

# Migration to Russia

September 11, 2021

```
[1]: #Import the ibm_db Python library
import ibm_db
# Db2 Service Credentials
dsn_hostname = "54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.
↳databases.appdomain.cloud" # e.g.: "54a2f15b-5c0f-46df-8954-7e38e612c2bd.
↳c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud"
dsn_uid = "qpc41961"          # e.g. "abc12345"
dsn_pwd = "PR69nCPpho9ZTSFf" # e.g. "7dBZ3wWt9XN6$o0J"

dsn_driver = "{IBM DB2 ODBC DRIVER}"
dsn_database = "BLUDB"          # e.g. "BLUDB"
dsn_port = "32733"             # e.g. "32733"
dsn_protocol = "TCPIP"         # i.e. "TCPIP"
dsn_security = "SSL"           #i.e. "SSL"

[2]: #Create database connection

dsn = (
    "DRIVER={0};"
    "DATABASE={1};"
    "HOSTNAME={2};"
    "PORT={3};"
    "PROTOCOL={4};"
    "UID={5};"
    "PWD={6};"
    "SECURITY={7};").format(dsn_driver, dsn_database, dsn_hostname, dsn_port,
↳dsn_protocol, dsn_uid, dsn_pwd,dsn_security)

try:
    conn = ibm_db.connect(dsn, "", "")
    print ("Connected to database: ", dsn_database, "as user: ", dsn_uid, "on_
↳host: ", dsn_hostname)

except:
    print ("Unable to connect: ", ibm_db.conn_errormsg() )
```

Connected to database: BLUDB as user: qpc41961 on host: 54a2f15b-5c0f-46df-89

54-7e38e612c2bd.clogj3sd0tgtu0lqde00.databases.appdomain.cloud

```
[3]: selectQuery = "select * from migration"

#Execute the statement
selectStmt = ibm_db.exec_immediate(conn, selectQuery)
ibm_db.fetch_both(selectStmt)
```

```
[3]: {'COUNTRY': 'AZERBAIJAN',
      0: 'AZERBAIJAN',
      'REGION': 'Post-Soviet states',
      1: 'Post-Soviet states',
      '2011': 22316,
      2: 22316,
      '2012': 22287,
      3: 22287,
      '2013': 23453,
      4: 23453,
      '2014': 26323,
      5: 26323,
      '2015': 24326,
      6: 24326,
      '2016': 24109,
      7: 24109,
      '2017': 25602,
      8: 25602,
      '2018': 26690,
      9: 26690,
      '2019': 34619,
      10: 34619,
      '2020': 32135,
      11: 32135,
      'TOTAL': 261860,
      12: 261860}
```

```
[4]: while ibm_db.fetch_row(selectStmt) != False:
      print (" Country", ibm_db.result(selectStmt, 0), " in 2011 is:", ibm_db.
      ↪result(selectStmt, "2011"))
```

```
Country ARMENIA in 2011 is: 32747
Country BELARUS in 2011 is: 10182
Country KAZAKHSTAN in 2011 is: 36474
Country KYRGYZSTAN in 2011 is: 41562
Country MOLDOVA, REPUBLIC in 2011 is: 19578
Country TAJIKISTAN in 2011 is: 35087
Country TURKMENISTAN in 2011 is: 4524
Country UZBEKISTAN in 2011 is: 64493
Country UKRAINE in 2011 is: 43586
```

