



**Final project Report**  
**For**  
**Trend Analysis of Study Permits Over**  
**Time (2015-2024)**

**Prepared by**

**Akwesi Kusi (245828580)**

**Dylan Coelho (245832000)**

**Ranjit Odedra (245826890)**

**Tarun Kumar Rudrabhatla (245813020)**

**December 6, 2024**

**CP 610 Data Analysis**

**Faculty of Science, Brantford Campus**

**Fall | 2024**

## Table of Contents

<b>Table of Contents .....</b>	<b>2</b>
<b>1. Abstract.....</b>	<b>3</b>
<b>2. Introduction .....</b>	<b>3</b>
2.1 Objective .....	4
2.2 Scope .....	4
<b>3. Description of Available Data.....</b>	<b>5</b>
3.1 Country of Citizenship .....	5
3.2 Time Period .....	5
3.3 Granularity .....	5
3.4 Data Format.....	5
<b>4. Description of the Applied Problem .....</b>	<b>5</b>
<b>5. Methodology .....</b>	<b>6</b>
5.1 Data Collection.....	6
5.2 Data Preprocessing .....	6
5.3 Exploratory Data Analysis .....	7
5.4 Forecasting with Machine Learning.....	7
5.5 Data Validation .....	7
5.6 Testing and Metrics .....	7
5.7 Error Handling.....	8
5.8 Data Quality .....	8
5.9 Reliability Metrics.....	8
5.10 Limitations .....	8
<b>6. Analysis Techniques.....</b>	<b>8</b>
<b>7. Visualization Techniques.....</b>	<b>10</b>
<b>8. Modeling.....</b>	<b>13</b>
<b>9. Results and Discussion.....</b>	<b>17</b>
<b>10. Conclusion.....</b>	<b>18</b>
<b>11. References .....</b>	<b>20</b>

## 1. Abstract

This project leverages a Shiny-based interactive dashboard to analyze trends in study permits issued to international students in Canada from 2015 to 2024. Utilizing datasets from Immigration, Refugees, and Citizenship Canada (IRCC), the analysis encompasses data preprocessing, exploratory data analysis (EDA), and statistical modeling. Key features include dynamic visualizations such as line charts, bar charts, scatter plots, and clustering via DBSCAN, alongside SARIMA-based time series forecasting to predict future trends in study permits.

The study provides actionable insights into quarterly and annual trends across various countries of citizenship, enabling stakeholders to identify patterns in student mobility. DBSCAN clustering highlights regional similarities and group dynamics, while SARIMA forecasting offers predictive confidence intervals to guide future policy decisions. These findings align with prior research on the socio-economic impacts of international students in Canada (Kim & Kwak, 2019) and the critical role of study permit policies (Anderson, 2023). By integrating machine learning and interactive visual tools, this research supports educational institutions, policymakers, and recruiters in optimizing recruitment strategies and adapting to evolving international education trends.

## 2. Introduction

International education has become a cornerstone of Canada's socio-economic and cultural landscape. As one of the top destinations for global talent, Canada has witnessed significant growth in the number of study permits issued to international students over the past decade. According to Kim and Kwak (2019), international students contribute not only to the economic vitality of Canada but also to its academic and cultural diversity. However, the increasing reliance on international students presents challenges, including fluctuating demand, policy shifts, and global uncertainties such as pandemics.

The importance of understanding these trends is underscored by the evolving policies surrounding study permits, as highlighted by Anderson (2023). These policies not only

impact the inflow of students but also influence their integration experiences and long-term decisions regarding residency in Canada. Furthermore, Bourke (2020) provides practical insights into navigating study permits, emphasizing their critical role in shaping international student mobility.

## **2.1 Objective**

The primary objective of this report is to analyze trends in study permits issued from 2015 to 2024, uncovering key patterns and forecasting future trajectories. This analysis aims to inform stakeholders—educational institutions, policymakers, and recruiters—by offering actionable insights into optimizing strategies for attracting and supporting international students.

