

# Analysis and Conclusion

Anjila, Talal, Amit, Neda

## Research Question

The main research question we had was how socioeconomic factors such as population size, median family income, and geographical location influence real estate prices across various cities and provinces.

Our main research question was divided up into different factors that may affect housing. Neda researched the relationship between population, affordability by cities, and housing. Anjila researched the relationship between income and housing. Talal researched the relationship between geographical location and housing. Amit researched the relationship between age, education and homeownership. Below is the breakdown of the research question per person.

### Population - Neda

- How does housing affordability vary across different Canadian cities?
- Is there a correlation between city population size and average house price?
- How does population density (people per square kilometre) influence housing prices and affordability across Canadian cities?

### Income - Anjila

- How does median family income influence housing prices across Canadian cities?
- Which cities have the greatest disparity between median family income and average housing prices? What is the distribution of house prices within a highest/lowest disparity city?

### Geographical Location - Talal

- How do housing prices differ among provinces and territories in Canada, and what are the contributing factors? Is there a correlation between geographical location (e.g., elevation, climate, coastal, inland) and housing prices in Canada?
- Is there a correlation between the density of population and bedrooms/bathrooms in homes across Canadian cities? Which cities have the highest average price per bedroom? What is the relationship between the number of bedrooms and home ownership in urban versus rural areas?
- How do housing prices vary between areas with high population (urban) versus lower population (rural) areas?

### Demographic Data - Amit

- How does age and education level affect home ownership in the top 5 cities?

# Summaries for each Research Question

## Population - Neda

This analysis investigates the key factors influencing housing affordability across Canadian cities, focusing on the relationship between house prices, population size, density, and median income. By comparing these variables, the study seeks to uncover patterns that explain the varying levels of affordability between large, densely populated urban centers like Vancouver and Toronto and smaller cities such as Regina and Halifax. The goal is to better understand the socioeconomic and geographical elements contributing to housing affordability disparities across Canada's regions.

### How does housing affordability vary across different Canadian cities?

The main cleaned dataset used for this investigation was the merged house listing and population dwelling CSV files. It contains various information, including city, province, area type (urban/rural), population, population density, total dwellings, land area, median income, house prices, and the number of bedrooms and bathrooms in homes. For this specific question, I focused on comparing the average house prices in different Canadian cities to their respective median incomes to evaluate housing affordability.

To perform this analysis, I created two smaller DataFrames:

1-average\_house\_price\_city: This data frame was generated by grouping the data by city and calculating the average house price for each city.

