

World Population Changes In The Last 50 Years

This project studies the world's countries' populations over the last 50 years (1970-2022). We will study the changes in countries population and population density over the years and learn how these changes impacted/can impact these countries. Is there a-lot of over-populated countries and are we going to run out of space in this planet anytime soon? We'll find out at the end of this study

Plan of work

We are going to be using python and several of it's data analysis libraries to investigate and study a dateset of the populations of every country in the world the last 50 years. We have data on the years:

- 1970
 - 1980
 - 1990
 - 2000
 - 2010
 - 2015
 - 2020
 - 2022
- load the necessary libraries and datasets; preprocess and optimize them.
 - Study and understand the dataset in hands in order to be able to extract the most out of it.
 - Dig into the data, find interesting facts and conclusions about different countires and regions.
 - Sum up the findings and answering some questions.

Lets start then

Initilization: Data Preprocessing and Optimization

We'll start by loading the necessary python libraries and the dataset about the world populations and have a look at it.

In [1]:

```
import pandas as pd
import scipy.stats
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
! pip install plotly==5.10.0
import plotly.express as px
```

Requirement already satisfied: plotly==5.10.0 in d:\anaconda\lib\site-packages (5.10.0)
Requirement already satisfied: tenacity>=6.2.0 in d:\anaconda\lib\site-packages (from plotly==5.10.0) (8.0.1)

In [2]:

```
world_pops = pd.read_csv('C:/Users/ameer/Desktop/Practicum/World Population/world_population.csv')
```

In [3]:

```
world_pops.sort_values(by = '2022 Population', ascending = False).head()
```

Out[3]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km²)	Density (per km²)	Growth Rate	World Population Percentage	
	41	1	CHN	China	Beijing	Asia	1425887337	1424929781	1393715448	1348191368	1264099069	1153704252	982372466	822534450	9706961	146.8933	1.0000	17.88
	92	2	IND	India	New Delhi	Asia	1417173173	1396387127	1322866505	1240613620	1059633675	870452165	696828385	557501301	3287590	431.0675	1.0068	17.77
	221	3	USA	United States	Washington, D.C.	North America	338289857	335942003	324607776	311182845	282398554	248083732	223140018	200328340	9372610	36.0935	1.0038	4.24
	93	4	IDN	Indonesia	Jakarta	Asia	275501339	271857970	259091970	244016173	214072421	182159874	148177096	115228394	1904569	144.6529	1.0064	3.45
	156	5	PAK	Pakistan	Islamabad	Asia	235824862	227196741	210969298	194454498	154369924	115414069	80624057	59290872	881912	267.4018	1.0191	2.96

Everything looks fine at the first glance, and as expected: China, India and the USA are the world's most populous countries today. (The list above is sorted according to 2022 populations). Anyway, lets look for any problematic values in the dataset before we move on to our next step.

In [4]:

```
# checking for null values
print(world_pops.isnull().sum())
```

```
Rank      0
CCA3      0
Country   0
Capital   0
Continent 0
2022 Population
2020 Population
2015 Population
2010 Population
2000 Population
1990 Population
1980 Population
1970 Population
Area (km²)
Density (per km²)
Growth Rate
World Population Percentage
dtype: int64
```

Great, we have no nulls, no missing values. Lets have a look at the columns of our dataset:

In [5]:

```
world_pops.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 17 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   Rank                                234 non-null    int64
 1   CCA3                                234 non-null    object
 2   Country                            234 non-null    object
 3   Capital                            234 non-null    object
 4   Continent                          234 non-null    object
 5   2022 Population                    234 non-null    int64
 6   2020 Population                    234 non-null    int64
 7   2015 Population                    234 non-null    int64
 8   2010 Population                    234 non-null    int64
 9   2000 Population                    234 non-null    int64
10   1990 Population                    234 non-null    int64
11   1980 Population                    234 non-null    int64
12   1970 Population                    234 non-null    int64
13   Area (km²)                         234 non-null    int64
14   Density (per km²)                  234 non-null    float64
15   Growth Rate                        234 non-null    float64
16   World Population Percentage        234 non-null    float64
dtypes: float64(3), int64(10), object(4)
memory usage: 31.2+ KB
```

