

The Happiness Index

linked to population growth & location

in 2024

Joice and Timo

datasets used:

World Happiness Index 2024

<https://www.kaggle.com/datasets/jainaru/world-happiness-report-2024-yearly-updated>
by Jaina published on kaggle
country - region - happiness index

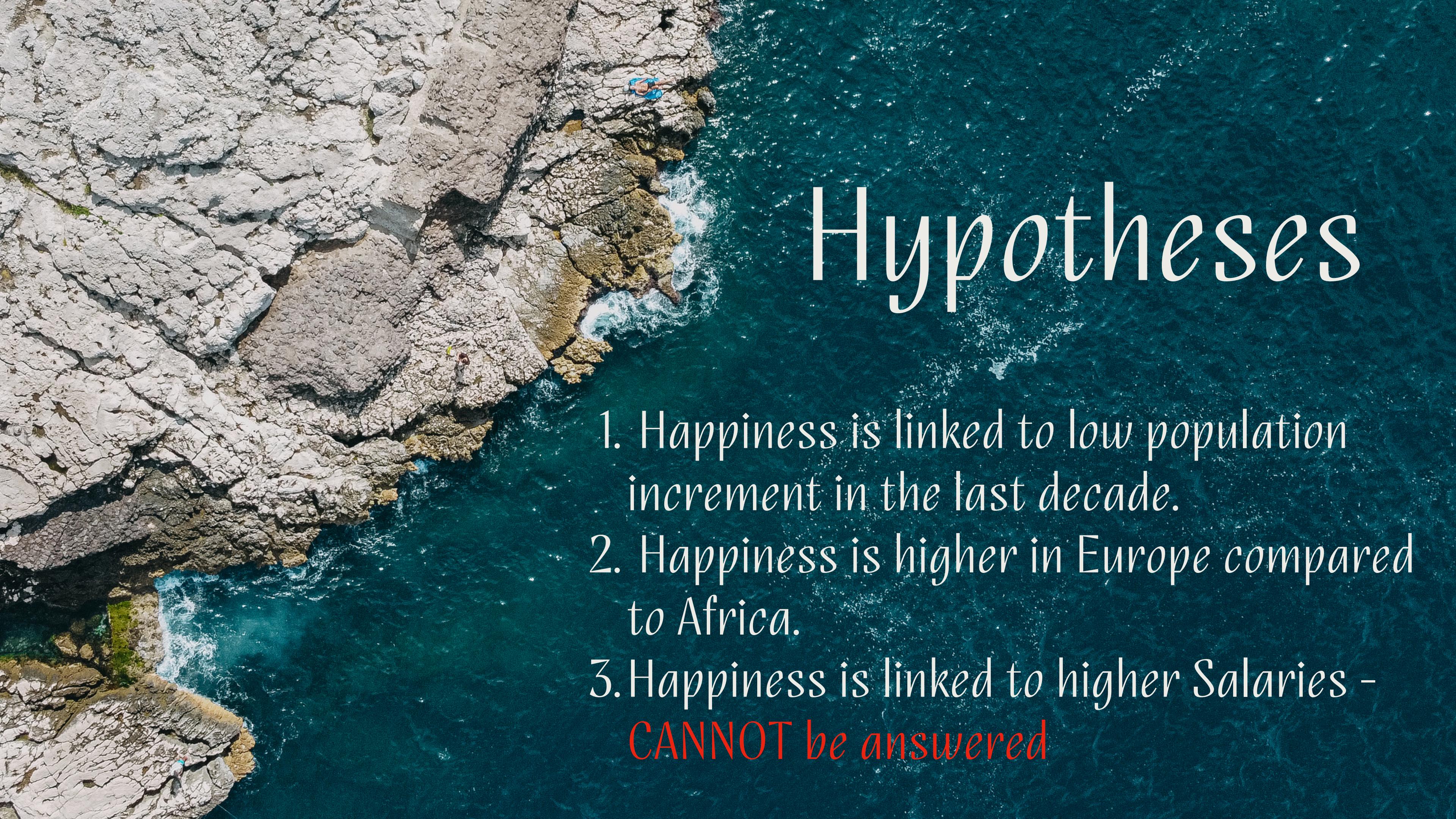
World Country Populations 2014 & 2024:

<https://www.worldometers.info/world-population/population-by-country/>
<https://m.worldometers.info/world-population/population-by-country/>

Dataset contains information on **absolute population counts by country for 2014 & 2024**
The data on this webpage is provided by <https://www.realtimestatistics.net/> - End Worldometers RTS Counters by Google

average annual wages per country 2010 & 2023:

https://en.wikipedia.org/wiki/List_of_countries_by_average_wage
Wikipedia, OECD = Organization for Economic Co-operation and Development

An aerial photograph of a rugged coastline. The foreground is dominated by light-colored, layered rock formations. Several small, bright blue figures of people are scattered across the rocks, providing a sense of scale. The water is a vibrant turquoise color, with white foam where waves break against the shore.

Hypotheses

1. Happiness is linked to low population increment in the last decade.
2. Happiness is higher in Europe compared to Africa.
3. Happiness is linked to higher Salaries -
CANNOT be answered

Data Acquisition, Enrichment, and Examination

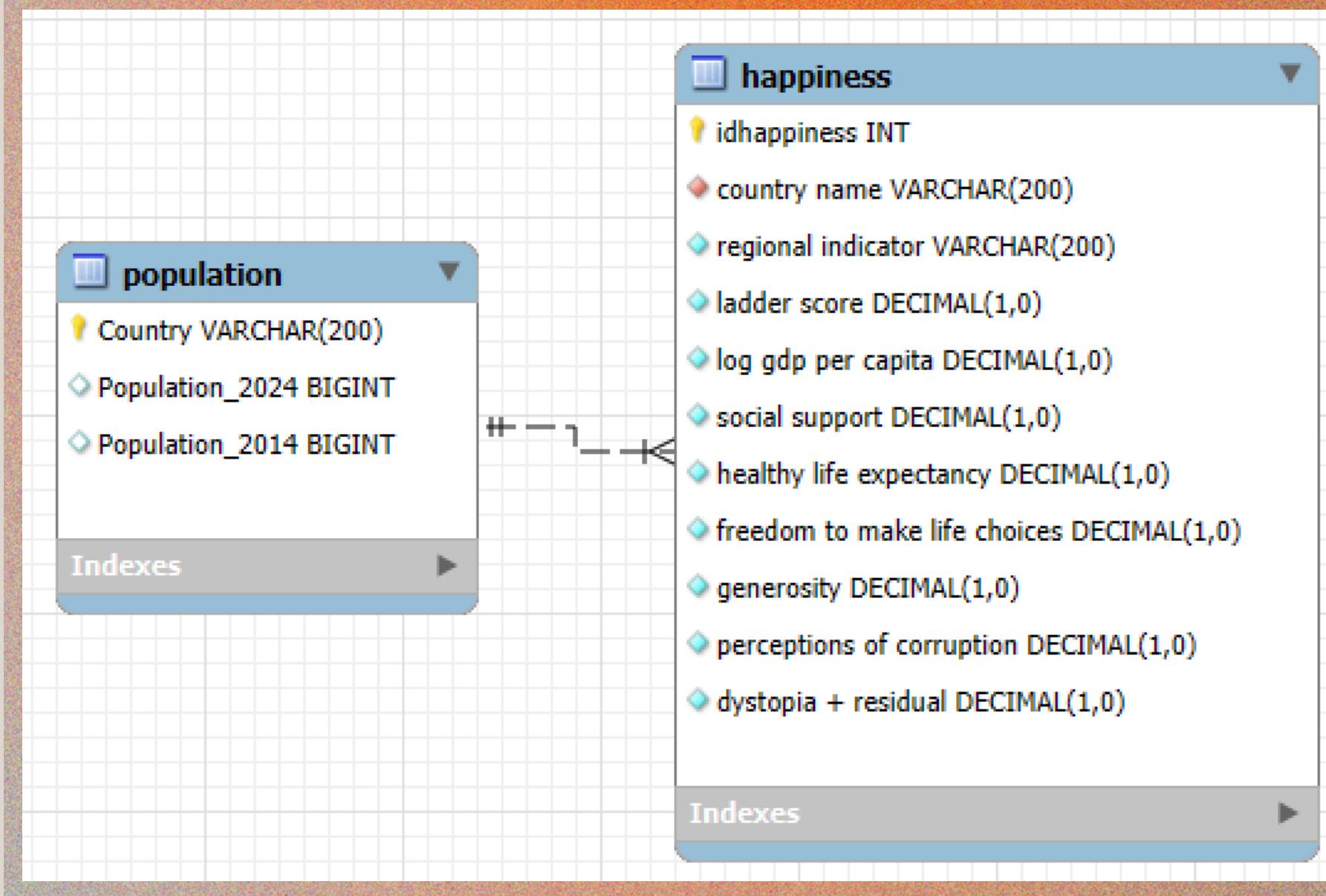
Primary Data Sources:

- Selected from 50 famous datasets relevant to the project
 - Core dataset: World Happiness Index 2024
 - Supplemental datasets:
 - Population Expansion Data 2014 & 2024 (Worldometer)
 - Average Annual Wages 2010 & 2023 (OECD)

Challenges Faced:

- Null Values: Required cleaning and handling missing data
- Reliable Global Salary Data: Difficult to obtain, leading to exclusion of wage component due to time constraints
- Data Alignment & Integration:
 - Supplemental datasets allowed correlation analysis
 - Explored relationships between population growth and happiness index

WORLD HAPPINESS vs. POPULATION COUNT mySQL schema



SQL Database Design & Data Transformation

Entity-Relationship Model (ER Model)

The database consists of two primary tables:

1. Happiness Table: Contains data on countries, their Happiness Index (Ladder Score), and regional classifications.
2. Population Table: Stores population figures for 2014 and 2024, used to compute growth percentages.

A **one-to-one relationship** exists between the two tables, linked by the Country Name field.

Key Data Transformation Challenges & Solutions

1. Header Name Differences & Alignment:

- The country name fields had inconsistent headers across tables (happiness.country name vs. population.Country).

2. Null Values & Data Integrity:

- Missing population values caused issues in growth calculations.

3. Duplicate Entries:

- Some countries appeared multiple times due to regional discrepancies.

4. Data Consistency & Standardization:

- Differences in country naming conventions led to mismatches during joins.

By addressing these challenges, the transformed dataset allowed smooth integration between happiness and population data, enabling effective analysis.

SQL Insights & Advanced Analysis

The following SQL queries were executed to derive insights:

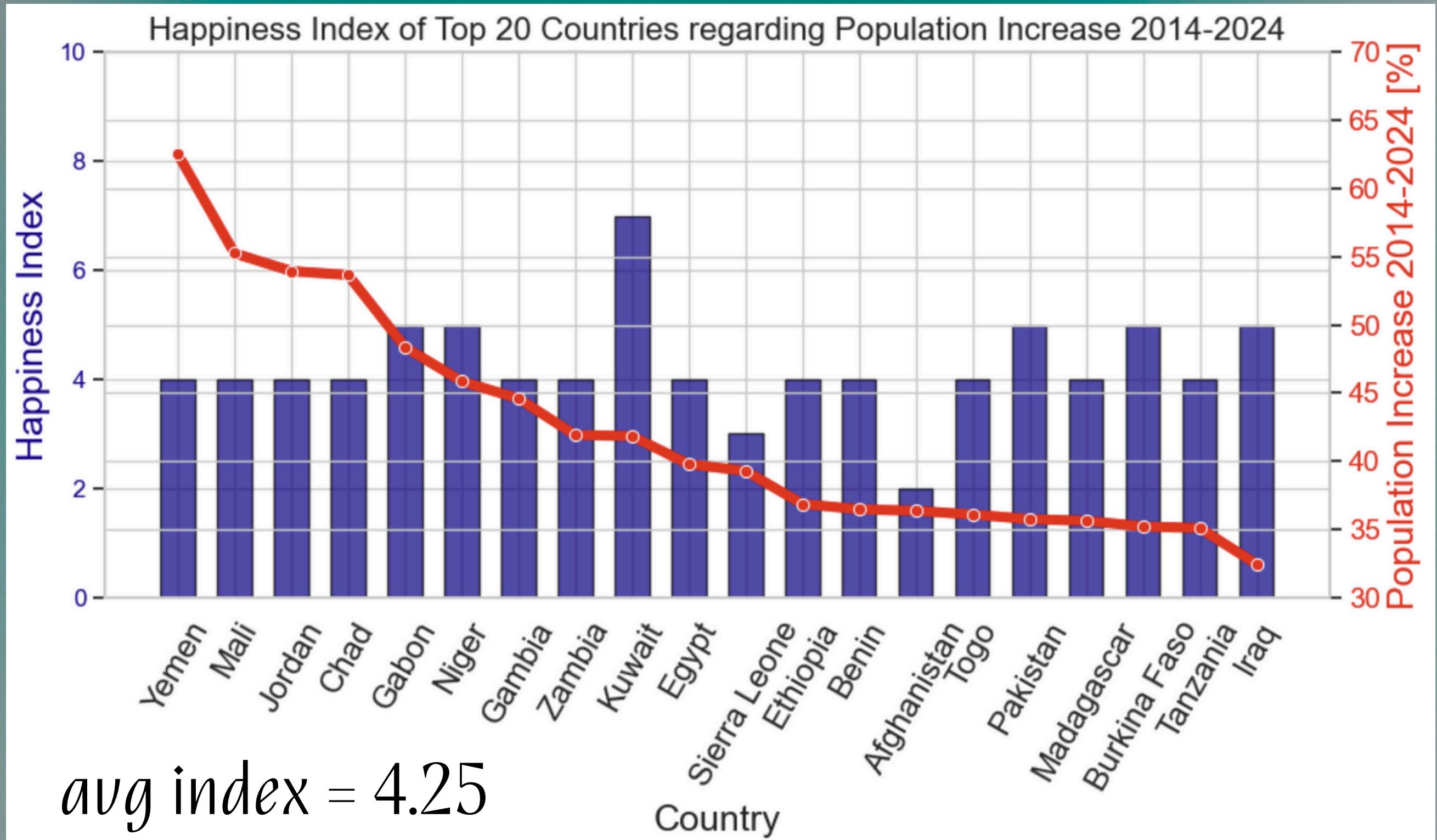
1. Happiness vs. Population Growth:

- A view (`happiness_increase_population`) was created to analyze whether rapid population growth/decrease correlates with higher/lower happiness scores.