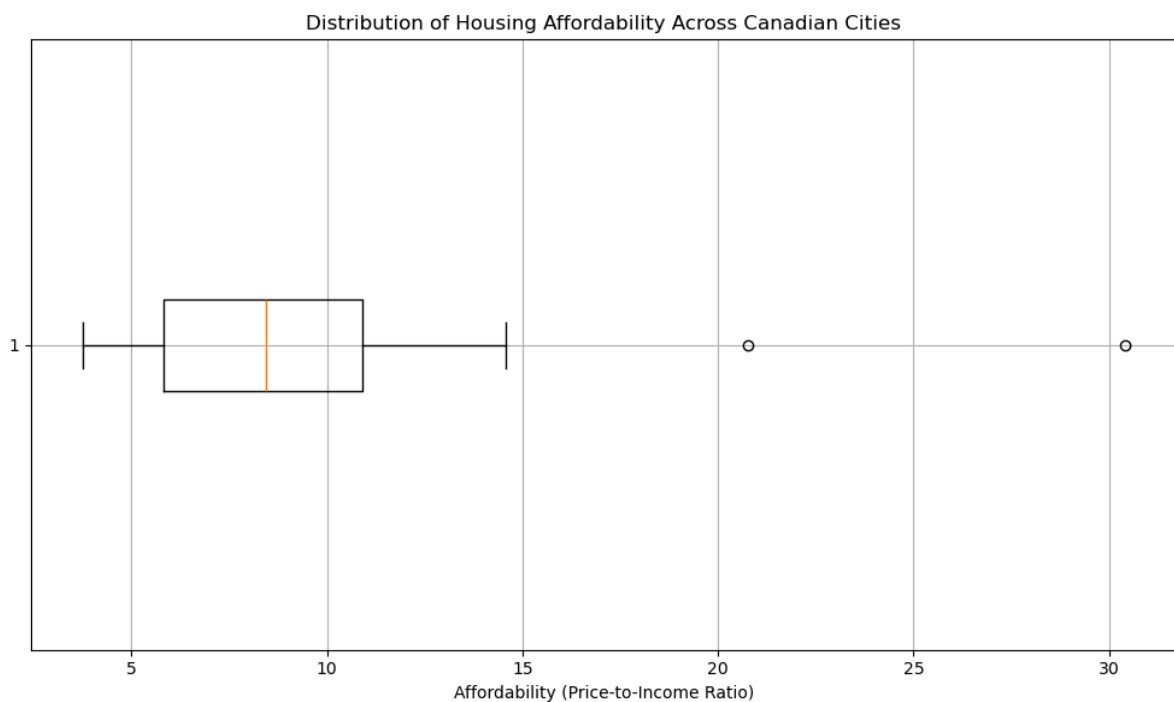
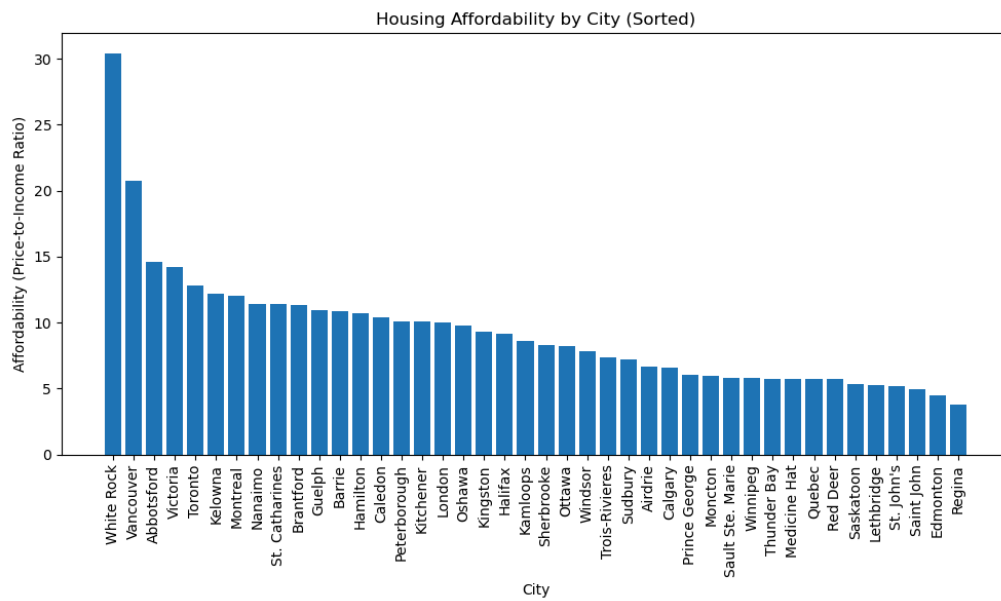
2-median\_income\_city: This data frame was produced by grouping the data by city to extract the median income for each city.

These two data frames allowed me to calculate the affordability ratio, defined as the house price-to-income ratio. I then used this ratio to compare housing affordability across various cities. Cities with higher ratios indicate lower affordability, where house prices are significantly higher than median incomes.

The results show that larger cities like Vancouver and Toronto have much lower affordability due to their higher house prices relative to income, while smaller cities like Regina or Halifax tend to have more affordable housing. This disparity suggests that housing affordability varies greatly across different Canadian cities, often depending on the size and economic activity in each area.

### Conclusion:

The bar chart illustrates that cities like White Rock, Vancouver, and Toronto have significantly higher price-to-income ratios, indicating lower housing affordability, while cities like Regina, Edmonton, and Saint John demonstrate more affordable housing. The box plot provides a broader overview of the distribution of affordability across Canadian cities, highlighting that most cities fall within a moderate affordability range, but a few outliers (such as Vancouver and White Rock) have notably high price-to-income ratios.



