


# The Dataset: Immigration to Canada from 1980 to 2013

## 1. Data set yang Digunakan

Dataset yang digunakan berasal dari "International Migration Flows to and from Selected Countries - The 2015 Revision." Dataset ini memuat data tahunan tentang arus imigrasi internasional yang dicatat oleh negara tujuan. Di dalamnya ada informasi tentang orang yang masuk (inflow) dan keluar (outflow) dari suatu negara, berdasarkan tempat lahir, kewarganegaraan, atau tempat tinggal sebelumnya atau berikutnya. Data ini mencakup warga negara asing maupun warga setempat, dan meliputi 45 negara. Dalam analisis ini, yang menjadi fokus adalah data imigrasi Kanada.



United Nations

Population Division

Department of Economic and Social Affairs

International Migration Flows to and from Selected Countries: The 2015 Revision

POPDB/MIG/FlowRev.2015

December 2019 - Copyright © United Nations. All rights reserved

Suggested citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). International Migration Flows to and from Selected Countries: The 2015 Revision. United Nations database, POPDB/MIG/FlowRev.2015).

Reporting country	Criteria	Classification	Coverage	Type	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Armenia	Residence	Immigrants	Both																			
Armenia	Residence	Immigrants	Both																			
Australia	Residence	Immigrants	Both		90860	85600	92340	100510	96360	93440	92450	97770	104770	120940	137470	143710	143660	140420	141880	149360	158260	176960
Australia	Residence	Immigrants	Both		184290	212690	185200	153570	153530	172550	186680	221620	253860	238050	234050	237240	220460	187940	197940	221920	253940	261330
Austria	Citizenship	Immigrants	Citizens																			
Austria	Citizenship	Immigrants	Foreigners																			
Austria	Citizenship	Immigrants	Citizens																			
Austria	Citizenship	Immigrants	Foreigners																			
Austria	Residence	Immigrants	Both																			
Austria	Residence	Immigrants	Both																			
Azerbaijan	Residence	Immigrants	Both																			
Azerbaijan	Residence	Immigrants	Both																			
Belarus	Residence	Immigrants	Both																			
Belarus	Residence	Immigrants	Both																			
Belgium	Citizenship	Immigrants	Citizens		13326	20326	21497	21090	20562	20481	21110	22263	16244	16076	15937	18002	13258	13616	14422	16442	16384	16250
Belgium	Citizenship	Immigrants	Foreigners		36887	36970	37207	36170	32747	30431	29509	31017	28881	24737	24373	24597	29412	30462	31745	30616	32710	
Belgium	Citizenship	Immigrants	Citizens		7834	7979	8479	8910	8643	8600	8663	8656	10253	10620	12193	13330	11713	10707	10182	9812	9638	
Belgium	Citizenship	Immigrants	Foreigners		39746	33907	29498	29477	29684	28609	29466	31468	31343	35084	39338	41793	43312	48344	51034	45614	47716	
Belgium	Residence	Immigrants	Both		54094	49296	44059	43657	47002	47042	48959	49750	48484	54169	62662	67400	66763	63749				
Bulgaria	Citizenship	Immigrants	Citizens																			
Bulgaria	Citizenship	Immigrants	Foreigners																			
Bulgaria	Citizenship	Immigrants	Citizens																			
Bulgaria	Citizenship	Immigrants	Foreigners																			
Bulgaria	Residence	Immigrants	Both																			
Bulgaria	Residence	Immigrants	Both																			
Canada	Citizenship	Immigrants	Citizens																			
Canada	Citizenship	Immigrants	Foreigners		143137	128641	121175	89185	86272	84346	99351	102075	161585	191550	216448	232799	254763	256635	224381	212853	226075	216036
Croatia	Citizenship	Immigrants	Citizens																			
Croatia	Citizenship	Immigrants	Foreigners																			
Croatia	Residence	Immigrants	Both																			
Croatia	Residence	Immigrants	Both																			
Cyprus	Citizenship	Immigrants	Citizens																			
Cyprus	Citizenship	Immigrants	Foreigners																			
Cyprus	Citizenship	Immigrants	Citizens																			
Cyprus	Citizenship	Immigrants	Foreigners																			
Cyprus	Residence	Immigrants	Both																			
Cyprus	Residence	Immigrants	Both																			
Czech Republic	Citizenship	Immigrants	Citizens																			
Czech Republic	Citizenship	Immigrants	Foreigners																			
Czech Republic	Citizenship	Immigrants	Citizens																			
Czech Republic	Citizenship	Immigrants	Foreigners																			
Czech Republic	Residence	Immigrants	Both																			
Czech Republic	Residence	Immigrants	Both																			
Denmark	Citizenship	Immigrants	Citizens		17979	18650	17961	16849	16890	17862	18666	19981	23893	25447	23528	22167	22557	22350	23819	23521	24355	24336
Denmark	Citizenship	Immigrants	Foreigners		11845	11077	10014	9122	8355	9171	9375	10066	10455	8273	8645	10185	9081	9814	10891	11188	12009	14023
Denmark	Citizenship	Immigrants	Citizens		14526	14513	15255	15556	15742	16012	16389	16239	16605	19180	21000	21445	21893	22921	23984	24041	22918	22694
Denmark	Citizenship	Immigrants	Foreigners		15282	12862	12606	11433	12900	19219	20052	18217	16756	16996	17729	19744	19539	19623	20469	36238	28914	29953

