Canada Immigration Data Visualization

By:
Imran Basha S
20th July 2023

Outline

- Executive Summary
- Introduction
- Methodology
  - Dataset
  - Data Cleaning & Preprocessing
  - Analysis Approach
- Exploratory Data Analysis
- Results & Analysis
- Key Findings
- Suggestions & Recommendations
- Conclusion
- Limitations & Future Work

Executive Summary

- We have an annual international immigration to Canada data(1980-2013) compiled by UN.
- The data spans across 195 countries, 6 continents, 22 regions, classifying each of the countries as developed or developing.
- After an initial data cleaning to remove redundant columns and tidying up of the dataset. A brief EDA was conducted to have a clear comprehension of the data, followed by Data Visualizations for each of the scenarios.
- It has been observed that Asia, majorly India & China have been the largest contributors of Immigrants.
- An increasing trend of immigration is observed in the recent decades.

Introduction

This Project studies the **international immigration to Canada** from across the world for a period from **1980 - 2013**.

The aim of this project is to understand the data, unearth some trends and insights regarding the pattern of immigration to Canada.

The Primary tool being **Data Visualization**, Matplotlib library is extensively used throughout.

Note: The project undertaken is purely for academic purpose and should not be considered for any purpose otherwise.

Methodology

Dataset

The dataset contains annual data on the flows of international immigrants as compiled by UN for the period 1980 - 2013.

The dataset spans across years from **1980 to 2013**.

It has various information regarding the Origin country of immigrants, continent, region, and whether that region is classifies as developed or developing.

The data contains information from **195 countries**.

The Dataset has **195 rows & 43 columns**

Data Source: The Canada Immigration dataset can be fetched from Here

Note: For detailed explanation please refer to the **Jupyter Notebook** file.

Data Cleaning & Preprocessing

- From the initial dataset **5 redundant columns** have been dropped, now the new dataset has 195 rows & 38 columns.
- 3 columns have been **renamed** for easy comprehension.
- A column named **‘Total’** has been added to **sum the immigrants** for each country for the whole period.
- For consistency data types of columns have been changed to ‘string’.
- There are **no missing** values.
- There are **no duplicates**.

Note: For detailed explanation please refer to the **Jupyter Notebook** file.

Analysis Approach

- Data Exploration followed by Data Cleaning & Data Preprocessing was done.
- Further Basic statistics and Exploratory Data Analysis (EDA) to have a better understanding of data was performed.
- Then, specific questions were explored to have more insights and understanding of trends.
- Data Visualization has been utilized prominently to derive inferences.
- The Analysis is based on:
  - Python version – 3.11.7
  - Numpy version – 1.26.2
  - Pandas version – 2.1.4
  - Matplotlib version – 3.8.0

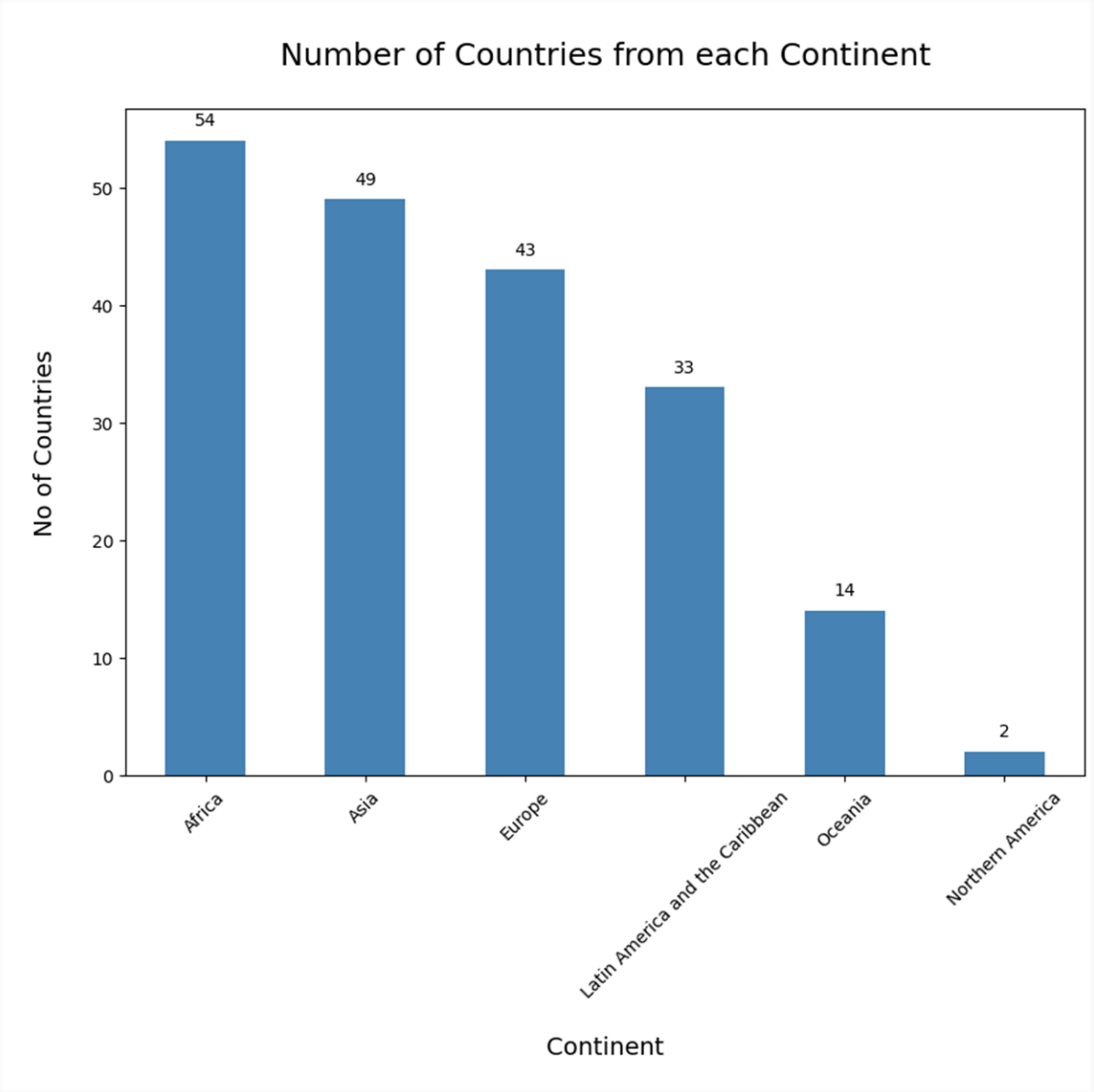
Exploratory Data Analysis

Data Exploration

- After data cleaning the dataset has **195 rows & 38 columns** of data.
- Totally there is data from **195 countries**(representing 1 row per country).
- **6 Continents** are included in the data.
- The whole data is categorized into **22 Regions**.
- Also, the countries are described as either **Developed or Developing** Regions.
  - 147 countries belong to Developing regions.
  - 48 countries belong to Developed regions.

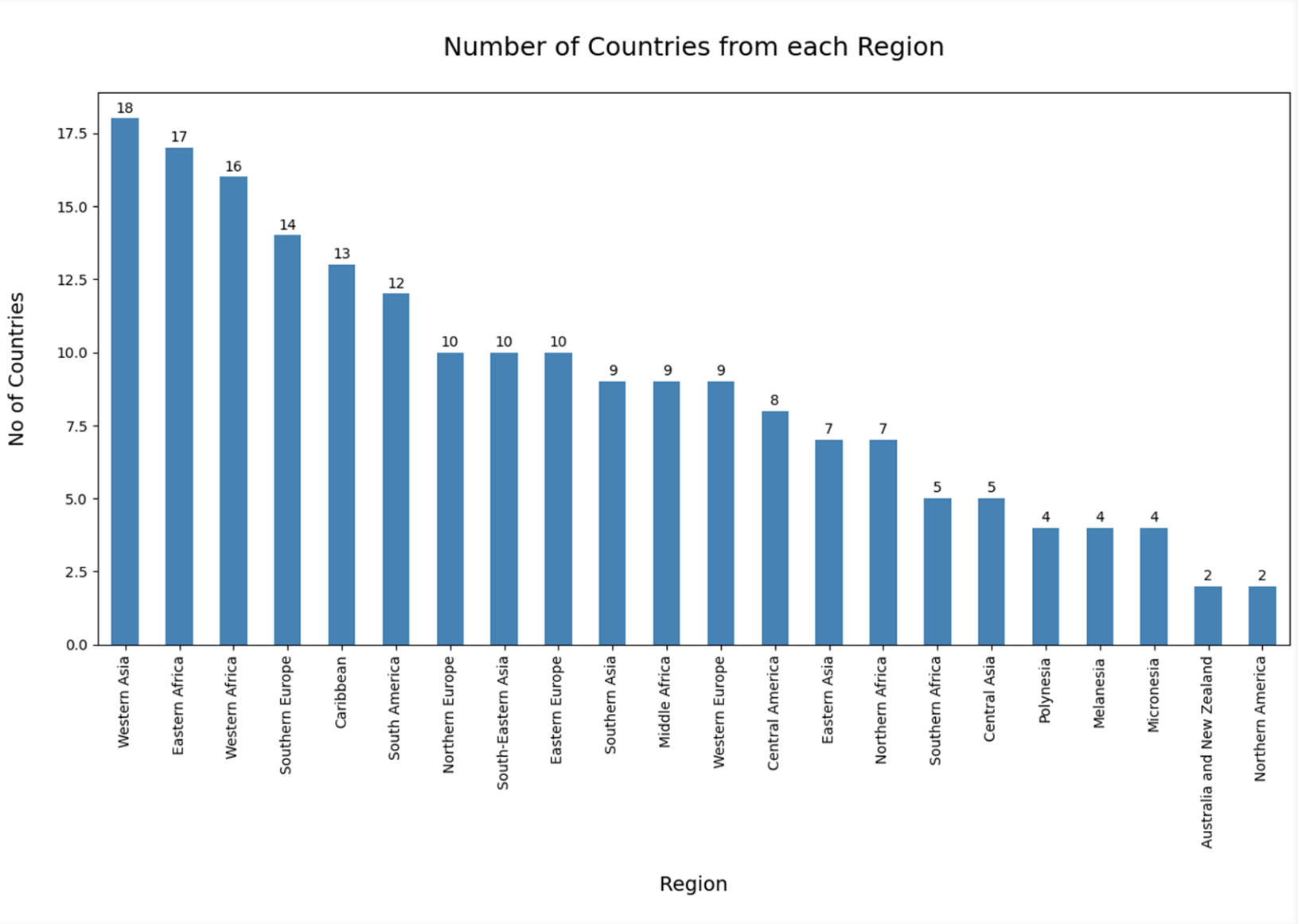
Note: For detailed explanation please refer to the **Jupyter Notebook** file.