### Is there a correlation between city population size and average house price?

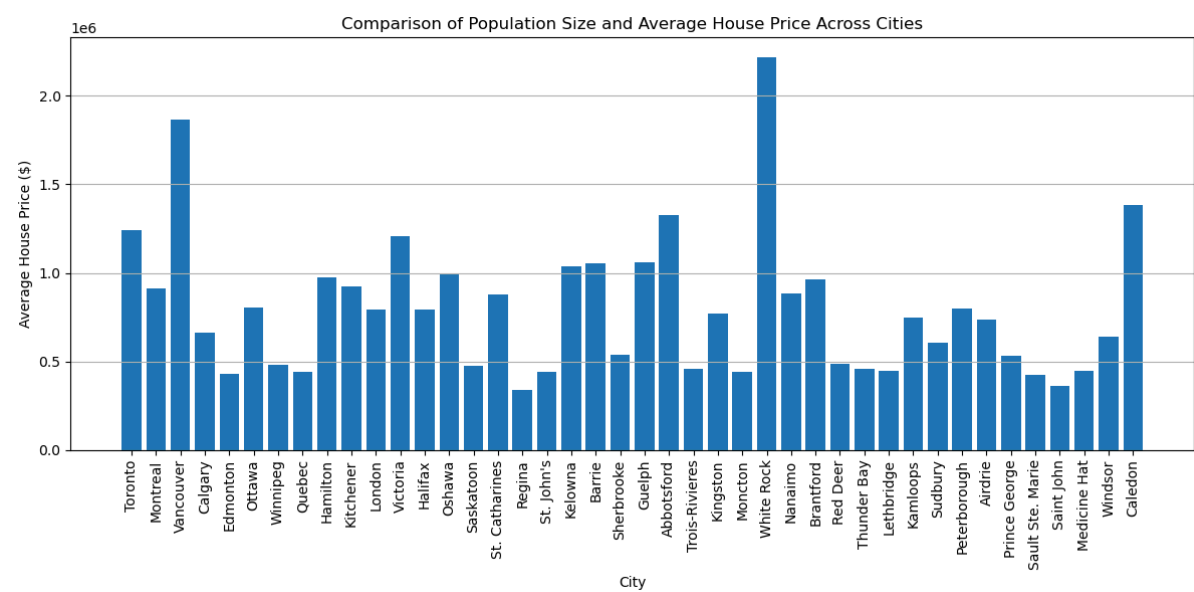
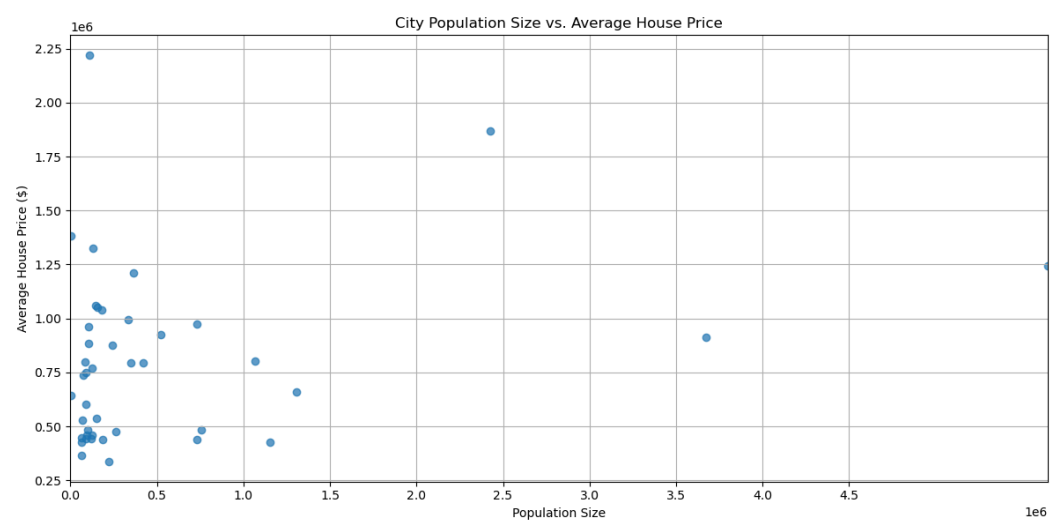
For this question, I analyzed whether there was a correlation between a city's population size and its average house price. This involved creating a new data frame that grouped cities by population size and average house price. The scatter plot created shows the relationship between the two variables to determine whether cities with larger populations tend to have higher house prices.

