World happiness report analysis

Overview across the world for the last 5 years

Table of Contents

Introduction

Introduction

This project explores the key factors that influence happiness across countries using data from the World Happiness Report (2020–2024).

Through exploratory analysis, statistical modeling, and visual storytelling, we aim to understand how economic, social, and personal freedom indicators relate to happiness levels worldwide.

Tools used include Excel, SQL (PostgreSQL), and Tableau to uncover patterns, compare regions, and identify outliers that challenge common assumptions.

Methodology

Data Details

The data comes from kaggle.com and comprises five CSV files:

2020.csv

2021.csv

2022.csv

2023.csv

2024.csv

Data Details

Each file has these columns and collects data for a specific year:

- **Country name**
- **Area**
- **Happiness Rank**
- **Happiness score**
- **Upperwhisker**
- **Lowerwhisker**
- **Economy (GDP per Capita)**
- **Social support**
- **Healthy life expectancy**
- **Freedom to make life choices**
- **Generosity**
- **Perceptions of corruption**

Data Cleaning & Preparation using Excel and SQL in PostgreSQL

Excel:

- Checked for missing or incorrect values (found that some countries had different names or typos)
- Ensured consistent formatting (e.g., country names, numerical columns)
- Checked for duplicates (nothing found)
- Added an extra column to group the countries in regions (Vlookups)
- Added a "year" column in each file to prepare for uploading the data to Pgadmin

Data Cleaning & Preparation using Excel and SQL in PostgreSQL

SQL (PostgreSQL via PgAdmin):

- Created a new database: HappinessDB
- Loaded the dataset into PostgreSQL from CSV and created a raw data table to include all years: happiness_raw
- Created normalised tables:
    - Country (Stores country names and regions)
    - Year (Stores years to reduce redundancy)
    - Happiness_report (Links countries and years with scores & ranks)
    - Happiness factors (Stores happiness indicators for each country-year)

Data Cleaning & Preparation using Excel and SQL in PostgreSQL

```sql
CREATE TABLE happiness_raw (
    country_name TEXT,
    region TEXT,
    year INT,
    happiness_rank INT,
    happiness_score FLOAT,
    upperwhisker FLOAT,
    lowerwhisker FLOAT,
    economy_gdp_per_capita FLOAT,
    social_support FLOAT,
    healthy_life_expectancy FLOAT,
    freedom_to_make_life_choices FLOAT,
    generosity FLOAT,
    perceptions_of_corruption FLOAT
);
```

```sql
CREATE TABLE country (
    country_id SERIAL PRIMARY KEY,
    country_name TEXT UNIQUE NOT NULL,
    area TEXT
);

CREATE TABLE year (
    year_id SERIAL PRIMARY KEY,
    year INT UNIQUE NOT NULL
);

CREATE TABLE happiness_report (
    report_id SERIAL PRIMARY KEY,
    country_id INT REFERENCES country(country_id),
    year_id INT REFERENCES year(year_id),
    happiness_rank INT,
    happiness_score FLOAT,
    upperwhisker FLOAT,
    lowerwhisker FLOAT
);
```

```sql
CREATE TABLE happiness_factors (
    factor_id SERIAL PRIMARY KEY,
    report_id INT REFERENCES happiness_report(report_id),
    economy_gdp_per_capita FLOAT,
    social_support FLOAT,
    healthy_life_expectancy FLOAT,
    freedom_to_make_life_choices FLOAT,
    generosity FLOAT,
    perceptions_of_corruption FLOAT
);
```

```sql
INSERT INTO country (country_name, region)
SELECT DISTINCT country_name, region FROM happiness_raw;
```