Country ABKHAZIA in 2011 is: 2429  
 Country AUSTRALIA in 2011 is: 83  
 Country AUSTRIA in 2011 is: 60  
 Country ALBANIA in 2011 is: 29  
 Country ALGERIA in 2011 is: 76  
 Country ANGUILLA in 2011 is: 17  
 Country ANGOLA in 2011 is: 109  
 Country ARGENTINA in 2011 is: 19  
 Country AFGHANISTAN in 2011 is: 604  
 Country BANGLADESH in 2011 is: 51  
 Country BELIZE in 2011 is: 32  
 Country BELGIUM in 2011 is: 70  
 Country BENIN in 2011 is: 23  
 Country BULGARIA in 2011 is: 371  
 Country BOLIVIA in 2011 is: 12  
 Country BOSNIA AND HERZEGOVINA in 2011 is: 115  
 Country BOTSWANA in 2011 is: 27  
 Country BRAZIL in 2011 is: 34  
 Country BRITISH TERRITORY IN THE INDIAN OCEAN in 2011 is: 20  
 Country BURUNDI in 2011 is: 22  
 Country UNITED KINGDOM in 2011 is: 166  
 Country VENEZUELA in 2011 is: 27  
 Country VIRGIN ISLANDS, BRITISH in 2011 is: 14  
 Country VIRGIN ISLANDS, USA in 2011 is: 94  
 Country VIETNAM in 2011 is: 3294  
 Country GHANA in 2011 is: 55  
 Country GUINEA-BISSAU. in 2011 is: 31  
 Country GUINEA in 2011 is: 18  
 Country GERMANY in 2011 is: 4520  
 Country GREECE in 2011 is: 614  
 Country GEORGIA in 2011 is: 7325  
 Country DENMARK in 2011 is: 23  
 Country EGYPT in 2011 is: 228  
 Country ZAMBIA in 2011 is: 78  
 Country ZIMBABWE in 2011 is: 37  
 Country ISRAEL in 2011 is: 1240  
 Country INDIA in 2011 is: 1390  
 Country INDONESIA in 2011 is: 17  
 Country JORDAN in 2011 is: 145  
 Country IRAQ in 2011 is: 206  
 Country IRAN, ISLAMIC REPUBLIC in 2011 is: 171  
 Country IRELAND in 2011 is: 31  
 Country SPAIN in 2011 is: 201  
 Country ITALY in 2011 is: 250  
 Country YEMEN in 2011 is: 56  
 Country CAMBODIA in 2011 is: 12  
 Country CAMEROON in 2011 is: 52  
 Country CANADA in 2011 is: 192

Country KENYA in 2011 is: 58  
Country CYPRUS in 2011 is: 65  
Country CHINA in 2011 is: 7063  
Country COLOMBIA in 2011 is: 30  
Country CONGO in 2011 is: 65  
Country CONGO, DEMOCRATIC REPUBLIC in 2011 is: 30  
Country KOREA, PEOPLE'S DEMOCRATIC REPUBLIC in 2011 is: 1948  
Country KOREA, REPUBLIC in 2011 is: 165  
Country CAT D'IVOIRE in 2011 is: 31  
Country CUBA in 2011 is: 57  
Country LAO PEOPLE'S DEMOCRATIC REPUBLIC in 2011 is: 75  
Country LATVIA in 2011 is: 1350  
Country LEBANON in 2011 is: 108  
Country LIBYAN ARAB JAMAHIRIYA in 2011 is: 22  
Country LITHUANIA in 2011 is: 790  
Country REPUBLIC OF MACEDONIA in 2011 is: 54  
Country MALAYSIA in 2011 is: 28  
Country MALTA in 2011 is: 12  
Country MOROCCO in 2011 is: 293  
Country MEXICO in 2011 is: 23  
Country MONGOLIA in 2011 is: 164  
Country NAMIBIA in 2011 is: 94  
Country NEPAL in 2011 is: 15  
Country NIGERIA in 2011 is: 128  
Country NETHERLANDS in 2011 is: 78  
Country NEW ZEALAND in 2011 is: 25  
Country NORWAY in 2011 is: 62  
Country UNITED ARAB EMIRATES in 2011 is: 34  
Country PAKISTAN in 2011 is: 70  
Country PALESTINE in 2011 is: 88  
Country PERU in 2011 is: 35  
Country POLAND in 2011 is: 187  
Country PORTUGAL in 2011 is: 55  
Country RWANDA in 2011 is: 11  
Country ROMANIA in 2011 is: 20  
Country SAUDI ARABIA in 2011 is: 13  
Country SENEGAL in 2011 is: 10  
Country SERBIA in 2011 is: 600  
Country SYRIAN ARAB REPUBLIC in 2011 is: 451  
Country SLOVAKIA in 2011 is: 63  
Country SLOVENIA in 2011 is: 14  
Country USA in 2011 is: 947  
Country SUDAN in 2011 is: 47  
Country THAILAND in 2011 is: 71  
Country TANZANIA in 2011 is: 27  
Country TOGO in 2011 is: 10  
Country TUNISIA in 2011 is: 107  
Country TURKEY in 2011 is: 1832