I grouped the dataset by city to calculate the population size and average house price to achieve this. The scatter plot shows a weak positive correlation between population size and house price, which was further confirmed by calculating the correlation coefficient. The

correlation indicates that larger cities like Toronto and Vancouver tend to have higher average house prices, though this is not always a strong or direct relationship. The correlation analysis shows that while larger cities may have higher housing prices due to higher demand, other factors, such as local economic conditions and real estate market trends, also play a role in determining house prices.

Conclusion:

The scatter plot and bar chart show a weak positive correlation between city population size and average house prices. While larger cities generally tend to have higher house prices, the relationship is not very strong, as some smaller cities also exhibit high house prices due to other influencing factors like local demand and housing supply.



## How does population density (people per square kilometre) influence housing prices and affordability across Canadian cities?

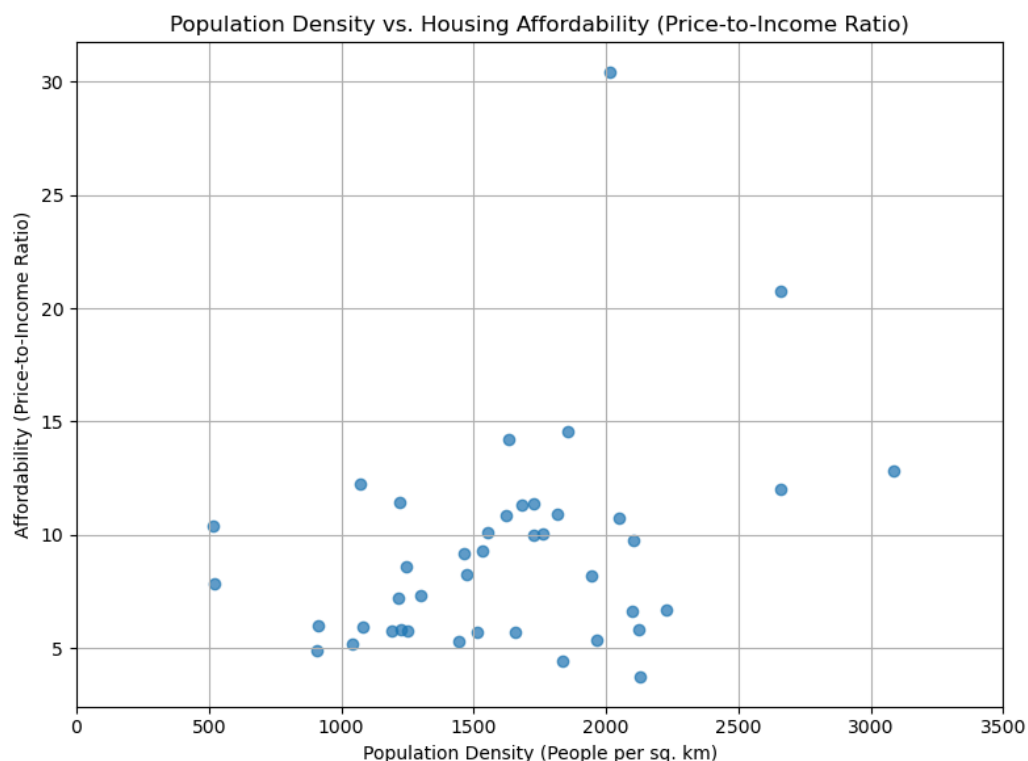
To investigate the effect of population density on housing prices and affordability, I used the population density data (people per square kilometre) from the dataset and compared it to the housing affordability ratio (price-to-income ratio). The goal was to determine if higher population densities are associated with higher house prices and lower affordability.