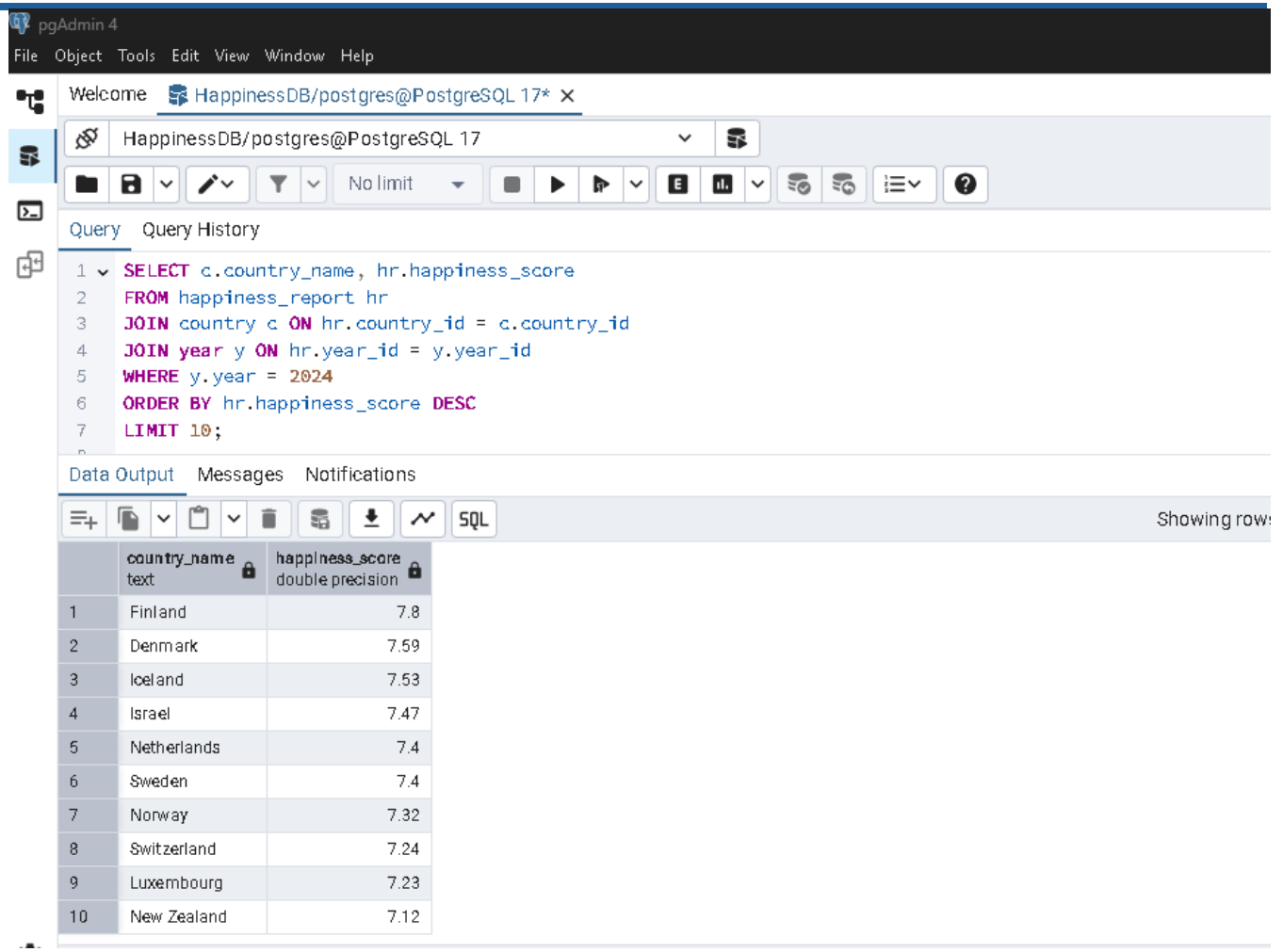
Exploratory Data Analysis using SQL in PostgreSQL and Excel

Used SQL to answer these questions:

- What are the top 10 happiest countries in 2024?
- What is the correlation between GDP and happiness?
- Are there regions with low GDP but high happiness score?