Number of Countries from Each Continent



We have data from 6 continents, and the countries belonging to those continents are shown.

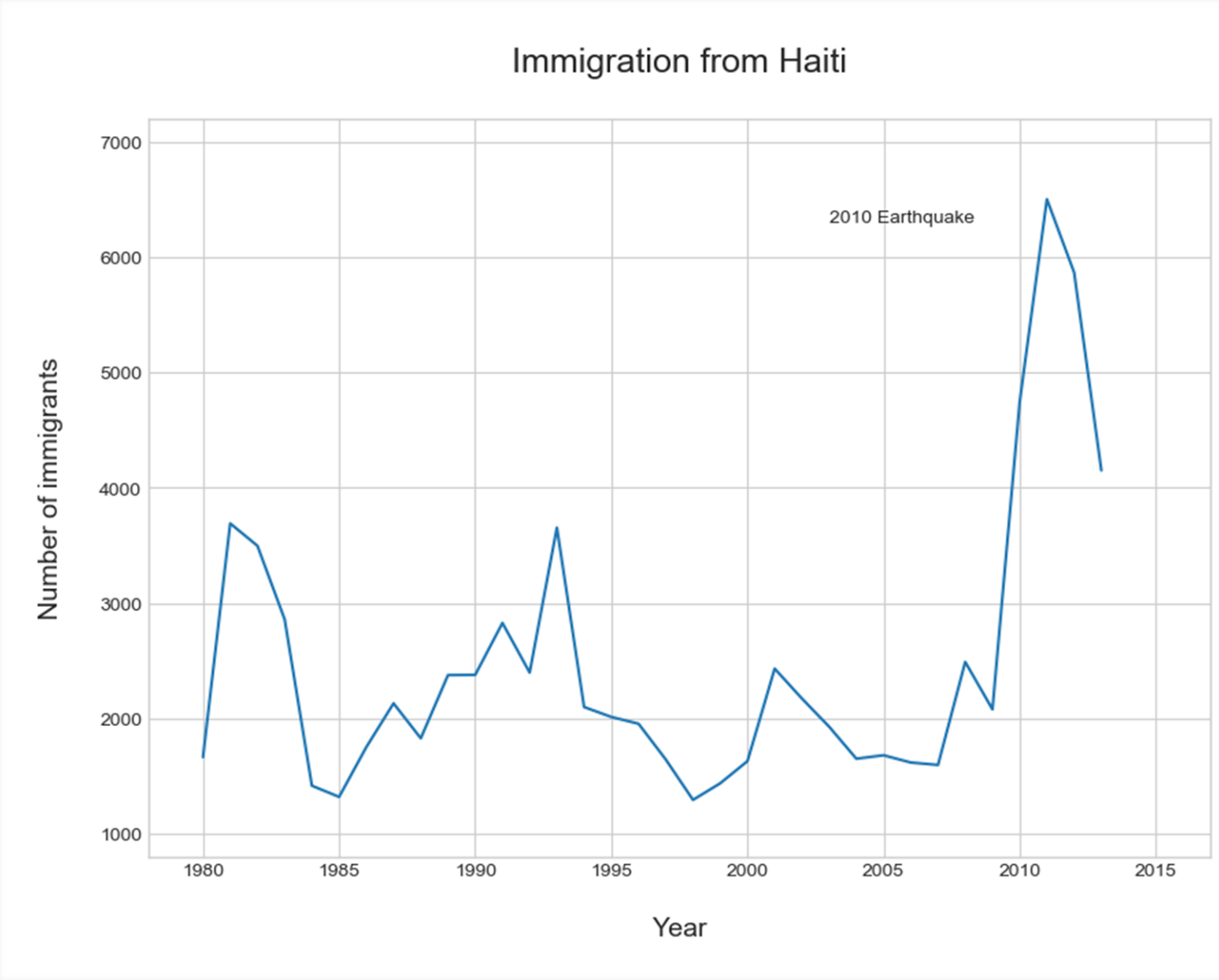
Number of Countries from Each Region



We have data belonging to 22 regions, and the countries belonging to those regions are shown.

Results & Analysis

1. Immigration from Haiti



Back drop

In 2010, Haiti suffered a catastrophic magnitude 7.0 earthquake.

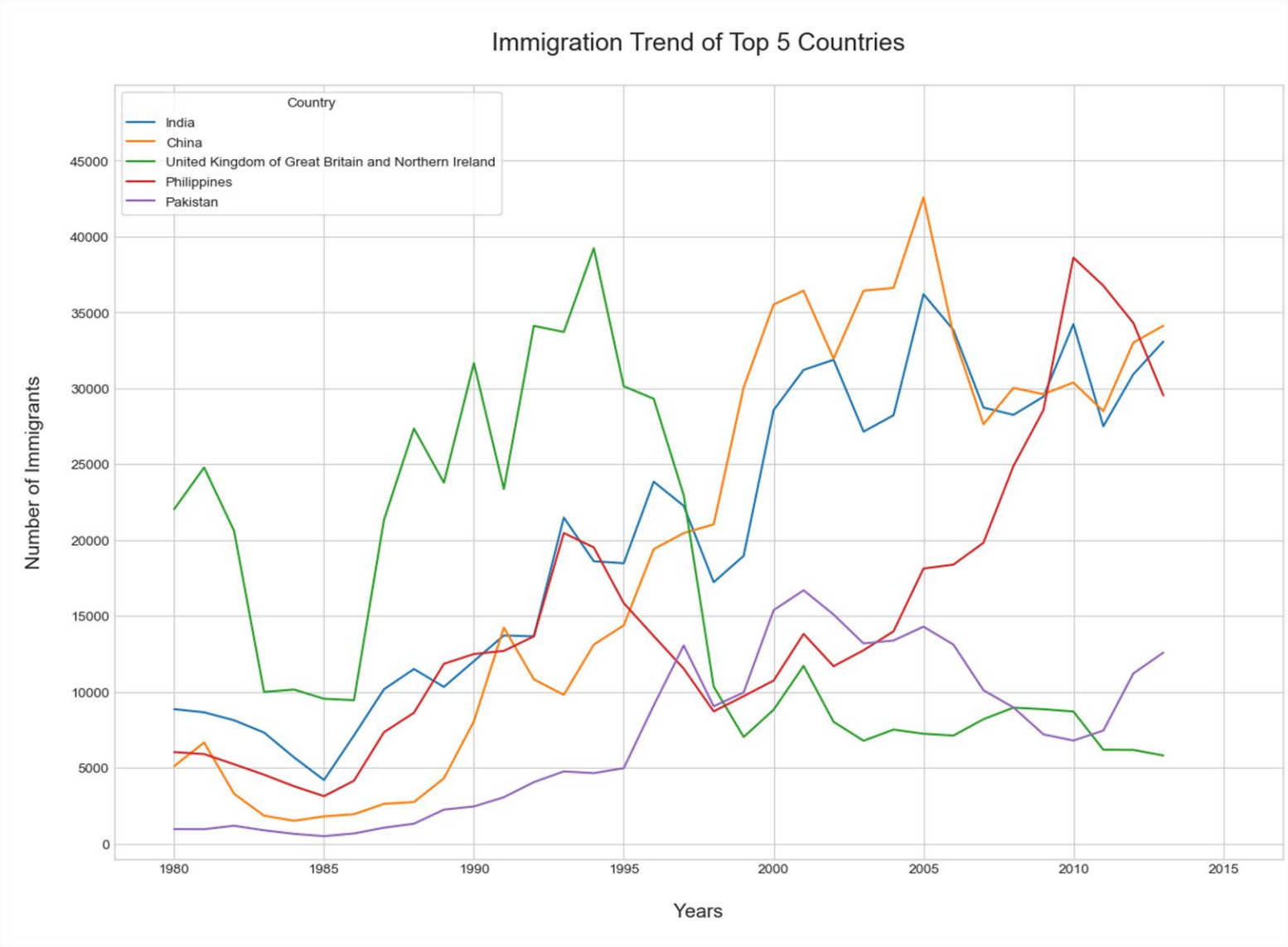
The quake caused widespread devastation and loss of life and about three million people were affected by this natural disaster.