Everything looks fine here as-well. Except for, the 'Area (km²)' and 'Density (per km²)' column names, they are too complicated to work with so we are going to change them to 'Area' and 'Density' and assume it is per km².

```
In [6]: # renaming some of our complicated column names
world_pops.rename(columns = {'Area (km²)': 'Area', 'Density (per km²)': 'Density'}, inplace = True)
world_pops.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  ---
0   Rank                                  234 non-null    int64
1   CCA3                                  234 non-null    object
2   Country                              234 non-null    object
3   Capital                              234 non-null    object
4   Continent                            234 non-null    object
5   2022 Population                      234 non-null    int64
6   2020 Population                      234 non-null    int64
7   2015 Population                      234 non-null    int64
8   2010 Population                      234 non-null    int64
9   2000 Population                      234 non-null    int64
10  1990 Population                      234 non-null    int64
11  1980 Population                      234 non-null    int64
12  1970 Population                      234 non-null    int64
13  Area                                  234 non-null    int64
14  Density                              234 non-null    float64
15  Growth Rate                          234 non-null    float64
16  World Population Percentage          234 non-null    float64
dtypes: float64(3), int64(10), object(4)
memory usage: 31.2+ KB
```

Okay, after inspecting the data thoroughly, everything works and looks good. We're good to go and start studying our populations!

Exploratory Data Analysis

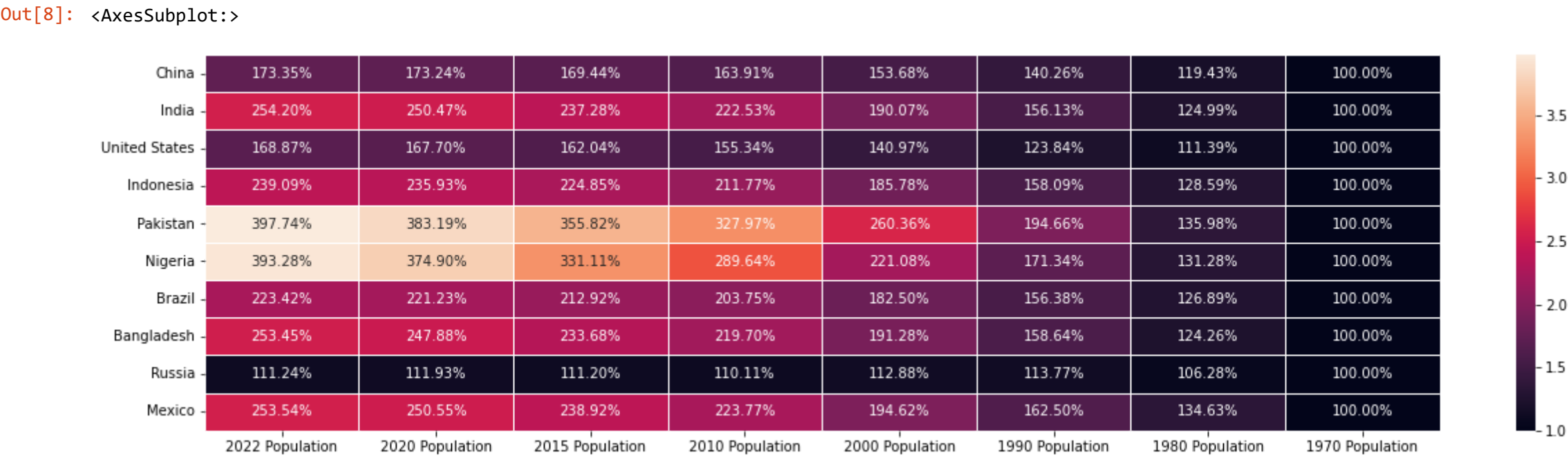
At this stage, we are going to study the most populous countries in 2022 and see how their populations have changed over the years, then we'll present the changes in a heatmap for a clearer understanding of how much each country have changed since 1970 up until 2022:

```
In [7]: # filtering the 10 most populous countries in 2022
top_10_pops = world_pops.sort_values(by = '2022 Population', ascending = False).head(10)
top_10_pops
```