Country URUGUAY in 2011 is: 11  
Country FINLAND in 2011 is: 266  
Country FRANCE in 2011 is: 322  
Country CROATIA in 2011 is: 41  
Country CHAD in 2011 is: 16  
Country MONTENEGRO in 2011 is: 55  
Country CZECH REPUBLIC in 2011 is: 157  
Country CHILE in 2011 is: 10  
Country SWITZERLAND in 2011 is: 59  
Country SWEDEN in 2011 is: 69  
Country SRI LANKA in 2011 is: 29  
Country ECUADOR in 2011 is: 20  
Country EQUATORIAL GUINEA in 2011 is: 60  
Country ESTONIA in 2011 is: 1588  
Country ETHIOPIA in 2011 is: 21  
Country SOUTH OSSETIA in 2011 is: 657  
Country JAPAN in 2011 is: 87

```
[5]: import pandas
import ibm_db_dbi
```

```
[6]: #connection for pandas
conn = ibm_db_dbi.Connection(conn)
```

```
[7]: selectQuery = "select * from migration"

#retrieve the query results into a pandas dataframe
dfmigr = pandas.read_sql(selectQuery, conn)

#print just the LNAME for first row in the pandas data frame
dfmigr.head(5)
```

```
[7]:
```

	COUNTRY	REGION	2011	2012	2013	2014	2015	2016	\
0	AZERBAIJAN	Post-Soviet states	22316	22287	23453	26323	24326	24109	
1	ARMENIA	Post-Soviet states	32747	36978	42361	46515	45670	43929	
2	BELARUS	Post-Soviet states	10182	16564	15748	17878	17741	14590	
3	KAZAKHSTAN	Post-Soviet states	36474	45506	51958	59096	65750	69356	
4	KYRGYZSTAN	Post-Soviet states	41562	34597	30388	28539	26045	28202	
	2017	2018	2019	2020	TOTAL				
0	25602	26690	34619	32135	261860				
1	46898	46442	71984	56511	470035				
2	21282	19045	18428	14536	165994				
3	71680	72141	86311	64493	622765				
4	41165	44408	53810	45676	374392				

```
[8]: years = list(map(str, range(2011, 2021)))
     years
```

```
[8]: ['2011',
      '2012',
      '2013',
      '2014',
      '2015',
      '2016',
      '2017',
      '2018',
      '2019',
      '2020']
```

```
[9]: %matplotlib inline

import matplotlib as mpl
import matplotlib.pyplot as plt
```

```
[10]: print(plt.style.available)
      mpl.style.use(['ggplot']) # optional: for ggplot-like style
```

```
['Solarize_Light2', '_classic_test_patch', 'bmh', 'classic', 'dark_background',
'fast', 'fivethirtyeight', 'ggplot', 'grayscale', 'seaborn', 'seaborn-bright',
'seaborn-colorblind', 'seaborn-dark', 'seaborn-dark-palette', 'seaborn-
darkgrid', 'seaborn-deep', 'seaborn-muted', 'seaborn-notebook', 'seaborn-paper',
'seaborn-pastel', 'seaborn-poster', 'seaborn-talk', 'seaborn-ticks', 'seaborn-
white', 'seaborn-whitegrid', 'tableau-colorblind10']
```

```
[11]: #dfmigr=[dfmigr.index == 'VIETNAM']
```

```
[12]: print(plt.style.available)
      mpl.style.use(['ggplot']) # optional: for ggplot-like style
```

```
['Solarize_Light2', '_classic_test_patch', 'bmh', 'classic', 'dark_background',
'fast', 'fivethirtyeight', 'ggplot', 'grayscale', 'seaborn', 'seaborn-bright',
'seaborn-colorblind', 'seaborn-dark', 'seaborn-dark-palette', 'seaborn-
darkgrid', 'seaborn-deep', 'seaborn-muted', 'seaborn-notebook', 'seaborn-paper',
'seaborn-pastel', 'seaborn-poster', 'seaborn-talk', 'seaborn-ticks', 'seaborn-
white', 'seaborn-whitegrid', 'tableau-colorblind10']
```

```
[13]: dfmigr.set_index('COUNTRY', inplace=True)
```

```
[14]: dfmigr.loc['VIETNAM']
```

```
[14]: REGION      Asia
      2011      3294
```

```
2012      3653
2013      3852
2014      3853
2015      4012
2016      3735
2017      3912
2018      3981
2019      6742
2020      6206
TOTAL     43240
Name: VIETNAM, dtype: object
```

