

Project 1: World Development Statistics

Global Analysis on Tuberculosis Death Rate

Nnene Geraldine

Problem Statement

The project aims to analyze the what impact Tuberculosis death rate has had on population and life expectancy across various countries. By integrating Tuberculosis death rate data, Life Expectancy data and Population data, we hope to single out with the greatest death rate of Tuberculosis in specific years and explore how it affects their population dynamics and life expectancy trends.

Procedure

1 - Data importation

- population.csv - Population by Country over time
- life_expectancy.csv - Life expectancy by country over time
- Tuberculosis_death_rates.csv - Global tuberculosis death rate over time

2 - Cleaning.

3 - Exploratory Data Analysis

4 - Data Visualization

Data Cleaning

1. Dropped some columns from the datasets (population.csv and life_expectancy.csv) to use 29 years interval(1990 to 2019)
2. Changed dataframe column names
3. Fixed incorrect data types
4. Dropped missing data and used mode to fill categorical missing data
5. Merged datasets and Saved them

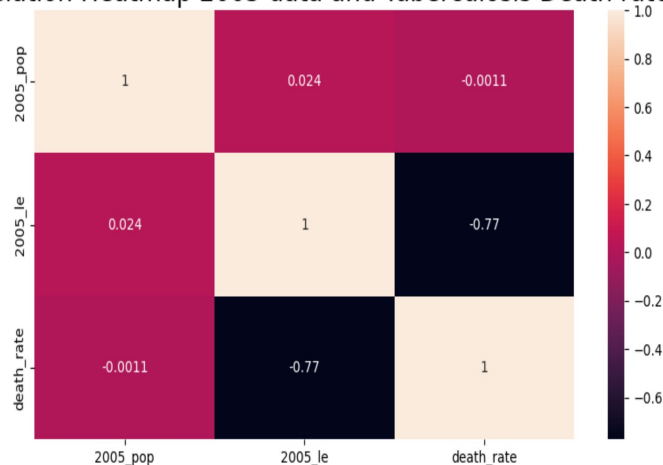
Exploratory Data Analysis

1. Got a Summary Statistics of the merged dataset
2. Applied customized standard deviation function from Part 1
3. Answered questions like;
 - **What country had the highest death rate due to Tuberculosis from 1990 to 2019?:** Central African Republic with a death rate of 336.76
 - **What country had the lowest death rate due to Tuberculosis from 1990 to 2019?:** Malta with a death rate of 0.13
 - **What are the smaller populated countries with highest death rates in 2005?:** Marshall Islands with a death rate of 32.76
 - **What are the larger populated countries with higher death rates in 2005? :** Pakistan with death rate of 87.94

Data Visualization

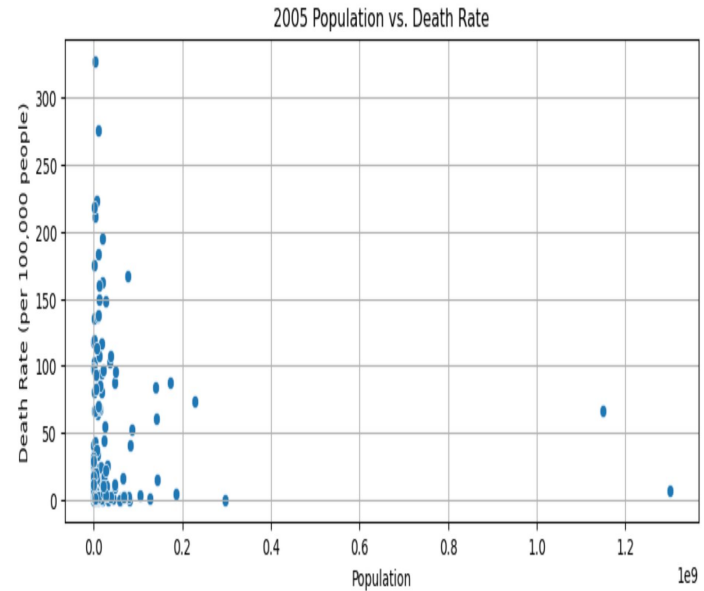
- The heatmap shows a negative correlation Between Tuberculosis Death rate and the Data of all countries (Life Expectancy and population) in 2005. Meaning increase in Life Expectancy or population in any country in 2005 Leads to a decrease in TB death rate.

Correlation Heatmap 2005 data and Tuberculosis Death rate



Data Visualization

- The scatter plot between population distribution of all Countries in 2005 and Death rate shows that most Countries with smaller population have a higher Death rate. The outliers detected at the right of the plot shows countries with very high population and low death rate which from our analysis indicates that this country is China.



Conclusion

- Some countries have experienced significant population growth, while others have remained relatively stable or declined.
- Outliers in population data suggest exceptional cases of population growth or decline, which may call for further investigation.
- While there might be a relationship between population size and Tuberculosis death rate, it's not a straightforward linear correlation.
- Correlation analysis reveals negative correlations between population, life expectancy, and Tuberculosis death rate. This indicates that an increase in population and life expectancy over time leads to a decrease in death rate. We have modern medicine to thank for this improvement.

Recommendations

- 1 - Conduct further statistical analysis to explore the factors influencing Tuberculosis mortality rates, considering variables beyond population size.
- 2 - Investigate socio-economic, healthcare, and environmental factors that may contribute to differences in life expectancy and disease prevalence among countries.
- 3 - Implement targeted interventions and public health strategies to address Tuberculosis burden in regions with high mortality rates.
- 4 - Continuously monitor population trends, health indicators, and disease outcomes to assess the effectiveness of interventions and inform future policies.

*Thank you for your keen
attention.*