**RUSSIAN DEMOGRAPHICS AND POPULATION ANALYSIS**

**Analysts**

*Archana Rajagopalan*

*Elena*

*Joe O*

*RK*

**DESCRIPTION**

After lot of discussion and brainstorming, we as a team finally decided to do the analysis of a country’s population. The dataset which we used has data from (1990 to 2017) which is approx. 27 years of data. The data contains the demographic features like natural population growth(npg),birth rate, death rate, population in the respective region and also the corresponding year.

**Data Content**

The data used for analysis: russiandemography.csv. The file contains 2380 rows and has 7 columns.

**Keys for the columns**

|  |  |
| --- | --- |
| **“year”** | *Year (1990 – 2017)* |
| **“region”** | *name of a federal subject of Russia. It could be oblast, republic, krai, autonomous okrug, federal city and a single autonomous oblast* |
| **“npg”** | *natural population growth by 1000 people. Calculating as the difference between birth rate and death rate* |
| **“birth rate”** | *number of births by 1000 people* |
| **“death rate”** | *number of deaths by 1000 people* |
| **“Migratory growth”** | *migratory population growth by 1000 people* |
| **“population”** | *calculating according to last census* |

Analysis Steps

1. Import the libraries numpy, pandas and matplotlib
2. Read the file used for analysis
3. Create a dataframe to hold the file contents
4. Get the top records from dataframe

**Analysis based on birth rate:**

* + 1. Get the birth rate mean group by year. This gives the average birth rate over the years.
    2. Get the birth rate mean group by region. This gives the average birth rate over the region.
    3. Based on the plots we will be able to identify which year the birth rate was highest and which region has the highest birth rate.
    4. The above graph will also be helpful when we do analysis based on region and the population.

**Analysis based on death rate:**

1. Get the death rate mean group by year. This gives the average death rate over the years.
2. Get the death rate mean group by region. This gives the average death rate over the region.
3. Based on the plots we will be able to identify which year the death rate was highest and which region has the highest death rate.
4. The above graph will also be helpful when we do analysis based on region and the population.

**Analysis based on the difference between birth rate and death rate:**

1. Get the difference mean group by year. This gives the average death rate over the years.
2. Get the difference mean group by region. This gives the average death rate over the region.
3. Based on the plots we will be able to identify which year the difference mean rate was highest and which region has the highest difference mean rate.
4. The above graph will also be helpful when we do analysis based on region and the population.

**Analysis based on the npg:**

1. Get the npg mean group by year and sort by ascending. This gives the average npg rate over the years.
2. Get the npg mean group by region and sort by ascending. This gives the average npg rate over the region.
3. Based on the plots we will be able to identify which year the npg mean rate was highest and which region has the highest npg mean rate.
4. The above graph will also be helpful when we do analysis based on region and the population.

**Analysis based on the top 10 population:**

1. Get the sum of the population over the regions and populate the graph based on ascending.
2. The result set will provide the top 10 highest population regions

**Analysis based on the top 10 based on death rate:**

1. Get the sum of the death rate over the regions and populate the graph based on ascending.
2. The result set will provide the top 10 highest death rate regions

**Analysis of 2 cities based on other factors with the highest peaks in population graph:**

1. Analyse and plot the graph of Moscow city and Krasnoyarsk Krai over the years based on factors like npg,birth rate,death rate and migratory growth

**Analysis of 2 cities based on other factors with the highest peaks in death rate graph:**

1. Analyse and plot the graph of Pskov Oblast and Tver Oblast city over the years based on factors like npg,birth rate,death rate and migratory growth

**Scatter plot analysis based on birth rate mean and npg and death rate and npg and apply the population correlation methods like “Pearson” and “Kendall”.**

Findings from Analysis

1. The population is on the increase during the year 1990 but it started to decrease in 1991 and 1992.It started to grow in the negative from 1993 to 2011. in 2012 it started to increase but started to decrease in 2017.
2. The npg graph over the region shows the max npg is on Chechen Republic and the least is at Pskov Oblast. There is no considerable change in the npg for the region Kamchatka krai.
3. The migratory growth graph shows that there is no migration happened or neglible on few regions or it could also mean
4. the file has null values.
5. Based on the graph for the population in the top 10 cities starts to decline in 2005 and gets its peak sometime around 2008 and gets down in 2011 and starts its peak sometime in 2014. after 2015 the rate is steady.
6. Moscow being the capital continues the population growth, but the decline happens after 2005 and has not increased in great number after that. But Krasnoyarsk Krai sees the population growth peak after 2010. This could be because of migration.
7. The comparison of Moscow with Krasnoyarsk krai region based on migration growth would give us more info.
8. The different methods of correlation produce different result values but overall birth rate and npg shows positive whereas

death rate and npg is negative.