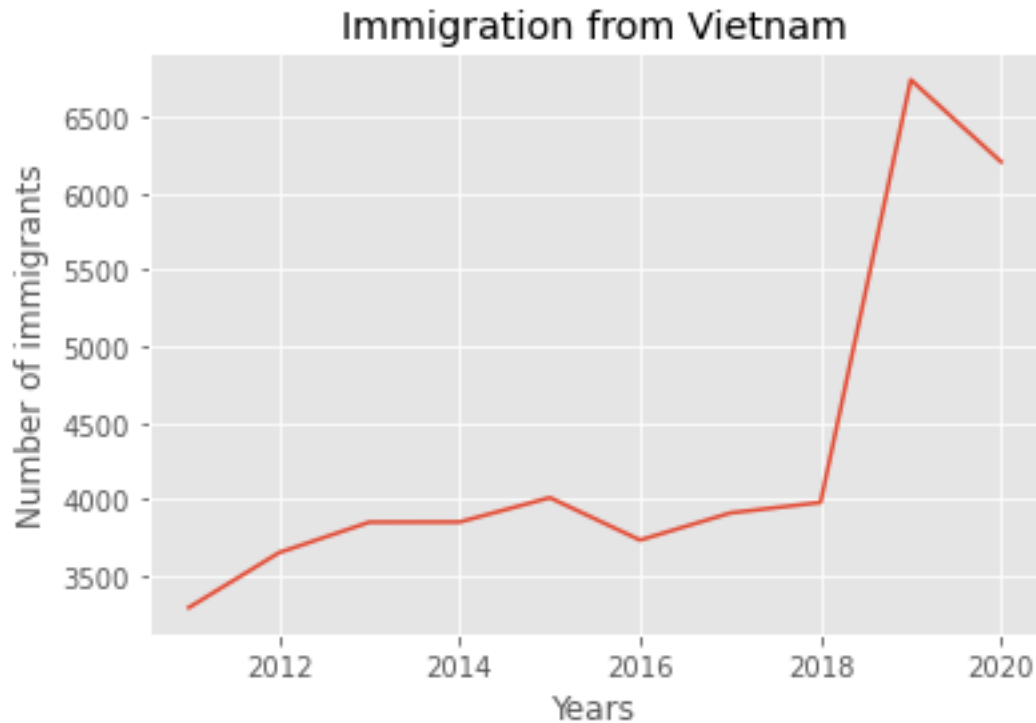
```
[15]: vietnam = dfmigr.loc['VIETNAM', years]
      vietnam.head()
```

```
[15]: 2011      3294
      2012      3653
      2013      3852
      2014      3853
      2015      4012
      Name: VIETNAM, dtype: object
```

```
[16]: vietnam.index = vietnam.index.map(int)
      vietnam.plot(kind='line')

      plt.title('Immigration from Vietnam')
      plt.ylabel('Number of immigrants')
      plt.xlabel('Years')

      plt.show()
```



```
[17]: # group countries by continents and apply sum() function

#print(type(dfmigr.groupby('REGION', axis=0)))

#

selectQuery = "select * from migration where REGION in ('Africa', 'Asia', 'Countries and regions of Europ', 'USA' )"

#retrieve the query results into a pandas dataframe
dfmigr = pandas.read_sql(selectQuery, conn)

#print just the LNAME for first row in the pandas data frame
dfmigr.head(5)
df_continents = dfmigr.groupby('REGION', axis=0).sum()
df_continents
```

```
[17]:
```

	2011	2012	2013	2014	2015	2016	\
REGION							
Africa	1565	2215	3004	2979	4235	4686	
Asia	16097	21297	23059	27508	28284	30496	
Countries and regions of Europ	11434	12411	14799	13074	12718	11973	



