



EM 624

Analyze People Migration Data

Instructor: Dr. Carlo Lipizzi

From: Yogendrasinh Sodha



Table of Content

1. Overview.....	2
2. Dataset Description	3
3. Dataset Preparation	4
4. Methodology	5
5. Data Representation	7
6. Conclusion.....	15



1. Overview

Human migration refers to any movement of people between locations, often across long distances or in huge groups. It is known that throughout prehistory and human history, humans have migrated widely. Human migration is the movement by people from one place to another, particularly different countries, with the intention of settling temporarily or permanently in the new location. It typically involves movements over long distances and from one country or region to another. Human migration has occurred throughout history and under a wide range of conditions. Tribal, national, class, and individual levels have all been involved. Climate change, politics, the economy, religion, or a simple desire for adventure have all been factors for migration.

Migrations are the essence of inhabiting the world and they are believed to have begun approximately 2 million years ago with the early expansions out of Africa by *Homo erectus*.

In more modern times, migrations have often been driven by necessity, that could include forced displacement (in various forms such as deportation, slave trade, trafficking in human beings) and flight (war refugees, ethnic cleansing).

This study analyzes the history of human migration from 2000- 2014 from data to get insights on the migration dynamics and attempts to answer the following questions:

1. What are the major flows of migration (meaning from where to where) ?
2. Where people from the more developed regions are migrating to ?
3. Where are people from the less developed regions are migrating to ?
4. What are the dynamics by income and geographic region ?
5. What are the migration patterns in India over the years ?

2. Dataset Description

The dataset used in this study was from United Nation migration data.

- The data is a collection of 289 columns and there are 2733 rows.
- There are 51 columns with data type as float and 238 columns with data type as object.
- There are many rows as the migration data contains data of every 5 years from 1990 to 2019 of every country and region defined by the United Nations. Dataset has more developed regions and least developed countries where according to the UN more developed regions comprise Europe, Northern America, Australia, New Zealand, and Japan whereas group of least developed countries, as defined by the United Nations General Assembly, currently comprises 47 countries: Afghanistan, Angola, Bangladesh, Benin, Bhutan, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Djibouti, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Haiti, Kiribati, Lao People's Democratic Republic, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Myanmar, Nepal, Niger, Rwanda, São Tomé and Príncipe, Senegal, Sierra Leone, Solomon Islands, Somalia, South Sudan, Sudan, Timor-Leste, Togo, Tuvalu, Uganda, United Republic of Tanzania, Vanuatu, Yemen and Zambia.
- There are rows with migrants' destinations to income based regions where income level is based on June 2018 Gross National Income per capita from the World Bank.



3. Dataset Preparation

1. The dataset UN_Migrations.csv was loaded into a dataframe using the pandas library in python and was stored in variable df.
2. The variables in dataset were :- Year, Sort order, Major area, region, country or area of destination, Notes, Code, Type of data (a), Unnamed:6, Unnamed:7, Unnamed:8, Country or area of origin, Unnamed:10, Unnamed:11,....., Unnamed:288.
3. As noticed there were many columns with unnamed names because the origin countries names were stored in 2 rows, so again the data frame was opened by passing parameter skiprows to skip the first row and all origin countries names were retrieved.
4. The dataset contained values like NaN, "..", and "-" As a result, these values were changed to 0.
5. The following variables were dropped from dataframe as they were not required in analysis :
 - 5.1. Sort order
 - 5.2. Notes
 - 5.3. Code
 - 5.4. Type of data (a)
 - 5.5. Other South
 - 5.6. Other North
6. The column with name as Major area, region, country or area of destination was renamed Destination for ease of future use.
7. Also, there was one duplicate column detected in visualization with the same name and data with a different case which was NORTHERN AMERICA was then dropped.