Out[7]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area	Density	Growth Rate	World Population Percentage	
	41	1	CHN	China	Beijing	Asia	1425887337	1424929781	1393715448	1348191368	1264099069	1153704252	982372466	9706961	146.8933	1.0000	17.88	
	92	2	IND	India	New Delhi	Asia	1417173173	1396387127	1322866505	1240613620	1059633675	870452165	696828385	557501301	3287590	431.0675	1.0068	17.77
	221	3	USA	United States	Washington, D.C.	North America	338289857	335942003	324607776	311182845	282398554	248083732	223140018	200328340	9372610	36.0935	1.0038	4.24
	93	4	IDN	Indonesia	Jakarta	Asia	275501339	271857970	259091970	244016173	214072421	182159874	148177096	115228394	1904569	144.6529	1.0064	3.45
	156	5	PAK	Pakistan	Islamabad	Asia	235824862	227196741	210969298	194454498	154369924	115414069	80624057	59290872	881912	267.4018	1.0191	2.96
	149	6	NGA	Nigeria	Abuja	Africa	218541212	208327405	183995785	160952853	122851984	95214257	72951439	55569264	923768	236.5759	1.0241	2.74
	27	7	BRA	Brazil	Brasilia	South America	215313498	213196304	205188205	196353492	175873720	150706446	122288383	96369875	8515767	25.2841	1.0046	2.70
	16	8	BGD	Bangladesh	Dhaka	Asia	171186372	167420951	157830000	148391139	129193327	107147651	83929765	67541860	147570	1160.0350	1.0108	2.15
	171	9	RUS	Russia	Moscow	Europe	144713314	145617329	144668389	143242599	146844839	148005704	138257420	130093010	17098242	8.4636	0.9973	1.81
	131	10	MEX	Mexico	Mexico City	North America	127504125	125998302	120149897	112532401	97873442	81720428	67705186	50289306	1964375	64.9082	1.0063	1.60

```
In [8]: plt.figure(figsize = (20, 5))
necc_data = top_10_pops[['2022 Population', '2020 Population', '2015 Population', '2010 Population', '2000 Population', '1990 Population', '1980 Population', '1970 Population']]
final_data = necc_data.iloc[:,0:].div(necc_data['1970 Population'], axis=0)
sns.heatmap(data = final_data, yticklabels = top_10_pops['Country'], annot = True, linewidths = 1, fmt='%.2%')
```



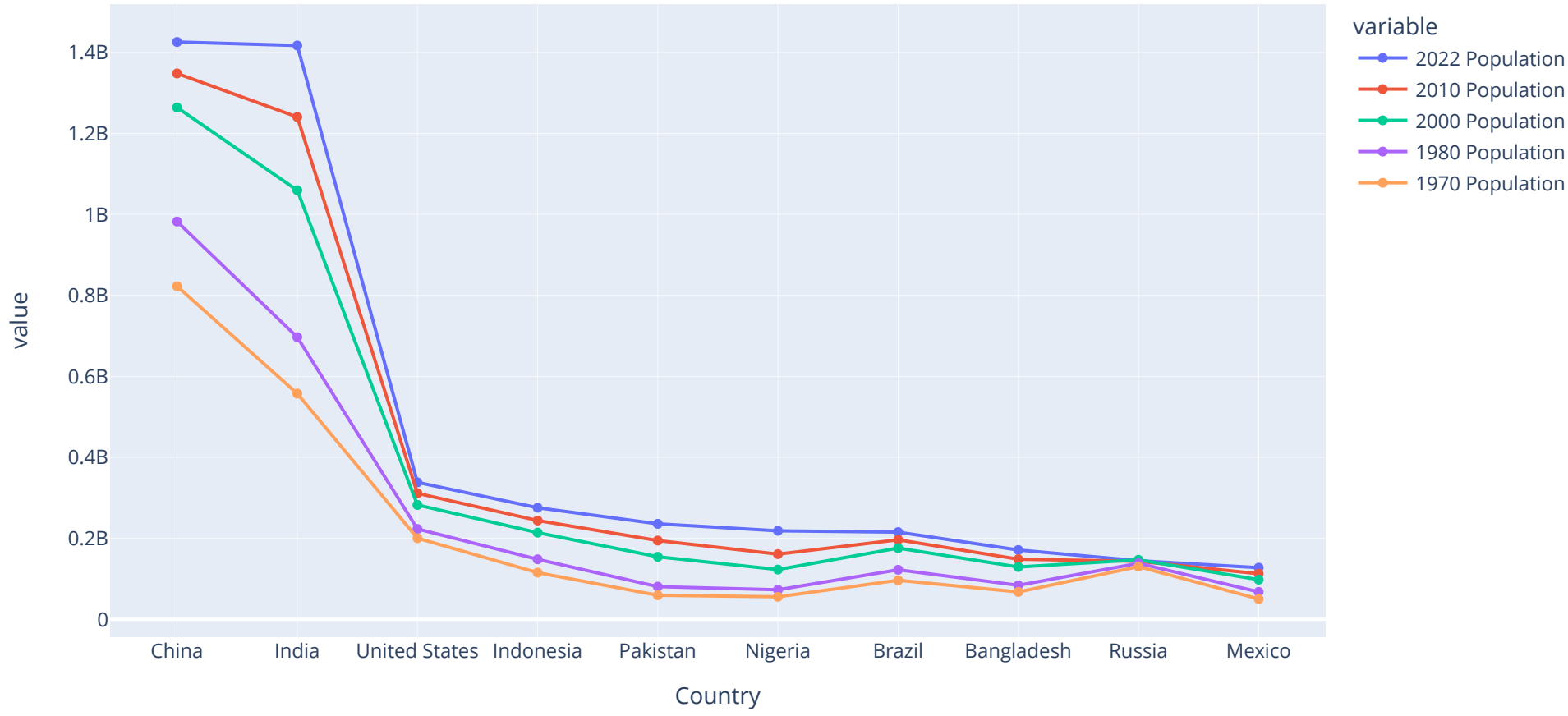
The heatmap above represents the changes in the most populous countries' populations over the past 50 years relative to the year 1970. This means, if a country has the value 200% in 2022, then it's population had doubled since 1970!

