

# Data

Data source: [UN Demographic Indicators Compact \(XLSX\)](#)

This dataset provides demographic insights such as population trends, mortality rates, fertility rates, and life expectancy across countries and regions over several decades.



---

## Objective of Our Analysis

The objective of our analysis is to conduct a detailed comparison of Turkey's demographic indicators on a yearly basis, contrasting them with those of regions categorized as "More Developed," "Less Developed," and "Least Developed." By examining these indicators over time, we aim to evaluate Turkey's development status relative to other global regions.

---

## Reason for Choosing This Dataset

The primary reason for selecting this dataset is its reliability and comprehensiveness, provided by the United Nations.



As a globally recognized institution, the United Nations collects and analyzes demographic, economic, and social data that is trusted internationally. Therefore, the data they provide is not only highly reliable but also widely accepted and verified.

The United Nations collects demographic data from national statistics offices and reliable sources, such as censuses and surveys. The data is then analyzed using standardized methods to ensure accuracy and comparability across regions and countries. This process allows the UN to create reliable demographic indicators used for global development analysis.

In addition to its trustworthiness, this dataset is exceptionally detailed and extensive. The data is presented on an annual basis, which allows us to observe demographic changes over time and provides a sufficient amount of data for conducting comparative analyses. This enables us to assess Turkey's development status by comparing it with "More Developed," "Less Developed," and "Least Developed" regions, offering a robust foundation for in-depth analysis.



---

## Accomplishments

By organizing, cleaning, and processing the data, we aimed to create a streamlined version that enables targeted analysis of specific indicators such as population trends, death rates, life expectancy, and other demographic parameters.

---

## Initial Cleaning Process

- We focused on Turkey, as well as the “More Developed,” “Less Developed,” and “Least Developed” groups by removing data related to other countries and regions.
  - Columns that were irrelevant to our objectives were removed, including:
    - Estimates (Variant)
    - Notes
    - ISO Codes (Alpha-2, Alpha-3, SDmx, Parent)
    - Population Annual Doubling Time (years)
    - Crude Birth Rate (births per 1,000 population)
    - All columns related to migration.
- 

## Simplifications

- Redundant columns for gender-separated data were aggregated into general metrics where possible. For example:
    - Life Expectancy and Deaths columns for males and females were replaced by general life expectancy and total deaths metrics.
    - Metrics such as population density, sex ratios, and mortality rates for specific age groups were excluded to streamline the data set.
- 

## Custom Metrics

- A new Death Rate (%) column was created by calculating the ratio of total deaths to the total population. We calculated this ratio using the Excel file.
- 

## Data Grouping

The cleaned data set was organized into the following thematic groups for analysis:

- **Change in Population:**  
Metrics include metrics like natural change, population growth rate, and birth counts.
  - **Child Demographics:**  
This section focuses on infant and under-five mortality rates, survival rates, and child-specific metrics.
  - **Death Rate:**  
This metric consists of the total population, total deaths, and the newly calculated death rate.
  - **Life Expectancy:**  
The data set contains life expectancy data that was directly provided in the original data set.
  - **Mortality:**  
It includes general mortality rates without detailed gender or age group segmentation.
  - **Population Dynamics:**  
The covers include total population data, which is broken down by year, as well as the distributions of male and female populations.
  - **Other Parameters:**  
The data includes indicators such as fertility rates, mean age at childbearing, and sex ratio at birth.
  - **Total Deaths:**  
The section contains data on total deaths and crude death rates.
- 

## Storing the Processed Data

[You can access these modified Excel files by clicking this link.](#)

To convert the data into **.RData** format, we performed the following steps:

1. First, we loaded the original data set into R from an Excel file using the **readxl** package:  

```
library(readxl)
```

```
file_path <- "data/change_in_population.xlsx"
```

```
change_in_population <- read_excel(file_path)
```
2. After processing and cleaning the data set, the modified data (total\_deaths in this case) was saved in the **.RData** format using the **save()** function. This format preserves the R objects, making it easier to load and analyze later.

```
save(change_in_population, file = "data/change_in_population.RData")
```

The final data set is now available in the **.RData** format, and you can access the final cleaned and grouped files stored in **.RData** format below:

[.RData Files.](#)