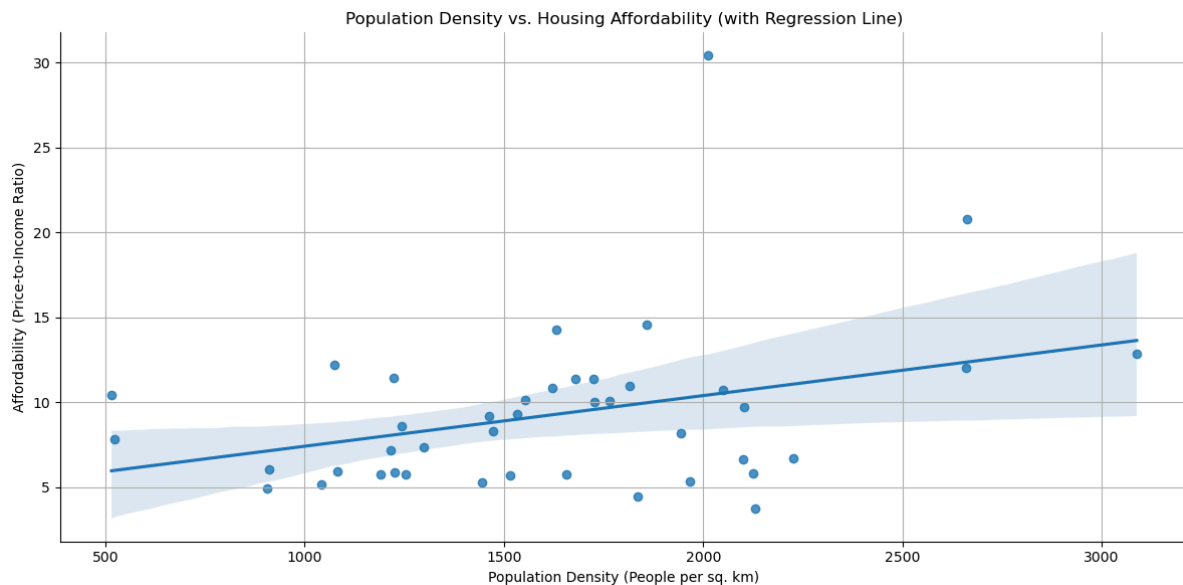
For this analysis, I created a scatter plot showing population density on the x-axis and housing affordability on the y-axis. The analysis shows a moderate positive correlation, indicating that cities with higher population densities generally tend to have lower housing affordability, meaning house prices are much higher than median incomes.

For example, cities like Vancouver and Toronto, which have high population densities, have higher affordability ratios, making housing less affordable for residents. In contrast, cities with lower population densities tend to have lower house prices relative to income, thus making housing more affordable. This relationship between population density and affordability highlights that as cities become more densely populated, housing demand increases, driving up prices and reducing affordability.

### Conclusion:

The scatter plot demonstrates a moderate positive correlation, where cities with higher population densities tend to have higher housing affordability ratios, meaning housing is less affordable in densely populated cities. The regression line in the second plot further confirms this trend, highlighting that denser cities tend to have more expensive housing relative to income.





## Income - Anjila

### Major Findings and Implications

- For the first graph in the Jupyter Notebook, I used these two DataFrames to graph median family income by average house prices to see if there is a correlation.
  - The equation of the regression line,  $y = 9.34x + -13766.22$ , suggests that for every unit increase in median family income, the house price increases by \$9.34, though there is a negative intercept of \$-13,766.22.
  - The R-squared value is 0.082, which means that about 8.2% of the variation in house prices can be explained by the median family income, indicating that the model does not explain most of the variation in house prices.
  - The p-value is 0.059 which means that the relationship between income and house prices is not statistically significant, but it's very close. There's a small chance that the observed relationship could be due to random chance rather than a true effect.
  - There is a **positive relationship between income and house prices** based on the slope of the line, the weak R-squared value and the **borderline p-value suggest that income alone is not a strong or significant factor** in predicting house prices in this data set. Other factors may be influencing house prices more strongly.
- I created two bar graphs, one has the top 10 cities with the highest disparities and the other is the top 10 cities with the lowest disparities. Calculated the mean disparity for the top 10 cities with the highest disparities and the top 10 cities with the lowest disparities.
  - By examining the relationship between housing prices and median income, I implicitly explored correlations between these variables. Although a formal correlation coefficient was not computed, the disparity ratio (price/income) serves as a measure of the relationship.
- Created two histograms of the distribution of houses by house price in Vancouver and Edmonton. It clearly shows the differences in housing availability in different price ranges.

**Vancouver has more expensive houses within minimal affordable housing and Edmonton shows a lot more affordable housing.**

### Answering Research Questions

#### **How does median family income influence housing prices across Canadian cities?**

The analysis demonstrates a clear relationship between housing prices and income through the disparity ratio (house price to income ratio). Cities with higher average housing prices tend to have a much larger disparity, indicating that housing affordability decreases as the housing market becomes more expensive. The average disparity for the top cities has a disparity ratio of 15 and the bottom cities have a disparity ratio of 5. The most expensive cities seem to be in the provinces of Ontario and British Columbia.

