	Manitoba	Manitoba	Land only	100	108870	1088.7
2020	Nova Scotia	Nova Scotia	House only	106.92	105815	989.67

Description of Columns (Part 1):

- **Year (int)**: The year the data was recorded, allowing tracking of housing and study permit trends over time.
- **Province (string)**: The Canadian province where the data was collected, showing geographical variation in housing and immigration data (e.g., Ontario, Nova Scotia).

- **Location (string)**: Specifies the type of housing data (land or house) for a more detailed view of housing prices in different regions within the province.
- **New Housing Price Index (float)**: An index reflecting the change in the price of new housing in the given year. A higher value indicates an increase in housing prices relative to a base year.
- **Avg. Housing Value (float)**: The average value of housing in the province or location for the respective year, offering insight into overall housing affordability.
- **Number of Study Permits Issued (int)**: The total number of study permits granted to international students in the province during the year, providing insight into immigration trends.
- **Study Permit to Value Impact Metric (float)**: A ratio showing the potential impact of international students on housing demand, calculated by dividing the number of study permits by the average housing value. A higher value suggests a stronger influence on housing demand.

Table 7: Final Combined Dataset Snapshot (Part 2)

Total Workers Count	Number of Class Titles	Most Common Class Title	Workers to Housing Impact Ratio	Combined Impact Ratio	Avg. Total CPI	Avg. Total CPI Seasonally Adjusted	Avg. Static Total CPI Change
25565	475	0011 - Legislators	262.42	663.55	128.37	128.38	1.44
6825	409	0012 - Senior government managers and officials	68.25	578.48	130.42	130.43	1.6
4250	391	0012 - Senior government managers and officials	39.75	514.71	136.96	136.95	0.73

Description of Columns (Part 2):

- **Total Workers Count (int)**: The total number of temporary workers in the province and year, indicating the workforce's size and its potential impact on local housing demand.
- **Number of Class Titles (int)**: The number of distinct job classifications for temporary workers in the region, providing a sense of job market diversity.
- **Most Common Class Title (string)**: The most frequently occurring job classification among temporary workers in the given year and province. This provides insight into the types of jobs that are most prevalent in the temporary workforce.
- **Workers to Housing Impact Ratio (float)**: A ratio of the total number of workers to the average housing value. A higher value indicates a greater potential impact of workers on the housing market.
- **Combined Impact Ratio (float)**: The combined impact of international students (study permits) and temporary workers on housing demand, derived from the averages of the Study Permit to Value Impact Metric and the Workers to Housing Impact Ratio.
- **Avg. Total CPI (float)**: The average Consumer Price Index for the given year, indicating the overall inflation rate in the economy.
- **Avg. Total CPI Seasonally Adjusted (float)**: The CPI adjusted for seasonal variations, providing a clearer picture of inflation trends that are not affected by regular seasonal fluctuations.
- **Avg. Static Total CPI Change (float)**: The CPI change that excludes seasonal adjustments, offering insight into the underlying inflationary trends.

Table 8: Final Combined Dataset Snapshot (Part 3)

Avg. CPI Median	Avg. Bank Interest Rate Weekly	Avg. Estimated Variable Mortgage Rate	Avg. Weekly Effective Household Interest Rate	Avg. Weekly Effective Business Interest Rate	Avg. Total CPI	Workers to Housing Impact Ratio
1.99	0.75	2.21	3.04	2.85	128.37	262.42
1.63	0.95	2.15	3.2	2.85	130.42	68.25
1.88	0.81	2.23	3.03	2.71	136.96	39.75

Description of Columns (Part 3):

- **Avg. CPI Median (float)**: The median value of the Consumer Price Index, which gives a central tendency of the inflation rate over the year, helping to understand the typical rate of inflation experienced across the economy.
- **Avg. Bank Interest Rate Weekly (float)**: The average interest rate that banks charge for loans, measured weekly. This indicates the general cost of borrowing for individuals and businesses.
- **Avg. Estimated Variable Mortgage Rate (float)**: The average rate of variable mortgage loans, which can change over time based on market conditions, providing insights into housing affordability.
- **Avg. Weekly Effective Household Interest Rate (float)**: The average interest rate on household loans, including both fixed and variable rates, measured weekly. It reflects the typical borrowing cost for households.
- **Avg. Weekly Effective Business Interest Rate (float)**: The average interest rate charged to businesses on loans, reflecting the cost of capital for commercial activities.
- **Avg. Total CPI (float)**: The average Consumer Price Index for the year, providing an overall picture of inflation trends in the economy.
- **Workers to Housing Impact Ratio (float)**: The ratio of temporary workers to the average housing value, which helps understand how the influx of workers may influence housing prices in the region.