## 2. Menampilkan Data set yang Digunakan

```
#Import Library yang digunakan
import pandas as pd
import matplotlib.pyplot as plt
import folium
from folium import plugins
from geopy.geocoders import Nominatim
!pip install openpyxl==3.0.9 folium geopy
!pip install --upgrade openpyxl
```

Requirement already satisfied: openpyxl==3.0.9 in  
/usr/local/lib/python3.10/dist-packages (3.0.9)  
Requirement already satisfied: folium in

```

/usr/local/lib/python3.10/dist-packages (0.17.0)
Requirement already satisfied: geopy in
/usr/local/lib/python3.10/dist-packages (2.4.1)
Requirement already satisfied: et-xmlfile in
/usr/local/lib/python3.10/dist-packages (from openpyxl==3.0.9) (1.1.0)
Requirement already satisfied: branca>=0.6.0 in
/usr/local/lib/python3.10/dist-packages (from folium) (0.8.0)
Requirement already satisfied: jinja2>=2.9 in
/usr/local/lib/python3.10/dist-packages (from folium) (3.1.4)
Requirement already satisfied: numpy in
/usr/local/lib/python3.10/dist-packages (from folium) (1.26.4)
Requirement already satisfied: requests in
/usr/local/lib/python3.10/dist-packages (from folium) (2.32.3)
Requirement already satisfied: xyzservices in
/usr/local/lib/python3.10/dist-packages (from folium) (2024.9.0)
Requirement already satisfied: geographiclib<3,>=1.52 in
/usr/local/lib/python3.10/dist-packages (from geopy) (2.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2>=2.9->folium)
(3.0.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->folium)
(3.4.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests->folium) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->folium)
(2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->folium)
(2024.8.30)
Requirement already satisfied: openpyxl in
/usr/local/lib/python3.10/dist-packages (3.0.9)
Collecting openpyxl
  Using cached openpyxl-3.1.5-py2.py3-none-any.whl.metadata (2.5 kB)
Requirement already satisfied: et-xmlfile in
/usr/local/lib/python3.10/dist-packages (from openpyxl) (1.1.0)
Using cached openpyxl-3.1.5-py2.py3-none-any.whl (250 kB)
Installing collected packages: openpyxl
  Attempting uninstall: openpyxl
    Found existing installation: openpyxl 3.0.9
    Uninstalling openpyxl-3.0.9:
      Successfully uninstalled openpyxl-3.0.9
Successfully installed openpyxl-3.1.5

# Mengubah path file Excel sesuai dengan lokasi di folder proyek
can_df = pd.read_excel('/content/Canada.xlsx',
                        sheet_name='Canada by Citizenship',
                        skiprows=range(20),
                        skipfooter=2

```

```

    )
# Menampilkan 5 baris pertama dari DataFrame
can_df.head()

{"type": "dataframe", "variable_name": "can_df"}

