**1**. **Project Overview**

The focus of this project is to analyze the impact of immigration on housing markets in Canada using machine learning techniques. To facilitate this, we have collected and processed several datasets that provide key insights into immigration trends, housing price indices, and housing completion trends.

**2.** **Datasets Used**

Three primary datasets were sourced from official data portals:

Canada Immigration Data: Monthly data on immigration counts by region.

<https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2410000501&pickMembers%5B0%5D=1.5&cubeTimeFrame.startMonth=01&cubeTimeFrame.startYear=1991&cubeTimeFrame.endMonth=12&cubeTimeFrame.endYear=2021&referencePeriods=19910101%2C20211201>

New Housing Price Index: Monthly data on housing price changes.

<https://open.canada.ca/data/en/dataset/324befd1-893b-42e6-bece-6d30af3dd9f1/resource/0d2c0c49-21d1-4291-8a11-8a9cfc7e711e>

Housing Completion Data: Quarterly data on completed housing units.

<https://open.canada.ca/data/en/dataset/440be770-8816-4956-9a9d-8c4c97e35242/resource/f953407b-894f-41c3-b21d-864137095df5>

Population estimates, quarterly

<https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901>

**3**. **Data Preprocessing and Merging**

Downloaded each immigration dataset (1992 till 2021) by region and merged them together.

Converted quarterly housing completions data to monthly intervals.

Converted quarterly population estimate data to monthly intervals.

Merged four datasets (immigration, housing prices, housing completions, Population estimates) by REF\_DATE and PROVINCE (GEO).

Ensured the date format was standardized to datetime.

Handled missing values and conducted initial data cleaning to remove any duplicates.

Data Types:

REF**\_**DATE: datetime format

GEO: categorical data

ImmigrationCount: integer data

HouseCount: integer data

HousePriceIndex: floating point data representing the house price index

PopulationCount: integer data

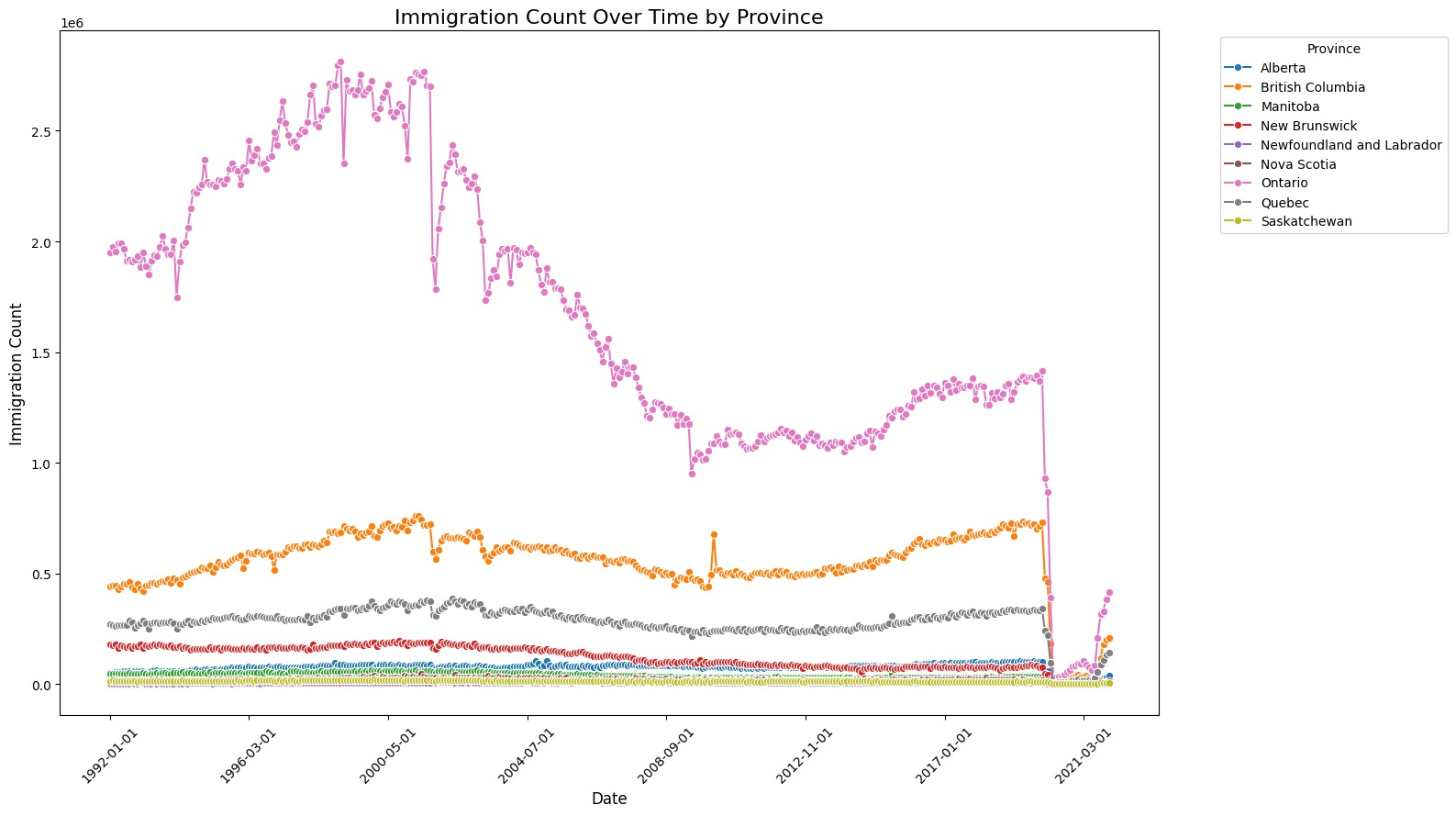
**4**. **Exploratory Data Analysis (EDA)**