8. Columns with migrant data figures were in string type for analysis migrant was converted to float
9. After the cleansing of data frame the first five rows of data frame looked like this :

	Year	Destination	Total	Afghanistan	Albania	Algeria	American Samoa	Andorra	Angola	Anguilla	...	Uruguay	Uzbekistan	Vanuatu	Venezuela (Bolivarian Republic of)
0	1990.0	WORLD	153011473.0	6823350.0	180284.0	921727.0	2041.0	3792.0	824942.0	2047.0	...	237486.0	1428020.0	5060.0	185946.0
1	1990.0	UN development groups	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
2	1990.0	More developed regions	82767216.0	119386.0	177986.0	867015.0	1027.0	3737.0	167381.0	540.0	...	56838.0	1078563.0	1017.0	114991.0
3	1990.0	Less developed regions	70244257.0	6703964.0	2298.0	54712.0	1014.0	55.0	657561.0	1507.0	...	180648.0	349457.0	4043.0	70955.0
4	1990.0	Least developed countries	11060221.0	0.0	0.0	5622.0	0.0	0.0	608108.0	0.0	...	286.0	2027.0	9.0	2510.0

5 rows × 235 columns

10. There was 235 columns with every :

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1974 entries, 0 to 1980
Columns: 235 entries, Year to Zimbabwe
dtypes: float64(234), object(1)
memory usage: 3.6+ MB
```

4. Methodology

1. All names of origin countries were the columns in the data frame whereas names of destination were stored in rows so it was difficult to plot data. Due to that origin countries were transposed to rows in column Origin.
2. A function typedf was created to modify the data frame based on requirements of origin and destination.



3. There were 232 countries with migrants in the data frame, so when the graph was plotted for every country from origin to destination with the number of migrants it was messy and difficult to analyze. Due to that function heatmap was created to filter the number of migrants which then plots heatmap taking the minimum number of migrants from origin to destination.