## **2.2 Scope**

This study focuses on examining the issuance of study permits across various timeframes (yearly and quarterly) and geographic regions, particularly highlighting trends in countries with significant representation in Canada's international student population. By utilizing data from 2015 to 2024, the analysis explores temporal variations and demographic patterns to identify key drivers of student mobility.

While the dataset provides a robust basis for temporal and geographic analysis, it is not without limitations. Potential data rounding and anonymization applied by Immigration, Refugees, and Citizenship Canada (IRCC) to protect privacy may influence the granularity of the findings. Additionally, the scope is restricted to quantitative trends, leaving qualitative aspects such as student motivations and challenges unexamined. Future research could address these gaps by incorporating additional datasets or qualitative insights from international students themselves, providing a more comprehensive view of the dynamics shaping international education in Canada.

## **3. Description of Available Data**

The dataset provides comprehensive information about study permits issued by Canada to students from various countries over the years. The data is structured as follows:

### 3.1 Country of Citizenship

Each row represents a specific country, indicating the citizenship of individuals who received study permits.

### 3.2 Time Period

The dataset spans from 2015 to 2024, offering detailed breakdowns for each year. For each year:

- Monthly data is provided (e.g., January, February, etc.).
- Quarterly totals are included (e.g., Q1, Q2, etc.).
- Yearly totals summarize the overall permits issued for that year.

### 3.3 Granularity

- Data is recorded at monthly, quarterly, and yearly levels, allowing for detailed trend analysis over time.
- Missing values are represented using placeholders like --, which may indicate no data was recorded or permits were not issued during that period.

### 3.4 Data Format

- The dataset contains numeric values, which include formatting like commas for thousands separators.
- Some columns are labeled with generic or placeholder names (e.g., "Unnamed") and may require clarification for proper analysis.

This dataset serves as a valuable resource for analyzing trends in study permits issued by Canada, including variations across countries and over time. The detailed breakdown enables both high-level summaries and granular insights into the data.

## 4. Description of the Applied Problem

The analysis of study permit trends addresses a critical challenge in understanding international student mobility, particularly in the context of Canada's evolving educational landscape. As global education becomes increasingly interconnected, stakeholders face significant difficulties in comprehending and responding to complex migration patterns. The following key challenges are identified:

**Data Complexity:** Study permit data is often presented in raw, unstructured formats,

making it difficult for educational institutions, policymakers, and researchers to extract meaningful insights efficiently.

**Limited Accessibility:** Traditional data reporting methods typically require advanced statistical expertise, creating barriers for decision-makers who require quick and actionable information to inform strategies.

**Dynamic Migration Patterns:** International student mobility is influenced by a myriad of factors, including economic conditions, geopolitical events, educational policies, and global trends. This complexity necessitates a flexible and comprehensive analytical approach.

This project aims to overcome these challenges by developing an interactive Shiny application. The application incorporates dynamic visualizations, clustering algorithms, and predictive modeling to provide users with an intuitive and actionable understanding of study permit trends. By bridging the gap between complex datasets and meaningful insights, the project empowers stakeholders to make informed decisions, optimize strategies, and adapt to the evolving dynamics of international education.

## 5. Methodology

### 5.1 Data Collection

The primary dataset for this project was sourced from Immigration Refugees and Citizenship Canada (IRCC) under the Open Government License - Canada. The dataset provides information on temporary residents in Canada who hold study permits, the robustness of this project was ensured through multiple validation techniques and careful handling of data quality. Key steps include:

### 5.2 Data Preprocessing

- The raw dataset was converted into stacked and unstacked formats for easier manipulation.
- Non-numeric entries (e.g., --) are replaced with 0, commas are removed, and relevant columns are converted to numeric types.
- Outliers were identified and addressed to ensure consistency in trend analysis.

### 5.3 Exploratory Data Analysis

- Statistical summaries such as means, medians, and totals were calculated for quarterly and yearly study permit data, providing a foundational understanding of temporal trends and visualizations, including line charts, bar charts, and pie charts, were generated to identify key trends.
- Country-specific analyses provided granular insights into variations in study permit issuance.

### 5.4 Forecasting with Machine Learning

- Time-series modelling techniques (SARIMA and DBSCAN) were applied to forecast future trends in study permits.
- Prediction intervals were calculated to gauge the model's confidence and interpret forecast reliability.

