***6/30/2023***

***Neha Tararr***

***Canadian Real Estate***

Table of Contents

[Introduction 3](#_Toc145590081)

[Canadian Real Estate 3](#_Toc145590082)

[Housing Affordability 3](#_Toc145590083)

[Process 3](#_Toc145590084)

[Driving Results 4](#_Toc145590085)

[Buyers versus Renters Analysis 4](#_Toc145590086)

[Process 4](#_Toc145590087)

[Driving Results 4](#_Toc145590088)

[Indicators 6](#_Toc145590089)

[Process 6](#_Toc145590090)

[Driving Results 6](#_Toc145590091)

[Dimensional Modeling (SQL Relational Diagram) 7](#_Toc145590092)

[Process 7](#_Toc145590093)

[Driving Results 8](#_Toc145590094)

[Feature Engineering (Python Coding) 9](#_Toc145590095)

[Process 9](#_Toc145590096)

[Driving Results 10](#_Toc145590097)

[Data Exploration and Visualizations (PowerBI) 10](#_Toc145590098)

[Process 10](#_Toc145590099)

[Driving Results 10](#_Toc145590100)

[Resources: 11](#_Toc145590101)

# Introduction

The Canadian real estate market is undergoing significant changes, driven by factors such as average household income, monthly mortgage payments, and the dynamics between buyers and renters. These factors play a crucial role in understanding market trends and making informed investment decisions. Additionally, key economic indicators, including inflation rate, interest rate, GDP, population, and unemployment rate, contribute to shifts in the real estate landscape. Leveraging SQL for dimensional modeling, Python for feature engineering, and Power BI for data exploration and visualization, investors can gain valuable insights and navigate the dynamic real estate market effectively.

# Canadian Real Estate

## Housing Affordability

* Analyze the relationship between monthly mortgage payments and monthly household income to assess housing affordability in Canada.
* Calculate the debt-to-income ratio to determine the proportion of income allocated to mortgage payments.
* Compare the affordability of housing across different provinces or major cities in Canada.
* Discuss trends in housing affordability over time and their implications for homeownership and rental markets.

### Process

During the process of uploading the 'Average Household Income' dataset and 'Average Scheduled Monthly Mortgage Payments' dataset into Power BI, an issue emerged regarding the clarity of their interrelation. While these datasets were deemed appropriate for the topic at hand, a challenge arose due to the representation of values in monetary data type ($), hindering the clear understanding of the percentage increase. To address this, the initial step involved converting the data type into percentage values, enabling a comprehensive depiction of the growth rate for both average monthly mortgage payments and average household income.

### Driving Results

Several steps were taken to prepare and analyze the data in Python. Firstly, unnecessary data from the 'Selected Metropolitan Areas' dataset was removed. The timeframe was then limited to the years 2013-2023, and quarterly data was aggregated to calculate annual values. To handle missing values, they were replaced with zeros. Lastly, a new column was added to display the rate of change between the 'Average Household Income' and 'Average Scheduled Monthly Mortgage Payments' datasets. This involved replacing null values with zeros, dropping irrelevant columns, and calculating the change rate between consecutive years. These steps ensured data quality and facilitated meaningful analysis.

## Buyers versus Renters Analysis

* Examine the reasons behind the choice of buying or renting, such as financial considerations, lifestyle preferences, or housing market conditions.
* Explore the role of real estate investment in building wealth and diversifying investment portfolios.

### Process

In the analysis of buying versus renting, several factors were explored, including financial considerations, types of housing, and rental verses buyer trend, 2006-2022, prevailing conditions of the housing market. Additionally, the role of real estate investment in the light of types of housing based on their Tenures was accumulation and the portfolio diversification was examined. However, during the analysis process, a problem emerged where the values inserted in certain columns were not being displayed. Furthermore, a challenge arose in comparing values between buyers and renters, as shown in, ‘Figure 2: Buyers verses Renters’. To overcome this obstacle, relationships between buyers’ and renters’ tables were examined.