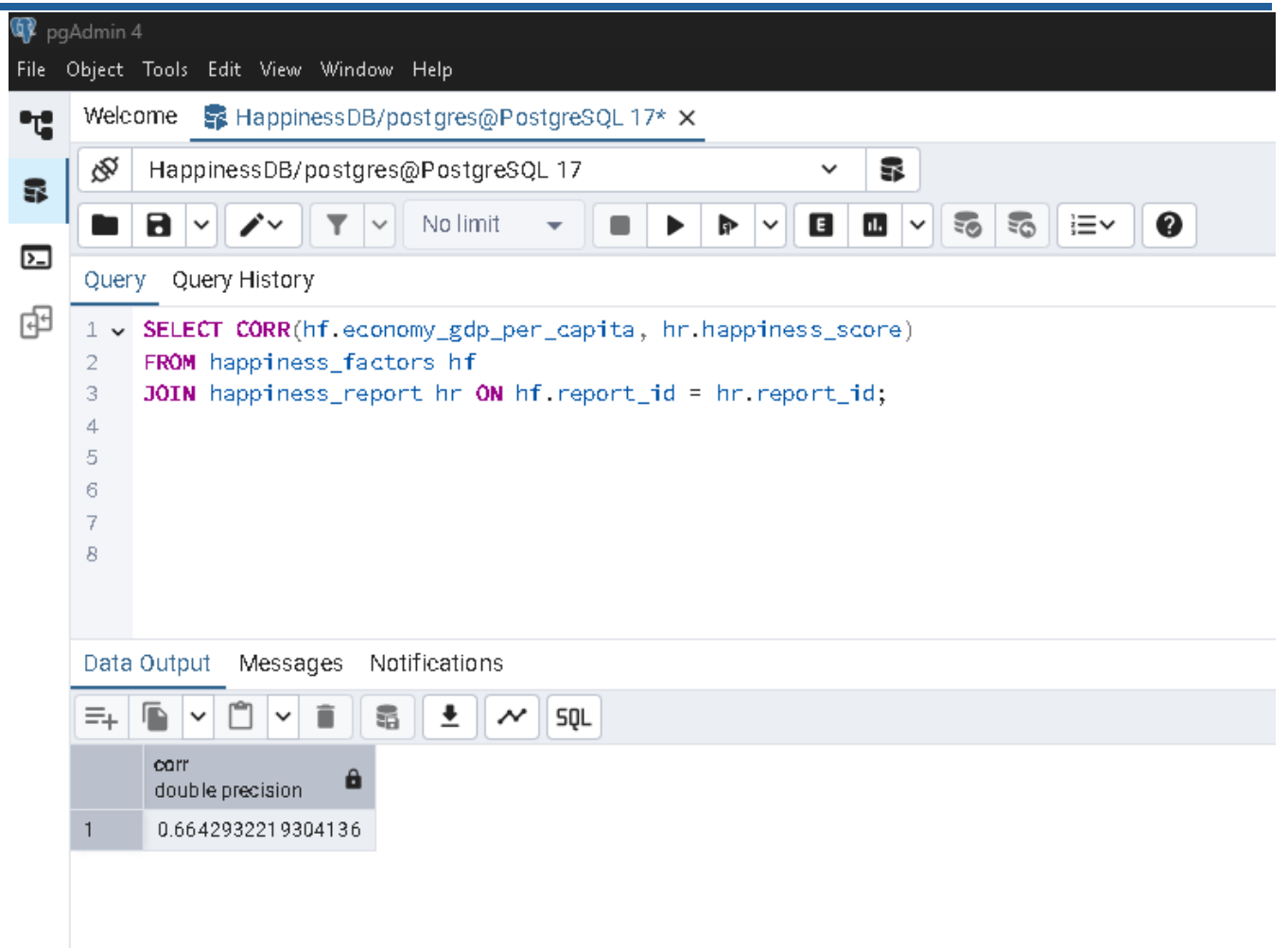
Exploratory Data Analysis using SQL in PostgreSQL

What are the top 10 happiest countries in 2024?



