Don’t Worry Be Happy

What does actually make people happy?

Mini Project by Rike and Victoria

PROJECT OVERVIEW

Does the the feeling of happiness depend on socioeconomic factors, some other variable or is it rather a character trait of an individual?

The World Happiness Report provides information on 147 countries and their respective happiness index scores - responders are asked to rate their own current lives on that 0 to 10 scale. This score is determined by several variables that include GDP per capita, healthy life expectancy, social support, freedom to make life choices, generosity, and corruption perception.

Hypothesis 1:

Economic prosperity influences happiness more than any other factor

Hypothesis 2:

If it is not just the income that makes people happy, there are also social factors that determine happiness

Hypothesis 3:

Happiness can also depend on other factors - like the amount of sunshine

*Dystopia - some hypothetical worst-country-in-the-world, with the lowest expectancy, highest corruption, etc. is introduced as benchmark for comparison with is a country ranked # 1 (in the latest ranks Finland)

DATA ACQUISITION, ENRICHMENT AND EXAMINATION

- Primary sources of data -

  https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2022

  https://www.kaggle.com/datasets/prasertk/sunshine-duration-by-city

---

- Did some cleaning, removed " * " from country-names, found out later, that some numbers were strings
- changed many country-names because the "pycountry_convert" function to get the countries continents did not know some of them
- Merged the tables, dropped all rows with null-values

142	Botswana*	3,471	0,187
143	Rwanda*	3,268	0,536
144	Zimbabwe	2,995	0,548

DATABASE DESIGN & DATA TRANSFORMATION ON

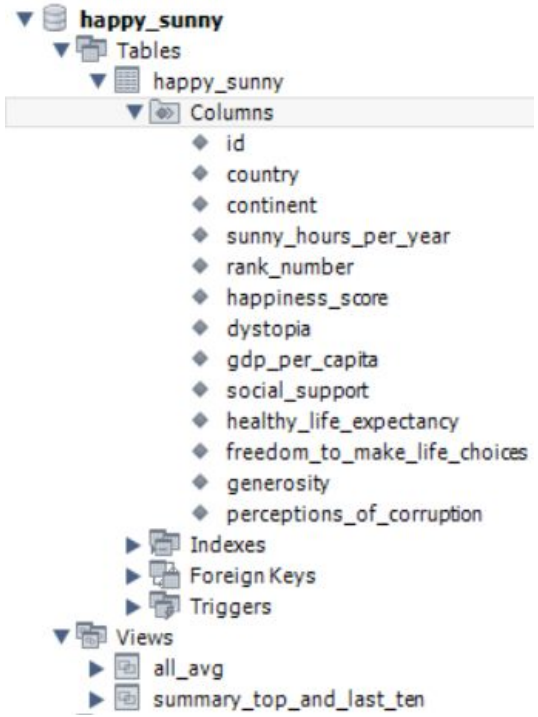
- Since we only had one entity, to create an ERD was not necessary

	Country	Continent	Year	RANK	Happiness score	Dystopia (1.83) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Afghanistan	Asia	3175.10	146.0	2.404	1.263	0.758	0.000	0.289	0.000	0.089	0.005

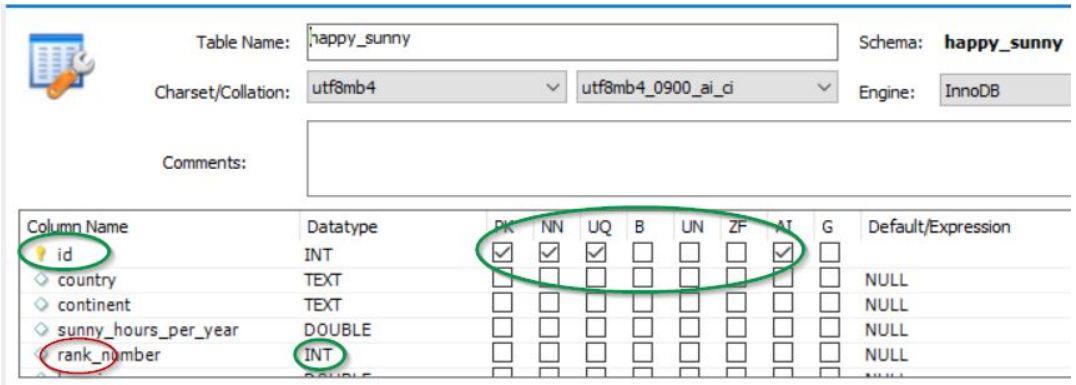
- After importing the CSV we struggled with changing the data-types first > some of our numbers were strings