Overall income across all the cities seems to be below 100k but the housing prices vary significantly across cities. The variability in house prices are more likely to do with the lack of affordable housing in cities with highest disparity ratio.

#### **Which cities have the greatest disparity between median family income and average housing prices? What is the distribution of house prices within a specific city?**

By calculating the affordability ratio, the top 10 cities with both the highest and lowest disparity ratios were identified. The disparity ratio measures how many times the average house price exceeds the median income. The highest disparity cities exhibit housing prices that are disproportionately higher than median incomes, while the cities with the lower disparity show a little bit more manageable relationship between income and housing prices.

Using the describe function, it was found that the average disparity for all cities in the dataset is 9, and the median disparity for all cities in the dataset is about 7.8. The fact that the median is lower than the mean suggests that there are some cities with very high disparities, meaning a few cities have extremely high disparities compared to the rest.

The highest disparity city is White Rock, BC with house prices being 30 times that of income. This city has a population of around 20,000. I wanted to use a city with a larger population to see the distribution of houses across house prices. The next city with the highest disparity is Vancouver, BC which has a population of 2.4 million. The disparity for Vancouver is 21.

The lowest disparity city is Regina, Saskatchewan with house prices being 4 times that of income. However, Regina has a population of around 200,000; the second lowest disparity city is Edmonton, Alberta, with a population of about 1.2 million. Edmonton, Alberta has a disparity of 4.

I created a histogram of house prices for Vancouver and Edmonton and it can be visually seen that Edmonton's distribution of house prices is a right-skewed distribution with many houses within that are less than 1 million dollars. Whereas Vancouver has no houses of less than 30,000 and many houses that are 2 million or more. The visualisation clearly shows a lack of affordable housing in Vancouver which may contribute to the high disparity.

## Conclusion

The most expensive city is Vancouver with a disparity ratio of 21 and the overall most expensive, not considering the population is White Rock, BC (with a disparity ratio of 30). The most affordable city is Edmonton, Alberta with a disparity ratio of 4. One reason for this difference may be the lack of affordable housing options in Vancouver, as seen on the histogram. Based off this information, it is safe to conclude that income does not impact the prices, the housing supply does.

## Geographical Location - Talal

We started by performing a **statistical summary** to understand key housing metrics across different Canadian provinces. This was done by grouping data by province and calculating:

- **Average housing price:** To see how prices differ by location.
- **Population density:** To check how crowded an area is and its potential effect on prices.
- **Median family income:** To explore if higher income areas have more expensive homes.
- **Total population:** To compare provinces' overall size and their housing markets.

The summary showed that provinces like **British Columbia** and **Alberta** have the highest housing prices, with substantial variation between provinces. This indicated that local factors like population density, economic activity, and income likely influence prices.

To explore the effect of **population density** and **income** on housing prices, we calculated correlation coefficients and created scatterplots:

- **Correlation between population density and housing price:** We found a weak positive correlation (0.11), indicating that areas with higher population density have slightly higher housing prices.
- **Correlation between median income and housing price:** This was also weak (0.05), suggesting that income alone does not strongly predict housing prices.

The conclusion was that while population density and income do affect housing prices, other factors (e.g., geographic location, and housing demand) likely play a larger role in determining prices.

We compared housing prices between **urban** and **rural** areas by creating a box plot and calculating correlation values. This showed:

- **Urban areas** have higher median housing prices than rural areas, and the range of prices is wider in urban settings, indicating more variability (from affordable to luxury homes).
- **Rural areas** have lower prices and a smaller range, suggesting more uniform, affordable housing.

The conclusion is that urban areas tend to have higher prices due to higher demand and diversity in housing types (from high-end properties to smaller, more compact homes).



To identify the cities with the **highest average price per bedroom**, we grouped the data by city, calculated the average price per bedroom, and then ranked the cities to find the top 5. The cities with the highest price per bedroom were:

1. White Rock
2. Vancouver
3. Caledon
4. Abbotsford
5. Toronto

#### **Key observations:**

- **Urban premium:** All the top cities, except Caledon, are large urban areas, where real estate prices are higher, especially for homes with fewer bedrooms.
- **Luxury and space:** Cities like **White Rock** and **Caledon** offer larger homes with more bedrooms and bathrooms, catering to wealthier families, while **Vancouver** and **Toronto** have smaller, more compact (but still expensive) homes due to high demand in dense urban settings.