### Driving Results

To address the issue of certain columns not displaying values, duplicate columns were created to ensure visibility of the data, as shown in ‘Figure 1: Duplication of Tables’. Additionally, when comparing values between buyers and renters, relationship management and establishing a connection between the two groups were implemented, enabling a comprehensive comparison and analysis. Two datasets were created: one to illustrate the increase between buyers and renters, and the other to facilitate a comparison of housing types, as shown in ‘Figure 3: Creating Relationships in PowerBI’.



Figure 1: Duplication of Tables

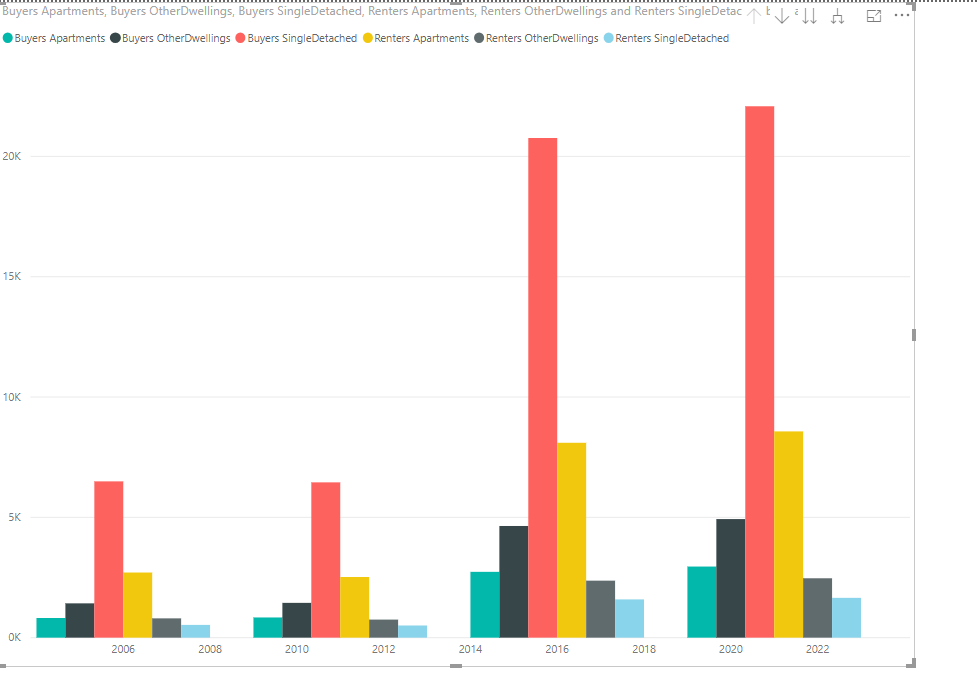


Figure 2: Buyers verses Renters

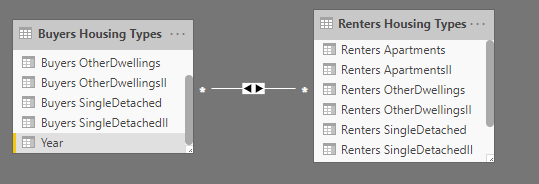


Figure 3: Creating Relationships in PowerBI

## Indicators

* Discuss key factors that influence the real estate market, such as interest rates, population growth, employability vs unemployability, and economic indicators.
* Consider macroeconomic factors, such as GDP growth, inflation, or demographic shifts, that could shape the future of the real estate market.
* Analyze the impact of interest rates on mortgage payments and discuss how changes in rates affect housing affordability.

### Process

The analysis of various datasets revealed several factors influencing the Canadian real estate economy. Firstly, the inflation rate exhibited an upward trend, explaining the increase in mortgage payments. Further investigation unveiled an increase in interest rates, which also contributed to the rise in mortgage payments. Additionally, Canada's GDP experienced a decline, indicating a slower pace of economic growth. Moreover, although population growth was not substantial, the country's unemployment rate showed a steady increase over time. Collectively, these datasets indicate challenges faced by the Canadian real estate economy. However, obtaining these datasets posed challenges due to variations in formats and timeframes. For instance, the interest rate table contained multiple values for each year, lacking an annual rate required for analysis, as shown in ‘Figure 4: Interest Rate Table’. Similarly, the population table was segmented into quarterly data instead of providing annual growth figures, as shown in ‘Figure 5: Population Table’.