```

### 3. Menampilkan Fitur Data set dan Tipe Data yang Digunakan

```

can_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 43 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Type        195 non-null    object
1   Coverage    195 non-null    object
2   OdName      195 non-null    object
3   AREA        195 non-null    int64
4   AreaName    195 non-null    object
5   REG         195 non-null    int64
6   RegName     195 non-null    object
7   DEV         195 non-null    int64
8   DevName     195 non-null    object
9   1980        195 non-null    int64
10  1981        195 non-null    int64
11  1982        195 non-null    int64
12  1983        195 non-null    int64
13  1984        195 non-null    int64
14  1985        195 non-null    int64
15  1986        195 non-null    int64
16  1987        195 non-null    int64
17  1988        195 non-null    int64
18  1989        195 non-null    int64
19  1990        195 non-null    int64
20  1991        195 non-null    int64
21  1992        195 non-null    int64
22  1993        195 non-null    int64
23  1994        195 non-null    int64
24  1995        195 non-null    int64
25  1996        195 non-null    int64
26  1997        195 non-null    int64
27  1998        195 non-null    int64
28  1999        195 non-null    int64
29  2000        195 non-null    int64
30  2001        195 non-null    int64
31  2002        195 non-null    int64
32  2003        195 non-null    int64

```

33	2004	195	non-null	int64
34	2005	195	non-null	int64
35	2006	195	non-null	int64
36	2007	195	non-null	int64
37	2008	195	non-null	int64
38	2009	195	non-null	int64
39	2010	195	non-null	int64
40	2011	195	non-null	int64
41	2012	195	non-null	int64
42	2013	195	non-null	int64

dtypes: int64(37), object(6)

memory usage: 65.6+ KB

## 4. Menampilkan Statistic Descriptive Data yang Digunakan

```
can_df.describe()
```

```
{"type": "dataframe"}
```

## Interpretasi Data yang digunakan

### Descriptive Statistics Overview

- Count:**
  - Semua kolom dari **AREA**, **REG**, **DEV**, dan setiap tahun dari **1980 hingga 1996** memiliki jumlah data (count) sebanyak 195. Ini menunjukkan bahwa dataset ini lengkap dan tidak ada nilai yang hilang untuk periode waktu tersebut.
- Mean:**
  - Rata-rata jumlah imigran yang datang ke Kanada bervariasi untuk setiap tahun. Misalnya, pada tahun 1996, rata-rata jumlah imigran mencapai **1138.71**, yang menunjukkan bahwa jumlah imigran meningkat secara bertahap dari tahun ke tahun.
  - Rata-rata terendah terjadi pada tahun **1983** dengan **508.39**, sedangkan rata-rata tertinggi terjadi pada tahun **1996**.
- Standard Deviation (std):**
  - Nilai deviasi standar menunjukkan variasi dalam jumlah imigran. Sebagai contoh, pada tahun 1996, deviasi standar adalah **3613.34**, yang menunjukkan bahwa ada variasi yang signifikan dalam jumlah imigran yang masuk pada tahun tersebut.
  - Deviasi standar tertinggi muncul pada tahun **1991** dan **1992**, menandakan adanya fluktuasi yang besar dalam jumlah imigran pada tahun-tahun tersebut.
- Min & Max:**
  - Nilai minimum pada setiap tahun menunjukkan bahwa ada beberapa tahun di mana tidak ada atau sangat sedikit imigran (terutama di tahun-tahun awal seperti 1983).

- Sebaliknya, nilai maksimum menunjukkan adanya puncak imigrasi yang signifikan, seperti pada tahun **1991** dan **1992**, di mana jumlah imigran mencapai **27359** dan **23795**.
5. **Percentiles (25%, 50%, 75%):**
- Persentil ke-25 (25%) menunjukkan bahwa 25% dari data berada di bawah nilai tertentu, yang berarti ada banyak negara dengan jumlah imigran yang sangat rendah pada tahun-tahun awal.
  - Persentil ke-50 (median) memberikan gambaran tentang jumlah imigran yang "standar" yang dapat diharapkan. Sebagai contoh, pada tahun **1996**, median adalah **993**, yang menunjukkan bahwa setengah dari negara memiliki jumlah imigran di bawah angka tersebut.
  - Persentil ke-75 menunjukkan bahwa 75% dari negara memiliki jumlah imigran di bawah nilai tersebut, yang memberi gambaran lebih lengkap tentang distribusi jumlah imigran.