4. Another graphIndia function was created to plot the number of migrants to India in the given years having most and least number of migrants from the country.

5. Data Representation

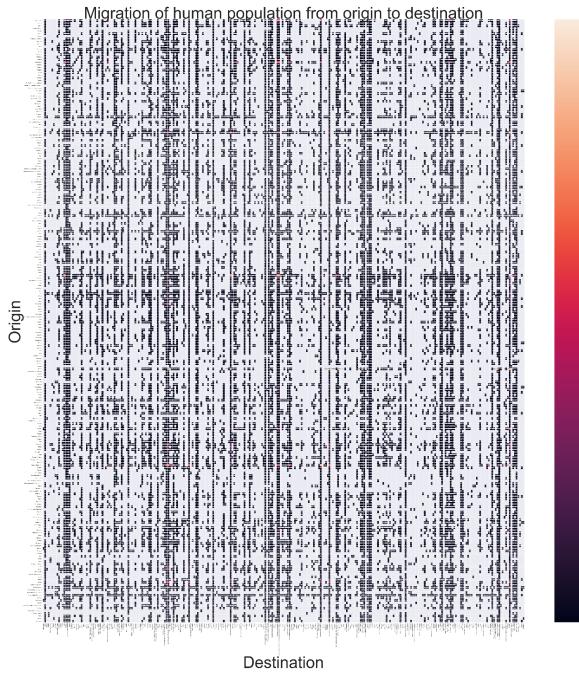


Fig 1 . Migrants from origin to destination.

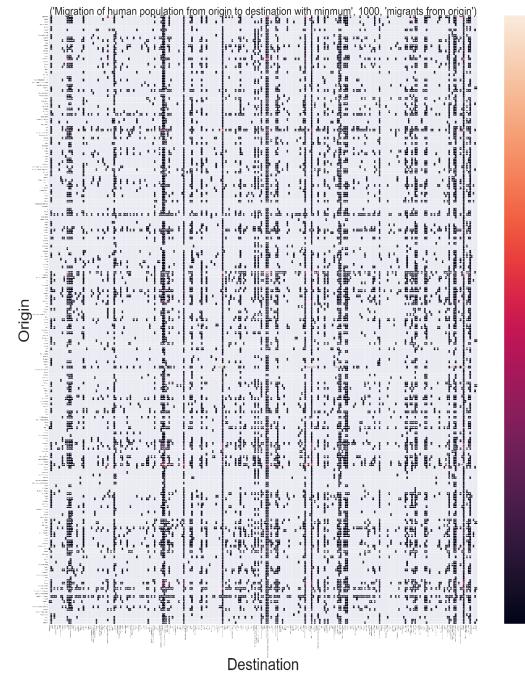


Fig 2. Minimum 1000 migrants from origin to destination.

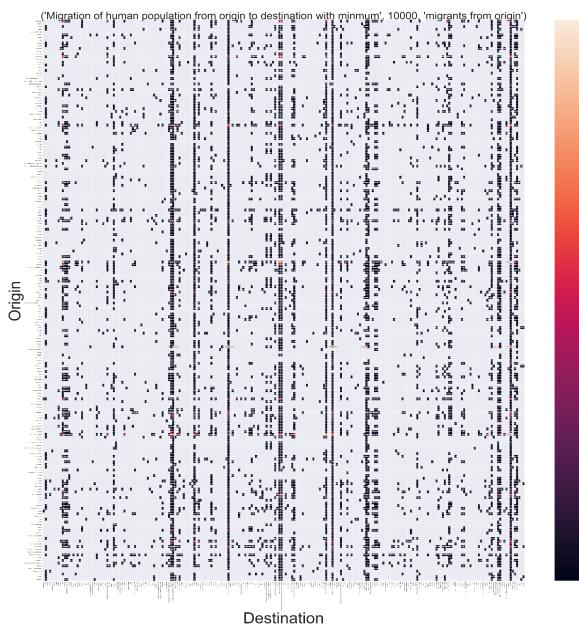


Fig 3. Minimum 10k migrants from origin to destination.

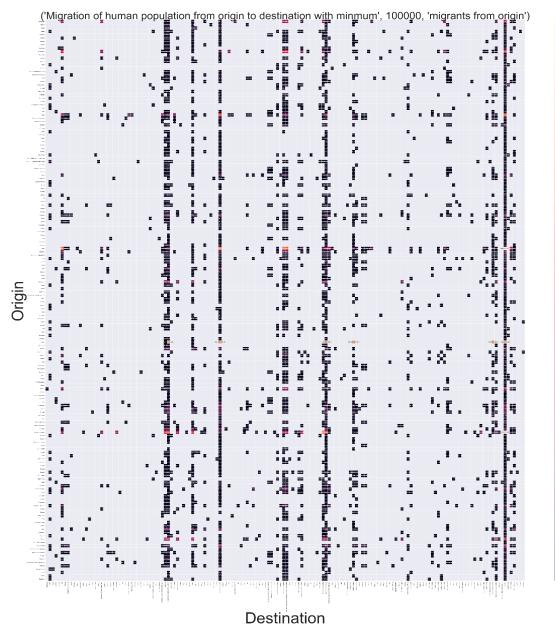


Fig 4. Minimum 100k migrants from origin to destination



Fig 5. Heatmap of minimum 10M migrants from origin to destination