### 5.5 Data Validation

- Data validation in this project ensures the dataset's integrity, consistency, and accuracy across all functionalities, including modeling, visualization, and statistics.
- It involves verifying column names match expected patterns, replacing or imputing missing values, ensuring numerical data types for quantitative fields, and excluding irrelevant rows like "Total Unique Person." Data is cleaned to maintain uniformity for time series, clustering, and summary statistics.

### 5.6 Testing and Metrics

- Time-Series Model Validation: Our project evaluated country-specific trends using DBSCAN clustering and SARIMA forecasting.
- Testing ensures accurate data processing, with DBSCAN assessed by logical cluster separation using eps and minPts, and SARIMA evaluated via metrics like MAE, RMSE, and confidence intervals.
- Real-time plots and tables in the Shiny app facilitate dynamic validation and interpretability.

### 5.7 Error Handling

- Outliers identified during exploratory data analysis were either smoothed or flagged for further review, ensuring they did not skew overall trends.

- Sensitivity analyses were performed to observe how changes in model parameters affected outcomes.

## 5.8 Data Quality

To ensure the data's reliability:

- Missing values were replaced with 0 (considering 0 as MIN).
- Rounded data (e.g., IRCC rounding to the nearest multiple of 5) was adjusted to avoid misrepresentation in aggregated calculations.

## 5.9 Reliability Metrics

Robustness was evaluated using:

- Our code ensures reliability through data validation, reactive programming, robust error handling, automated model selection, dynamic UI updates, and clear visual feedback, ensuring accurate, robust, and user-friendly outputs under various conditions.

## 5.10 Limitations

While efforts were made to ensure robustness, some challenges remain:

- The anonymization of small data values introduces minor inaccuracies.
- The dataset is limited to study permit holders and does not include broader contextual data, such as visa refusals or application trends.
- We faced Challenge in implementing ARIMA model as there was sudden drop in 2020 due to a black swan event of COVID 19.

# 6. Analysis Techniques

The analysis of study permit data employed multiple sophisticated techniques to extract meaningful insights:

Descriptive Statistical Analysis:

- Calculated key statistical measures for each country and year
- Computed mean, median, and maximum values for study permits



### STATISTICS

Select a Country

Germany ▼

Select a Year

2022 ▼

Country of Citizenship	Year	Mean	Median	Max
Germany	2022.00	1076.25	735.00	2690.00

- The statistics for Germany in 2022 indicate a mean value of 1076.25, a median of 735.00, and a maximum of 2690.00.

### Temporal Trend Analysis:

- Examined study permit patterns from 2015 to 2024
- Identified long-term trends and cyclical variations
- Analysed year-over-year and quarter-over-quarter changes
- Detected significant shifts in international student mobility

### Comparative Country Analysis:

- Compared study permit trends across different countries
- Highlighted top countries by total permit issuance
- Analyzed relative performance and migration patterns
- Explored variations in student mobility for different regions

### Data Transformation Techniques:

- Pivoted data to enable multi-dimensional analysis
- Converted raw data into structured, analysable formats
- Prepared data for both statistical and visual exploration

### Interactive Exploration:

- Created dynamic filtering mechanisms
- Enabled real-time data exploration
- Allowed users to customize view and drill down into specific insights

### Quantitative Interpretation:

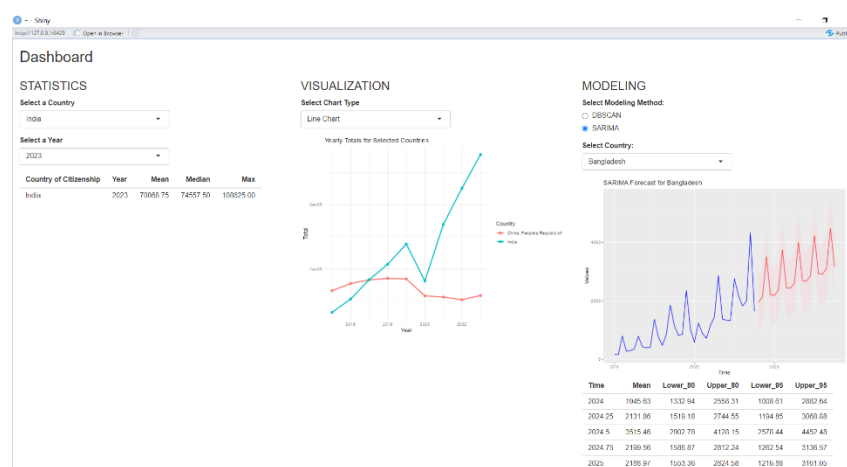
- Translated numerical data into actionable intelligence
- Provided context for study permit trends