- When trying to import CSV on Victoria's SQL-Workbench this time-consuming Error appeared

DATABASE DESIGN & DATA TRANSFORMATION ON



- Changed column names to avoid white spaces, added “id” as primary key
- Changed data-type of rank



*interesting fact: RANK is a keyword in SQL

SQL INSIGHTS AND ADVANCED ANALYSIS

```sql
CREATE VIEW summary_top_and_last_ten AS
WITH FirstTen AS (
    SELECT *
    FROM happy_sunny
    WHERE rank_number < 12
),
LastTen AS (
    SELECT *
    FROM happy_sunny
    ORDER BY rank_number DESC
    LIMIT 10
)

SELECT *
FROM FirstTen

UNION ALL

SELECT *
FROM LastTen

ORDER BY rank_number;
```

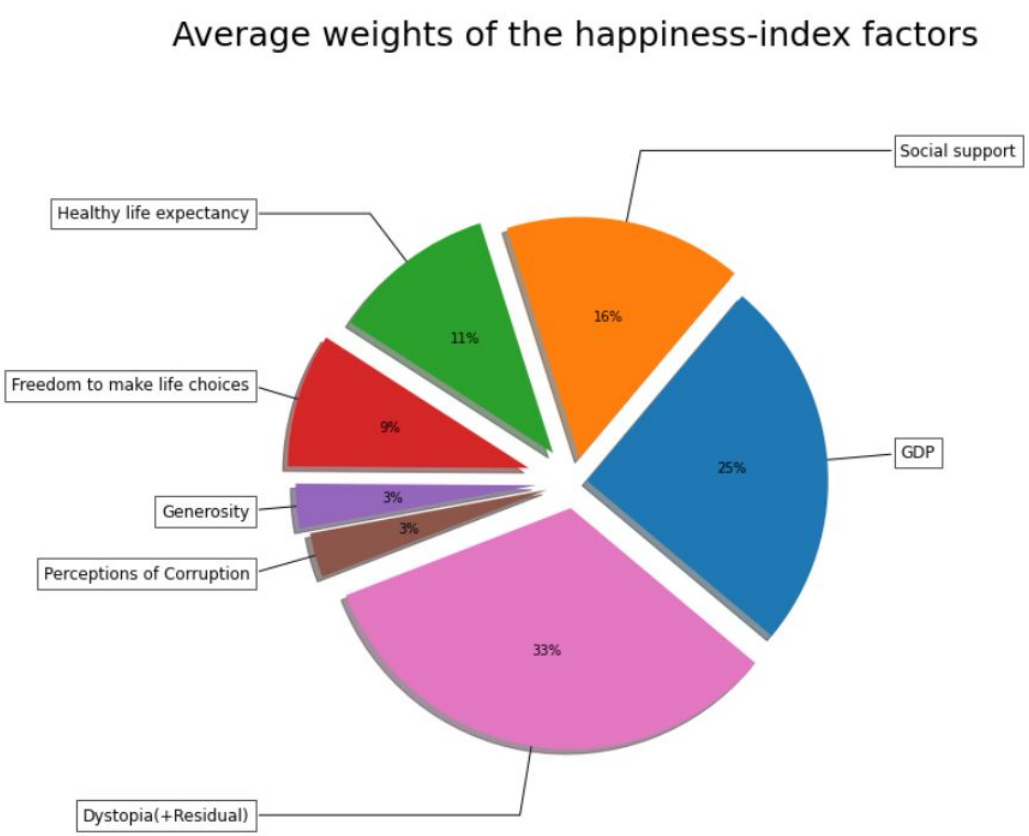
- Created some views to reuse them
  - Used CTEs for creation