1. The above figures of heatmap represent the number of migrants from origin to destination countries where in Fig 1 there were no minimum migrants considered, whereas in Fig 2,3,4,5 minimum migrants were 1000,10000,10k,10 Million respectively.
2. It could be noticed from figure 5 that the highest number of migrants were from Russia which were more than 76 million followed by Mexico, India and China.
3. Analyzing the Fig 5 it could be seen that the United States of America had the highest number of migrants from Mexico.
4. Major destination of Migrants from India and Mexico was in High income countries.

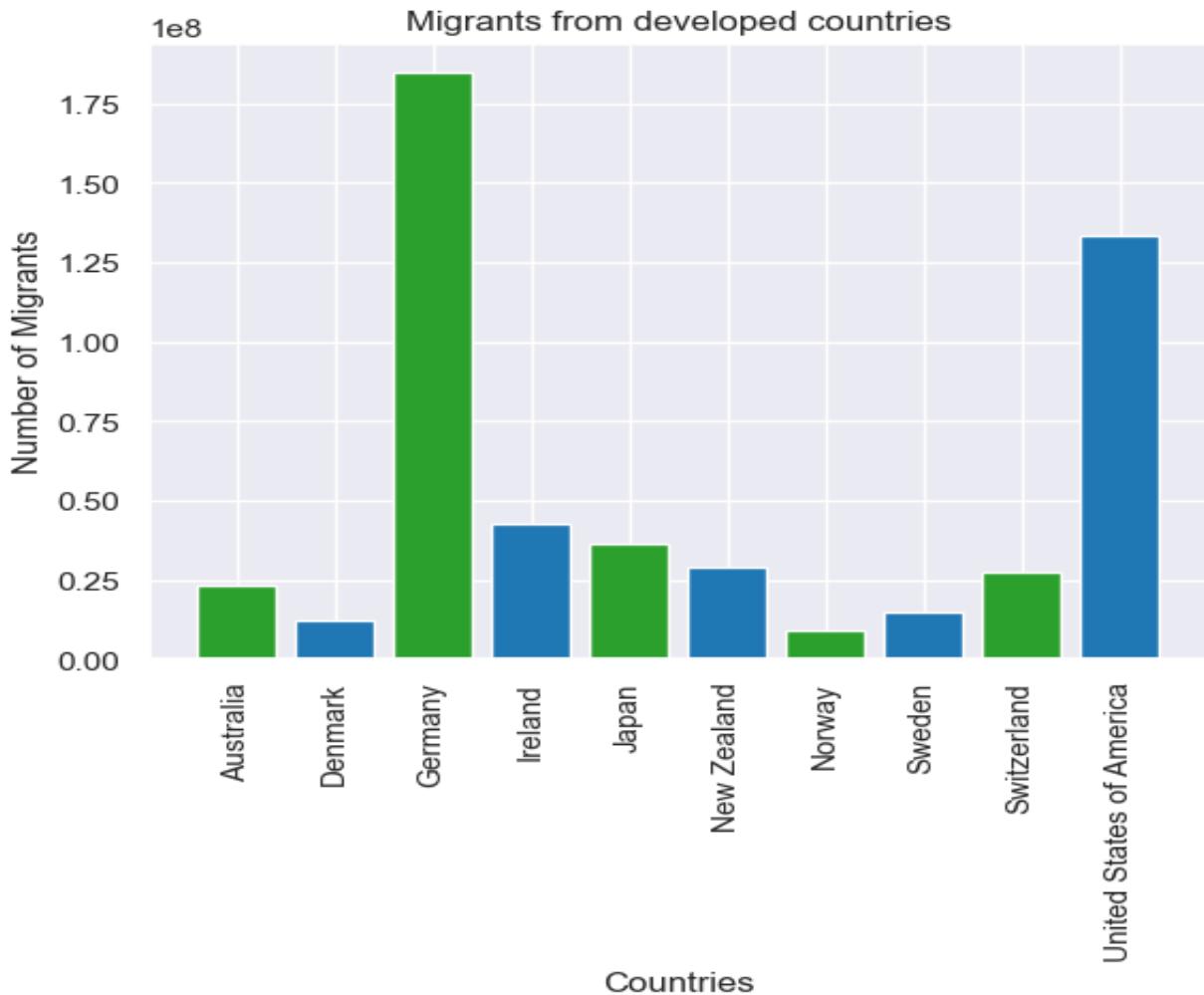


Fig 6. Bar chart showing Migrants from developed countries

1. The above graph depicts the most number of migrants from developed countries as defined by the United Nations.
2. Most number of migrants from developed countries were from Germany followed by the United States and Ireland and Japan having near to similar number of migrants.