Machine Learning Techniques: The following machine learning techniques were incorporated to analyze and predict housing prices based on various economic and demographic factors.

- **Principal Component Analysis (PCA)**:
  - **Type of Problem:** Unsupervised (Dimensionality reduction).
  - This technique helps reduce the complexity of the data by summarizing it into two main components, while still capturing 85% of the important information.
  - This allows us to focus on key factors like CPI, housing prices, and worker counts.
- **k-Means Clustering**:
  - **Type of Problem:** Unsupervised (Clustering).
  - We grouped provinces into four clusters based on their economic features, such as housing prices, CPI, and the number of workers.
  - This helps identify regions with similar housing trends.

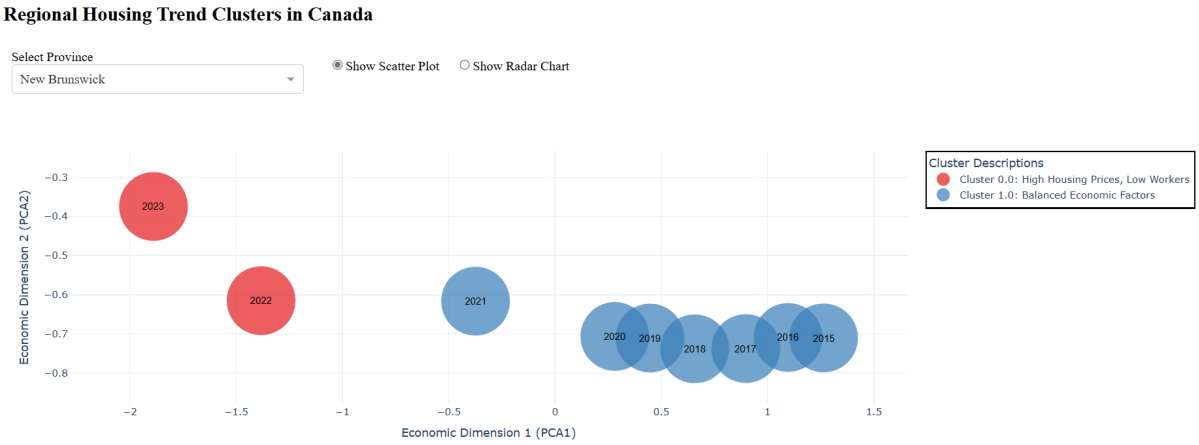

Figure 1: Clusters formed using PCA and k-Means based on housing trends and key economic metrics.

- **Time Series Analysis (ARIMA)**:
  - **Type of Problem:** Supervised (Regression).
  - We used this method to predict future housing prices based on past data.
  - We also included economic factors like CPI to make the predictions more accurate.
  - **Target Attribute:** Housing prices.

Machine Learning and Interactive Visualizations Integration: In our project, we combined machine learning models with interactive visualizations to make the results easier to explore and understand. Here's how they work together:

- Results from PCA and k-Means clustering are shown in interactive scatter plot. This lets users see how different provinces group based on factors like housing prices and economic conditions.
- Predictions from the ARIMA model are displayed in time series charts, showing both historical trends and future predictions for housing prices. This helps users understand how factors like CPI affect the housing market over time.
- Interactive dashboards allow users to filter data, like changing CPI or worker numbers, and automatically update the charts and model outputs to reflect the changes.

This approach makes the data more engaging and easier for stakeholders to explore and understand the insights from the machine learning models.

Integration of Large Language Model (LLM): The Qwen2.5-0.5B LLM [10] was integrated into the project to assist in analyzing complex relationships within the data and providing explanations for visualizations. This model is pre-trained with a transformer architecture, consisting of 0.49 billion parameters, and is optimized for generating structured outputs from large datasets. By leveraging the model's natural language processing capabilities, stakeholders would gain deeper insights into trends and patterns in housing prices and related economic metrics.

- **Dynamic Insights:** The LLM generated analysis for visualizations, such as bar plots, to help stakeholders interpret the relationships between housing prices, Consumer Price Index (CPI), and bank interest rates.
- **Trend Analysis:** The LLM provided explanations to highlight the interplay between economic factors and housing market trends, making the data more accessible for policymakers and urban planners.

- **Stakeholder Support:** Insights were presented in concise, human-readable formats to ensure clarity and provide actionable takeaways for decision-makers.

Key configuration parameters of the LLM:

- **Max Tokens**: 400. This controls the maximum length of the generated response. A lower value is used to provide succinct responses, focusing on the most relevant points.
- **Temperature**: 0.4. The temperature parameter influences the creativity of the model's output. A lower temperature produces more predictable responses, which are useful for tasks like generating concise summaries or factual analysis.
- **Top-p**: 0.9. Top-p sampling (nucleus sampling) focuses on the most probable next tokens, ensuring they are coherent and relevant outputs without the randomness that can come from the full probability distribution.
- **Max Context Length**: 32,768 tokens. This allows the LLM to analyze large datasets or historical trends without losing context, which is crucial for tasks involving extensive tabular data.