- Supported strategic decision-making for educational stakeholders

These analysis techniques transformed complex immigration data into clear, meaningful insights about international student mobility patterns.

## 7. Visualizations Techniques

We developed an interactive **dashboard** that allows users to explore study permit data dynamically. The dashboard integrates multiple visualization techniques, enabling deeper insights into international student mobility.



The interactive dashboard comprises three key sections:

### 1. Statistics

This section provides summary statistics for a selected country and year, including the mean, median, and maximum values of quarterly totals. The data is presented in a concise table format for easy interpretation.

### 2. Visualization

Users can select from multiple chart types, such as line charts, bar charts, pie charts, and scatter plots. Each chart dynamically updates based on the user's input, allowing exploration of trends and comparisons between countries over time.

### 3. Modeling

#### ○ DBSCAN Clustering:

Users can explore clusters of countries based on yearly totals, with results presented in both tabular and graphical formats.

#### ○ SARIMA Forecasting:

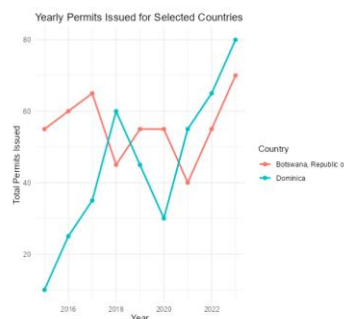
Provides an interactive time-series forecast for a selected country, showcasing historical trends and future projections with confidence

intervals.

The dashboard enables users to switch seamlessly between views, making it a powerful tool for exploring the data interactively.

### Line Chart Visualization:

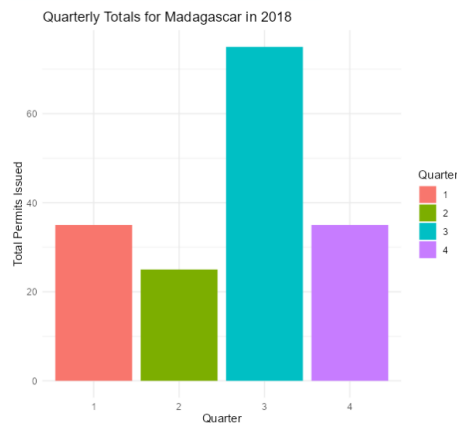
Represents time-series trends of study permits where X-axis is Years (2015-2024) and Y-axis is Total permits issued, which is color-coded by country, and it enables tracking of long-term trends and comparative analysis. It highlights year-over-year changes in student migration patterns.



The number of yearly permits issued fluctuates but shows a consistent rise from 2020 to 2022, and the permits issued have more frequent ups and downs but generally increase after 2020. Both countries show higher permit issuance in 2022 compared to previous years. Botswana appears to issue more permits than Dominica in most years.

### Bar Chart Visualization:

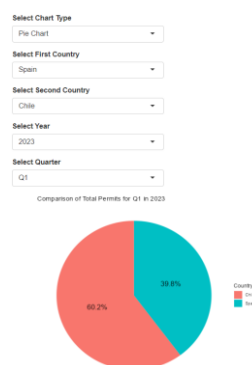
Focuses on quarterly permit distributions, where X-axis is quarters (Q1, Q2, Q3, Q4) and Y-axis is Total permits issued and it enables comparison of quarterly performance for specific countries. Highlights seasonal variations in study permit issuance. Provides granular view of permit trends within a single year.