Here's a couple things we learn from the heatmap above:

- Pakistan, Nigeria showed the greatest change in population where their population in 2022 represents almost 400% (or x4) of their population in the 1970
- Russia impressively managed to keep a stable population for 50 straight years with only 11% increase in population (It even had population drops from 1990 to 2010)
- Despite the fact that USA and China are the most populous countries, their population increase was among the lowest.

Enough talk about percentages, lets talk numbers:

```
In [9]: px.line(top_10_pops, x = 'Country', y = ['2022 Population', '2010 Population', '2000 Population', '1980 Population', '1970 Population'], markers = True)
```

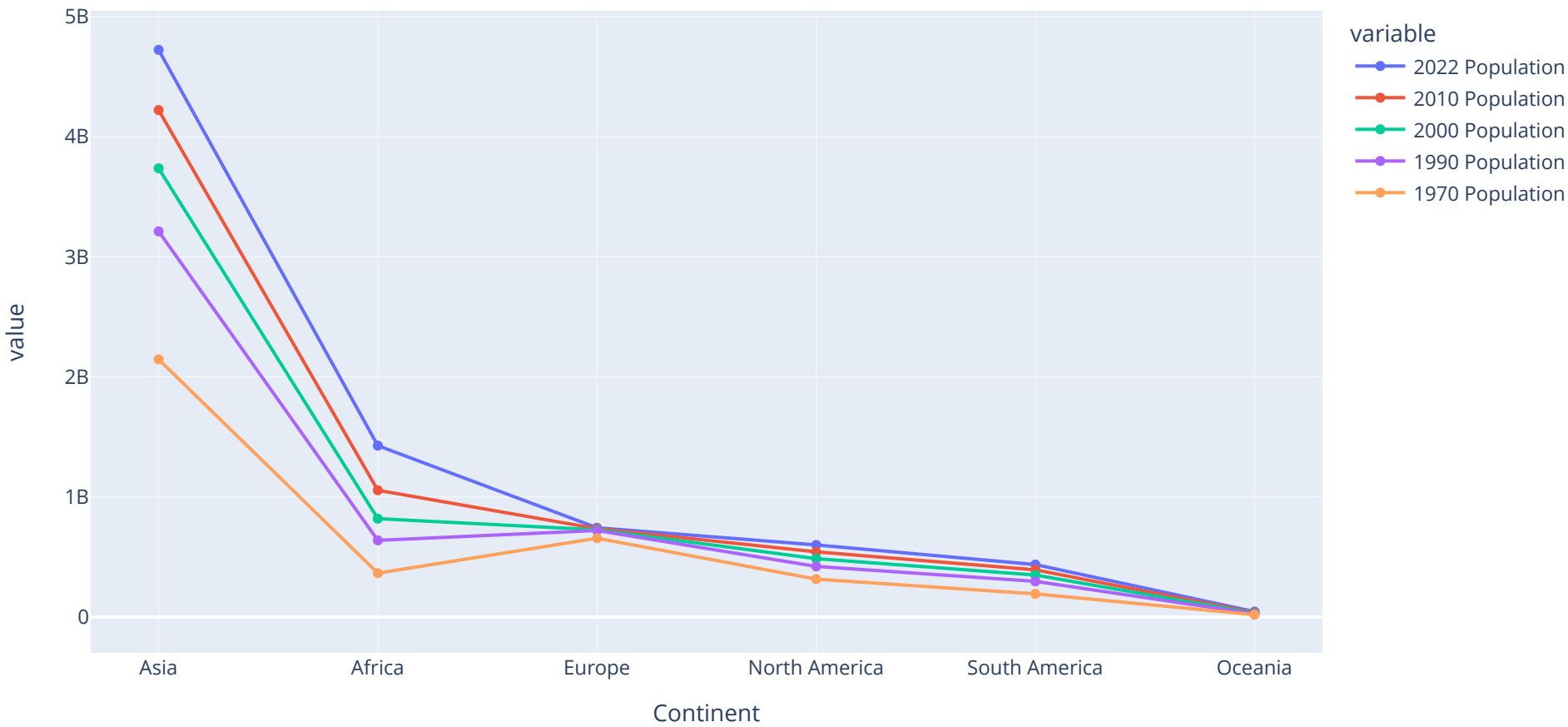


The order of the most populous countries was the very same since 1970, with no major changes over years 1970-2010. Interestingly, India was catching up to China, becoming closer in population each decade, and in 2022 we can see that China's population and India's population are almost equal (~1.4 Billion each).

Everything that we noticed in the heatmap we presented above, can be seen here too. Russia appearing as a one dot in the graph above indicating almost no change in their population, Nigeria and Pakistan quadrupling their populations from 55.5M to 218.5M and 59.2M to 235.8M respectively and India's massive (254%) population increase from 557 million up to 1.4 billion!

Next: We'll learn which continents are the most populous and how different are continent's population sizes from each other.

```
In [10]: continent_pops = world_pops[['Continent', '2022 Population', '2010 Population', '2000 Population', '1990 Population', '1970 Population', 'Area']]
continent_pops_grouped = continent_pops.groupby('Continent').sum()
continent_pops_grouped = continent_pops_grouped.reset_index().sort_values('2022 Population', ascending = False)
px.line(continent_pops_grouped, x = 'Continent', y = ['2022 Population', '2010 Population', '2000 Population', '1990 Population', '1970 Population'], markers = True)
```