## Interpretasi Tren Imigrasi

- **Peningkatan Jumlah Imigran:** Secara umum, ada tren peningkatan jumlah imigran yang masuk ke Kanada dari tahun ke tahun. Hal ini dapat dilihat dari rata-rata yang semakin meningkat dan nilai maksimum yang meningkat tajam pada tahun-tahun tertentu.
- **Variabilitas dalam Data:** Tingginya deviasi standar menunjukkan bahwa ada variasi yang besar dalam jumlah imigran dari tahun ke tahun, yang bisa disebabkan oleh banyak faktor, termasuk perubahan kebijakan imigrasi, kondisi sosial ekonomi di negara asal, dan peristiwa global yang memengaruhi migrasi.
- **Pola Musiman:** Data ini juga menunjukkan pola musiman, di mana terdapat lonjakan imigrasi pada tahun-tahun tertentu. Misalnya, puncak yang terlihat pada tahun-tahun tertentu bisa jadi terkait dengan krisis atau peluang kerja yang ada di Kanada.

## 6. Perbandingan dan Tren Jumlah Imigran dari Vietnam , China, Algeria, India dan Zambadia Tahun 2000 - 2013

```
can_df.drop(['AREA', 'REG', 'DEV', 'Type', 'Coverage'], axis=1,
inplace=True)

can_df.rename(columns = {'OdName': 'Country', 'AreaName': 'Continent',
'RegName': 'Region'}, inplace=True)

can_df.head(2)

{"type": "dataframe", "variable_name": "can_df"}

can_df['Total'] = can_df.iloc[:,4:].sum(axis=1)
can_df.head(5)

{"type": "dataframe", "variable_name": "can_df"}
```

```

print(can_df.columns)
# Filter untuk memilih negara yang diinginkan
countries_of_interest = ['Algeria', 'China', 'India', 'Zambia']
can_df_filtered =
can_df[can_df['Country'].isin(countries_of_interest)]

# Ambil kolom tahun dari 2000 sampai 2013
# Pastikan menggunakan nama kolom yang benar, jika formatnya integer
hilangkan 'str()'
years = [2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013]
can_df_trend = can_df_filtered[['Country'] + years]

# Mengatur index untuk memudahkan plotting
can_df_trend = can_df_trend.set_index('Country').T

# Memvisualisasikan data
plt.figure(figsize=(12, 6))
for country in countries_of_interest:
    plt.plot(can_df_trend.index, can_df_trend[country], marker='o',
label=country)

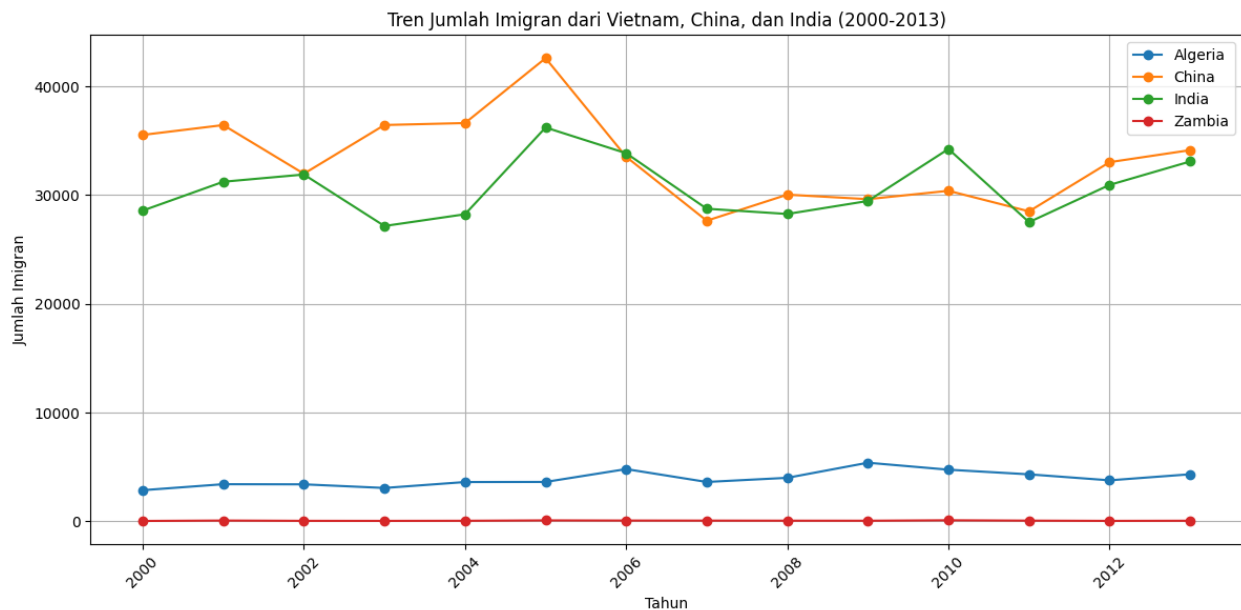
# Menambahkan judul dan label
plt.title('Tren Jumlah Imigran dari Vietnam, China, dan India (2000-
2013)')
plt.xlabel('Tahun')
plt.ylabel('Jumlah Imigran')
plt.xticks(rotation=45)
plt.legend()
plt.grid()

# Menampilkan plot
plt.tight_layout()
plt.show()