### Driving Results

To ensure data consistency and relevance, certain adjustments were made to specific datasets. For instance, the interest rate table contained multiple values for each year, necessitating the summation and averaging of these values. Similarly, in the population dataset, extra columns were eliminated, retaining only the data from quarter 4. This decision was made as quarter 4 provided the most recent and representative population figures, rendering quarters 1, 2, and 3 irrelevant for the analysis.

A screenshot of a computer

Description automatically generated with low confidence

Figure 4: Interest Rate Table

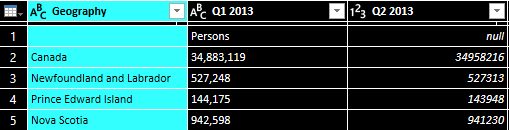


Figure 5: Population Table

## Dimensional Modeling (SQL Relational Diagram)

* Create a Fact Table.
* Create a virtual key in the Fact Table and Dimensional Table.
* Define Primary and Foreign keys.
* Add Primary key to the Fact Table and define relationships with dimensional tables via foreign key.

### Process

The process of establishing SQL Dimensional Modeling proved to be challenging due to the diverse range of tables sourced for Canada’s Real Estate research. The lack of commonality among the tables presented obstacles, as there were no pre-existing relationships, foreign or primary keys, or a fact table that linked to the dimensional tables. Consequently, the manual creation of relationships became necessary to bridge the gaps and establish a cohesive framework for analysis.

### Driving Results

A table named "Capstone" was created, and CSV files were added as separate tables within it. Initially, the tables lacked relationships, so virtual key was introduced to establish connections. This allowed the creation of a relational diagram to visualize the relationships. Primary key was assigned to the Fact table, linking them to the foreign keys in the Dimensional tables, as shown in ‘Figure 6: Primary and Foreign Keys’. Virtual key was also added to ensure uniqueness without creating explicit relationships. The virtual key from the Fact table became the primary key, and the dimensional tables referenced it as foreign keys. This process resulted in the formation of a star schema among all the tables, as shown in ‘Figure 7: Star Schema’.

