# Project Overview

The purpose of this project is to analyze the world population data and gain insights into population trends and patterns. The objectives of the analysis are to identify the countries with the largest and smallest populations, explore changes in population growth rates over time, and examine the age distribution of the world population.

| | country name | 2021 population | 2020 population | area | density sq km | growth rate | population ratio | rank |
|---|---|---|---|---|---|---|---|---|
| 0 | China | 1446364682 | 1439323776 | 9706961 | 149 | 0.34 | 18.34 | 1 |
| 1 | India | 1399367516 | 1380004385 | 3287590 | 424 | 0.97 | 17.69 | 2 |
| 2 | United States | 333777266 | 331002651 | 9372610 | 36 | 0.58 | 4.23 | 3 |
| 3 | Indonesia | 277625249 | 273523615 | 1904569 | 145 | 1.04 | 3.51 | 4 |
| 4 | Pakistan | 227103913 | 220892340 | 881912 | 255 | 1.95 | 2.86 | 5 |

# Data Source

The data for this project was sourced from https://www.kaggle.com . The data includes population estimates and projections by country from 2020 to 2021.

# Data Preparation

The data was cleaned and transformed in python notebook before being imported into Power BI. The steps taken to prepare the data included removing null values, filtering out unnecessary columns, and creating calculated columns to compute population growth rates and other metrics.

we shall try to handle our data types

```
[ ]  def replace(data, col_name):

         data[col_name] = data[col_name].str.replace(',' , '')
         data[col_name] = data[col_name].str.replace("/sq_km", "")
         data[col_name] = data[col_name].str.replace("sq_km", "")
         data[col_name] = data[col_name].str.replace("%", "")
         data[col_name] = pd.to_numeric(data[col_name])
```

```
[ ]  replace(df, '2021 population')
     replace(df, '2020 population')
     replace(df, 'area')
     replace(df, 'growth rate')
     replace(df, 'population ratio')
     replace(df, 'density sq km')
```

What are the important questions we are trying to answer?

1) what is the difference between 2020 population and 2021?

2)what are the most crowded countries (5)? and the least (5)?

3)what does density column describe?

4)Is there a relation between the area of the country and its population?

5) Is there a relation between how the country is developed (educated) and its population?

6) what is there a relation between density and area ?

7) countries having high density in 2020, did it solve the problem in 2021 or not?

8)comparing continents population (Africa, Aisa,….).

9)Trying to figure out correlations between our features and the population to see what affects it more.

10)is the given growth rate applied on the difference between 2020 and 2021 populations?

Power Bi questions:

1) what is the difference between 2020 population and 2021?