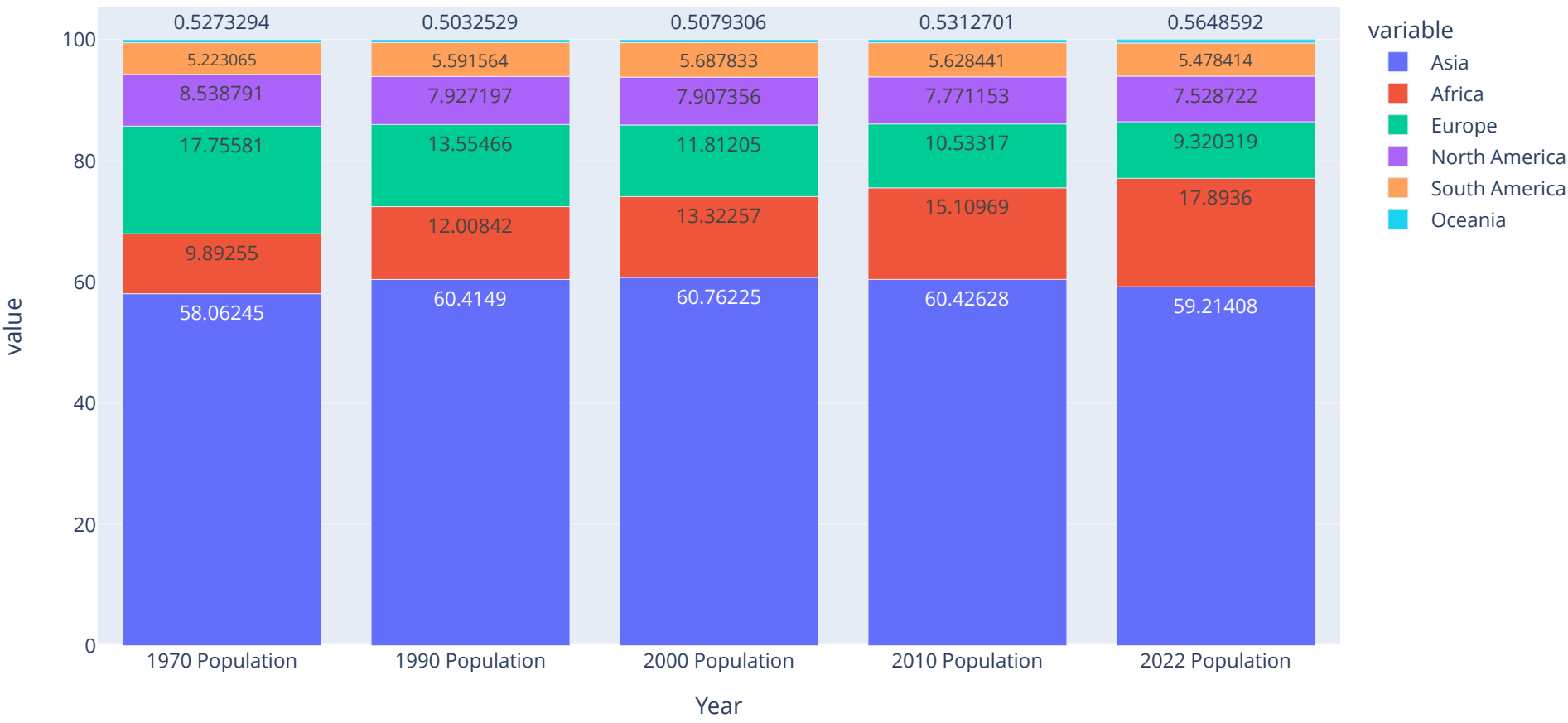


Asia was always the most populous continent, as expected, thanks to 2 massive contributors; China and India. Europe was the second most populous continent up until the 2000s (If you filter the graph above to keep only population data about 1990 and 2000 we can see that Africa surpassed Europe at this stage) and then Africa took the second place and kept growing fast ever since, while Europe's growth on the other hand was way slower.

Asia and Africa growth rate is currently very high relative to the rest of the continents, we can tell this by the size of the gaps between each point at the graph; we see big gaps for Africa and Asia while the gaps for the rest of the continents are way smaller.

Next: Lets see the impact of each continent on the population of the world in the period of time given.

```
In [11]: continent_pops_pivot = continent_pops.pivot_table(values = ['2022 Population', '2010 Population', '2000 Population', '1990 Population', '1970 Population'], columns = 'Continent', aggfunc = 'sum')
continent_pops_pivot = continent_pops_pivot.reset_index()
continent_pops_pivot = continent_pops_pivot.rename(columns = { 'index':'Year'})
continent_pops_pivot_perc = continent_pops_pivot.copy()
for v in range(1,7):
    for i in range(5):
        continent_pops_pivot_perc.iloc[i,v] = continent_pops_pivot.iloc[i,v]/continent_pops_pivot.iloc[i,1:].sum()*100
continent_pops_pivot_perc
px.bar(continent_pops_pivot_perc, x = 'Year', y = ['Asia', 'Africa', 'Europe', 'North America', 'South America', 'Oceania'], text_auto = True)
```



Apearently, the world population distribution was somewhat stable over the past 50 years. There is an exception though; We see that in 1970, Europe represented 17.75% of the world population while Africa represented a mere 9.8% percent. But this changed over the years, Africa now represents 17.89% of the 2022 world population while Europe only represents 9.3% of the world population, a switch has happened between these continents and we could notice it on the previous graph as well.

Next: We studied populations per country, then per continent and we found out which countries were the most populous, which continents were the most populous but we haven't talked about population density yet, are the most populous countries the densier countries as well? Lets learn this next.