Trend Analysis of Housing Prices with CPI, Bank Interest Rates, and Housing Market Metrics

LLM Generated Response:

Based on the following aggregated data for Canada, analyze the trends and interactions between the housing market, CPI, and mortgage interest rates for each year: Year: 0 2015 1 2016 2 2017 3 2018 4 2019 5 2020 6 2021 7 2022 8 2023 Name: Year, dtype: int64, Average CPI: 0 126.57 1 128.37 2 130.42 3 133.38 4 135.98 5 136.96 6 141.61 7 151.24 8 157.11 Name: Avg_Total_CPI, dtype: float64, Average bank interest rate: 0 0.88 1 0.75 2 0.95 3 1.66 4 2.00 5 0.81 6 0.50 7 2.16 8 4.99 Name: Avg_Bank_Interest_Rate_Weekly, dtype: float64, Average housing price: 0 82.736581 1 84.937009 2 100.797436 3 101.724017 4 102.147009 5 104.249573 6 114.040342 7 121.292735 8 121.233077 Name: Avg_Housing_Value, dtype: float64. Please focus on understanding and anlysing the relationship and trends between these metrics over the years and provide explaiantion. The data provided is for Canada and shows the average CPI and bank interest rates over the years. The average housing price is also provided. The relationship between the average CPI and bank interest rates can be seen in the trend line of the bank interest rate. The average bank interest rate has been increasing over the years, from 0.88 in 2015 to 4.99 in 2023. This indicates that the interest rate has increased significantly over time. The relationship between the average CPI and housing prices can be seen in the trend line of the average housing price. The average housing price has been increasing over the years, from 82.736581 in 2015 to 121.292735 in 2023. This indicates that the price of houses has been increasing over time. Overall, the data suggests that the housing market in Canada has been experiencing a steady increase in both the average CPI and the average bank interest rate. However, the price of houses has also been increasing over time. It is important to note that these trends and interactions are subject to change over time, and the data may not be accurate for all years.

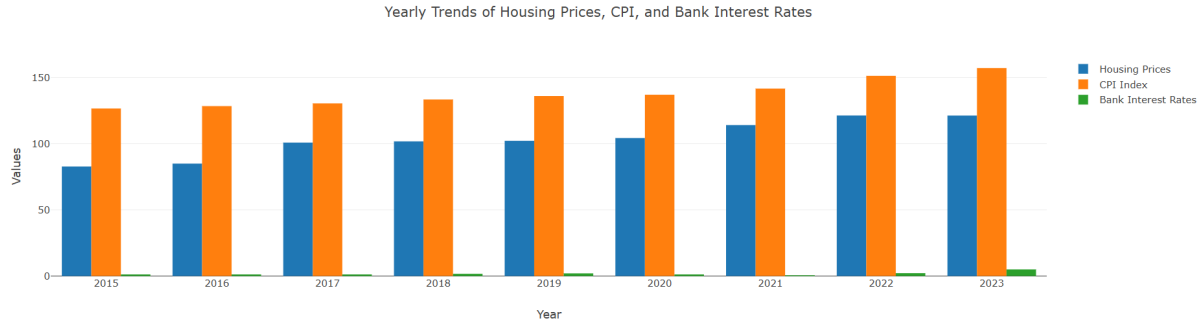

Figure 2: Bar plot comparing housing prices, CPI, and interest rates across Canada with LLM's insights.

The integration of the Qwen2.5-0.5B LLM enhanced the interpretability of the project's findings by complementing visualizations with meaningful and concise explanations, ensuring that the results were easy to comprehend and actionable for a diverse group of stakeholders.

2.2 Justification of Machine Learning Approach

To better understand the complex relationships between housing prices and various economic factors, we applied a combination of machine learning techniques. These methods helped simplify the data, identify patterns, and make reliable predictions. The chosen techniques were selected for their ability to make complex data easier to interpret, reveal insights specific to different regions, and predict future trends.

Principal Component Analysis (PCA): PCA was used to simplify the large dataset by reducing it to two key components, which together captured 85% of the important information. This allowed us to better understand the main factors affecting housing prices and economic conditions without losing any essential data. By transforming complex data into these two components, we could easily visualize and analyze the relationships between various variables, making the data much easier to interpret.

k-Means Clustering: After applying PCA, we used k-Means clustering to group Canadian provinces into four distinct clusters based on their housing trends and economic factors. This clustering helped us identify regions with similar housing behaviors, which is useful for tailoring policies or strategies to specific areas. By grouping provinces in this way, we could focus on the most relevant patterns for different regions, making our analysis more targeted and actionable.