3. Least number of migrants from developed countries were from Norway.

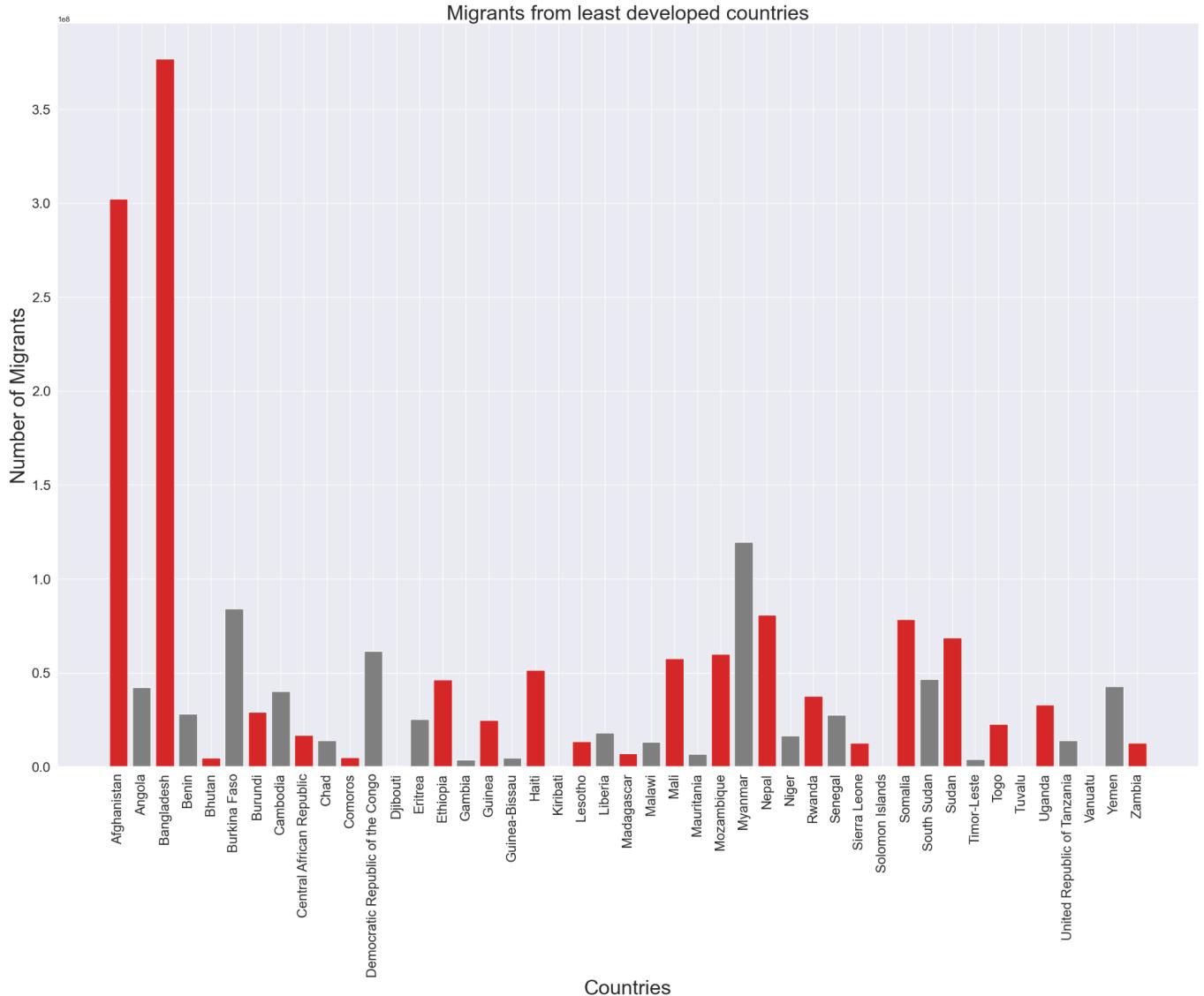


Fig 7. Bar chart showing Migrants from developed countries

1. The above bar graph represents the number of migrants from least developed countries as defined by the United Nations.
2. It could be analyzed from the graph that the most number of migrants were from Bangladesh then followed by Afghanistan, Myanmar and Burkina Faso country in South Africa.

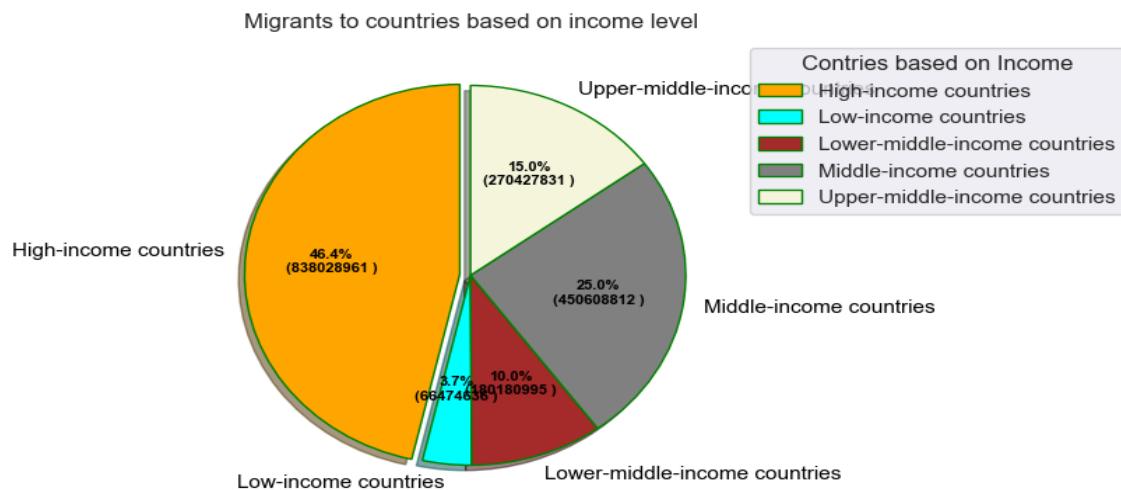


Fig 8. Pie chart showing migrants to countries based on income

The above pie chart in Fig 8 shows migrants to countries based on their income level. As it could be seen that high - income countries have the highest flow of migrants contributing to more than 46% followed by Middle Income countries with 25% which is surprising as Upper middle income countries had only 15% of migrants.