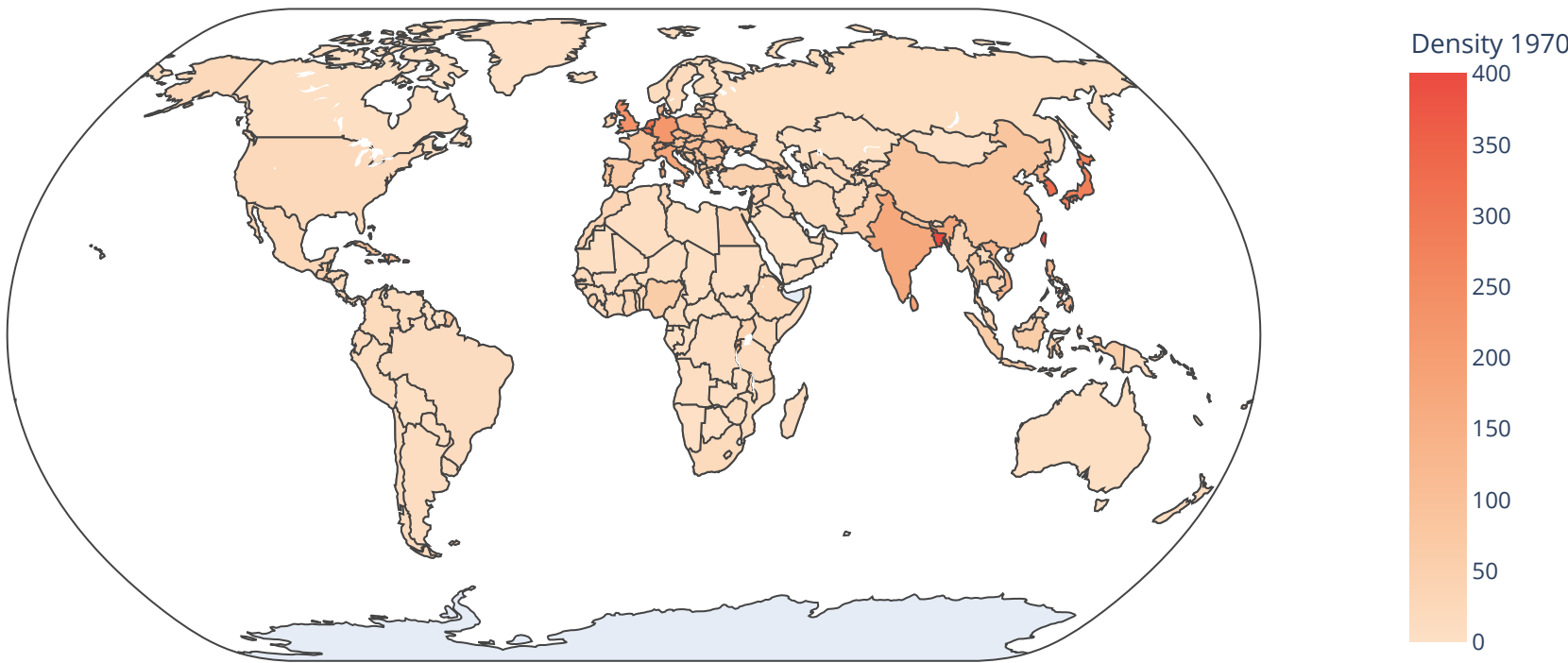
```
In [12]: world_pops = world_pops.rename(columns = {'Density':'Density 2022'})
world_pops['Density 1970'] = world_pops['1970 Population']/world_pops['Area']
print('Average density in 1970 was {:.2f} person per km², Average density in 2022 is {:.2f} person per km²'.format(world_pops['Density 1970'].mean(), world_pops['Density 2022'].mean()))
```

Average density in 1970 was 227.63 person per km², Average density in 2022 is 452.13 person per km²

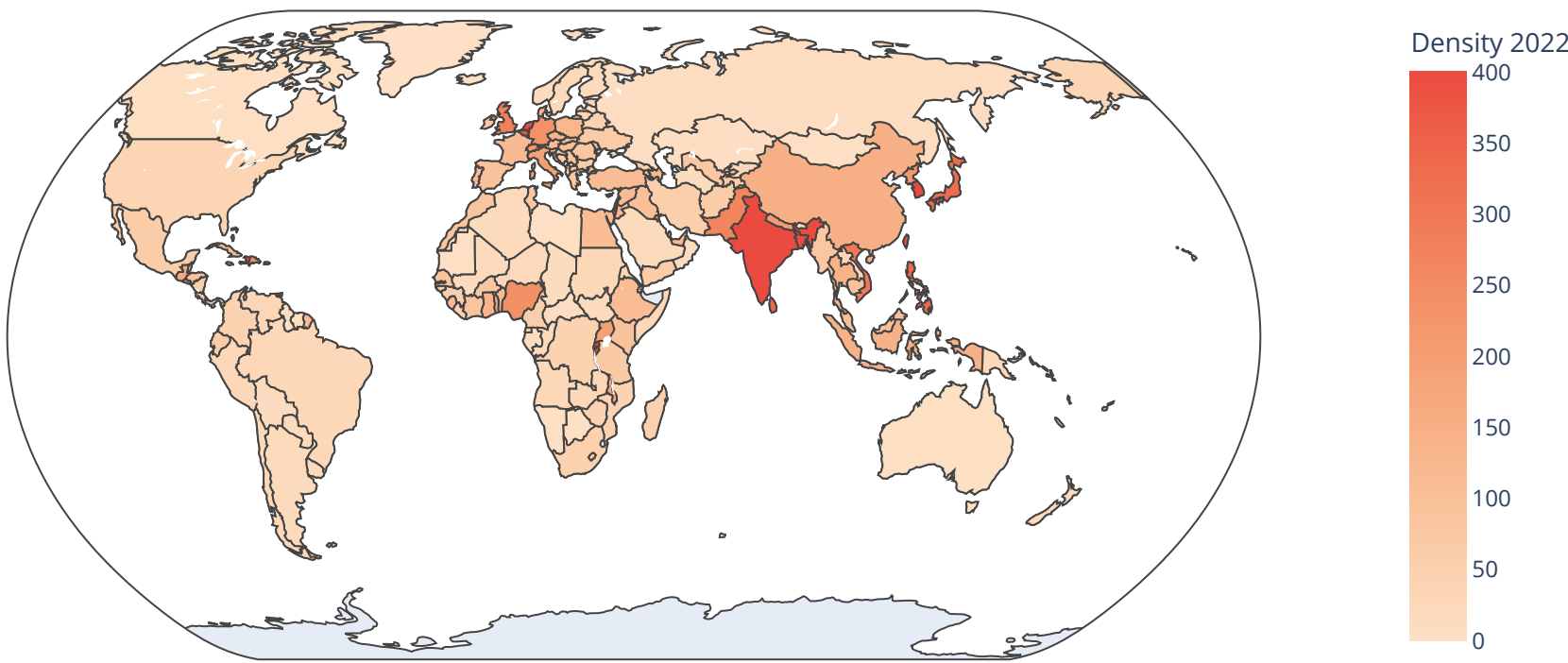
The average density of the world literally doubled in the last 50 years, meaning more people live in the same area nowadays. But this doesn't mean this is the case in all countries. Apearently, some countries had it worse than others, lets have a look a the change visually:


```
In [13]: fig1 = px.choropleth(world_pops,
                             locations = 'Country',
                             locationmode = 'country names',
                             color = 'Density 1970',
                             hover_name = 'Country',
                             range_color = [0,400],
                             title = '1970 Countries Density',
                             projection = 'natural earth',
                             color_continuous_scale = 'peach'
                             )
fig2 = px.choropleth(world_pops,
                     locations = 'Country',
                     locationmode = 'country names',
                     color = 'Density 2022',
                     hover_name = 'Country',
                     range_color = [0,400],
                     title = '2022 Countries Density',
                     projection = 'natural earth',
                     color_continuous_scale = 'peach'
                     )
fig1.show()
fig2.show()
```