Time Series Analysis (ARIMA): To predict future housing prices, we used the ARIMA model, which looks at past trends to forecast future values. We also included the Consumer Price Index (CPI) as an external factor to ensure the predictions reflected inflation trends, leading to more accurate forecasts. The ARIMA model helped us identify potential challenges in the housing market and allowed us to predict where housing prices might go in the future, supporting decision-making for planners and policymakers.

2.3 Justification of Visualization Modules

The visualizations in this project were selected to provide clear, accessible, and actionable insights into the relationship between housing prices and economic factors. These visual tools help to explore trends, detect patterns, and highlight key data points. Below are the justifications for each of the chosen visualizations, as well as how to interpret them in the context of the project.

Scatter Plots

- **Purpose:** Scatter plots were used to explore the relationships between variables such as housing prices, study permits, and worker counts. These relationships are key to understanding how factors like immigration and labor supply affect housing demand and prices.
- **Rationale:** Scatter plots allow for the immediate visualization of correlations and outliers, making it easier to identify how closely related two variables are. This helps in detecting trends or anomalies that may require further investigation.
- **Interpretation:** In the scatter plot showing the relationship between study permits and housing prices, a positive correlation suggests that as the number of study permits increases, housing prices tend to rise, highlighting the impact of international students on housing demand. Outliers, such as provinces with a sudden spike in housing prices despite low study permits, may indicate other influencing factors.

Figure 3: Scatter Plot showing the relationship between study permits and housing prices.

Line Graphs

- **Purpose:** Line graphs were used to show the temporal trends of housing prices and key economic indicators over time. Understanding the progression of these factors helps to predict future housing trends and assess the long-term impact of external variables.
- **Rationale:** Line graphs are effective for displaying changes over time, making it easy to identify year-over-year trends, seasonal fluctuations, and long-term growth or decline in housing prices.
- **Interpretation:** The line graph showing the yearly trend of housing prices allows stakeholders to see how prices have evolved over time. For example, an upward trend could indicate a growing demand for housing, while a downward trend might suggest a cooling market. The presence of cyclical patterns may also highlight seasonal effects or external economic factors influencing prices.

Analyzing the Impact of Temporary Residents on Canada's Housing Market

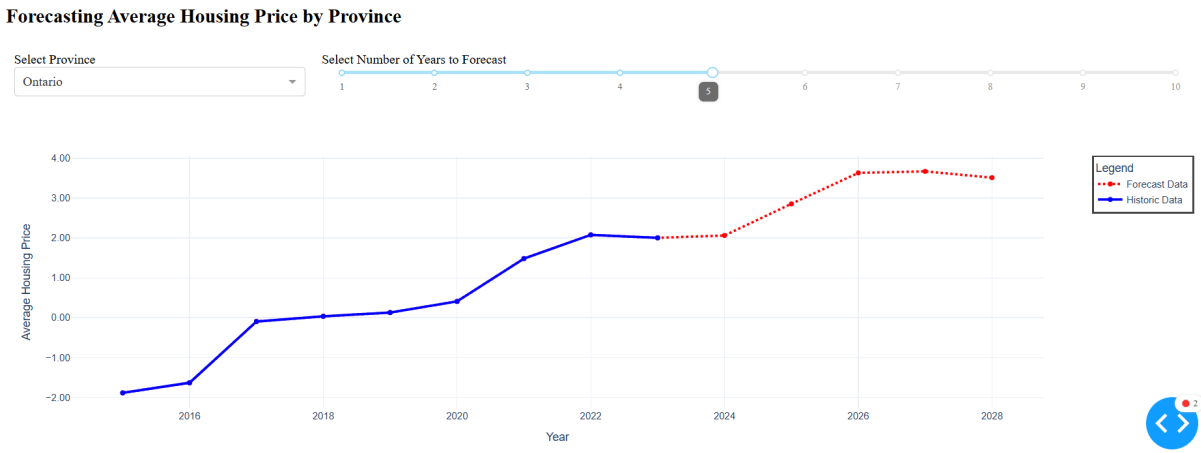

Figure 4: Line Graph showing the yearly trend of housing prices.

Bar Charts

- **Purpose:** Bar charts were used to compare housing prices and other metrics (e.g., study permits and worker counts) across provinces. These comparisons help to assess regional differences in housing markets and economic conditions.
- **Rationale:** Bar charts are great for comparing categories side by side. In this case, they enable the comparison of housing prices across different provinces, making it easy to spot regions with higher or lower costs.
- **Interpretation:** The bar chart comparing average housing prices across provinces helps identify which regions are experiencing the highest housing prices. A larger bar indicates a province with higher average housing prices, which could be due to various factors such as population density or economic development. Smaller bars indicate provinces where housing is more affordable, and this can guide region-specific strategies or policies.