```

Index([	'Country',	'Continent',	'Region',	'DevName',
1980,				
	1981,	1982,	1983,	1984,
1985,				
	1986,	1987,	1988,	1989,
1990,				
	1991,	1992,	1993,	1994,
1995,				
	1996,	1997,	1998,	1999,
2000,				
	2001,	2002,	2003,	2004,
2005,				
	2006,	2007,	2008,	2009,
2010,				

```
2011, 2012, 2013, 'Total'],
dtype='object')
```



## Imigran setiap Tahun dari China Tahun 2000-2013

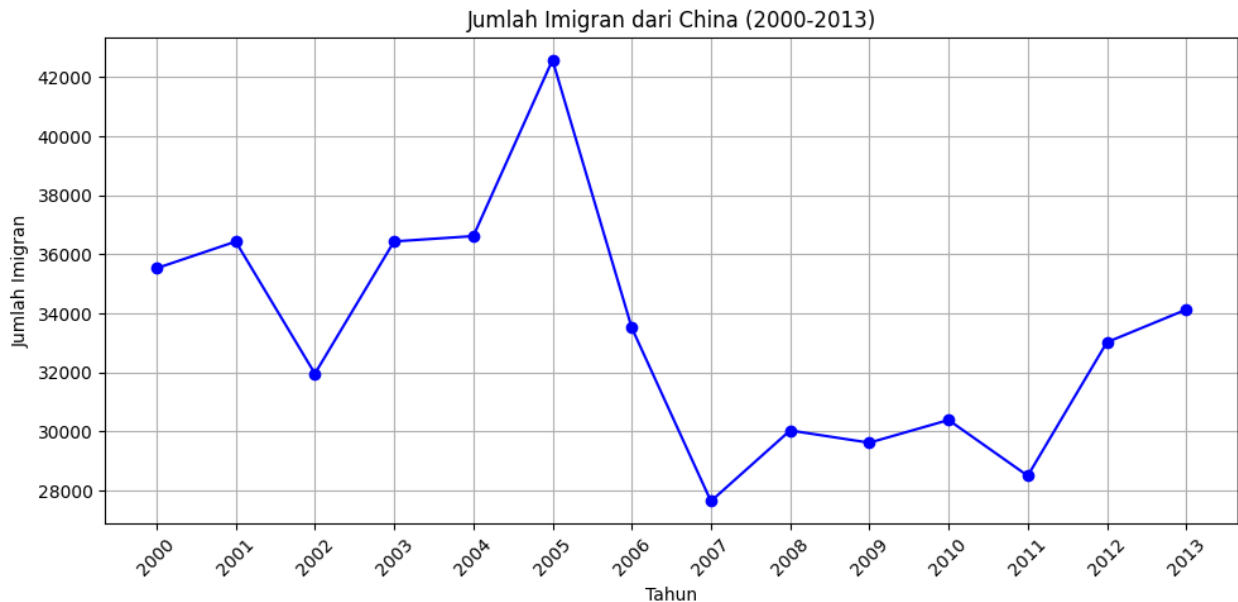
```
# Filter untuk mengambil data hanya dari China
china_data = can_df[can_df['Country'] == 'China']

# Ambil kolom tahun dari 2000 sampai 2013
years = [2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013]
china_immigrants = china_data[years].values.flatten() # Ambil data
tahun sebagai array

# Memvisualisasikan data
plt.figure(figsize=(10, 5))
plt.plot(years, china_immigrants, marker='o', color='blue')

# Menambahkan judul dan label
plt.title('Jumlah Imigran dari China (2000-2013)')
plt.xlabel('Tahun')
plt.ylabel('Jumlah Imigran')
plt.xticks(years, rotation=45) # Label tahun dirotasi
plt.grid()

# Menampilkan plot
plt.tight_layout()
plt.show()
```