- Mostly basic querying was needed for efficient data extraction from the database

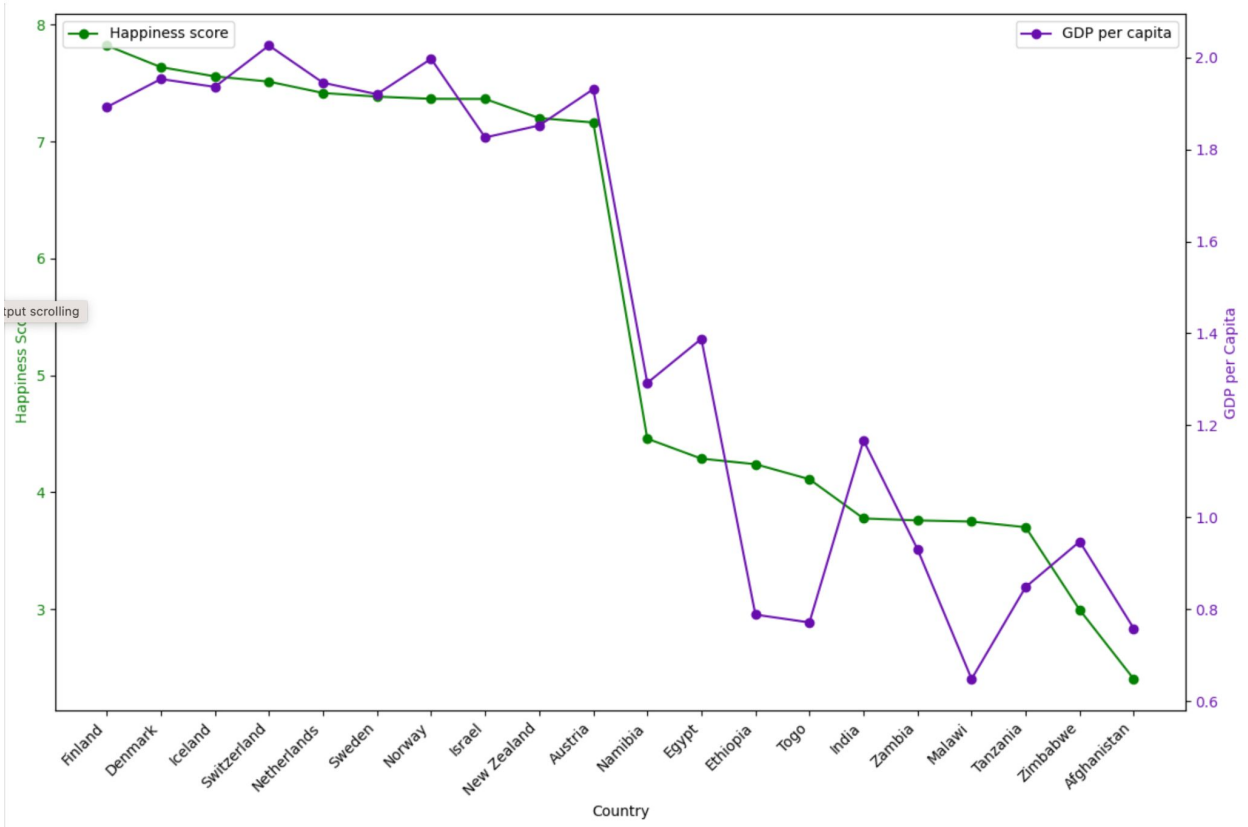
```sql
USE happy_sunny;

SELECT country, happiness_score, gdp_per_capita
FROM happy_sunny
ORDER BY rank_number;
```