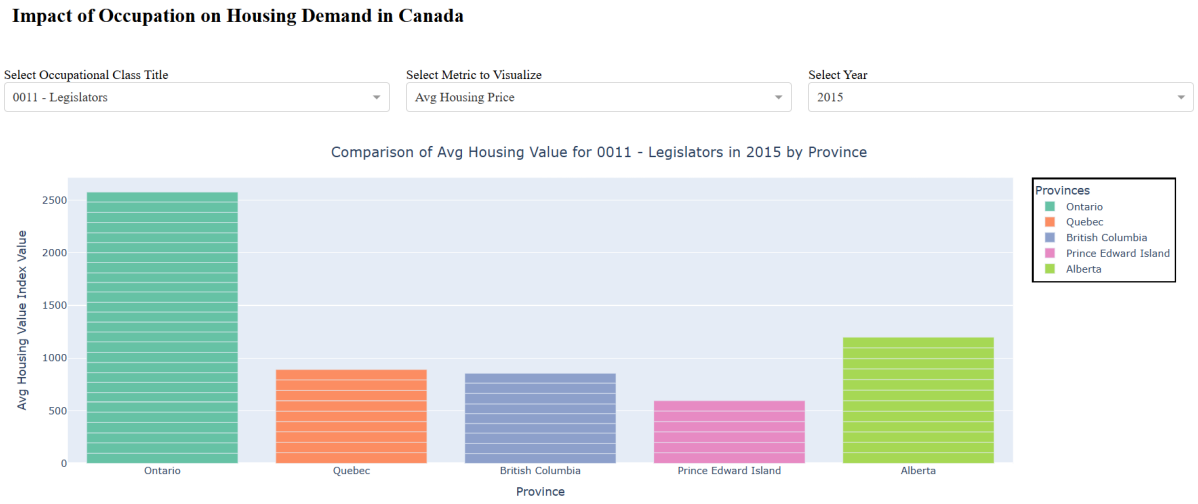

Figure 5: Bar Chart comparing average housing prices across provinces.

Radar Charts

- **Purpose:** Radar charts were used to compare multi-dimensional data across regions. This type of chart is ideal for showing how various factors, such as study permits and worker counts, collectively influence housing prices in different provinces.
- **Rationale:** Radar charts provide a compact way to visualize complex, multi-dimensional relationships between several factors. This makes it easier to see which factors are most influential in each region and how they compare with others.
- **Interpretation:** In the radar chart visualizing the combined impact of study permits and workers on housing prices, the chart shows how each region's study permits and worker data contribute to housing prices. Regions with larger values in the "study permits" and "workers" areas will have more significant impacts on housing demand, which could drive up prices. This allows urban planners and policymakers to target specific factors in each region to address housing challenges.

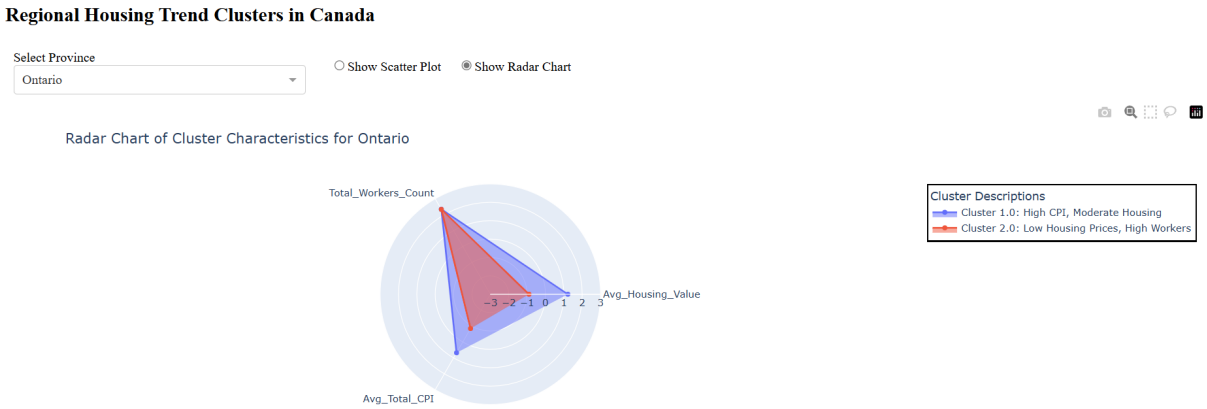

Figure 6: Radar Chart visualizing the combined impact of study permits and workers on housing prices.