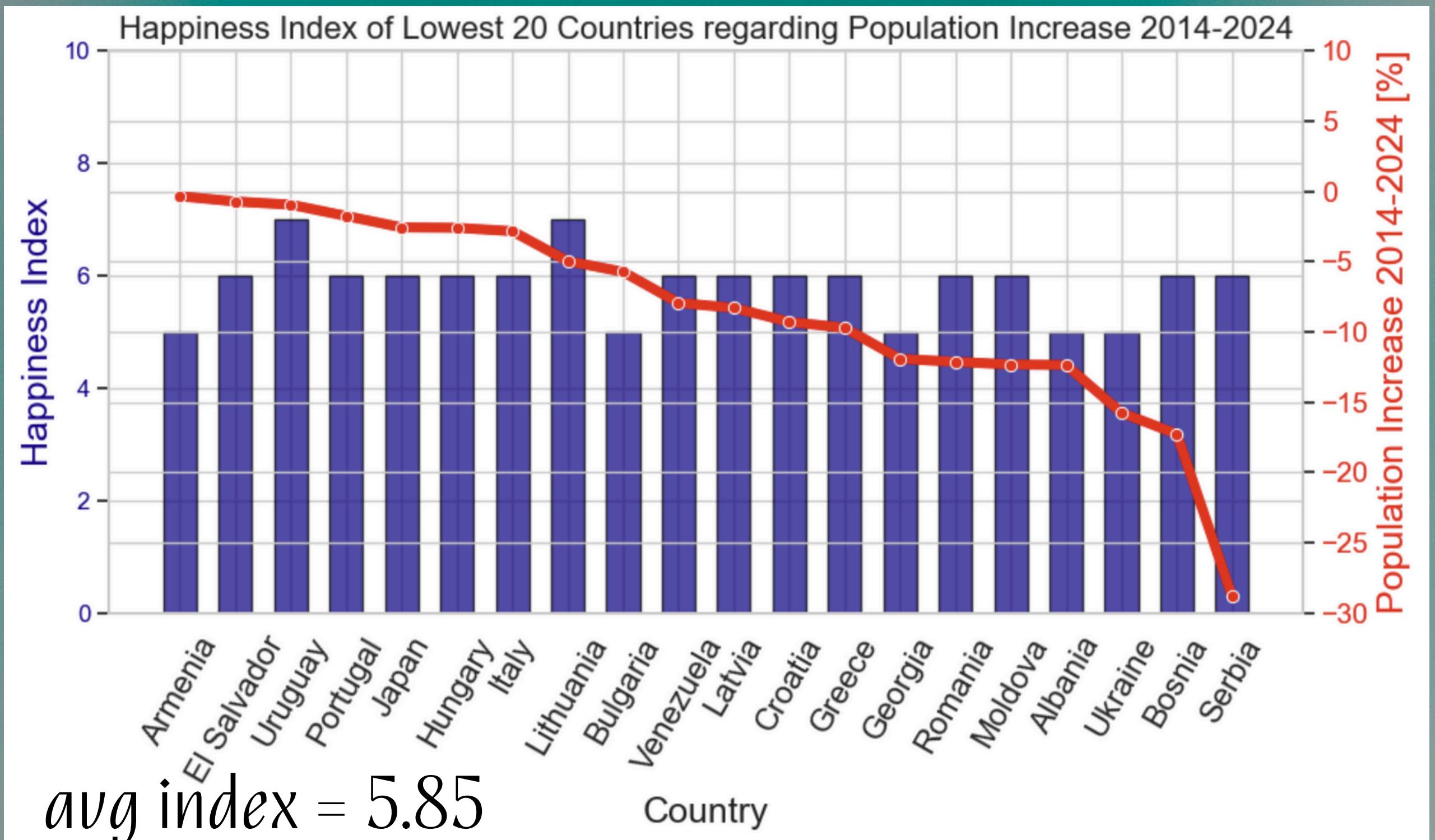
2. Regional Happiness Insights:

- European and Middle Eastern & African (MEA) happiness trends were separately analyzed using CASE statements to segment countries based on their regional indicators.
- Views such as `european_happiness` and `african_happiness`(MEA) were created to facilitate further analysis.