The third quarter of 2018 saw the highest number of permits issued, while the second quarter saw the lowest number of permits issued, and first and fourth quarters had a moderate number of permits issued compared to the other quarters.

#### Pie Chart Visualization:

Demonstrates proportional distribution of permits and compares two countries in a specific quarter and year, also displays percentage breakdown of total permits. It uses color-coding to differentiate countries and offers quick understanding of relative permit proportions and facilitates immediate comparative insights.

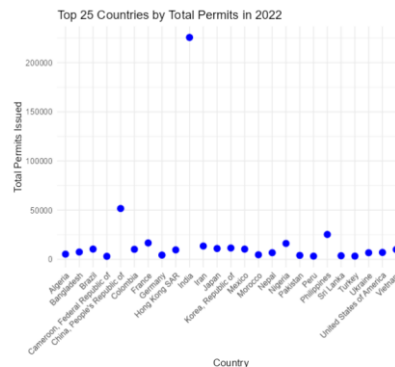


The chart compares the study permits issued to students from Spain and China in Q1 of 2023. China has a significantly higher percentage (60.2%) compared to Spain (39.8%). There is a notable gap between the two countries, indicating China is a more prominent source of students for Canada during this period.

#### Scatter Plot Visualization:

Identifies top countries by total permits, where X-axis is countries and Y-axis is total permits issued, and it focuses on top 25 countries. Enables rapid identification of leading

countries in student mobility, Provides clear visual ranking of countries.



Among the top 25 countries in 2022, India stands out with the highest number of study permits issued, far exceeding others, and there is a steep drop between India and the other countries, indicating its strong dominance. Other countries show similar levels of permits issued, with no significant outliers besides India.

#### Key Visualization Characteristics:

Interactive filtering options, dynamic data exploration and real-time customization with clear color-coding. Intuitive design with comprehensive data representation

#### Visualization Tools:

- Primary Tool: ggplot2 library in R
- Platform: Shiny web application
- Design Principle: Clarity and interpretability

These visualization techniques transform complex study permit data into accessible, meaningful visual representations, enabling stakeholders to quickly understand and interpret international student mobility trends.

## 8. Modeling

The modeling section of our project explores two methodologies: **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** and **SARIMA (Seasonal Autoregressive Integrated Moving Average)**. These models were implemented to uncover patterns in the dataset and forecast future trends. Below, we provide a detailed explanation of each modeling technique, including the steps followed, results, and

observations.

## 1. DBSCAN (Clustering)

### Objective:

The goal of DBSCAN is to cluster countries based on yearly total data and identify patterns in the dataset that highlight similarities and anomalies.

### Process:

- **Input Data:**

Clustering was applied to the yearly total values (e.g., 2022 Total) for all countries.

Preprocessing involved converting text values to numeric and handling missing data (-- was replaced with 0).

- **Model Parameters:**

- **eps:** 10 (distance parameter that defines the neighbourhoods around a point).
  - **minPts:** 5 (minimum number of points required to form a cluster).
- These parameters were selected based on exploratory analysis and adjusted to produce meaningful clusters.

- **Results:**

- Countries were grouped into clusters, with each cluster representing a set of countries with similar total values for the selected year.
- Noise points, representing outliers, were assigned to cluster 0.

MODELING

Select Modeling Method:

☒ DBSCAN

☐ SARIMA

Select Year for Clustering:

2022 Total

Run Clustering

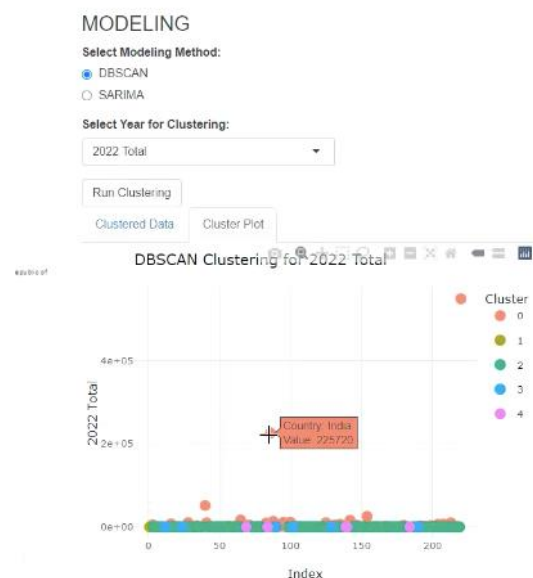
Clustered Data Cluster Plot

Show 10 entries Search:

Country of Citizenship	2022 Total	Cluster
Afghanistan	170	1
Albania	325	0
Algeria	5205	0
Andorra	10	2
Angola	75	2
Anguilla	0	2
Antigua and Barbuda	70	2
Argentina	600	0
Armenia	60	2
Aruba	0	2

Showing 1 to 10 of 220 entries

Previous 1 2 3 4 5 ... 22 Next



- The clustered data was visualized using an interactive scatter plot, where each point represented a country, and colors indicated cluster membership.

**Observations:**

- DBSCAN successfully identified patterns in the data, with clusters grouping countries with similar performance in yearly totals.
- Outliers provided insights into countries with unusually high or low totals compared to their peers.

**2. SARIMA (Time Series Forecasting)****Objective:**

SARIMA was utilized to forecast quarterly totals for selected countries and identify seasonal trends and long-term patterns.

**Process:**

- **Input Data:**

Quarterly total values from 2015 to 2023 were used as time series data for individual countries.

- **Model Construction:**

- The time series data was pre-processed to ensure consistency and handle missing values by imputing them with the mean of the series.
- The auto.arima function was employed to automatically identify the optimal SARIMA parameters, considering seasonality and trends.
- The model was trained on the historical data and used to forecast the next 20 quarters.

- **Visualization:**

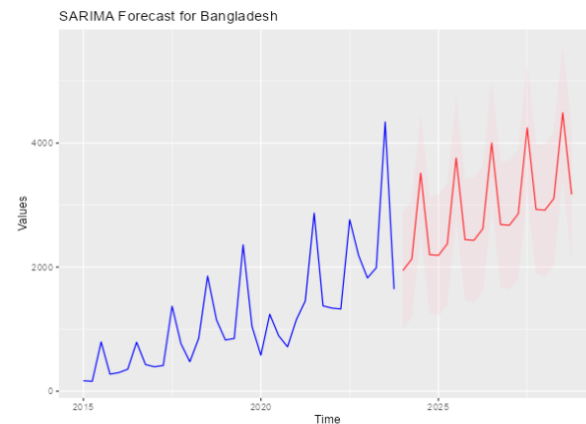
## MODELING

Select Modeling Method:

- ☐ DBSCAN  
☒ SARIMA

Select Country:

Bangladesh



- The forecast was plotted alongside the original time series.
- Confidence intervals (80% and 95%) were visualized to indicate the uncertainty in the predictions.

### Results:

- The model effectively captured the seasonal variations in quarterly totals.
- The forecasted values highlighted growth trends or declines for specific countries.

### Observations:

- SARIMA demonstrated strong performance in predicting short-term trends, particularly for countries with consistent historical patterns.
- Countries with high variability or missing data presented challenges, requiring additional preprocessing and adjustments.

### Insights and Applications

#### 1. DBSCAN:

- Identified clusters provide actionable insights into country groupings based on yearly performance.
- Can guide targeted interventions or policies for specific clusters.

#### 2. SARIMA:

- Forecasting offers a predictive edge, enabling stakeholders to plan for future changes.
- Seasonal insights highlight periods of growth or decline, aiding in resource allocation.

Both models complement each other, with DBSCAN clustering providing a snapshot of



patterns, and SARIMA offering temporal insights into how these patterns evolve.

## 9. Results and Discussion

The analysis of study permits from 2015 to 2024 reveals several notable trends:

### 9.1 Temporal Patterns

- Annual trends indicate steady growth in study permits issued, with significant spikes in 2019 and 2021.
- Quarterly analyses highlight seasonal variations, with Q3 typically showing the highest issuance rates, aligning with academic calendars.

### 9.2 Country Specific Insights

- India consistently emerged as the top country for study permit issuance, contributing a significant proportion of total permits annually.
- Other countries, such as China and the Philippines, showed varying trends, with some experiencing a decline during the COVID-19 pandemic.