2.4 Evaluation Metrics and Quality Assessment

The evaluation of the project's effectiveness was done using a combination of quantitative metrics and qualitative feedback. Each metric was selected to assess different aspects of the proposed approach, from clustering quality and dimensionality reduction to forecasting accuracy and user experience. The following outlines the evaluation metrics used in the project:

Clustering Quality (Silhouette Score): To evaluate the quality of the k-Means clustering, we used the Silhouette Score. This metric measures how similar each data point is to its own cluster compared to other clusters. A higher score indicates that the clusters are well-defined and that the provinces are grouped based on similar characteristics. The Silhouette Score ensures that the clustering is meaningful, with clear distinctions between the groups based on housing prices and economic factors.

Explained Variance (PCA): The explained variance metric quantifies the amount of information preserved after dimensionality reduction. In this project, we assessed how much of the original data's variance was retained by the first two principal components. A higher percentage of explained variance indicates that the reduced dataset still captures the key characteristics of the original data, ensuring that important patterns and relationships are maintained.

Forecast Accuracy (RMSE): To assess the accuracy of the ARIMA time-series forecasting model, we used the Root Mean Squared Error (RMSE). This metric measures the difference between the predicted housing prices and the actual outcomes, providing a quantifiable measure of prediction error. Lower RMSE values indicate more accurate forecasts, which is crucial for understanding future housing trends. By comparing predicted values to actual results, we ensured the reliability and usefulness of the forecasts.

User Feedback on Visualizations: In addition to quantitative metrics, we collected qualitative feedback from the Professor and Teaching Assistant regarding the usability and clarity of the visualizations. This feedback was used to assess how effectively the visualizations conveyed insights into the housing market and economic factors. Their input helped ensure that the visualizations were intuitive and useful for decision-making, contributing to the refinement of the interface and the improvement of the overall user experience.

2.5 Tools, Libraries, and External Resources

Programming Tools and Libraries: Python has been used as the primary programming language for data processing, analysis, and modeling. The key libraries utilized include:

- **pandas**: For handling and analyzing structured data which helped with the manipulation and cleaning of large datasets [11].
- **NumPy**: For numerical computations, especially useful for working with arrays and large datasets of numbers [12].
- **Plotly**: For creating interactive visual graphs to display trends and patterns in the data [13].
- **scikit-learn**: For implementing machine learning models, such as clustering and dimensionality reduction using PCA [14].
- **statsmodels**: For time series forecasting, models like ARIMA have been included to predict future trends based on historical data [15].
- **transformers**: For integrating large language models to generate insights, summaries, and alerts based on data analysis [16].
- **Google Colab**: For collaborative coding and data preprocessing which enabled real-time work sharing and version control [17].
- **Dash**: A framework for building interactive web applications with Python, used for developing interactive data visualizations [18].

Novelty: This project introduces the use of Large Language Models (LLMs) to analyze complex data and generate easy-to-understand insights. Key features include:

- Leveraging LLMs to automatically summarize data trends and provide actionable insights for decision-makers.
- Developing interactive dashboards and visualizations to explore housing affordability and the impact of temporary residents, enhanced with explainable clusters.
- Applying machine learning models along with economic data to predict housing market trends.

3 Functionality

All of the proposed functionalities were successfully implemented, with each stage enhancing the complexity and depth of the analysis.

- **Complete Demo Link:** https://tinyurl.com/trhademorecording
- **Deployed Dashboard Link:** https://trha-dashboard.onrender.com/

The system's functionality was developed in three stages, gradually increasing the complexity and depth of the analysis.

3.1 Minimum Functionality

- **Data Loading, Cleaning, and Pre-processing**: Load the datasets, clean them, and ensure consistency, taking care of missing values.
- **Basic Time-Series Visualizations**: Visualization to show the trend of housing prices and permit holders over time to bring out the temporal relationship.
- **Correlation Analysis**: Correlation analysis to understand the influx of temporary residents' relationships with housing price fluctuation.

3.2 Expected Functionality

- **Advanced Machine Learning Models**: Implement **ARIMA** for time series forecasting to predict housing prices using historical data. Utilize **Random Forest** for feature importance analysis to evaluate how different factors, such as permit type, country of origin, and region, impact housing prices.
- **Clustering Models**: Utilize clustering algorithms to segment regions by housing price trends and influx of residents to determine regions with similar market behaviors.
- **Advanced Visualizations**: Create advanced visualizations including time-series plots, correlation heatmaps, and clustering results, that are interactively filtered by region, time, and permit type.