#### **Conclusion**

Overall, the analysis showed clear trends in the Canadian housing market: urban areas are more expensive, and although population density and income have some impact on housing prices, they are not the sole determining factors. The top cities with the highest price per bedroom reflect the broader trend of urbanization driving up housing costs.

#### **Demographic Data - Amit**

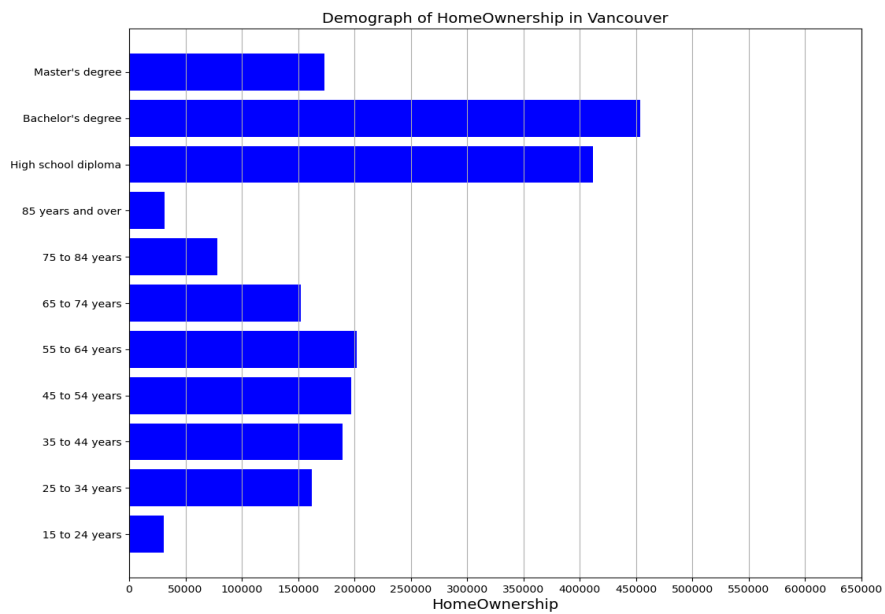
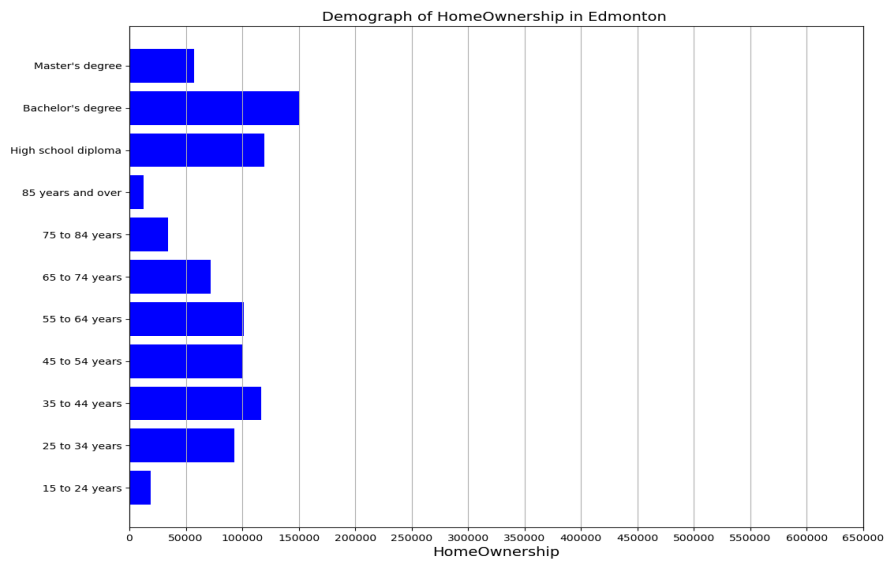
By calculating the Homeownership, the top 5 cities with both the highest and lowest homeownership among which education and age groups were identified. The homeownership disparity cities exhibit housing prices that are disproportionately higher among big cities like Toronto and lower for less populated cities like Edmonton. Age and Education affects homeownership as mostly "High School Diploma" people have more homeownership compared to Bachelor's and Masters degree in most cities in Canada. Homeownership is highest among the age group of 55 to 64 years old, followed by 45 to 54 years old.