As part of Canada's humanitarian effort, the Government of Canada stepped up its effort in accepting refugees from Haiti.

Inference:

We can clearly see a spike in immigrants from 2010, from Haiti. Reemphasizing the point presented in the case background earlier.

2. Immigration Trend of Top 5 Countries



Inference:

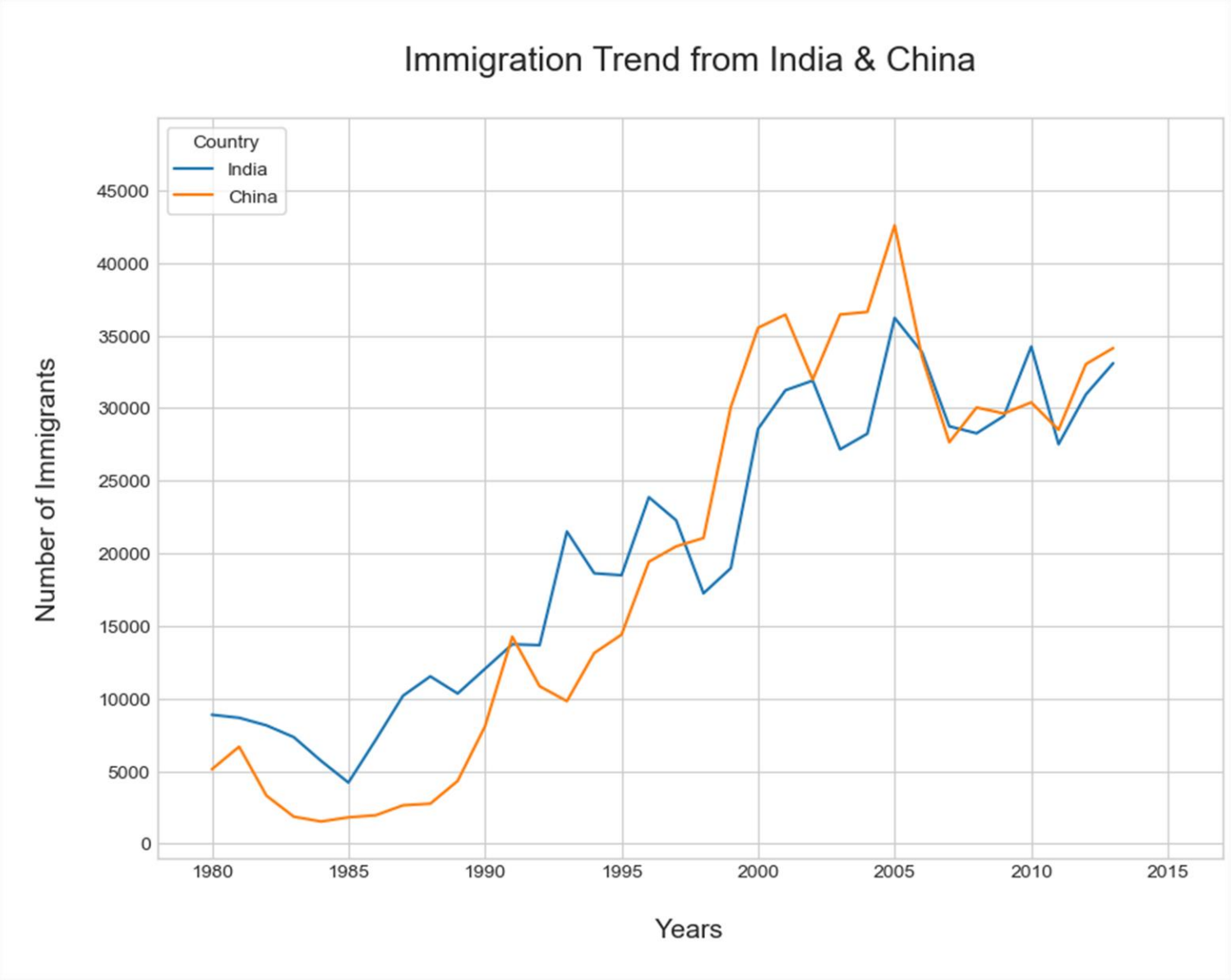
1.We can see **India & China** contribute the **most** of the immigrants.

2.Also we can observe that, India & China represented by Blue & Orange respectively, have **similar pattern** of immigration.

3.We see a **spike** in immigrants from **UK** during the year 1994 and then a gradual declining trend after 1996.

4.There is a sudden spike in Immigrants from **Philippines** after 2004.

3. Immigration Trend from India & China

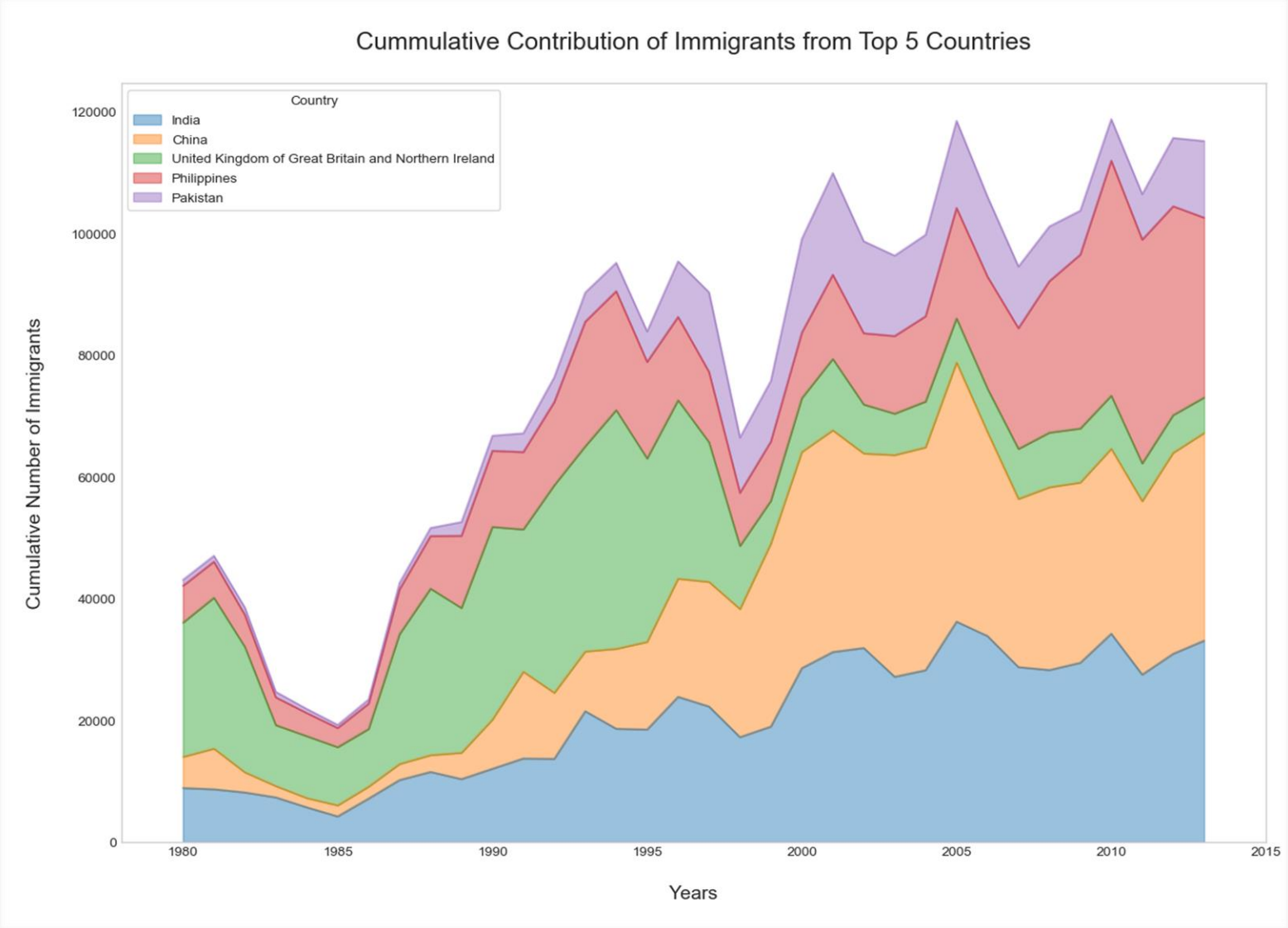


Inference:

1.China & India contribute the most of the immigrants and both have very **similar trends**.

2.Overall, there has been a gradual increasing trend, both spiking the most in 2005.