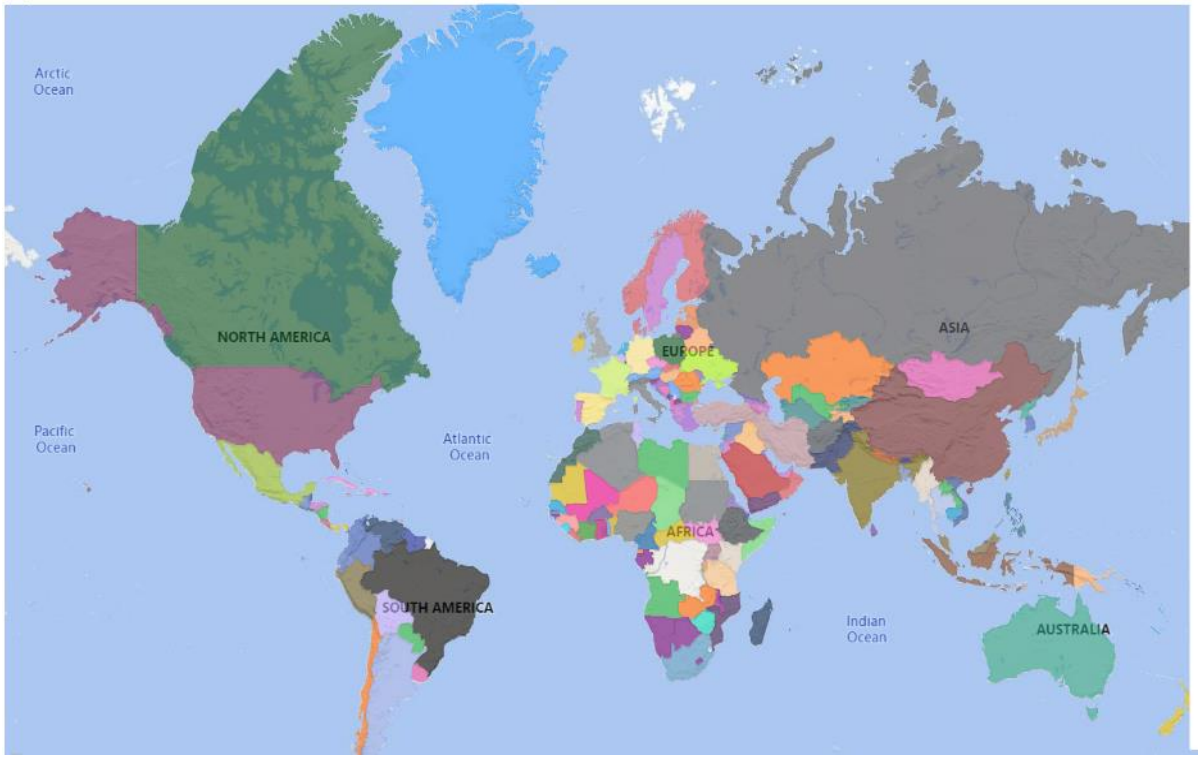
2)what are the most crowded countries (5)? and the least (5)?