#### Answering Research Questions

##### **How do age and education level impact homeownership rates in the top Canadian cities?**

In Edmonton, homeownership is highest among 35 to 44-year-olds. This may indicate that Vancouver has homeowners who are older while Edmonton has younger homeowners. In general, 3 age groups between 35 to 64 seem to have the highest rate of homeownership. For both cities, home ownership seems to be highest among people who have a Bachelor's degree.

The gap between homeownership between a Bachelor's degree and a High School Diploma for Edmonton and Vancouver is less than 50000. Edmonton has a smaller gap in Education level compared to Vancouver



Write-up summarizes major findings and implications at a professional level (5 points)

Each question in the project proposal is answered with precise descriptions and findings (5 points)

Findings are strongly supported with numbers and visualizations (5 points)

Each question response is supported with a well-discerned statistical analysis from lessons (e.g., aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis) (5 points)

## **Final Conclusion for the Analysis of Real Estate Prices and Socioeconomic Factors Across Different Canadian Regions:**

This analysis has provided valuable insights into the factors influencing real estate prices and housing affordability across Canadian cities and regions. The study primarily focused on examining the relationships between population size, population density, median family income, geographical location, and demographic factors such as age and education, all of which contribute to variations in housing prices and homeownership trends.

**Housing Affordability and Population:** The research revealed that affordability varies greatly across Canadian cities. Larger urban centers, such as Vancouver and Toronto, exhibit significantly lower affordability due to higher house prices relative to median incomes. In contrast, smaller cities, like Regina and Halifax, offer more affordable housing. Population density was a stronger predictor of housing prices than population size. Cities with higher population density tend to have less affordable housing, as demand increases with crowding, driving prices upward.

**Income and Housing Prices:** Although there is a positive relationship between median family income and house prices, income alone does not serve as a strong predictor of housing prices. The analysis showed that only a small percentage of the variation in housing prices can be attributed to income, highlighting that other factors, such as population density and geographical location, play more significant roles. High-disparity cities, like Vancouver, showcase the challenges of high housing prices and low affordability, while cities such as Edmonton demonstrate more balanced housing markets.

**Geographical and Urban-Rural Divide:** Geographical location plays an important role in housing prices. Coastal cities and those with favourable climates, such as Vancouver and Toronto, generally have higher prices than inland or colder regions. Urban areas consistently exhibit higher housing prices and more variability in price ranges than rural areas. In rural regions, prices tend to be lower and more uniform, contributing to a more affordable and stable housing market.

**Demographics and Homeownership:** Age and education also have a significant impact on homeownership rates. Middle-aged individuals (35-64 years old) are more likely to own homes, with ownership rates peaking among those aged 55 to 64. Education levels similarly influence homeownership, with bachelor's degree holders showing higher ownership rates than those with only a high school diploma. Urban centers like Vancouver and Edmonton display these trends, with younger homeowners being more prevalent in Edmonton and older, more established homeowners in Vancouver.

In summary, the Canadian housing market is shaped by a complex interplay of factors, with population density, geographical location, and demographic elements such as age and education influencing housing prices and affordability. Urban centers, particularly Vancouver and Toronto, present the greatest challenges in terms of affordability due to high demand and limited space, while smaller cities and rural areas offer more affordable housing options. The disparities observed between income levels and housing prices underline the growing divide in housing affordability across the country, posing challenges for policymakers and potential homeowners.

This analysis provides a comprehensive overview of the Canadian housing market and highlights the need for more nuanced approaches to housing policy, considering the diverse economic and social factors that affect affordability in different regions.

## **Data sets used in this Analysis**

Housing Listings Data:

<https://www.kaggle.com/datasets/jeremylarcher/canadian-house-prices-for-top-cities/data>

Population Data:

<https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810001101&geocode=A000011124>

Age and Education Homeownership Data:

<https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=Edmonton&DGUIDlist=2021A00053520005,2021A00054806016,2021A00053506008,2021S0503933,2021S0503835&GENDERlist=1&STATISTIClist=1&HEADERlist=37,36,20>