4. Cumulative Contribution of Immigrants from Top 5 Countries



Inference:

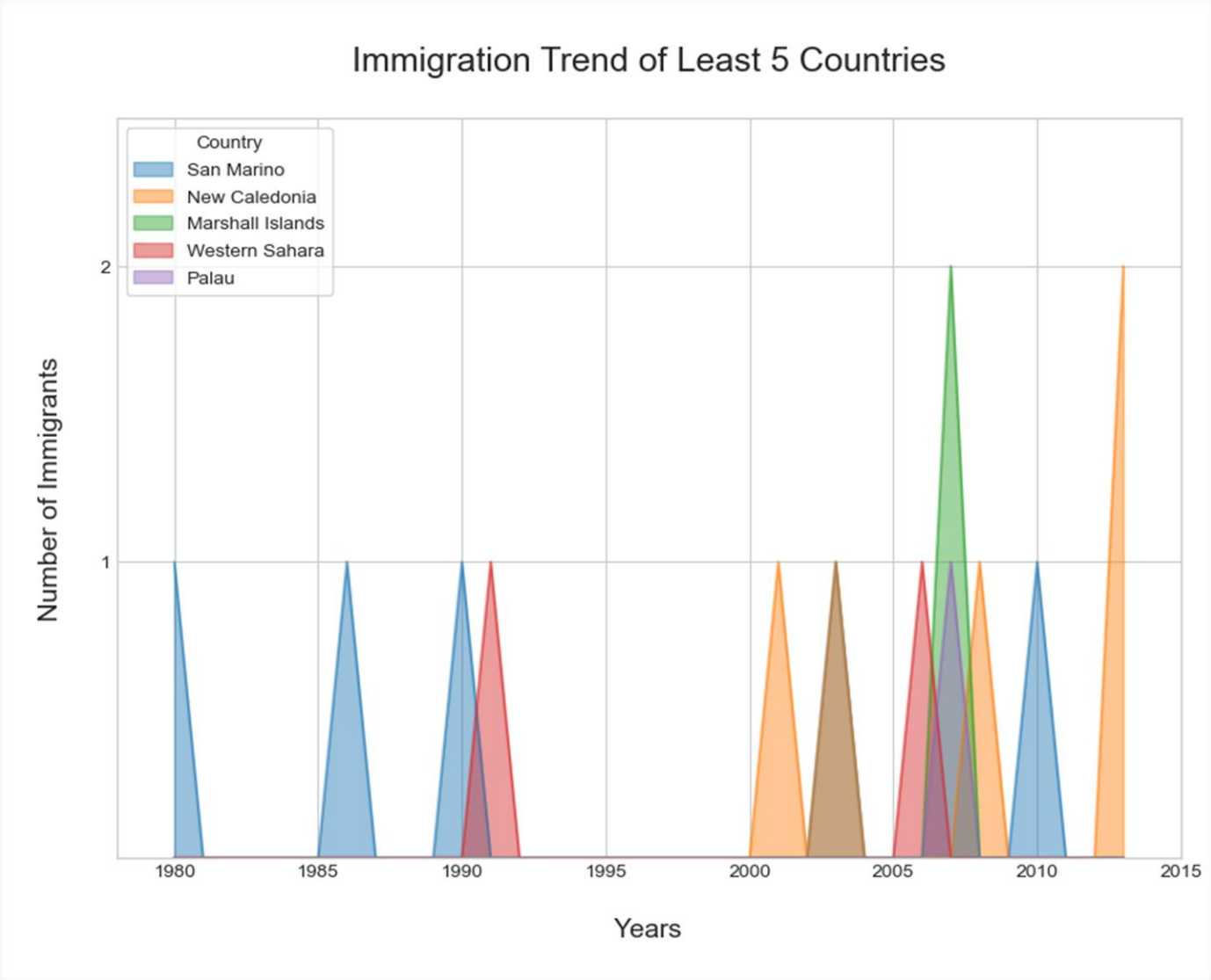
1.The Plot here show the cumulative immigrants total added through each of the years.

2.We need to look at the cumulative area, which gives us an estimated outlook of total immigrants from each of the countries.

3.We can see China tops here, closely followed by India.

4.It can be also be observed that **UK's** overall contribution of immigrants has decreased after 1998, and, it seems vice versa in the case of **Philippines.**

5. Immigration Trends of 5 Least Immigrant Contributing countries

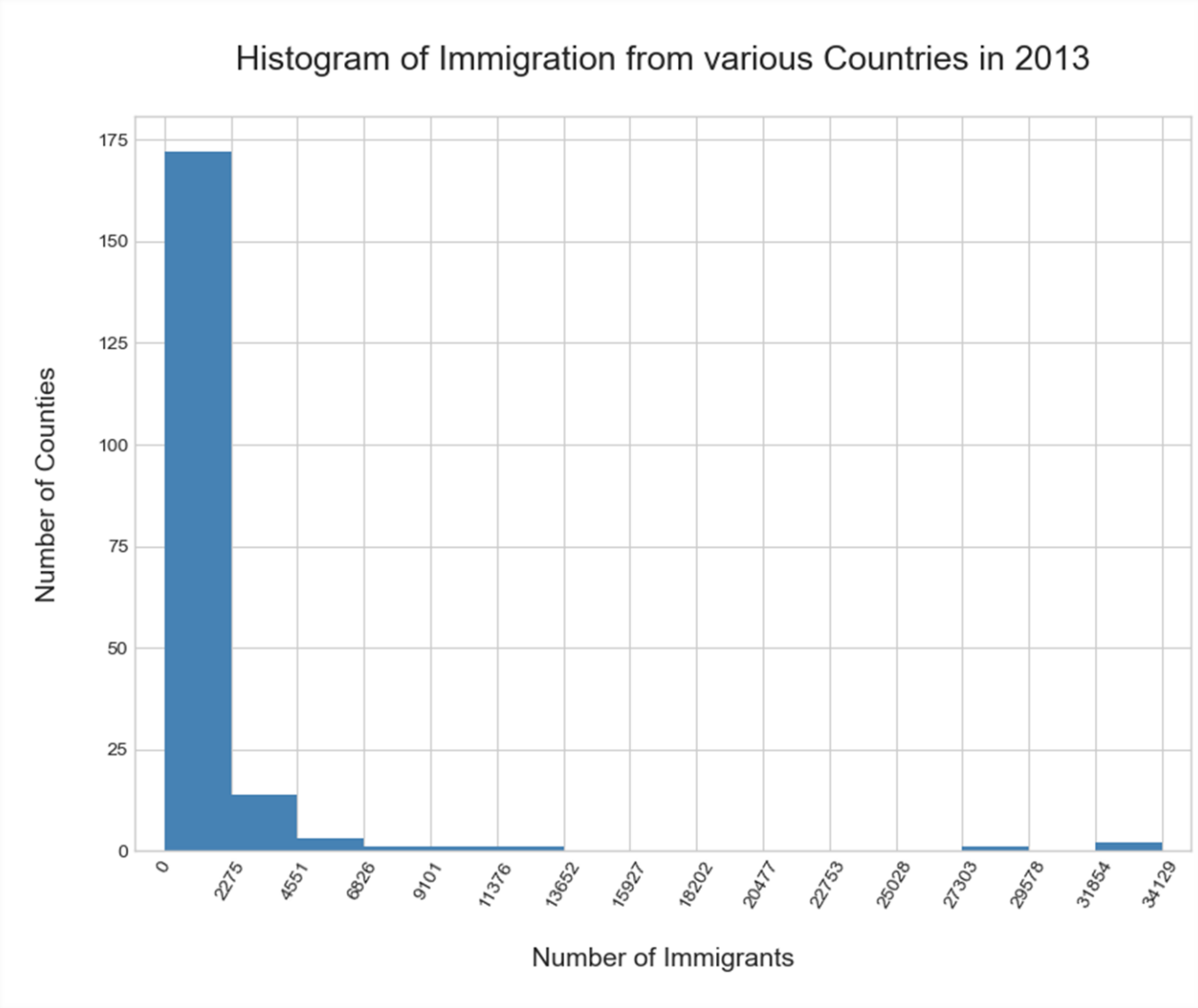


Inference:

1.It can be observed that the least contributing countries, add in values in single digits.

2.Contributing at most **2** no's at some point, while also having **0** additions sometimes.

6. Histogram of Immigration from Various Countries in 2013



Inference:

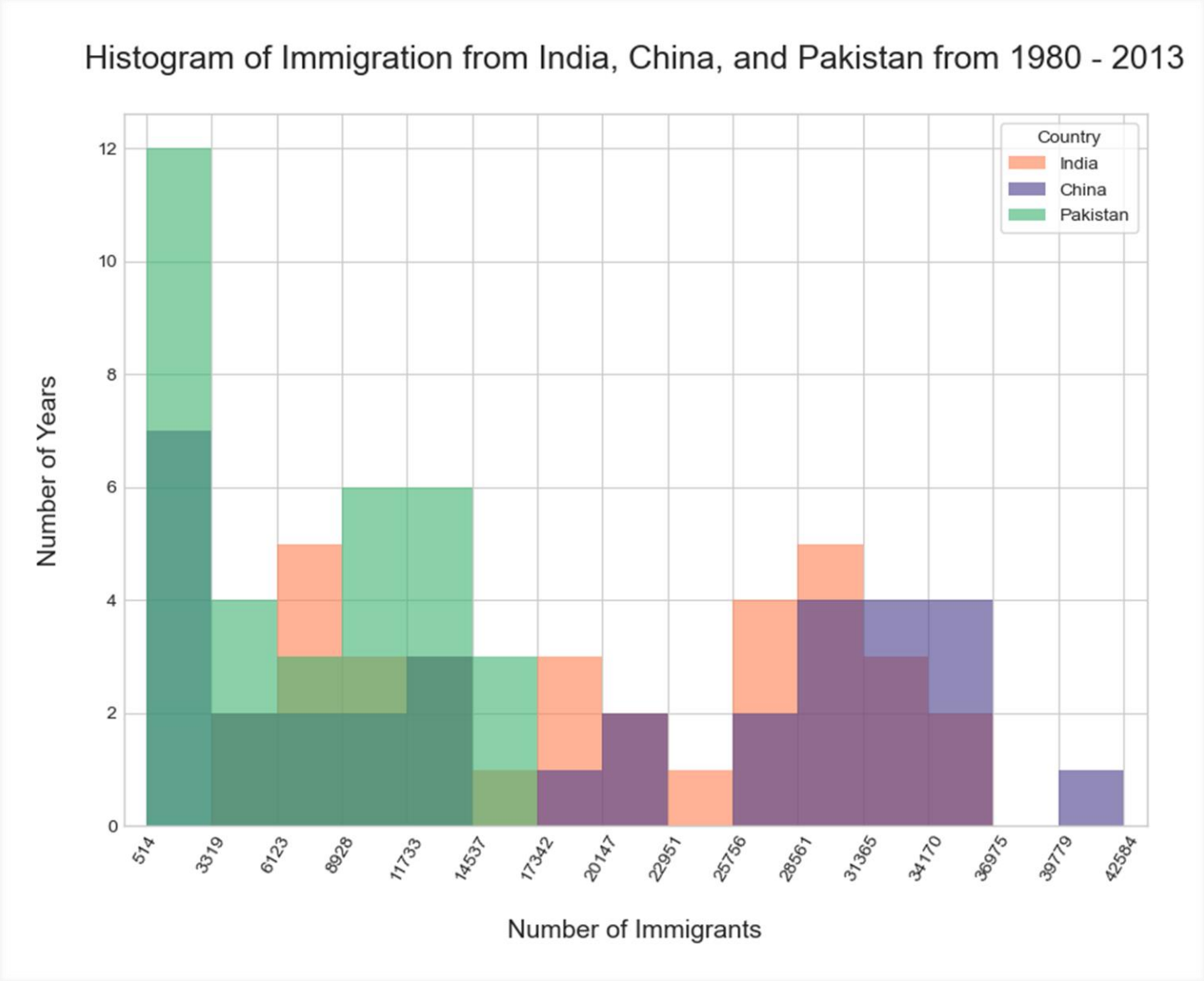
1.The histogram breaks up the data of year **2013** in to 15 bins

2.172 countries have contributed immigrants in the range **0 to 2275.26**

3.14 countries have contributed immigrants in the range **2275.27 to 4550.53**

4.3 countries contributed between **4550.54 and 6825.8**

7a. Histogram of Immigration from India, China, and Pakistan from 1980 - 2013

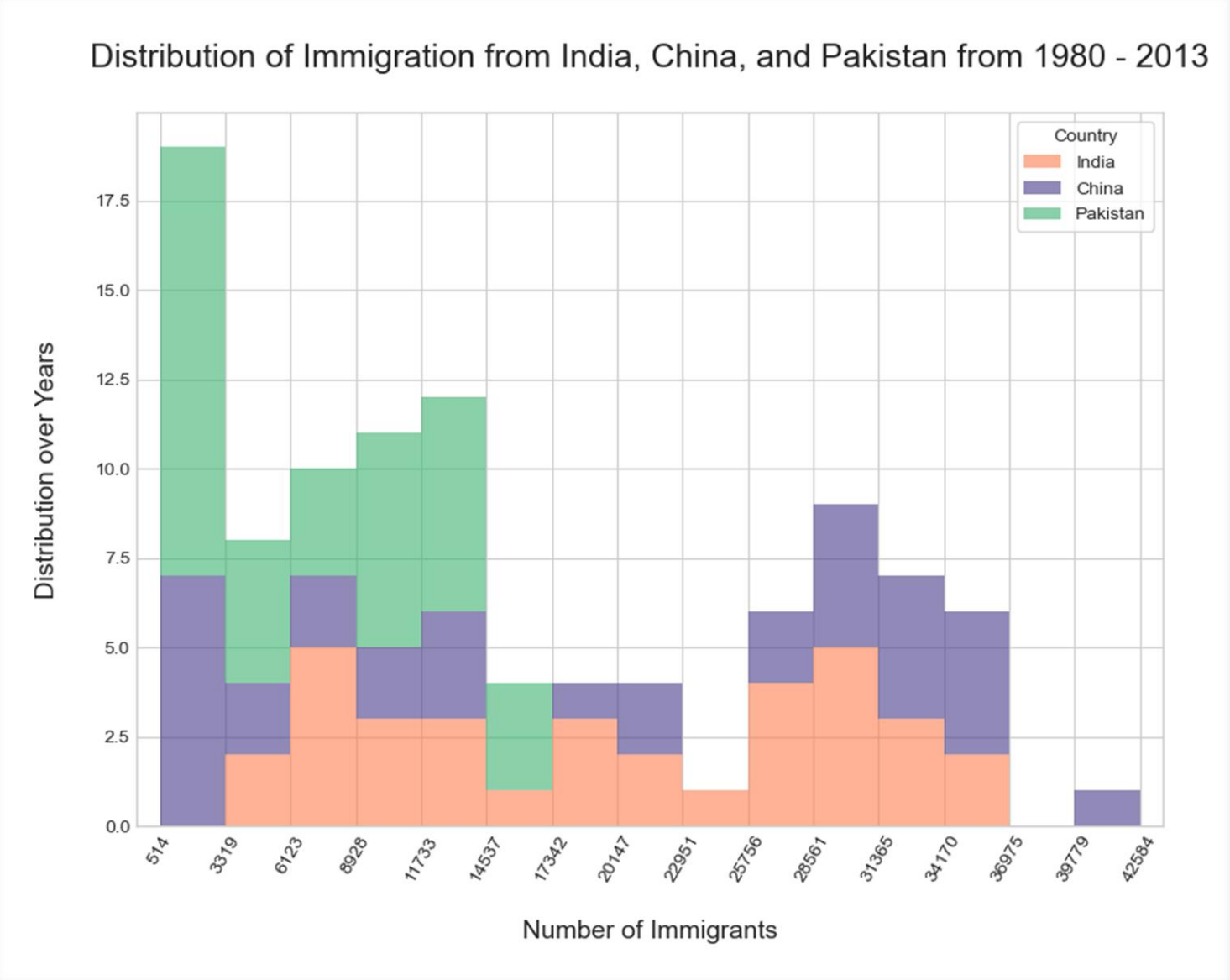


Inference:

1.The figure shows an unstacked Histogram. To understand it , the bin-frequency table provided below it can be utilized in conjunction.

2.This can be used to make a comparative guess between different countries. However, the overlapping area limits it utility in this specific case.

7b. Distribution of Immigration from India, China, and Pakistan from 1980 - 2013



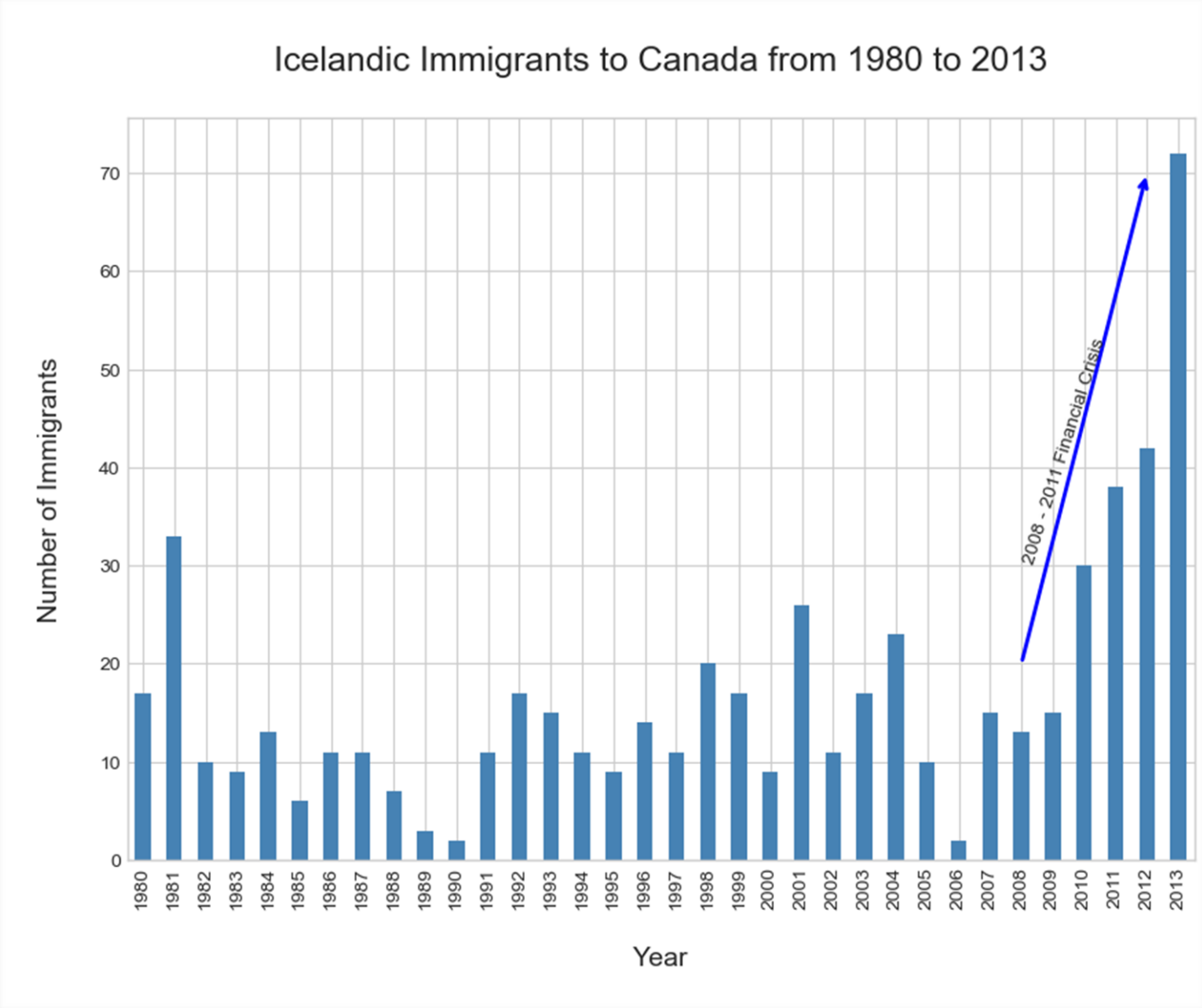
Inference:

1.The figure above shows a stacked Histogram. To understand it , the bin-frequency table provided below it can be utilized in conjunction.