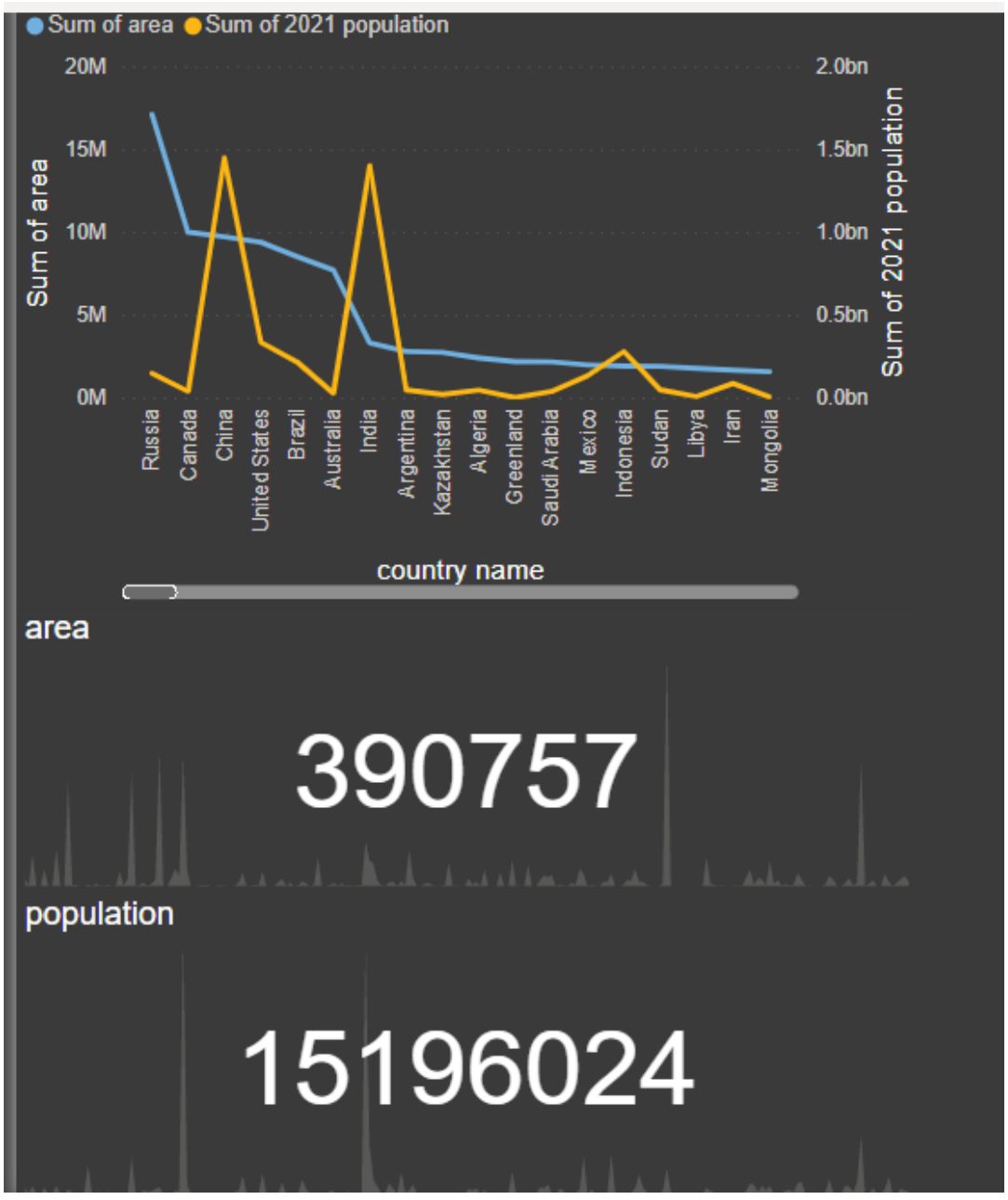
3)Is there a relation between the area of the country
and its population?