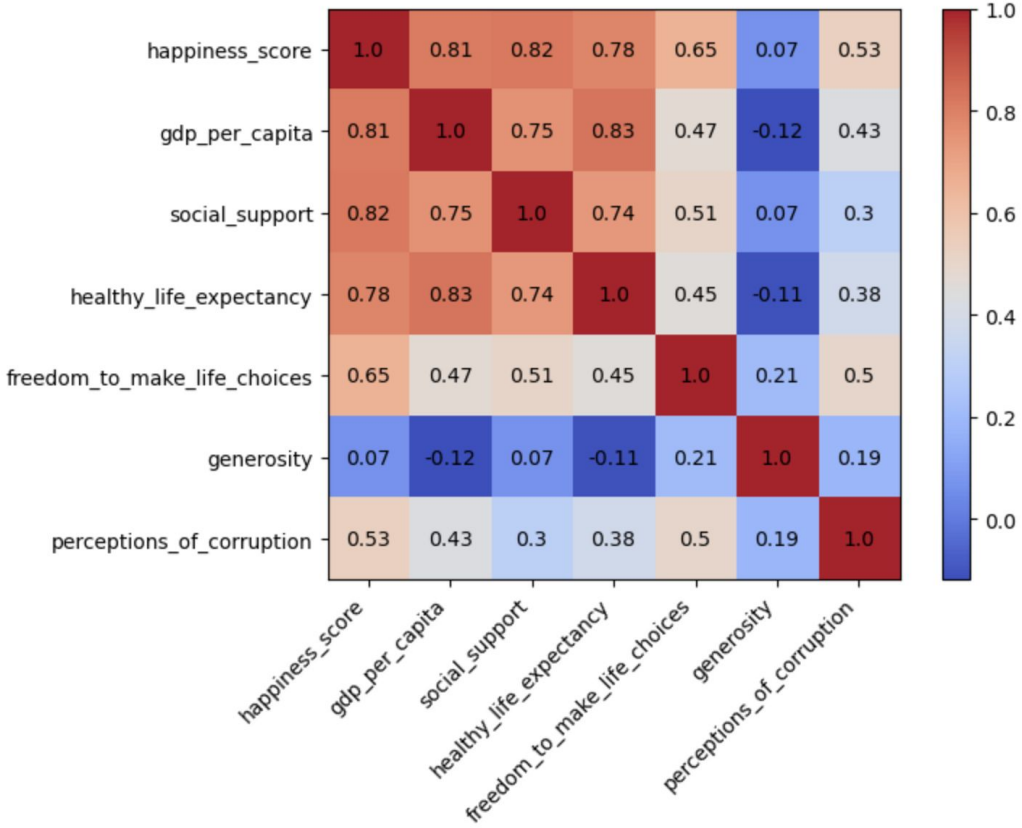
VISUALIZATION & KEY INSIGHTS



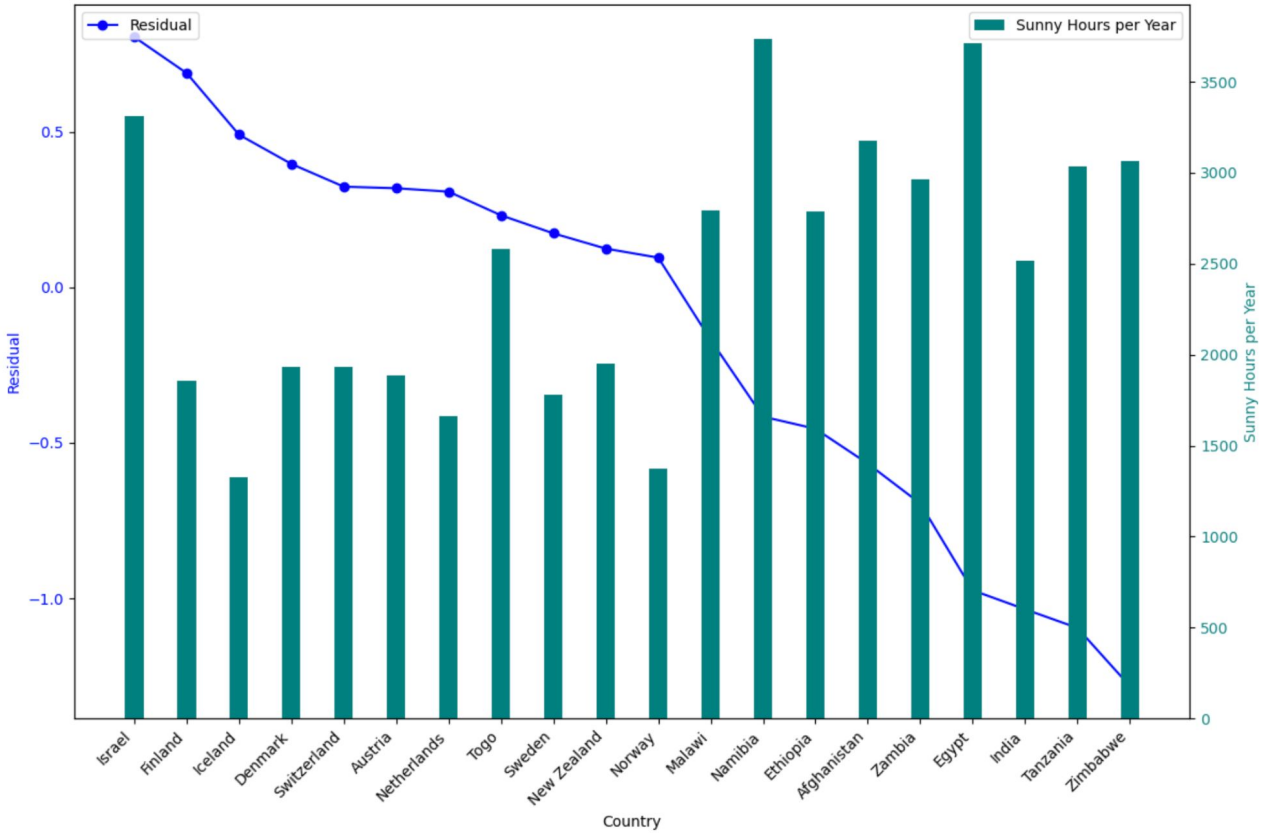
H1: ECONOMIC PROSPERITY DEFINES HAPPINESS



H2: SOCIAL FACTORS HAVE SIGNIFICANT IMPACT



H3: BUT WHAT ABOUT THE SUNSHINE?



CONCLUSIONS

Hypothesis 1: Confirmed

The analysis of the World Happiness Report 2022 dataset reveals significant insights into the factors contributing to happiness across different nations, with major factor being economic factors like GDP

Hypothesis 2: Confirmed

Along with GDP, social support, and health are playing crucial roles in determining overall happiness scores.

Hypothesis 3: Falsified

The dystopia residual, which reflects the unexplained components of happiness, also tends to be higher in high-ranking countries, but showed little correlation with the annual amount of sunshine per country

MAJOR OBSTACLES

- Formatting and merging data from different sources
- Decoding a csv file to the SQL was a major challenge
- Losing time on being stuck with a certain challenge - at some point the similar experience of other team could help a lot!

```
:
import pandas as pd
#Remove non-ASCII characters
file_path = "/Users/alexkabanovsky/Downloads/df_happy_sunny_11.csv"
df = pd.read_csv(file_path, encoding="utf-8")
def remove_non_ascii(text):
    # If value is not string (e.g., NaN), return it as is
    if not isinstance(text, str):
        return text
    return "".join([char for char in text if ord(char) < 128])
df_cleaned = df.applymap(remove_non_ascii)
cleaned_file_path = "/Users/alexkabanovsky/Downloads/df_happy_sunny_cleaned.csv"
df_cleaned.to_csv(cleaned_file_path, encoding="utf-8", index=False)
print("Non-ASCII characters removed and saved to new file.")
```

```
Non-ASCII characters removed and saved to new file.
```