2.All the three countries have contributed in the range **0 to 3318.66** on 19 instances.

3.And, in the range **3318.67 to 6123.33** on 8 instances.

8a. Icelandic Immigrants to Canada from 1980 - 2013

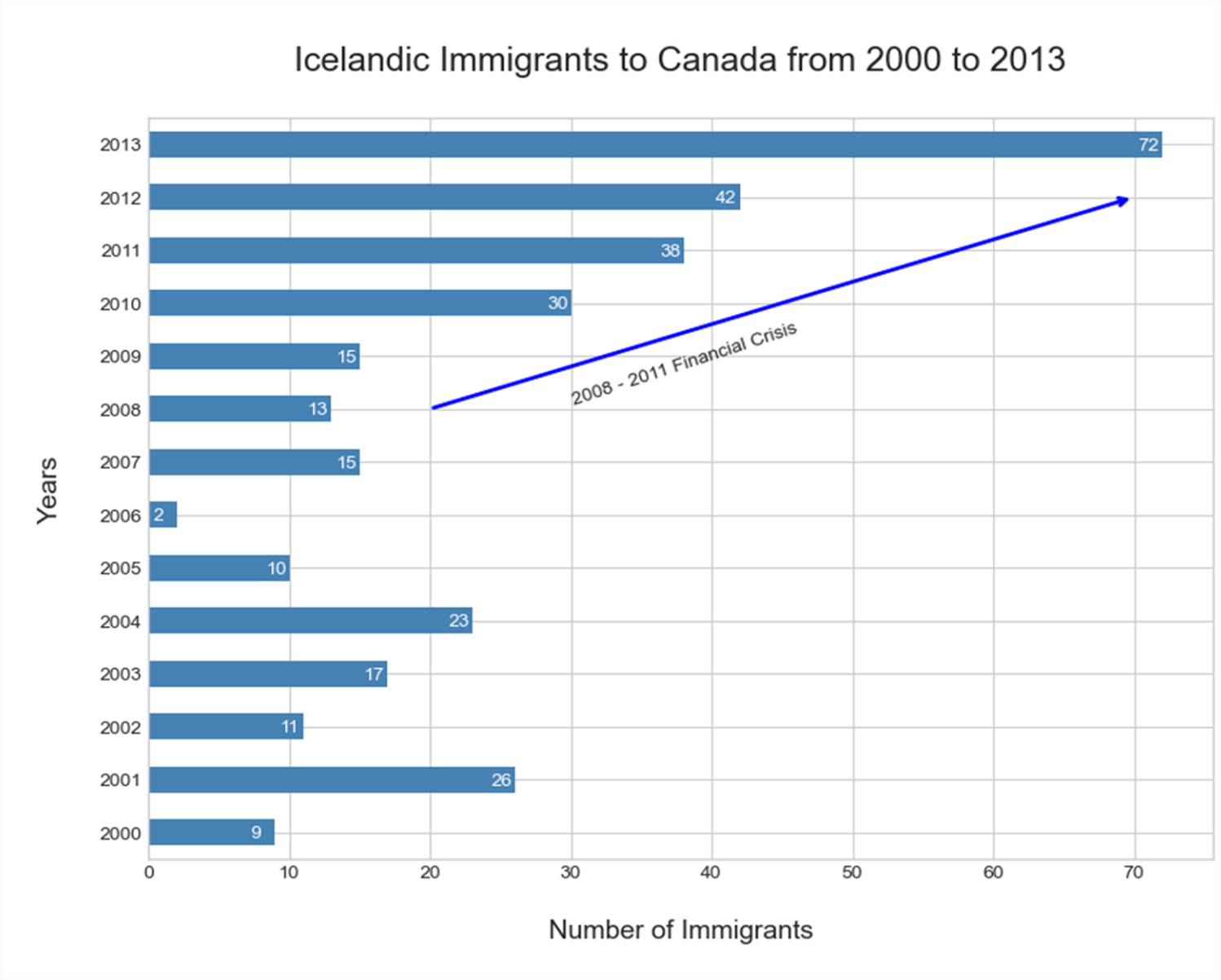


Inference:

1.In line with the back drop of the 2008 - 2011 Icelandic Financial crisis, we can see a corresponding spike in influx of immigrants, after 2008.

2.This is also indicated by the arrow pointer in the figure.

8b. Icelandic Immigrants to Canada from 2000 - 2013

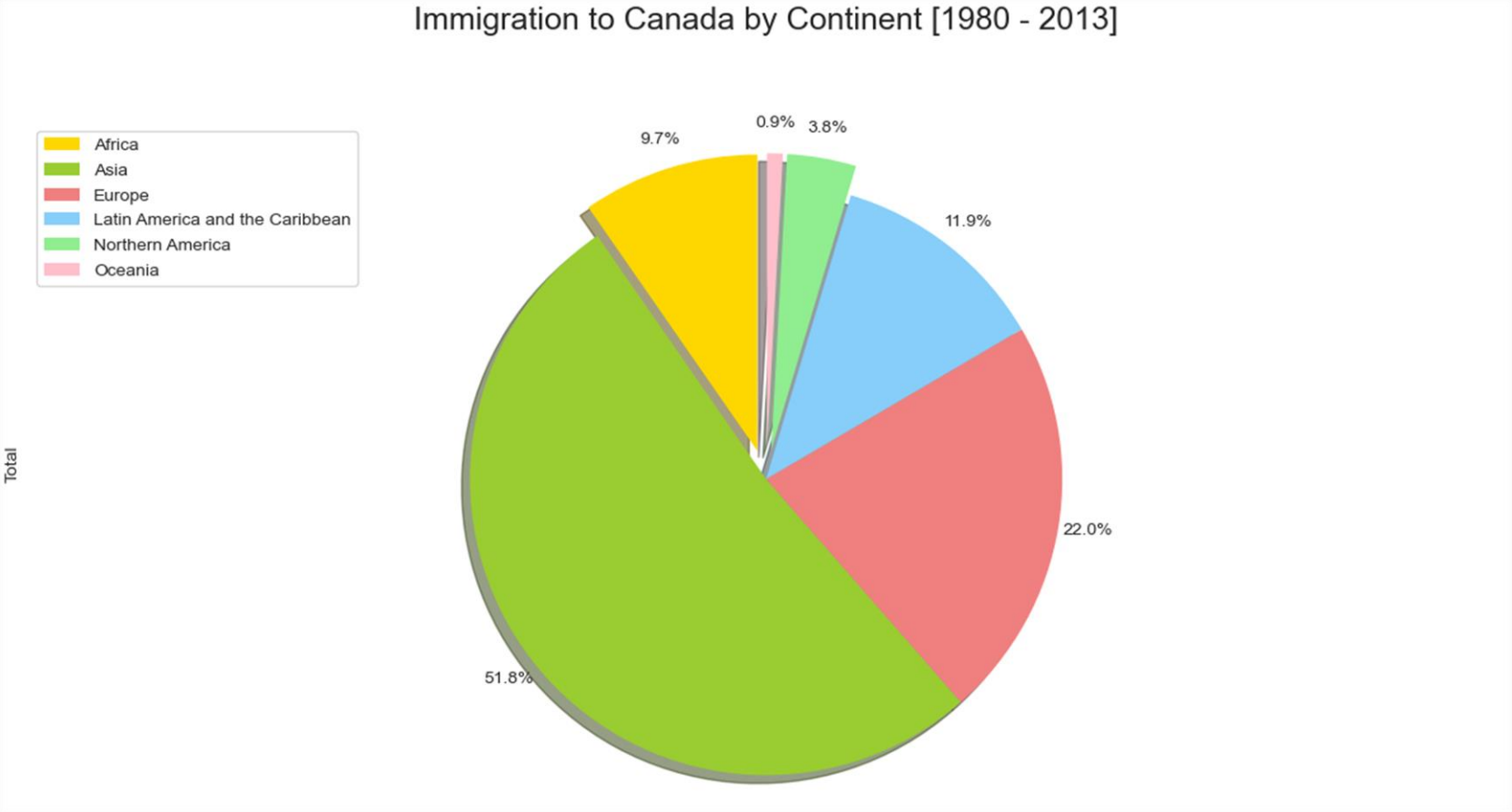


Inference:

1.For better visualization of previous figure, horizontal bar graph is used here.

2.It depicts the same point as before in much more clearer way.