Countries with highest population increase show medium happiness

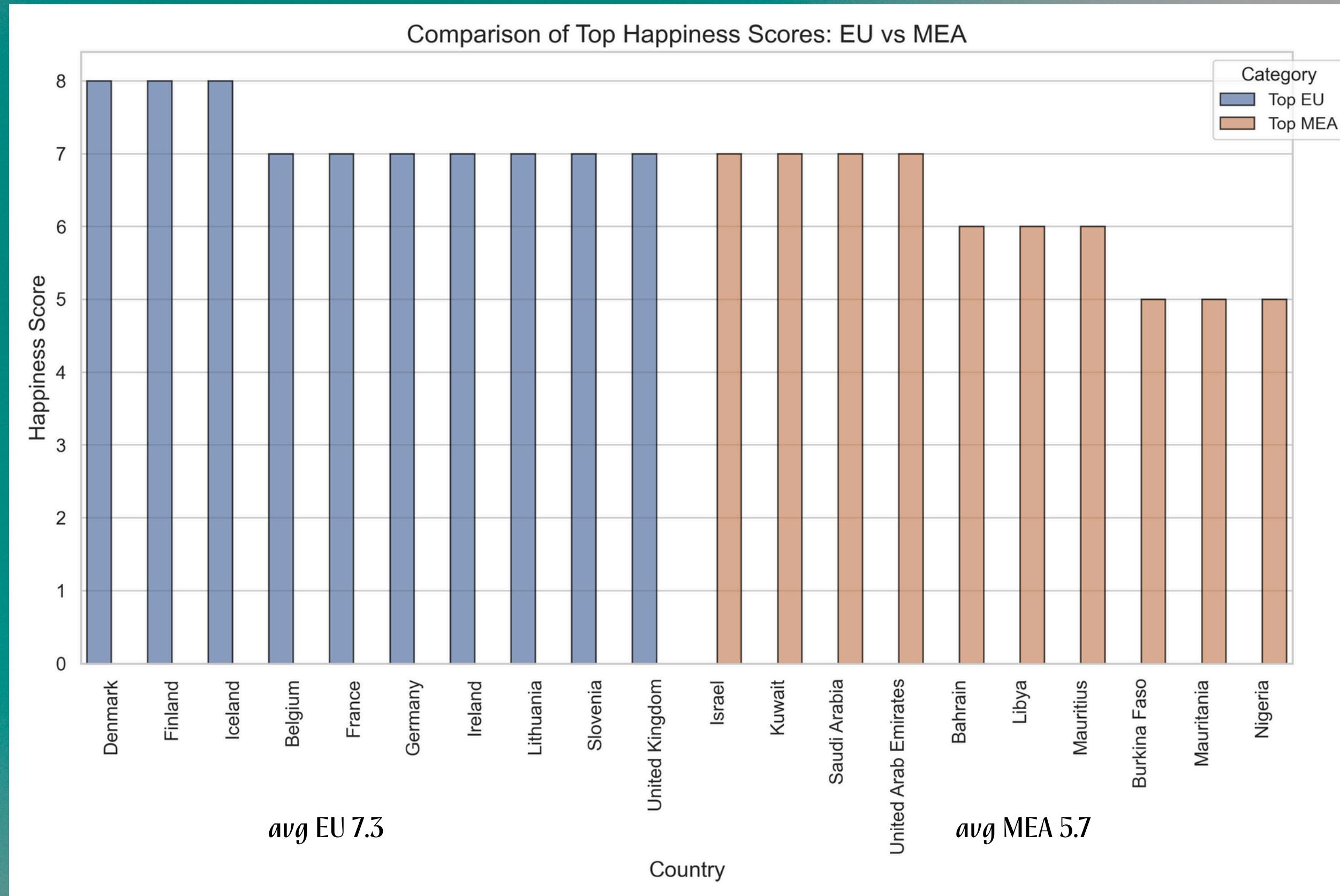


Countries with highest population decrease show higher happiness

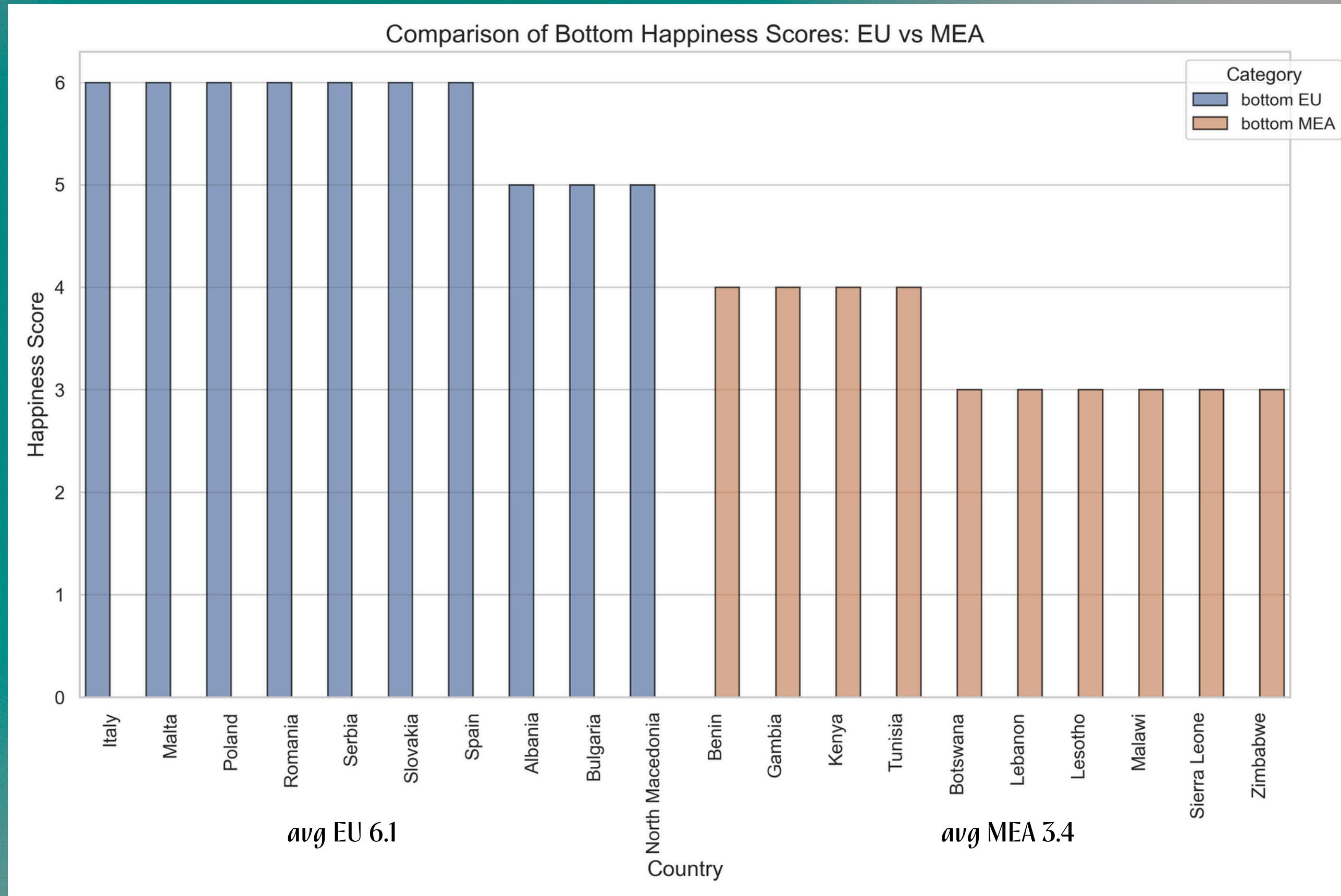


country name	ladder score	Increase
Armenia	5.0	-0.34015
El Salvador	6.0	-0.71367
Uruguay	7.0	-0.93913
Portugal	6.0	-1.74370
Japan	6.0	-2.55651
Hungary	6.0	-2.58767
Italy	6.0	-2.82848
Lithuania	7.0	-4.95887
Bulgaria	5.0	-5.72418
Venezuela	6.0	-7.92769
Latvia	6.0	-8.29156
Croatia	6.0	-9.28640
Greece	6.0	-9.71017
Georgia	5.0	-11.91744
Romania	6.0	-12.13059
Moldova	6.0	-12.31934
Albania	5.0	-12.35783
Ukraine	5.0	-15.75629
Bosnia	6.0	-17.26894
Serbia	6.0	-28.85565

Top 10 countries in both regions



Bottom 10 countries in both regions



Major Obstacle

The biggest challenge was to upload the csv data tables into MySQL.

*Reasons: insufficient cleaning and formating of data tables
and novelty of the uploading process for us.*

*As this was the most time consuming step: next time,
we'll invest more time in data table cleaning and formating.*

Conclusions & Implications

Our data sets are not particularly well suited for creation of a database and relevant statistical analysis.

Nevertheless, we could formulate hypotheses and find answers to them:

- 1.) happiness is higher when population increase is lower*
- 2.) higher happiness in european compared to african (MEA) countries*

Outlook: would be great to include the wage development into the analyses.

The Happiness Index

Thank you

*Joice and Timo
Ironhack Data Analyst Class January 2024*