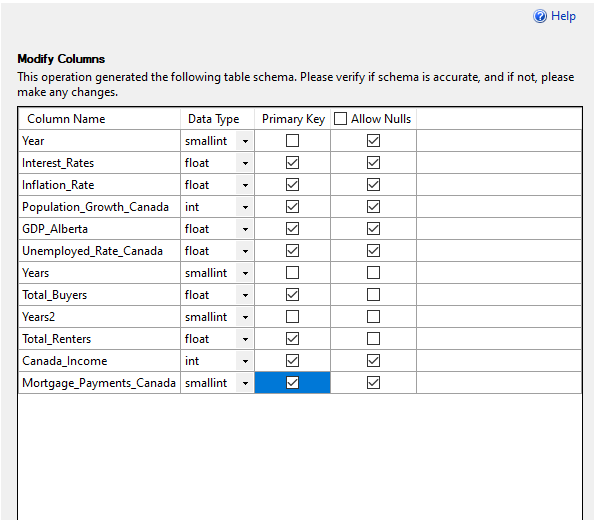


Figure 6: Defining Primary and Foreign Keys

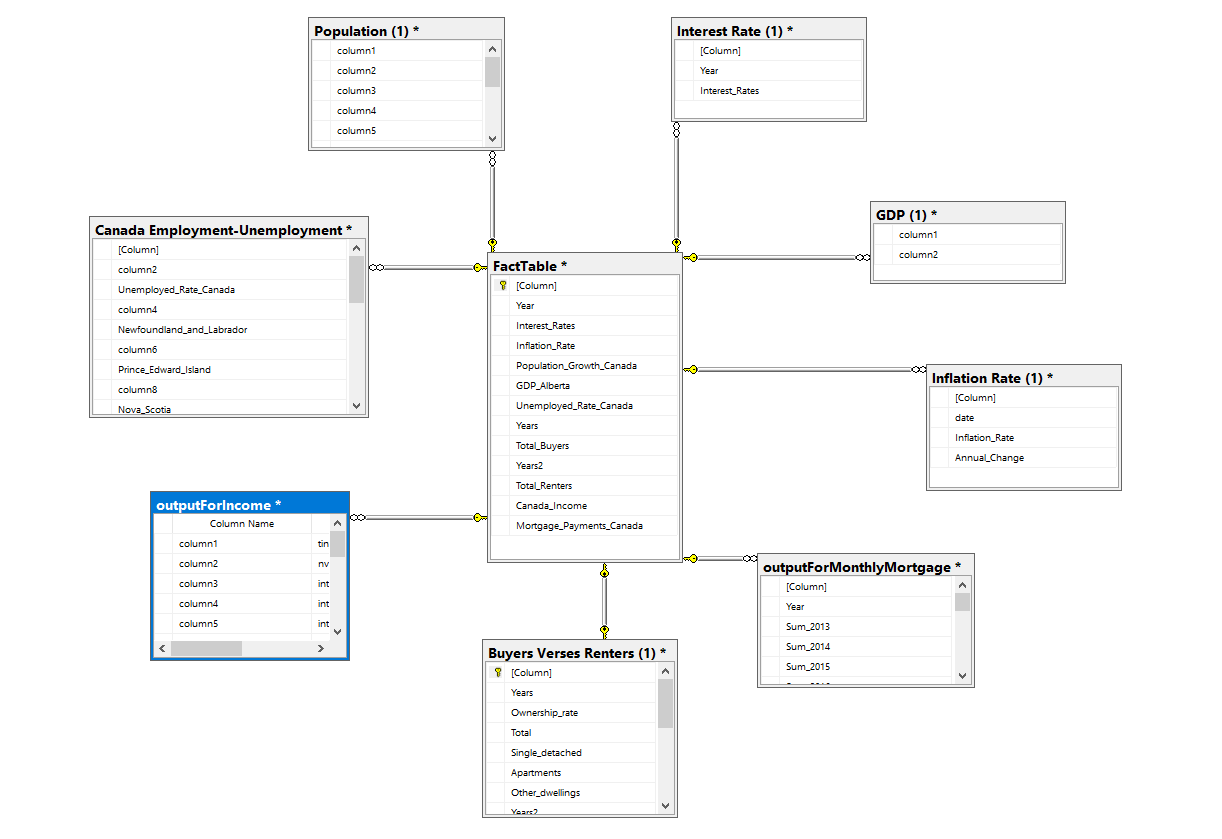


Figure 7: Star Schema

## Feature Engineering (Python Coding)

* Calculations, aggregations, filtering, merging, or creating visualizations using libraries like Pandas.
* Data Loading and Exploration.
* Data Cleaning and Preprocessing.

### Process

A dataset was created using Python's Pandas library. However, there were connectivity issues between Pandas and my device, which needed to be resolved. Additionally, the dataset contained numerous null values that could impact the code's functionality. The timeframe of the dataset did not align with the requirements of the analysis, necessitating adjustments. Furthermore, the tables were formatted in dollar data type, which hindered the estimation of changes between each year.

### Driving Results

Due to connectivity issues with Pandas, I opted to utilize PowerShell to display the coding results successfully. To ensure data integrity, null values were removed from the dataset. Additionally, the timeframes were carefully restricted to align with the analysis requirements. To calculate the change rate between the current year and the previous year, new columns were created using aggregated functions. Subsequently, the original columns were deleted to streamline the dataset and maintain focus on the desired variables.

## Data Exploration and Visualizations (PowerBI)

* Visualizations.
* Data exploration.
* Arranging and renaming the Columns.
* Cleaning unneeded information.
* Creating aggregating functions to extract annual values.

### Process

With the cleaned datasets at hand, the next step involved renaming the columns and addressing any empty values within Power BI. To ensure coherence and consistency across the data, new custom columns were created as needed. The selection of appropriate aggregated functions, for these custom columns, was determined by the data types and the specific questions being addressed. It was also noted that some column headers were undefined, as shown in ‘Figure 8: Column Headers’. Additionally, there were instances where the data types were distorted and required attention for accurate analysis.