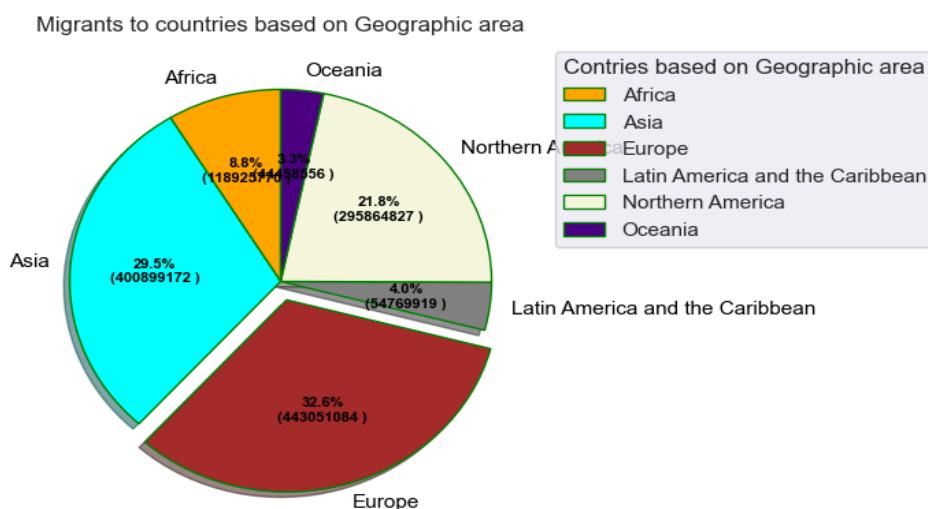


Fig 9. Pie chart showing migrants to countries based on Geographic area

The above pie chart in Fig 9 shows migrants to countries based on Geographic Area. As it could be seen that European countries have the highest flow of migrants contributing to more than 32% followed by Asia with 29.5% and the least number of migrants were in Oceania with 3.3%.

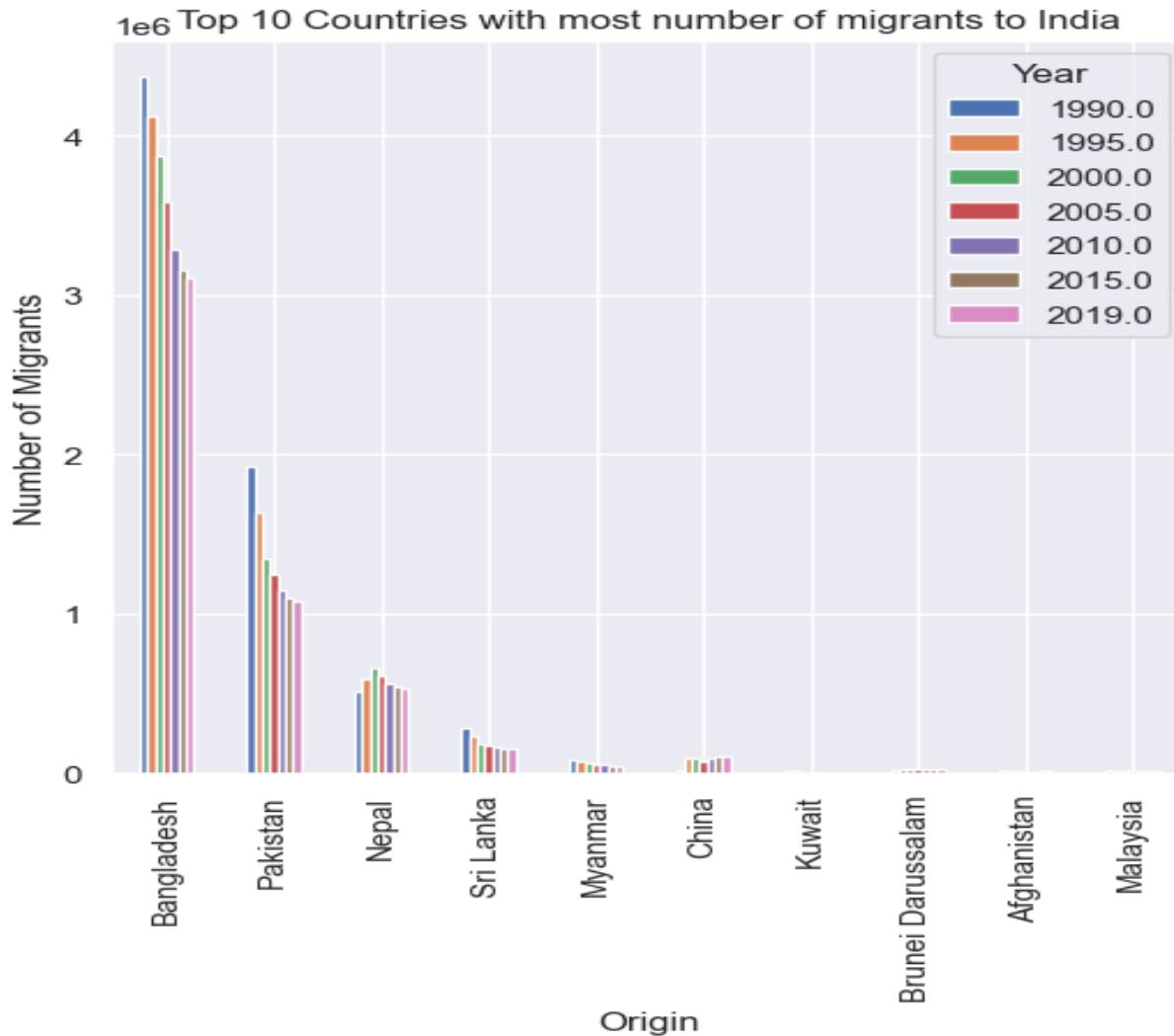


Fig 10. Grouped bar chart showing migrants to India

The above bar graph represents the most number of migrants from top 10 countries to India over the given years. Bangladesh has the most migrants to India in every given year. As it is the neighboring country of India whereas the second highest country with migrants to India is also neighboring country which is Pakistan.

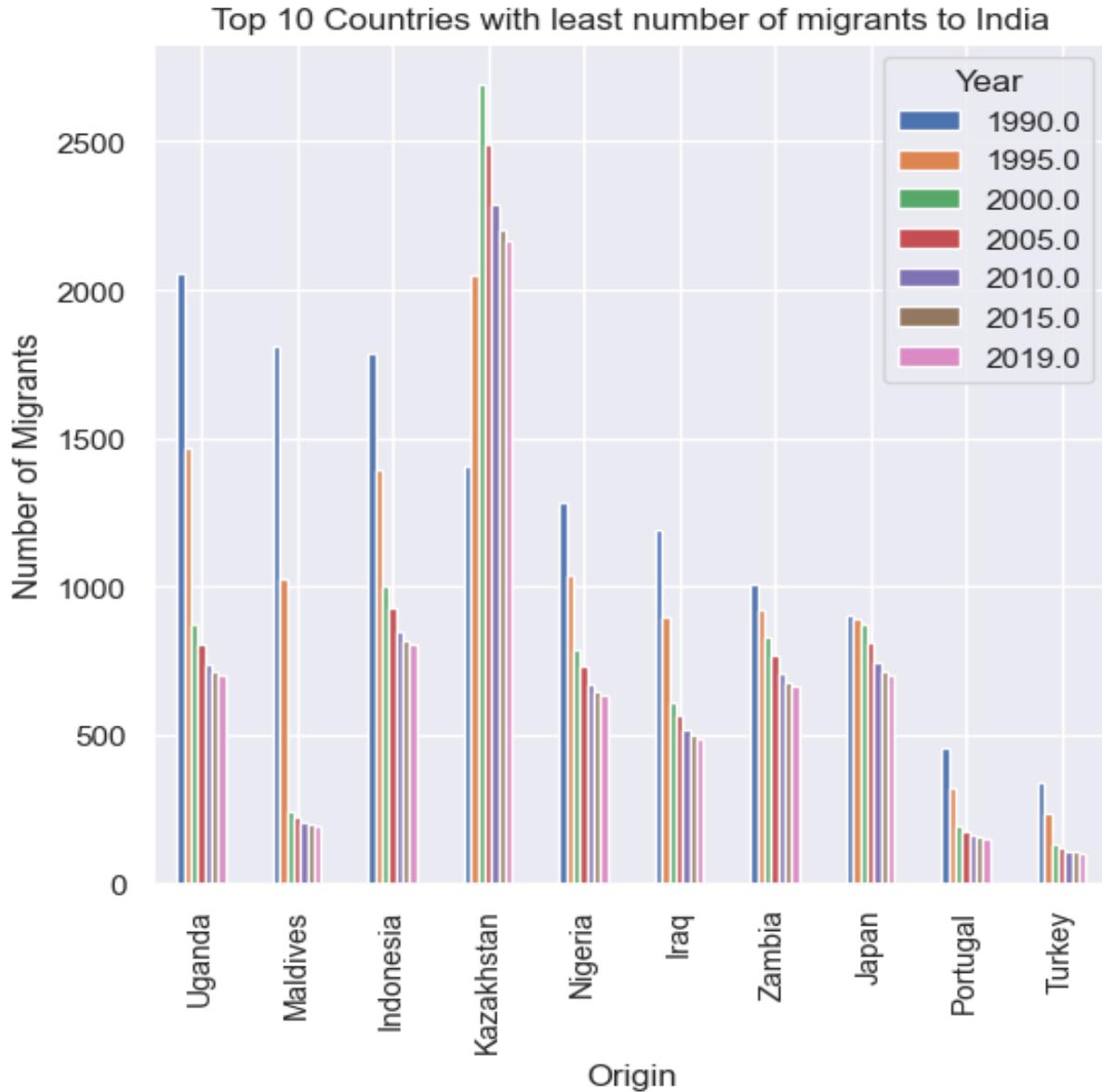


Fig 11. Grouped bar chart showing least migrants to India

The above bar graph represents the top 10 countries with the least number of migrants to India over the given years. Turkey has the fewest migrants to India. Also, it could be noticed that in the span of every 5 years migrants to India are decreasing continuously.



6. Conclusion

From the information obtained from this study, it can be seen that countries that are developed tend to have more migrants from the surrounding neighboring countries. As can be seen from the heatmap, the majority of immigrants to the United States came from Mexico. In India, the majority of immigrants came from nearby nations when the migration pattern was examined. It could be concluded that migration to developed countries and high income countries had the highest number of migrants as they offer more job opportunities and ease of living.