4)comparing continents population (Africa, Aisa,….).

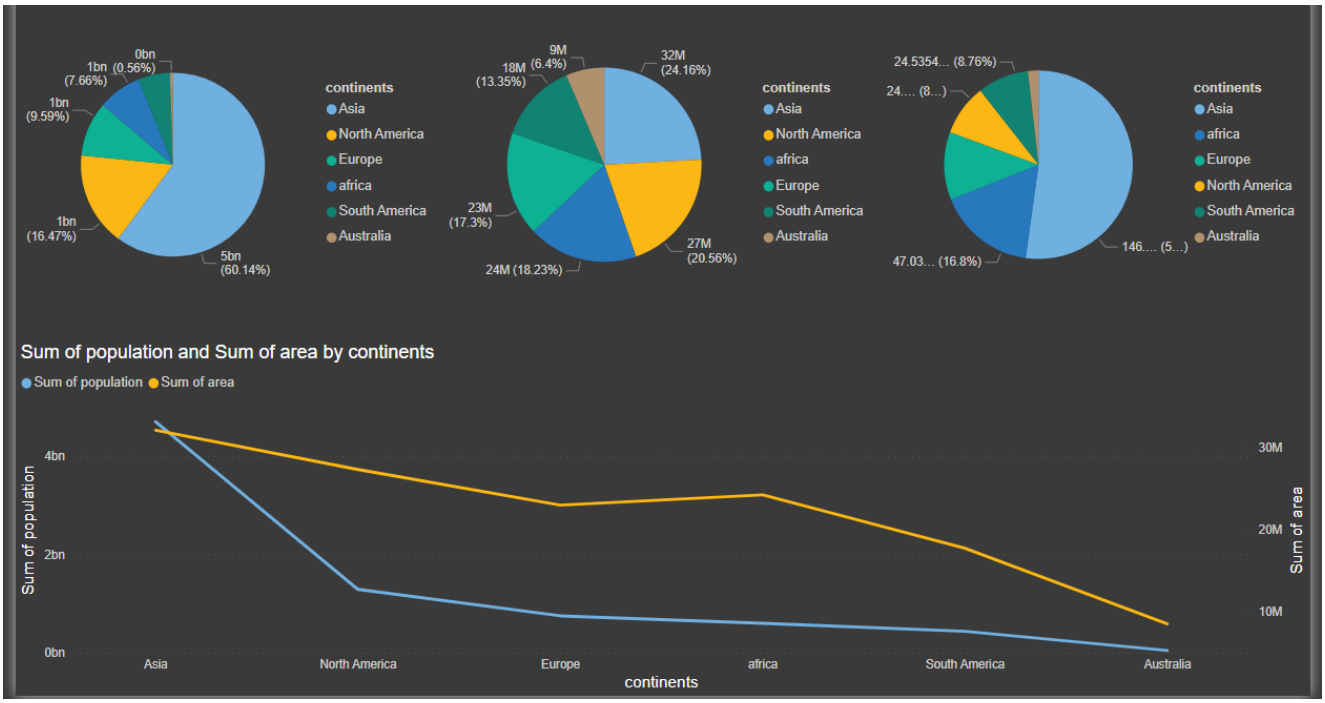
# Visualizations

The visualizations used in this project include a map showing population density by country