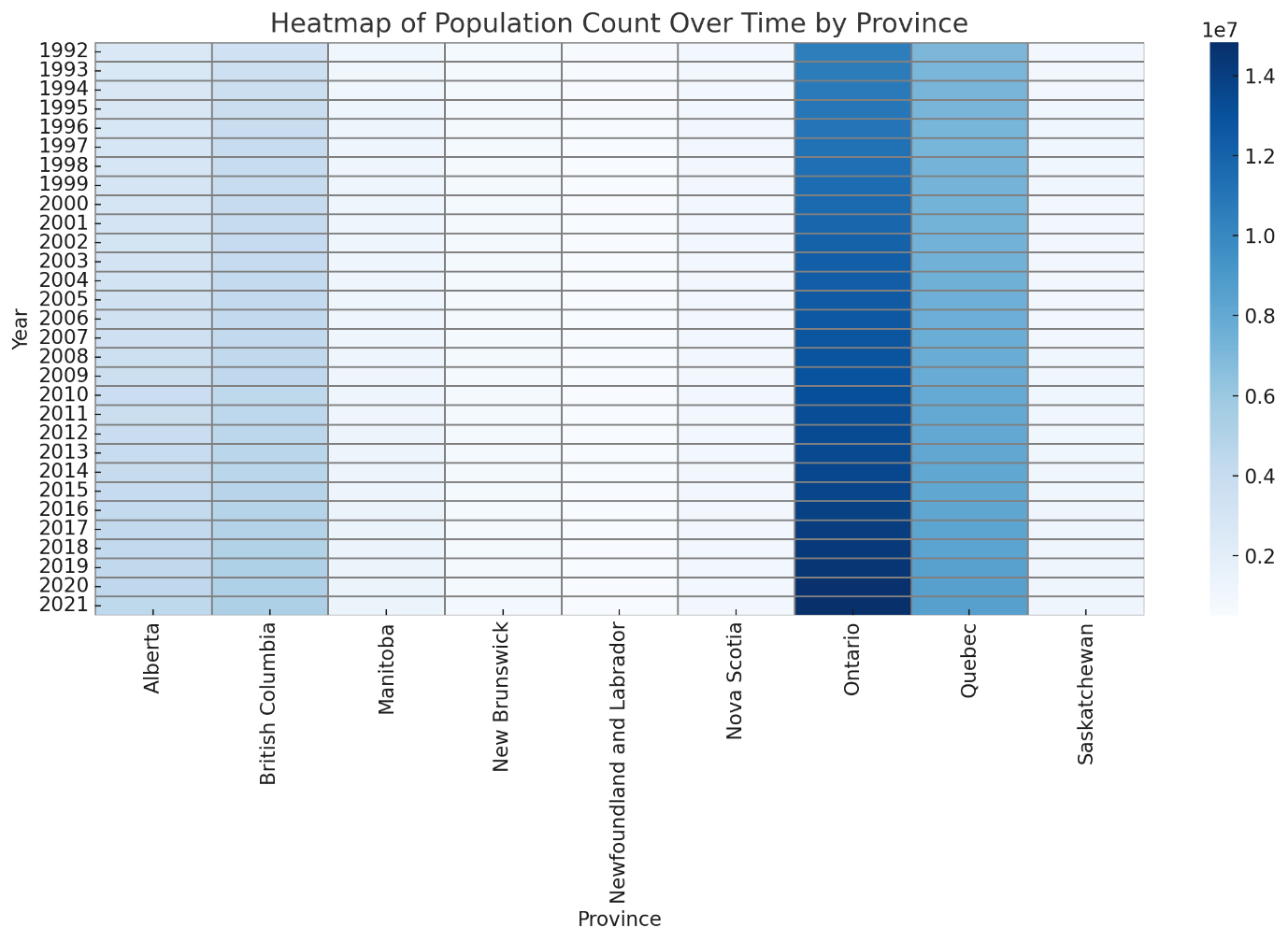
EDA was conducted on the combined dataset to gain initial insights:

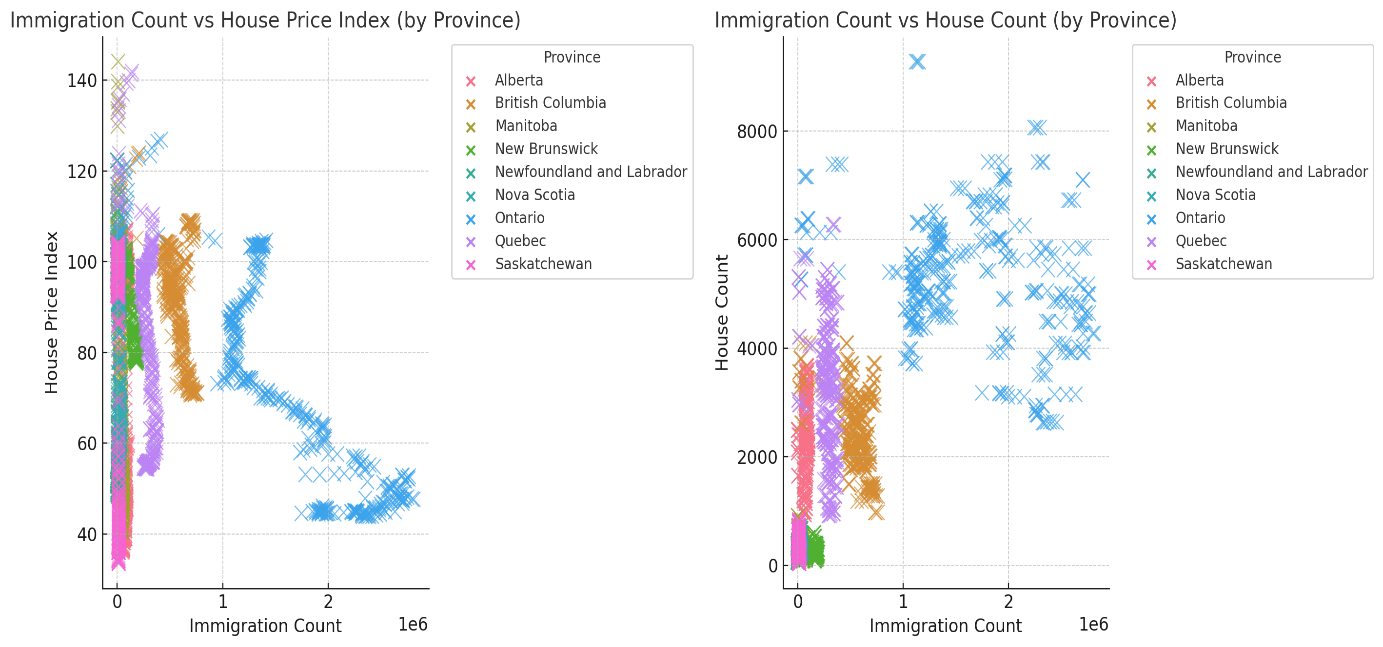
Descriptive Statistics:

We explored basic statistics such as mean, median, and standard deviations for key variables like immigration counts, housing counts, housing price index, and population count across provinces.

Visualization:



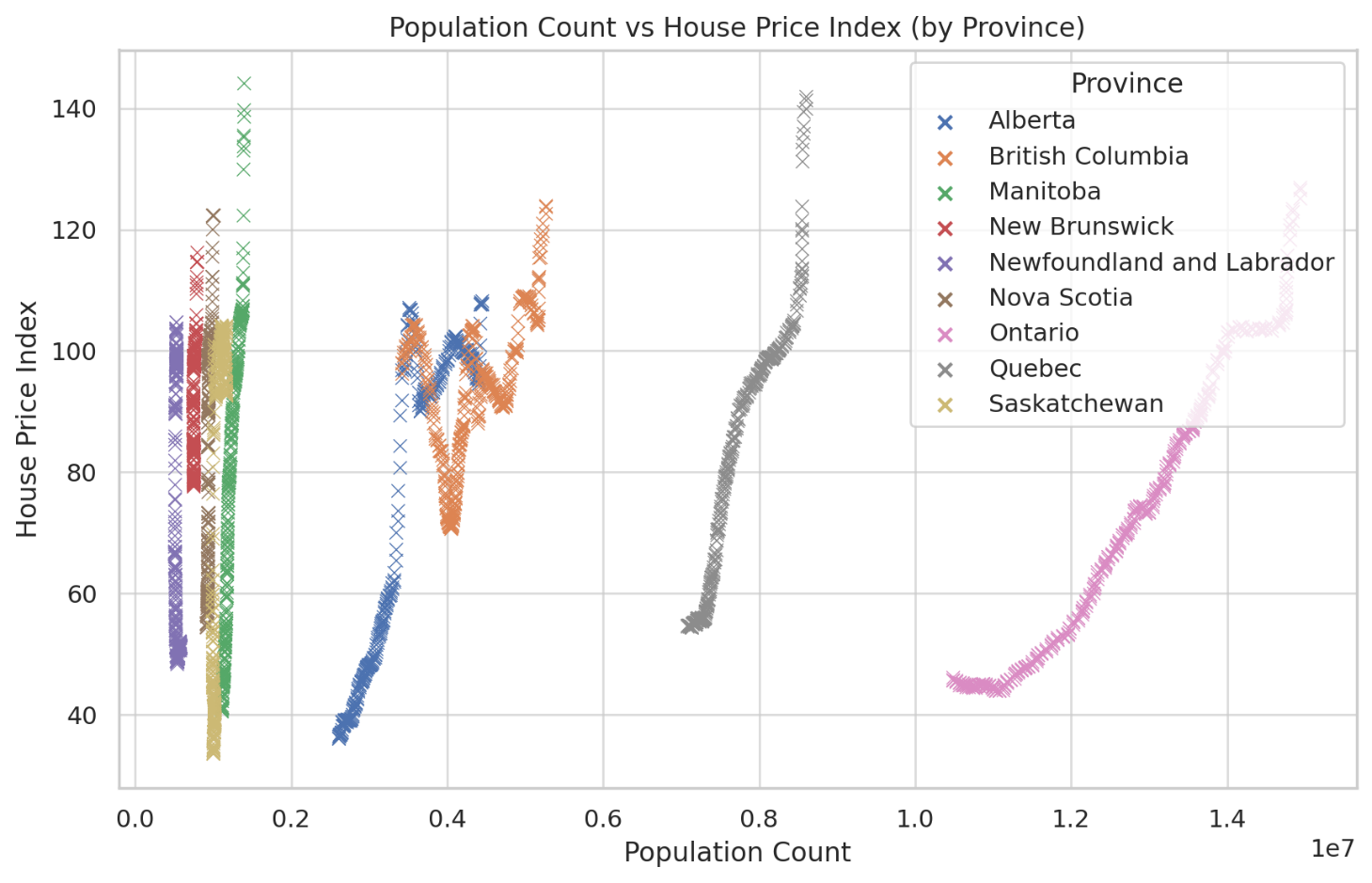




1)The plot shows no clear or strong pattern between immigration count and house prices, with data points scattered across the plot. The correlation between **immigration count** and **house price index is** indicating a **weak negative relationship**. This suggests that as immigration count increases, house prices may slightly decrease, but the relationship is too weak to be considered impactful.

2)**strong positive relationship**. This means that as immigration increases, the number of houses also tends to increase.

There seems to be a positive relationship, with house counts increasing as immigration counts rise. This indicates that housing supply adjusts in response to increasing immigration, likely due to rising demand for housing.



It suggests a weak positive relationship, as the points are quite scattered without a strong trend.

Provinces with larger populations, such as **Ontario** and **British Columbia**, show sharp increases in house prices over time.

**Smaller provinces**, such as **Newfoundland and Labrador**, **Nova Scotia**, and **New Brunswick**, show slower and more stable population growth, which aligns with more stable house prices in these regions.

The results

show that immigration and population, on average, is not the main factor driving price increases.

**Software Tools Used:**

* **Python** – Used for data cleaning, merging datasets, and conducting exploratory data analysis.
* **Pandas & NumPy** – Utilized for handling and manipulating large datasets, including immigration and housing data.
* **Matplotlib & Seaborn** – Employed to visualize the relationship between immigration, housing prices, and housing starts in Canada.
* **VS Code & Google Colab** – Used as the primary integrated development environment (IDE) for coding, debugging, and organizing project files.

**Challenges Encountered:**

* Difficulty in finding comprehensive and reliable datasets specifically related to the rental market.
* Identifying direct relation between immigration trends and housing market changes, as there are several external factors involved.