### Driving Results

The datasets were successfully loaded into Power BI, allowing the creation of a dataset with established relationships in the module view. To address null values, new custom columns were generated, while the outdated tables were removed from the Power Query Editor. Header definitions were assigned to each column to enhance clarity. Any remaining null values were replaced with zero to ensure data completeness. Furthermore, column names were renamed, and their data types were adjusted to align with their respective values within the Power Query Editor. To provide meaningful insights, certain columns were removed and replaced with newly created custom columns that presented aggregated values such as sums, averages, and other relevant metrics based on the nature of the table and data. For example, in the Interest Rate Table, obsolete columns containing multiple values for a single year were removed. The year was transformed into a grouping by criterion, and aggregated functions were applied, as shown in ‘Figure 9: Interest Rate Table’.

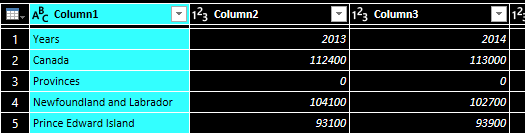


Figure 8: Column Headers

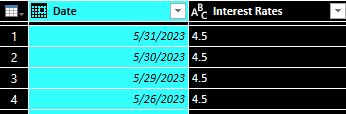


Figure 9: Interest Rate Table

# Resources:

[**New Mortgage Loans: Average Monthly Payments | CMHC (cmhc-schl.gc.ca)**](https://www.cmhc-schl.gc.ca/en/professionals/housing-markets-data-and-research/housing-data/data-tables/mortgage-and-debt/average-scheduled-monthly-payments-new-mortgage-loans) **- Average Scheduled Monthly Payments for New Mortgage Loans – average mortgage people were paying from 2012 – 2023**

[**Projections: Households by Tenure and Ownership Rate to 2036 (cmhc-schl.gc.ca)**](https://www.cmhc-schl.gc.ca/en/professionals/housing-markets-data-and-research/housing-data/data-tables/household-characteristics/number-households-tenure-ownership-rate-1976-2036) **– Buyers verses Renters**

[**2006 – 2020 Housing Stats (average before-tax household income) | CMHC (cmhc-schl.gc.ca)**](https://www.cmhc-schl.gc.ca/en/professionals/housing-markets-data-and-research/housing-data/data-tables/household-characteristics/real-average-total-household-income-before-taxes) **- Average Household Income (2006-2020)**

[**Canada Inflation Rate 1960-2023 | MacroTrends**](https://www.macrotrends.net/countries/CAN/canada/inflation-rate-cpi) **– Inflation Rates**

[**Canadian interest rates and monetary policy variables: 10-year lookup - Bank of Canada**](https://www.bankofcanada.ca/rates/interest-rates/canadian-interest-rates/?lookupPage=lookup_canadian_interest.php&startRange=2013-06-26&rangeType=dates&dFrom=2013-06-26&dTo=2023-05-31&rangeValue=1&rangeWeeklyValue=1&rangeMonthlyValue=1&ByDate_frequency=daily&submit_button=Submit) **– Interest Rates**

[**Gross domestic product (GDP) at basic prices, by industry, provinces and territories (statcan.gc.ca)**](https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610040201&pickMembers%5B0%5D=2.2&pickMembers%5B1%5D=3.1&cubeTimeFrame.startYear=2013&cubeTimeFrame.endYear=2022&referencePeriods=20130101%2C20220101) **– GDP**

[**Population estimates, quarterly (statcan.gc.ca)**](https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901&cubeTimeFrame.startMonth=01&cubeTimeFrame.startYear=2013&cubeTimeFrame.endMonth=10&cubeTimeFrame.endYear=2023&referencePeriods=20130101%2C20231001) **– Population**

[**Employment and unemployment rate, annual (statcan.gc.ca)**](https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410037501&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=4.1&pickMembers%5B2%5D=5.1&cubeTimeFrame.startYear=2013&cubeTimeFrame.endYear=2022&referencePeriods=20130101%2C20220101) **– Employment verses Unemployment**

[**Housing Data Tables | CMHC (cmhc-schl.gc.ca)**](https://www.cmhc-schl.gc.ca/en/professionals/housing-markets-data-and-research/housing-data/data-tables) **- website for my datasets**