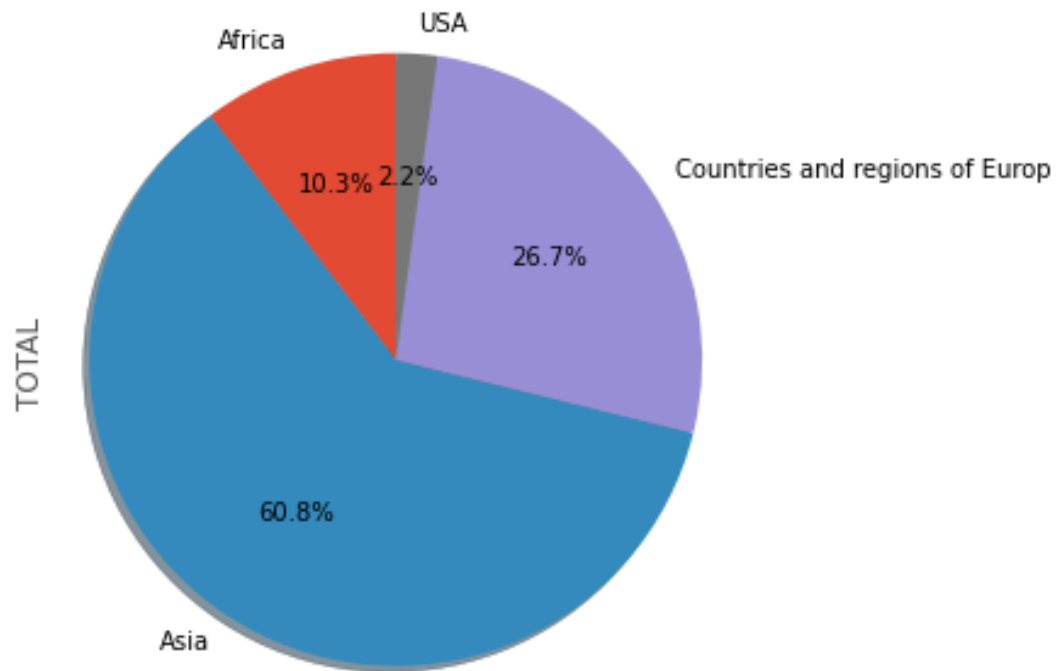
USA	947	1122	954	989	1084	1137
	2017	2018	2019	2020	TOTAL	
REGION						
Africa	5565	5026	9134	8268	46677	
Asia	30254	23824	45725	27904	274448	
Countries and regions of Europ	12277	11148	11965	8582	120381	
USA	1240	960	857	526	9816	

```
[18]: df_continents['TOTAL'].plot(kind='pie',
                                   figsize=(5, 8),
                                   autopct='%1.1f%%', # add in percentages
                                   startangle=90,      # start angle 90° (Africa)
                                   shadow=True,        # add shadow
                                   )

plt.title('Immigration to Russia by Continent [2011 - 2020]')
plt.axis('equal') # Sets the pie chart to look like a circle.

plt.show()
```

## Immigration to Russia by Continent [2011 - 2020]



```
[43]: #Visualize the trend of total immigranton to Russia (all countries combined) ↵  
↪for the years 2011-2020  
#and regression line on the scatter plot.  
  
import pandas as pd # primary data structure library  
# we can use the sum() method to get the total population per year  
df_tot = pd.DataFrame(dfmigr[years].sum(axis=0))  
  
# change the years to type int (useful for regression later on)  
df_tot.index = map(int, df_tot.index)  
  
# reset the index to put in back in as a column in the df_tot dataframe
```

```
df_tot.reset_index(inplace = True)

# rename columns
df_tot.columns = ['year', 'total']

# view the final dataframe
df_tot
```