9a. Immigration to Canada by Continent (1980-2013)



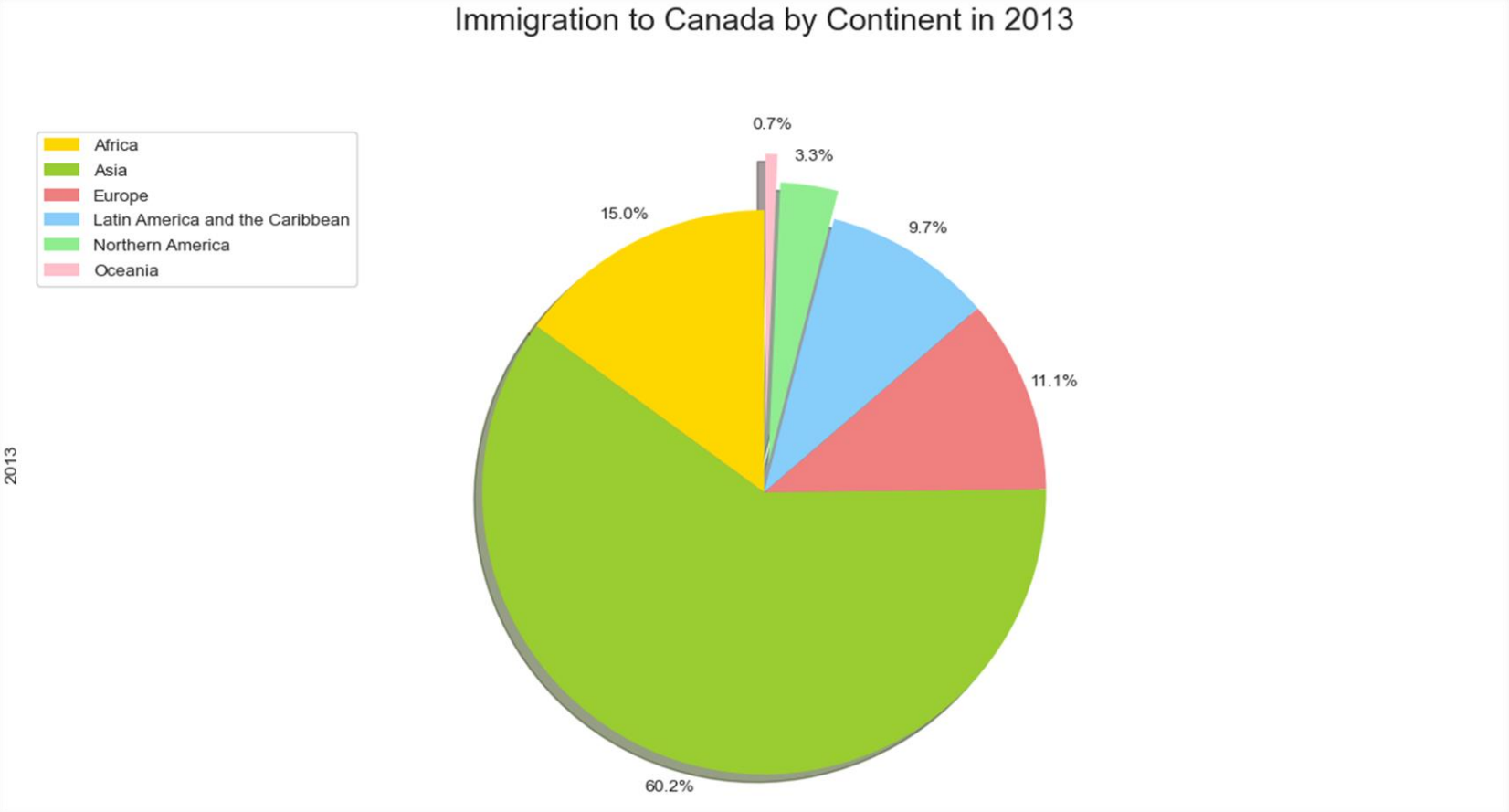
Inference:

1.The figure shows the contribution from each of the continents towards the total immigration during the period 1980 - 2013.

2.We can see **Asia** contributes a majority share - 51.8%

3.Followed by **Europe** - 22% and **Latin America** - 11.9%

9b. Immigration to Canada by Continent in 2013



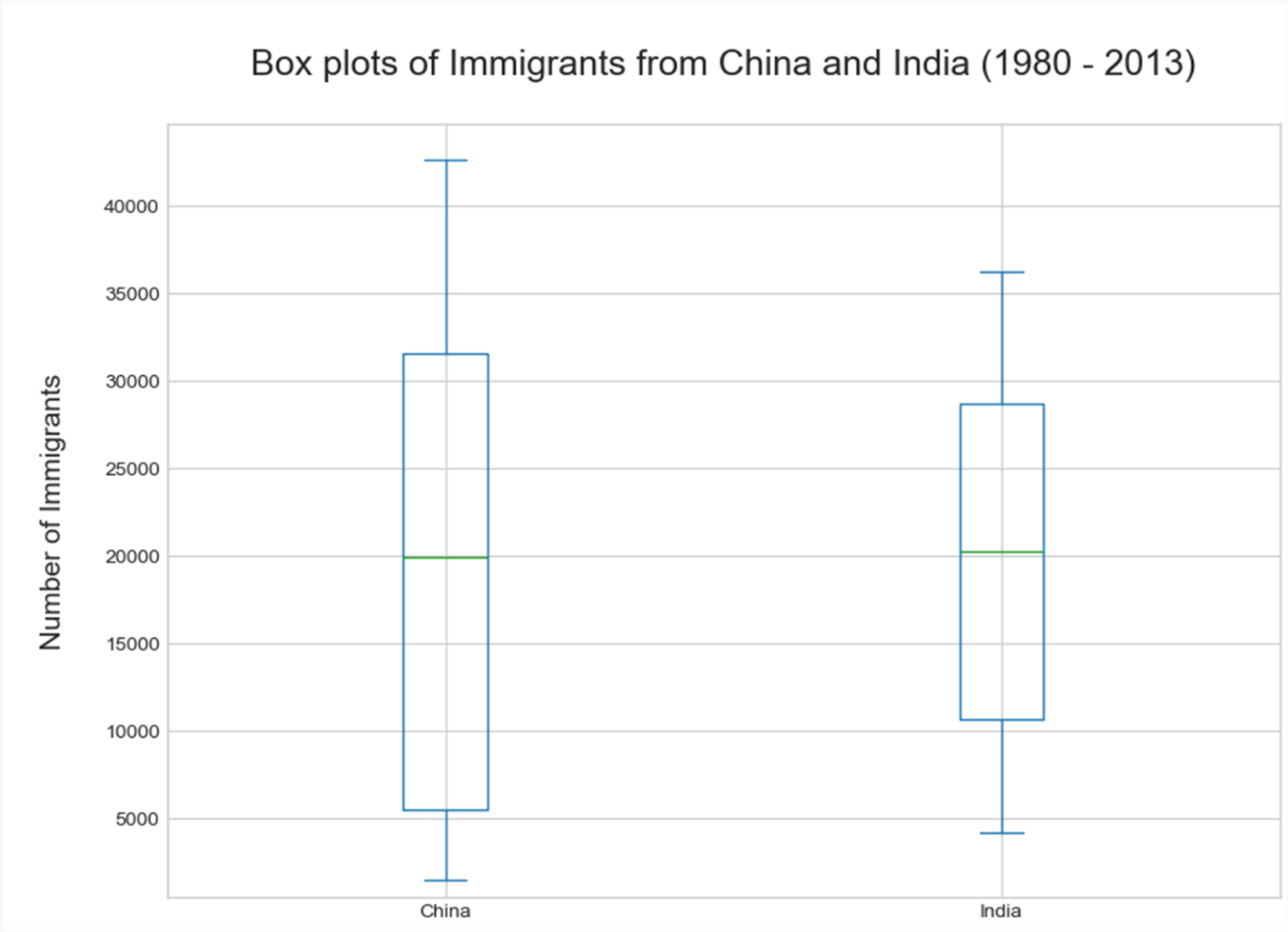
Inference:

1.The figure above shows the contribution from each of the continents towards the immigration during the year **2013**.

2.The major share is from **Asia** as before.

3.Oceania contributes less than 1% as seen earlier.