3.3 Bonus Functionality

- **Anomaly Detection Visuals**: Employ models such as **Isolation Forest** that detect unusual housing price fluctuations and visualize abnormal trends in housing prices.
- **External Data Integration**: Integrate external data sources relating to news, policy changes, or economic indicators and update them in real time through web scraping or APIs to put the housing trends into context.
- **Generative AI Integration**: Use Generative AI or Large Language Models (LLMs) for suggestions and recommendations (e.g., news articles, policy changes).
- **Deploying the Dashboard**: Make the dashboard widely accessible by deploying it to a cloud service.

Overall, the functionality was successfully implemented, with each stage contributing to a more detailed and comprehensive analysis. Starting from basic data processing and visualization, the system evolved to include advanced machine learning models, clustering techniques, and real-time data integration. The final stage incorporated innovative features like anomaly detection and generative AI, enhancing the system's ability to provide valuable insights. Nonetheless, the deployed dashboard would allow stakeholders to easily interact with the results and make informed decisions based on historical data, providing a solid foundation for future improvements and extensions.

4 Individual Contributions of Team Members

Each team member contributed to different aspects of the project, ensuring smooth execution and successful results. Below are the individual contributions:

- **Ovaiz Ali:**
  - Designed and implemented the machine learning models (PCA, k-Means, and ARIMA).
  - Integrated the Large Language Model (LLM) for generating insights and summaries.
  - Prepared and cleaned datasets for analysis, ensuring consistency and accuracy.
  - Developed the dashboard's interactive visualizations and functionalities.
  - Responsible for loading, cleaning, and preprocessing all datasets.
  - Developed machine learning models for feature importance analysis, time-series forecasting, clustering, and anomaly detection.
  - Evaluated models, performed hyperparameter tuning, and improved accuracy.
  - Ensured model accuracy, tested system functionality, and optimized overall performance.

- Collaborated on the preparation of the final report and presentation.

- **Bhavya Mukesh Dave:**
  - Contributed to the design and testing of visualization modules, ensuring clarity and usability.
  - Assisted in data preprocessing and creation of derived metrics for analysis.
  - Conducted exploratory data analysis (EDA) to identify trends and patterns.
  - Collaborated on the report writing, proofreading, and presentation preparation.
  - Assisted with data exploration and initial cleaning.
  - Enhanced dashboard visualizations for feature importance, clustering, and anomaly detection.
  - Assisted with testing and optimization processes.
  - Collaborated on the preparation of the final report and presentation.

5 Conclusions and Future Work

5.1 Conclusions and Learnings

This project showed how visual analytics, machine learning, and large language models (LLMs) can be used together to understand the impact of temporary residents and economic factors on housing prices in Canada. Here are the main takeaways from the analysis:

- Temporary residents, like international students and foreign workers, have a significant effect on housing prices, especially in urban areas.
- Economic factors such as the Consumer Price Index (CPI) and interest rates are closely connected to housing price trends, making them important for predicting the housing market.
- Grouping provinces into clusters revealed different housing market behaviors across Canada, helping us understand regional differences in housing prices.
- Combining visual analytics and machine learning provides valuable insights that can help policymakers, urban planners, real estate investors, and researchers make better decisions.
- Adding data on bank interest rates and CPI, based on our professor's suggestion, added more depth to the analysis and will be useful in addressing complex issues like housing affordability.

5.2 Challenges Encountered

Throughout the project, several challenges arose, and we found ways to resolve them based on:

- Our professor's suggestion to add bank interest rates and CPI data gave an extra boost to an all-encompassing dimension to our analysis, that will be critical in teasing different factors when solving such complex societal challenges as housing affordability.
- In the integration of KMeans clustering with visualizations, we experienced slow responses. The process of loading large datasets into memory for clustering caused performance bottlenecks, especially in interactive visualization tools. This resulted in delays in real-time insight presentation and made the system less responsive.
- For these reasons, we optimized the machine learning models by using a lighter, parameter-efficient LLM model that allowed for faster processing on local machines. Additionally, we improved the efficiency of the KMeans clustering by using dimensionality reduction techniques - such as PCA - before applying the clustering algorithm, which helped reduce memory usage and improve response times in the visualizations.

5.3 Future Work

If we had more time and resources, there are a few areas where we could improve and expand this project:

- **Real-Time Data Integration**: Integrating real-time data sources such as news feeds and policy changes to provide a more dynamic responsive predictive model.
- **Enhanced Forecasting Models**: Studying other advanced machine learning models like LSTM could provide even better results in the accuracy and robustness of the prediction of house prices.
- **Expanded Geographic Scope**: Adding data on rural areas will extend the knowledge base of the Canadian housing market, giving a wider overview of national trends.
- **Broader Stakeholder Engagement**: Conducting user studies to refine the usability of the dashboard would, in turn, enable us to incorporate feedback into future versions to ensure it meets the needs of various stakeholders.
- **Optimizing Clustering and Visualization Integration**: Addressing the performance issues with KMeans clustering and improving the efficiency of data handling could enhance the responsiveness of the visual analytics system.