1970 Countries Density



2022 Countries Density



If we define the term 'densely populated country' to be a country with an average of 500 person per a km² then looking at the maps above, we can see that in 1970 barely any country was considered densely populated; except maybe for Bangladesh and Taiwan who were very close to this number.

On the other hand, looking at the population density of today's world, we can see that a-lot of other countries have joined the list of densely populated countries; India, Vietnam, Philippines, Rawanda, Burundi, Israel, Palestine, Lebanon and a couple other islands and small countries are all considered now to be densely populated.

Being a high density country can impact the daily lives of the citizens of these countries greatly, and this can be either way;

1. In a good way:
- Everything is usually available within a small radius.
 - Large scales of human resources and workforce
 - Exposure to wide range of people and opportunities
2. In a bad way:
- Higher traffic congestion
 - Very competitive employment market
 - Less space
 - Faster spread of infectus diseases

Fortunately, we've still got time before we become a densely populated planet. Therefore, we should work on not becoming this type of planet or otherwise we'll have to find other solutions, like moving to other planets; which doesn't seem possible at the moment!

Summary

We have talked about countries, continents and their populations and population density;

1. Most populous countires:
- China @ 1.4B
 - India @ 1.4B
 - USA @ 338M
 - Indoensia @ 275M
 - Paksitan @ 235M
 - Nigeria @ 218M
2. Most populous continents:
- Asia (1) - Overall most populous continent

- Africa (2) - Became the second most populous continent between 1990 and 2000
- Europe (3) - Second most populous country up until the late 1990s

3. High density countries: countries that have an average of around 500 person per km2 or more

- India
- Taiwan
- Lebanon
- Rwanda
- Bangladesh
- South Korea
- Japan

and much more, with new countries joining the list every other decade. Interestingly, most of these countries are Asia countries.

Are we, humans, going to over-populate the planet earth anytime soon? **Depends.**

- If the average growth rate becomes as high as India's or Nigeria's growth rate for instance then over-population is inevitable and will happen sooner than later.
- If the average growth rate becomes as low as Europe's growth rate then we definitely aren't going to over-populate this planet anytime soon; heck, we might even risk going extinct at this rate.
- If the average remains as is, we'll probably have enough time to think of solutions to avoid over-populating the earth, like finding and traveling to distant habitable planets.

To sum up, as we saw in the world map above, our planet has plenty of available space for us to populate at the time being, but we should always remember that we're not the only species that live on this planet and while we need to build up and populate the earth we also need to consider the needs for other, weaker perhaps, species to populate as well. Every species has their own role in preserving the life-cycle in our planet earth, thus preserving the planet it self, so we should make sure to preserve this diversity and keep it **perfectly balanced, as all things should be** (Thanos).

The End

By Ameer Shehadey