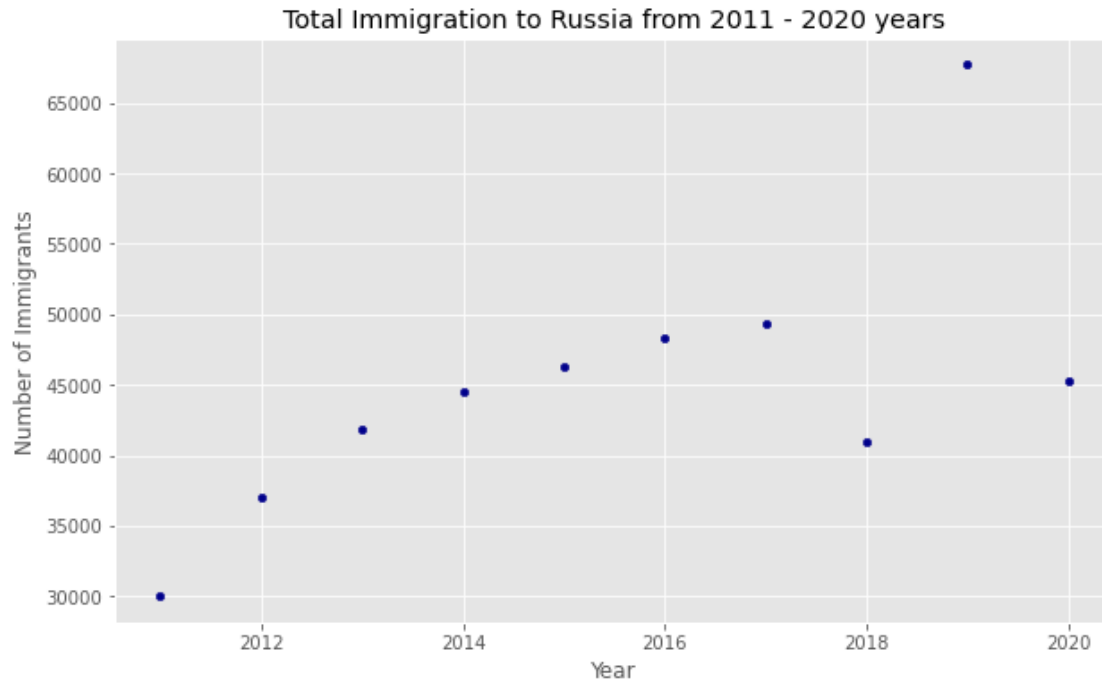
```
[43]:
```

	year	total
0	2011	30043
1	2012	37045
2	2013	41816
3	2014	44550
4	2015	46321
5	2016	48292
6	2017	49336
7	2018	40958
8	2019	67681
9	2020	45280

```
[44]: df_tot.plot(kind='scatter', x='year', y='total', figsize=(10, 6),
    ↪color='darkblue')

plt.title('Total Immigration to Russia from 2011 - 2020 years')
plt.xlabel('Year')
plt.ylabel('Number of Immigrants')

plt.show()
```



```
[24]: import numpy as np
```

```
[46]: #So let's try to plot a linear line of best fit, and use it to predict the
      ↪number of immigrants in 2021.
      #get the equation of line of best fit
      #deg: Degree of fitting polynomial. 1 = linear
      x = df_tot['year']      # year on x-axis
      y = df_tot['total']     # total on y-axis
      fit = np.polyfit(x, y, deg=1)

      fit
```

```
[46]: array([ 2.20378182e+03, -4.39659005e+06])
```

```
[47]: df_tot.plot(kind='scatter', x='year', y='total', figsize=(10, 6),
      ↪color='darkblue')

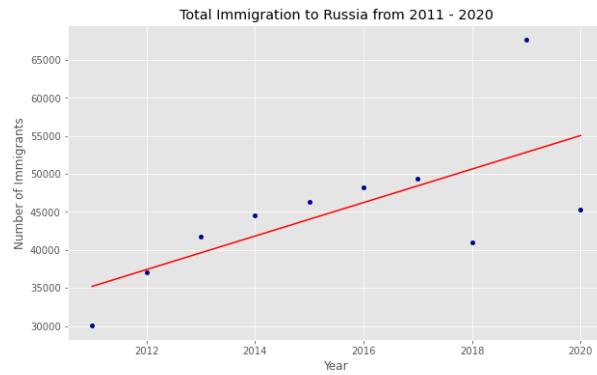
      plt.title('Total Immigration to Russia from 2011 - 2020')
      plt.xlabel('Year')
      plt.ylabel('Number of Immigrants')

      # plot line of best fit
      plt.plot(x, fit[0] * x + fit[1], color='red') # recall that x is the Years
      plt.annotate('y={0:.0f} x + {1:.0f}'.format(fit[0], fit[1]), xy=(2000, 15000))
```

```
plt.show()
```

```
# print out the line of best fit
```

```
'No. Immigrants = {0:.0f} * Year + {1:.0f}'.format(fit[0], fit[1])
```



```
[47]: 'No. Immigrants = 2204 * Year + -4396590'
```

```
[48]: #Immigrants = 2204 * Year - 4396590  
      #Immigrants = 2204 * 2021 - 4396590  
      #Immigrants = 57,694
```

```
[ ]:
```