10. Box plots of Immigrants from China and India (1980-2013)



Inference:

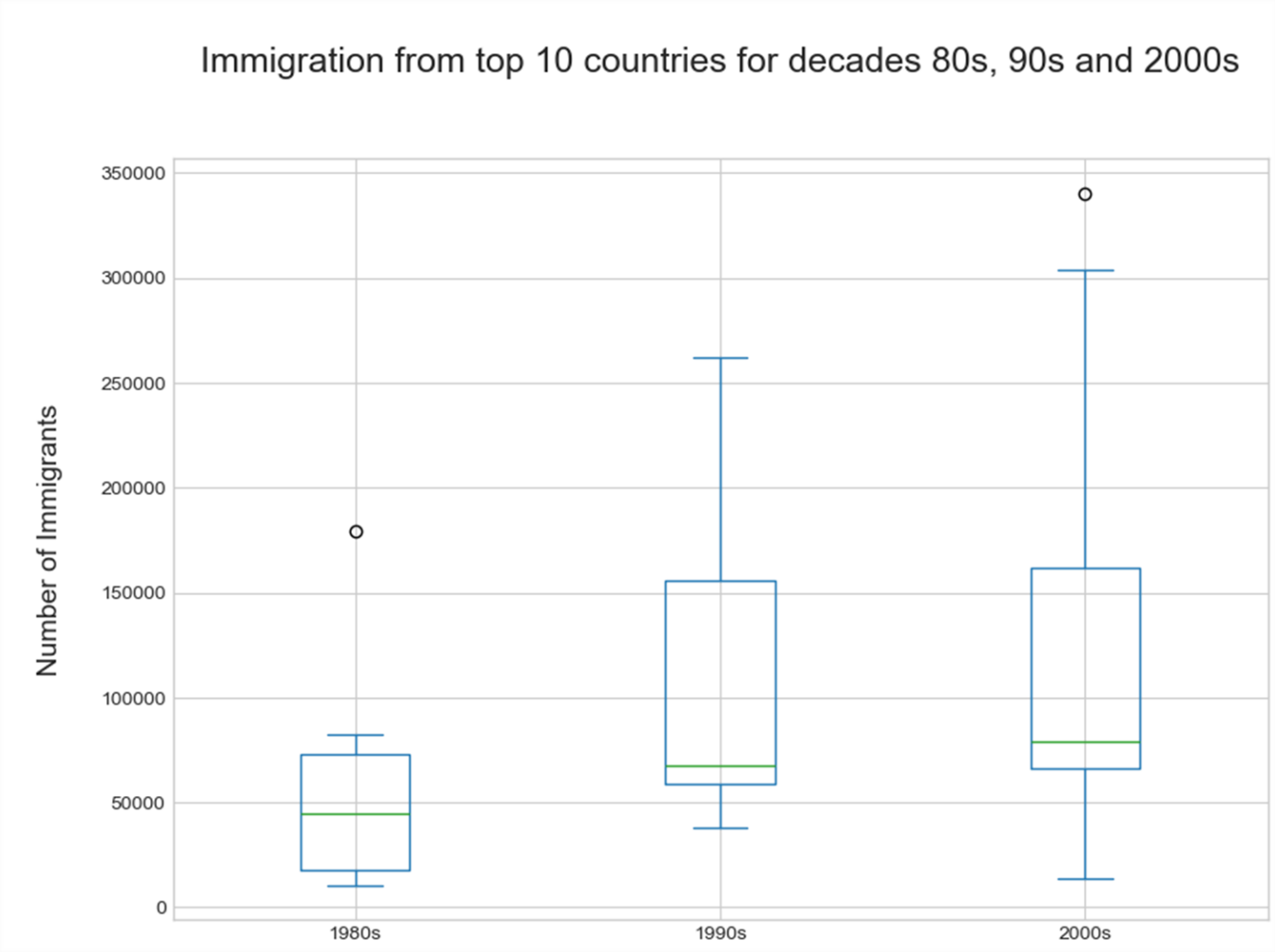
1.The figure shows the distribution of data points of yearly immigration numbers from China & India for 34 years.

2.The **median** value for both the countries is almost the same.**(~20000)**

3.The maximum ever is **42584 for China** and **36210 for India**.

4.The range of data points is **more spread out for China** as compared to India.

11. Immigration from Top 10 Countries for Decades 80s, 90s, and 2000s



Inference:

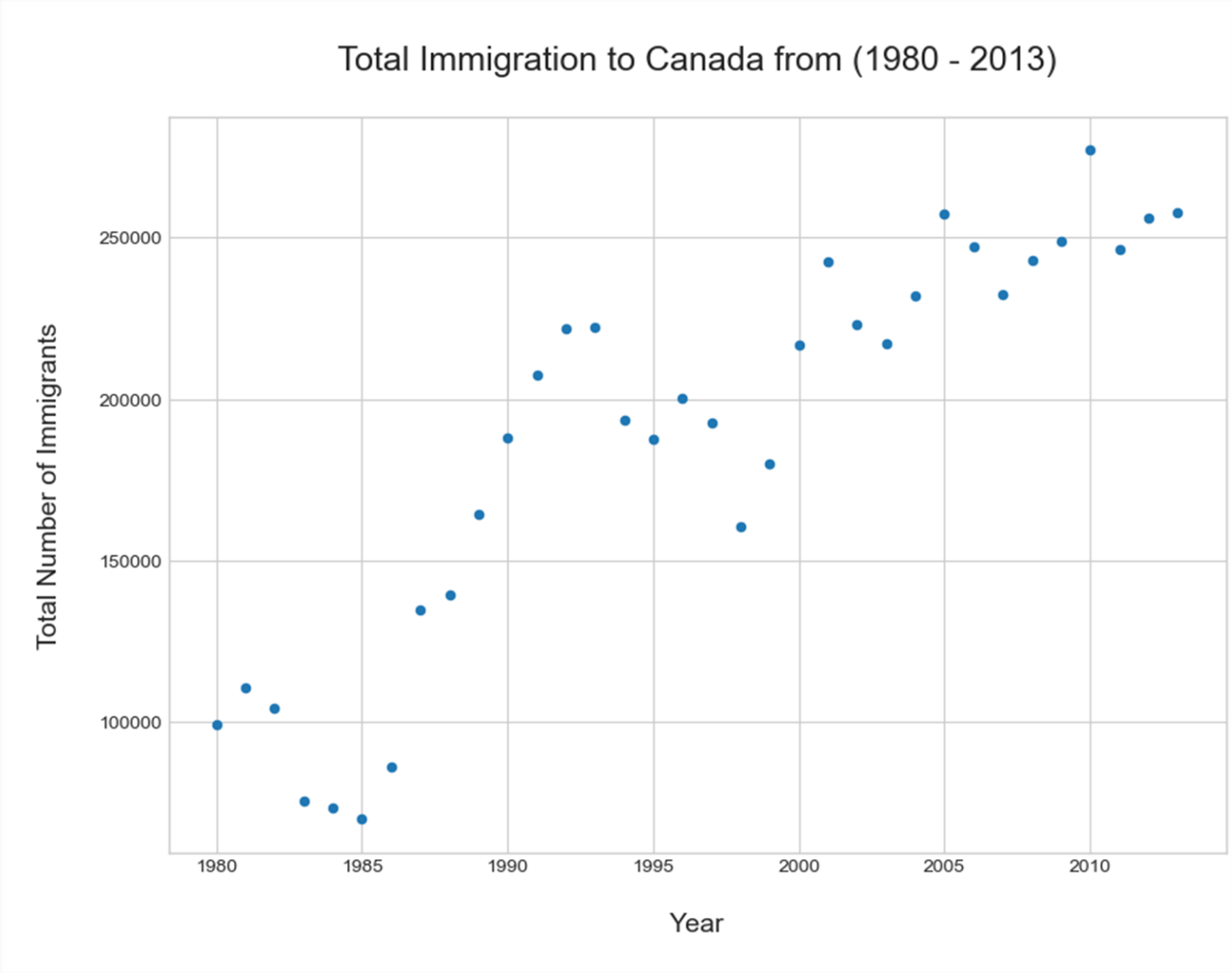
1.The figure shows distribution of data points for the top 10 countries across 3 decades.

2.The immigration **during 80's** seems **less** and also the spread of data is close.

3.The immigration **during 90's & 2000's** has data points more spread out. The **minimum value is highest (38189) during 90's** compared to all three decades.

4.The **outliers** seen, would likely be contributed by India or China.

12. Total Immigration to Canada from (1980-2013)



Inference:

1.The Total immigration for each of the years shows an upward trend as years go by.

2.The highest ever was in 2010 and lowest was in 1985.

Key Findings

Key Findings

- The **minimum** immigration was observed in 1985 and **maximum** in 2010.
- India, China, United Kingdom, Philippines, and Pakistan are the top 5 countries from where Canada receives immigrants.
- Both **India & China** have a **similar** patterns of immigration.
- It is observed that a natural calamity, in the case of Haiti(2010) or an economic crisis (Iceland 2008 - 2011) or a political unrest **increases the influx** of immigrants.
- **Asia** contributes the largest about 51.8% of immigrants to Canada.
- **Oceania** send the least number of immigrants, 0.9%
- The 80’s decade had the least migration and then from 90’s decade an **increasing trend** in migration is observed.

Suggestions & Recommendations

Suggestions & Recommendations

- Canada receives a majority of immigrants from Asia, a comprehensive immigrant management protocol in this regard may be developed if needed.
- Any natural calamity, economic or political crisis has seen an increased influx of immigrants. Such a pattern may be anticipated in the event of an incident. Supplementary efforts may be amplified accordingly, to accommodate or handle the insurge, in sync with Canada’s policies on humanitarian grounds.
- Since there has been a trend of increasing immigration since recent decades, rehabilitation and integration programmes may be suggested.

Conclusion

Conclusion

We have clearly observed that recent decades have an increasing trend of immigration, and this can be expected to rise, considerably from Asia.

Hence, measures can be taken to evaluate & utilize such a situation in a good stride for the country. Protocols & policies may be improvised for the benefit of all and moreover help can be sought under United Nations comprehensive programmes which cater to such occurrences.

Limitations & Future Work

Limitations & Future Work

Limitations:

The dataset has data only till 2013, supplementary recent data may help uncover more relevant insights.

Further, it has to noted that the project undertaken is purely for academic purposes, the results or any inferences may not be cited or utilized for other purpose.

Future Work:

As the data is from 195 countries, involving many continents & regions, projects specific to each of the scenarios can be conducted to have a more relevant insights.

With the recent data in place, the data can be further utilized to predict the future occurrences.

Thank You!