## 7.Geospasial imigrasi ke Canada

```
# Menambahkan kolom 'Total' yang merupakan jumlah total imigran per
negara dari tahun 2000-2013
```

```
years = list(range(2000, 2014))
can_df['Total'] = can_df[years].sum(axis=1)
```

```
# Filter negara-negara dengan jumlah imigran terbesar
top_countries = can_df.nlargest(15, 'Total')
```

```
# Daftar negara dengan koordinat manual berdasarkan kolom 'Country'
```

```
country_coordinates = {
    'India': [20.5937, 78.9629],
    'China': [35.8617, 104.1954],
    'Philippines': [12.8797, 121.7740],
    'Pakistan': [30.3753, 69.3451],
    'United States': [37.0902, -95.7129],
    'United Kingdom': [55.3781, -3.4360],
    'Iran': [32.4279, 53.6880],
    'South Korea': [35.9078, 127.7669],
    'Sri Lanka': [7.8731, 80.7718],
    'France': [46.6034, 1.8883],
    'Mexico': [23.6345, -102.5528],
    'Jamaica': [18.1096, -77.2975],
    'Bangladesh': [23.6850, 90.3563],
    'Vietnam': [14.0583, 108.2772],
    'Germany': [51.1657, 10.4515],
    'Italy': [41.8719, 12.5674],
    'Lebanon': [33.8547, 35.8623],
    'Poland': [51.9194, 19.1451],
    'Russia': [61.5240, 105.3188],
```



```

'Egypt': [26.8206, 30.8025],
'Haiti': [18.9712, -72.2852],
'Portugal': [39.3999, -8.2245],
'Ukraine': [48.3794, 31.1656],
'Netherlands': [52.1326, 5.2913],
'Colombia': [4.5709, -74.2973],
'Japan': [36.2048, 138.2529],
'Brazil': [-14.2350, -51.9253],
'Turkey': [38.9637, 35.2433],
'Afghanistan': [33.9391, 67.7100],
'Indonesia': [-0.7893, 113.9213],
'Israel': [31.0461, 34.8516],
'Peru': [-9.1900, -75.0152],
'Greece': [39.0742, 21.8243],
'Romania': [45.9432, 24.9668],
'Venezuela': [6.4238, -66.5897],
'Chile': [-35.6751, -71.5430],
'Sweden': [60.1282, 18.6435],
'Malaysia': [4.2105, 101.9758],
'South Africa': [-30.5595, 22.9375],
'Guyana': [4.8604, -58.9302],
'Hungary': [47.1625, 19.5033],
'Argentina': [-38.4161, -63.6167],
'El Salvador': [13.7942, -88.8965],
'Belgium': [50.8503, 4.3517],
'Dominican Republic': [18.7357, -70.1627],
'Iraq': [33.2232, 43.6793],
'Trinidad and Tobago': [10.6918, -61.2225],
'Ecuador': [-1.8312, -78.1834],
'Hong Kong': [22.3193, 114.1694],
'Jordan': [30.5852, 36.2384],
'Nepal': [28.3949, 84.1240],
'Somalia': [5.1521, 46.1996],
'Barbados': [13.1939, -59.5432],
'Morocco': [31.7917, -7.0926],
'Syria': [34.8021, 38.9968],
'Switzerland': [46.8182, 8.2275],
'Spain': [40.4637, -3.7492],
'Thailand': [15.8700, 100.9925],
'Sudan': [12.8628, 30.2176],
'Norway': [60.4720, 8.4689]
}

# Buat peta dunia yang terpusat pada Kanada
world_map = folium.Map(location=[56.130, -106.35], zoom_start=2)

# Plot setiap negara ke peta
for idx, row in top_countries.iterrows():
    country = row['Country']
    if country in country_coordinates:

```

```
        folium.CircleMarker(
            location=country_coordinates[country],
            radius=row['Total'] / 100000, # Ukuran lingkaran sesuai
dengan jumlah imigran
            popup=folium.Popup(f"{country}: {row['Total']}
immigrants", parse_html=True),
            color='blue',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        ).add_to(world_map)

# Tampilkan peta
world_map

<folium.folium.Map at 0x7e4a7f7c3310>
```