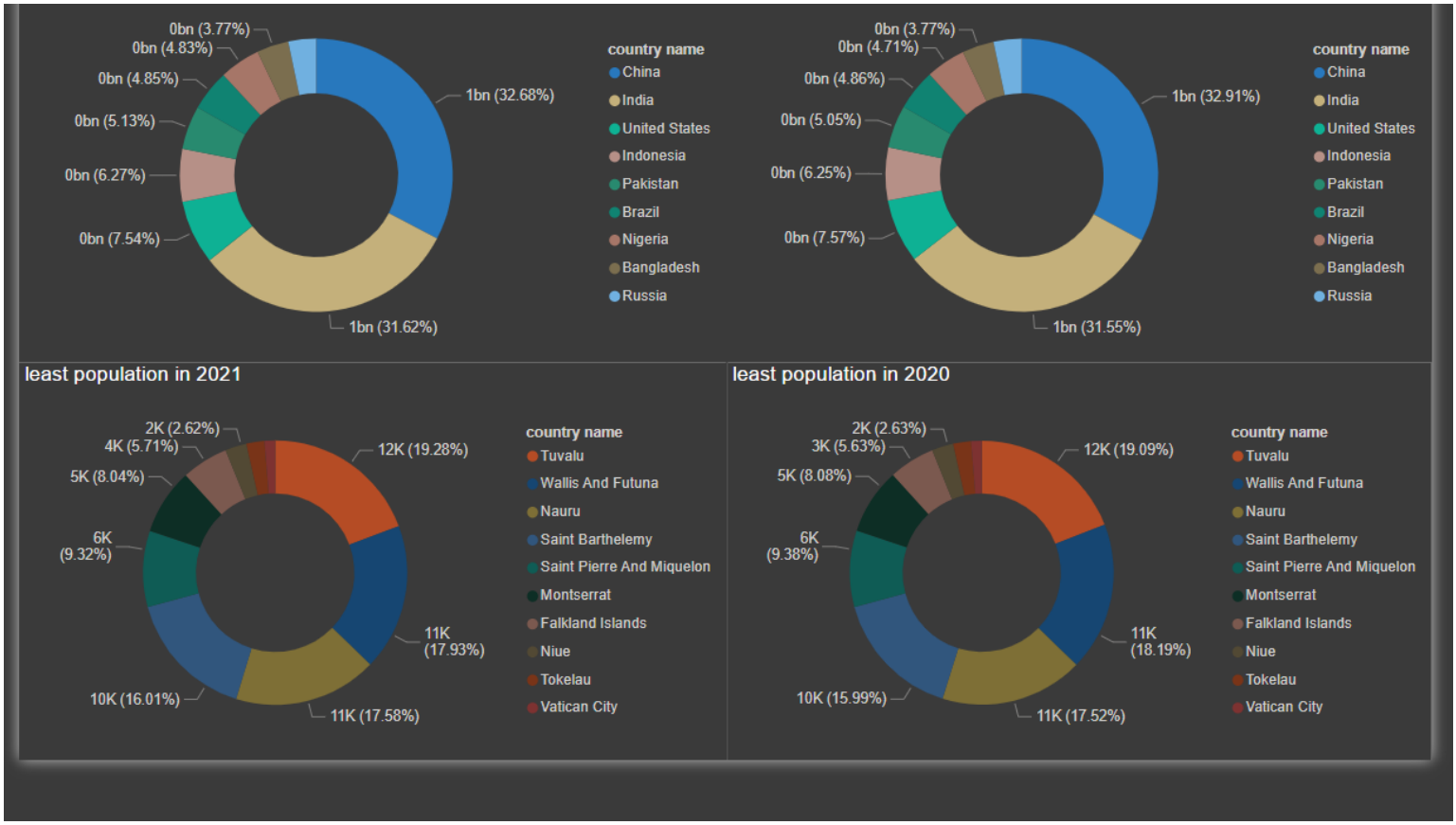





A line chat that describes the relation between the countries area and their populations

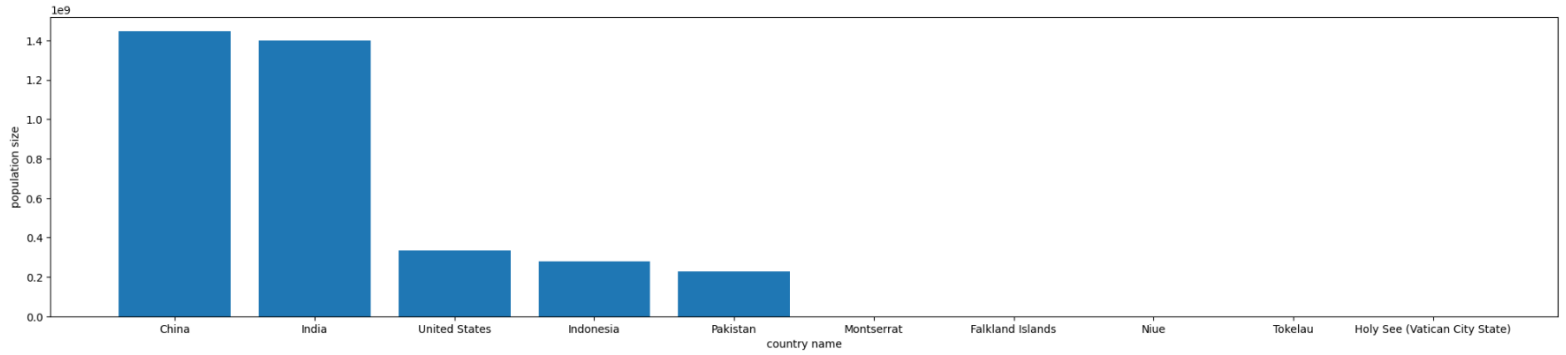



We also added a page that give you insights about continents.




Pie charts that compares 2020 and 2021 populations

```
Text(0, 0.5, 'population size')
```




Bar graphs that compares between countries with high and low populations.

And many other graphs.

## Insights and Findings

The analysis revealed that China and India have the largest populations, while several small island nations have the smallest populations.2020 and 2021 populations are very close and have only slightly difference between them.

## Limitations and Future Work

One limitation of the analysis is that it relies on population estimates and projections, which may be subject to error. Future work could include exploring the impact of migration on population trends, or examining the relationship between population growth and economic development.

# Conclusion

Overall, this analysis provides valuable insights into the world population trends and patterns. The visualizations and findings can be used to inform policy decisions and planning efforts related to population growth, aging, and other demographic factors.

## Content created by ApplAi

**Directed By:**

- Director of Machine Learning and Training Department: Eslam Shouman

**Team leader:**

- Ammar Thabet

**Technical Team:**

- yara ihab
- marina atef
- ahmed yasser
- mohamed samy

## Date of creation

- **30 April 2023**