Analyzing the Effect of Population Growth on Temperature Changes

1. Introduction

1.1 Motivation for Analyzing Population Growth and Temperature Changes

In recent decades, global warming and climate change have emerged as significant challenges affecting ecosystems, human health, and economies worldwide. The rising concern over increasing global temperatures necessitates understanding the various contributing factors. Among these factors, population growth is often debated for its potential impact on climate change through increased carbon emissions, deforestation, and urbanization. This project aims to explore the relationship between population growth and temperature changes on a global scale, providing insights that can help in formulating effective climate policies.

1.2 Main Research Question

The central question this project seeks to answer is:

- "Is population growth related to temperature changes on a global scale?"

2. Datasets

For this project, we have identified two primary datasets to analyze the relationship between population growth and temperature changes:

2.1 World Population Growth

Dataset

The World Population Growth Dataset provides detailed data on global population trends from 1951 to 2023. Key attributes include total population count, annual growth rate, population density (individuals per square kilometer), and yearly population increases.

- **Metadata URL**: World Population Growth Dataset
- **Data URL**: Available for download via Kaggle API (only 2020 data is used in this project)
- **Usability: 10.00**
- **Data Type**: CSV
- **Description**: This dataset provides comprehensive data on world population growth, including metrics for different countries and regions over various years.
- **License**: Open Data Commons Public Domain Dedication and License (PDDL) v1.0

For this project, the population is calculated for each decade on a global scale, and the average growth rate is also computed. To Match this dataset with second dataset, Dataset is narrowed from 1970 to 2020 and after cleaning data we only have Decade, Population and growth rate column for every decade.

2.2 Temperature Change

The FAOSTAT Temperature Change domain provides annual updates on mean surface temperature changes by country for the period 1961–2019. It includes monthly, seasonal, and annual mean temperature anomalies relative to the baseline period 1951–1980. The data, sourced from NASA's GISTEMP, also includes the standard deviation of temperature changes based on this baseline.

- **Metadata URL**: Temperature Change Dataset
- **Data URL**: Available for download via Kaggle API (only 2020 data is used in this project)
- **Usability: 10.00**
- **Data Type**: CSV
- **Description**: This dataset includes historical temperature change data globally, allowing for analysis of temperature trends over time.
- **License**: Attribution 3.0 IGO (CC BY 3.0 IGO)

For this project amount of Temperature change is calculated for every Decade from 1970 to 2020 and in at the end we only need Decade and Temperature Change Rate.

3. Data Pipeline

3.1 Data Collection and Preparation

Data collection involves downloading the datasets from Kaggle using the provided APIs. The datasets are then preprocessed to ensure compatibility for analysis. Thisincludes cleaning the data, handling missing values, and ensuring consistent formatting across datasets.Figure 1 in below show the pipeline as a flowchart.

Figure 1: Data pipeline
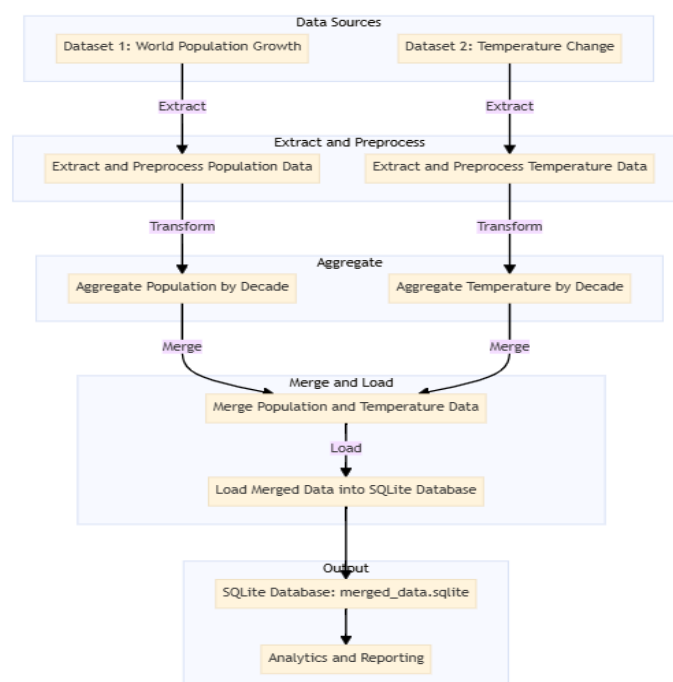
3.2 Transformation or Cleaning
Since the population dataset only provides data for each decade by country, I needed to aggregate this data to a global scale. I summed the populations of all countries to get the world population for each decade and did the same for the growth rate. For this dataset, I only needed the columns for the decade, population, and growth rate.

To align the "Temperature Change" dataset with the population dataset, I calculated the average global temperature change for each decade. This made the temperature data comparable to the population data. Finally, I merged both datasets into a new database to facilitate data visualization and analysis.

3.3 problems

The "Temperature Change" dataset only contains data up to 2019, while the population dataset extends to 2023. To address this discrepancy and avoid limiting my comparison to the 2010 decade, I treated the data from 2011 to 2019 as a single decade for comparison with the 2020 population data.

4. Results

4.1 Output Data Description

The final dataset from the pipeline includes global population, population growth rate, and temperature changes for each decade from 1970 to 2020. The data is stored in an SQLite table with columns for "Decade," "Population," "Population Growth Rate," and "Temperature Change." Each row represents a decade, showing the total global population at the start of the decade, the average annual growth rate, and the average global temperature change relative to the 1970-2020 baseline.

4.2 Data Structure and Quality

The data is structured in a table with four columns and six rows, each representing a decade. The dataset is mostly complete and consistent, though the population growth rate for the 1970 decade is missing (NaN). The data is accurate, sourced from reputable institutions, and up-to-date with the most recent information from 2020.

4.3 Critical Reflection

The dataset provides valuable global insights into population and temperature trends. Aligning the timeframes for both datasets allows for meaningful comparisons. However, there are some issues, such as the missing growth rate for the 1970 decade and the approximation made by treating 2011-2019 as a single decade for temperature data. Additionally, any biases in the original datasets could affect the results. Despite these issues, the data pipeline effectively integrates and processes the datasets to offer a comprehensive view of global trends, but conclusions should be drawn with caution considering the identified gaps and assumptions.