Hence, this project showed how combining visual analytics, machine learning, and large language models (LLMs) can help us understand the factors affecting housing prices in Canada. We focused on how temporary residents and economic factors like interest rates impact the housing market. The key takeaway is that these technologies together can give us useful insights to make better decisions about housing trends. In the future, improving the models and using real-time data will make this system even more valuable for decision-makers.

References

[1] N. A. Mallees, "Immigration is making Canada's housing more expensive, the government was warned 2 years ago," *CBC News*, Jan. 2024. Available: https://www.cbc.ca/news/politics/ircc-immigration-housing-canada-1.7080376. [Accessed: Oct. 6, 2024].

[2] S. Pomeroy, "Temporary foreign workers and students are adding pressure to the rental market," *CBC News*, Jan. 2024. Available: https://www.cbc.ca/news/politics/immigration-housing-crisis-costs-1.7088878. [Accessed: Oct. 6, 2024].

[3] P. Zimonjic, "Bank of Canada says government efforts to curb housing crisis will help 'gradually'," *CBC News*, Jan. 2024. Available: https://www.cbc.ca/news/politics/immigration-bank-of-canada-housing-1.7093426. [Accessed: Oct. 6, 2024].

[4] Immigration, Refugees, and Citizenship Canada (IRCC), "Strengthening temporary residence programs for sustainable volumes," *Government of Canada*, Sept. 2024. Available: https://www.canada.ca/en/immigration-refugees-citizenship/news/2024/09/strengthening-temporary-residence-programs-for-sustainable-volumes.html. [Accessed: Oct. 6, 2024].

[5] Government of Canada, "Temporary residents: Study permit holders – Monthly IRCC updates - Canada - Study permit holders by country of citizenship and year in which permit(s) became effective," *Open Government Portal*, 2023. Available: https://open.canada.ca/data/en/dataset/360024f2-17e9-4558-bfc1-3616485d65b9. [Accessed: Oct. 6, 2024].

[6] Government of Canada, "Temporary residents: Temporary foreign worker program (TFWP) and international mobility program (IMP) work permit holders – Monthly IRCC updates - Canada - International mobility program work permit holders by province/territory, intended occupation (4-digit NOC 2011) and year in which permit(s) became effective," *Open Government Portal*, 2023. Available: https://open.canada.ca/data/en/dataset/360024f2-17e9-4558-bfc1-3616485d65b9/resource/8af07cf1-2d7a-4428-a27c-4d6e719b18cf. [Accessed: Oct. 6, 2024].

[7] Government of Canada, "New housing price index," *Open Government Portal*, 2023. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810020501. [Accessed: Oct. 6, 2024].

[8] Bank of Canada, "Consumer Price Index (CPI)," *Bank of Canada*, 2023. Available: https://www.bankofcanada.ca/rates/price-indexes/cpi/. [Accessed: Oct. 6, 2024].

[9] Bank of Canada, "Canadian Interest Rates," *Bank of Canada*, 2024. Available: https://www.bankofcanada.ca/rates/interest-rates/canadian-interest-rates/. [Accessed: Oct. 6, 2024].

[10] "Qwen2.5-32B," Hugging Face, 2024. Available: https://huggingface.co/Qwen/Qwen2.5-32B. [Accessed: Dec. 6, 2024].

[11] "pandas: Powerful Python data analysis toolkit," *pandas*, 2024. Available: https://pandas.pydata.org/. [Accessed: Dec. 6, 2024].

[12] "NumPy: The fundamental package for scientific computing with Python," *NumPy*, 2024. Available: https://numpy.org/. [Accessed: Dec. 6, 2024].

[13] "Plotly: Interactive graphing library," *Plotly*, 2024. Available: https://plotly.com/python/. [Accessed: Dec. 6, 2024].

[14] "scikit-learn: Machine Learning in Python," *scikit-learn*, 2024. Available: https://scikit-learn.org/stable/. [Accessed: Dec. 6, 2024].

[15] "statsmodels: Statistical modeling in Python," *statsmodels*, 2024. Available: https://www.statsmodels.org/. [Accessed: Dec. 6, 2024].

[16] "transformers: State-of-the-art Natural Language Processing for Pytorch and TensorFlow 2.0," *transformers*, 2024. Available: https://huggingface.co/transformers/. [Accessed: Dec. 6, 2024].

[17] "Google Colab," Google, 2024. Available: https://colab.research.google.com/. [Accessed: Dec. 6, 2024].

[18] "Dash: A Python framework for building analytical web applications," *Plotly*, 2024. Available: https://dash.plotly.com/. [Accessed: Dec. 6, 2024].