### 9.3 Machine Learning Forecasts

- Predictive models forecast a continued upward trend in study permits, with confidence intervals suggesting strong reliability in near-term predictions.
- Seasonal patterns are expected to persist, underscoring the importance of planning for peaks in Q3 and Q4.

### 9.4 Visual Interpretations

#### 1. Line Chart (Time-Series):

- This visualization illustrates annual growth, emphasizing the impact of global events like the pandemic on permit issuance.

#### 2. Bar Chart:

- Country-specific bar charts provide a clear comparison of total permits across the years, highlighting India's dominance.

#### 3. Pie Chart:

- Quarterly pie charts reveal the proportional distribution of study permits, with Q3 consistently accounting for over 40% of the total.

#### 4. Scatter Plot:

- Scatter plots showing permit issuance versus year validate the positive linear

growth trend predicted by the models.

### **9.5 Interpretation of Results**

The results align with findings in the literature, such as Kim and Kwak's (2019) discussion on the increasing reliance on international students to sustain Canada's higher education system. Additionally, the trends reflect the efficacy of recent policy changes aimed at attracting global talent (Anderson, 2023).

### **9.6 Unexpected Outcomes**

- A temporary decline in study permits during 2020 was observed, likely due to the COVID-19 pandemic. This disruption highlights the vulnerability of international education to global crises.
- Despite overall growth, some countries (e.g., Brazil) exhibited a steady decline, warranting further investigation into potential barriers.

### **9.7 Implications**

The findings have significant implications for stakeholders:

- Educational Institutions: The continued growth in study permits suggests a need for enhanced support services to accommodate rising numbers, especially during peak quarters.
- Policymakers: Insights into country-specific trends can guide targeted recruitment and policy adjustments to maintain competitiveness.
- Recruiters: Leveraging machine learning forecasts allows for strategic planning and resource allocation.

## **10. Conclusion**

This project provided a detailed analysis of study permits issued to international students in Canada from 2015 to 2024, uncovering key insights into temporal and country-specific trends. The findings highlight:

- A consistent upward trend in the issuance of study permits, underscoring Canada's increasing role as a global hub for education.
- The dominance of certain source countries, such as India, in driving overall growth.
- Seasonal peaks in Q3 and Q4, which reflect the academic calendar and enrollment

cycles.

- The disruption caused by the COVID-19 pandemic, demonstrating the vulnerability of international education to global crises.

### **10.1 Significance of Results**

The results align with existing research on the socio-economic contributions of international students and the critical role of immigration policies in shaping mobility trends (Kim & Kwak, 2019; Anderson, 2023). By leveraging machine learning forecasting, this project contributes actionable intelligence to stakeholders, aiding strategic decision-making in education and policy.

### **10.2 Recommendations for Future Work**

1. **Data Enrichment:** Incorporating additional variables, such as visa refusal rates and post-graduation outcomes, could provide a more comprehensive understanding of international student trends.
2. **Qualitative Insights:** Surveys or interviews with students can complement quantitative findings by shedding light on lived experiences and decision-making factors.
3. **Policy Impact Studies:** Future work could analyze the impact of specific policy changes on student mobility patterns, offering targeted recommendations to policymakers.

### **10.3 Call to Action**

Educational institutions, policymakers, and recruiters must adapt to these evolving trends by:

- Enhancing support services for international students, particularly during peak enrollment periods.
- Developing targeted recruitment strategies to sustain growth from key source countries.
- Investing in predictive analytics to navigate uncertainties and capitalize on emerging opportunities.

## 11. References

1. Anderson, T. (2023). International student policy in Canada: A critical analysis. *Comparative and International Education*.
2. Kim, A., & Kwak, M. (2019). *International students in Canada*. University of Toronto Press.
3. Bourke, P. (2020). *Studying in Canada: A guide for international students*. Self-Published.
4. de Moissac, D., Graham, J. M., Prada, K., Gueye, N. R., & Rocque, R. (2020). Mental health status and help-seeking strategies of international students in Canada. *Canadian Journal of Higher Education*