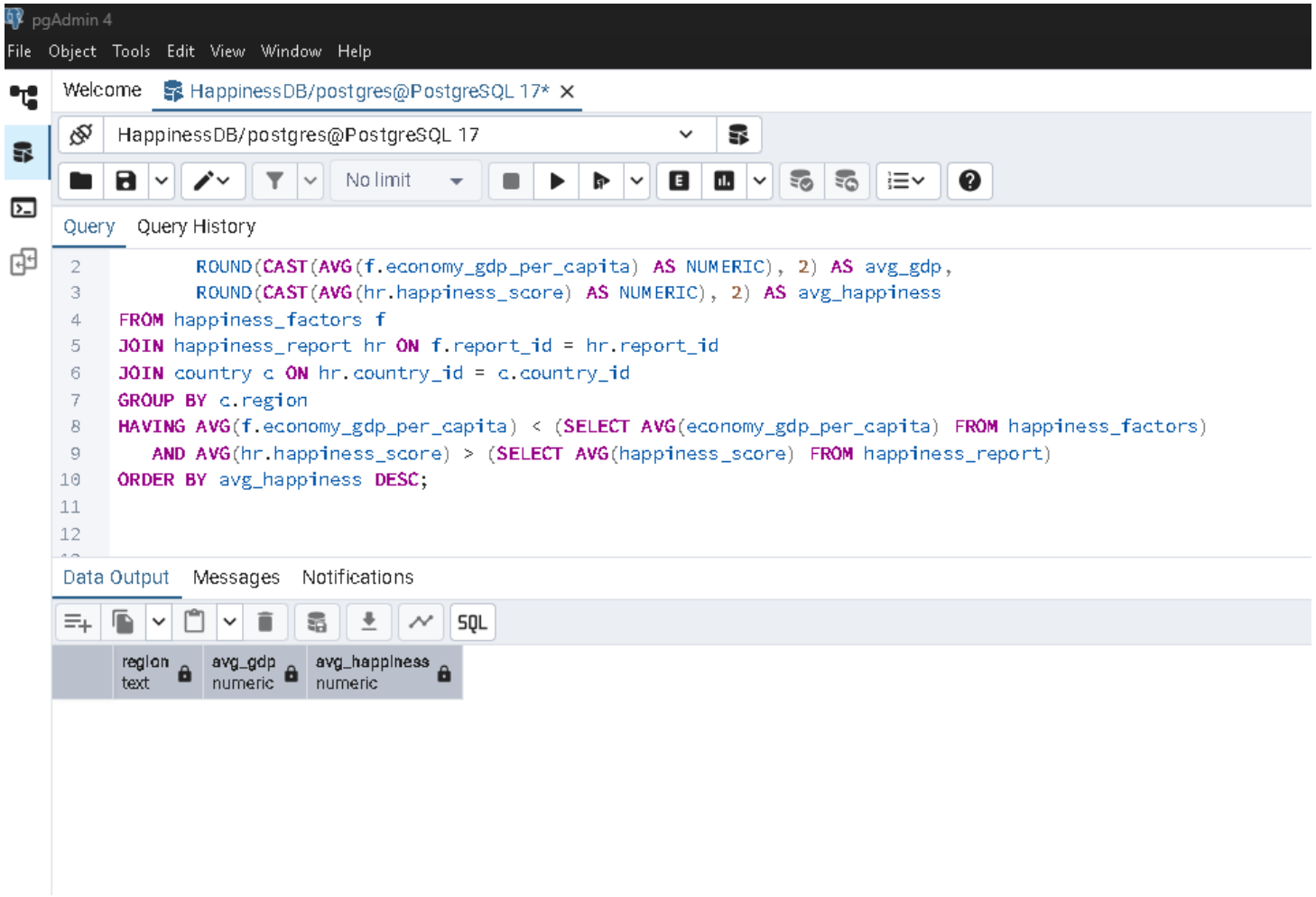
Exploratory Data Analysis using SQL in PostgreSQL

What is the correlation between GDP and happiness?



Exploratory Data Analysis using SQL in PostgreSQL

Are there regions with low GDP but high happiness score?

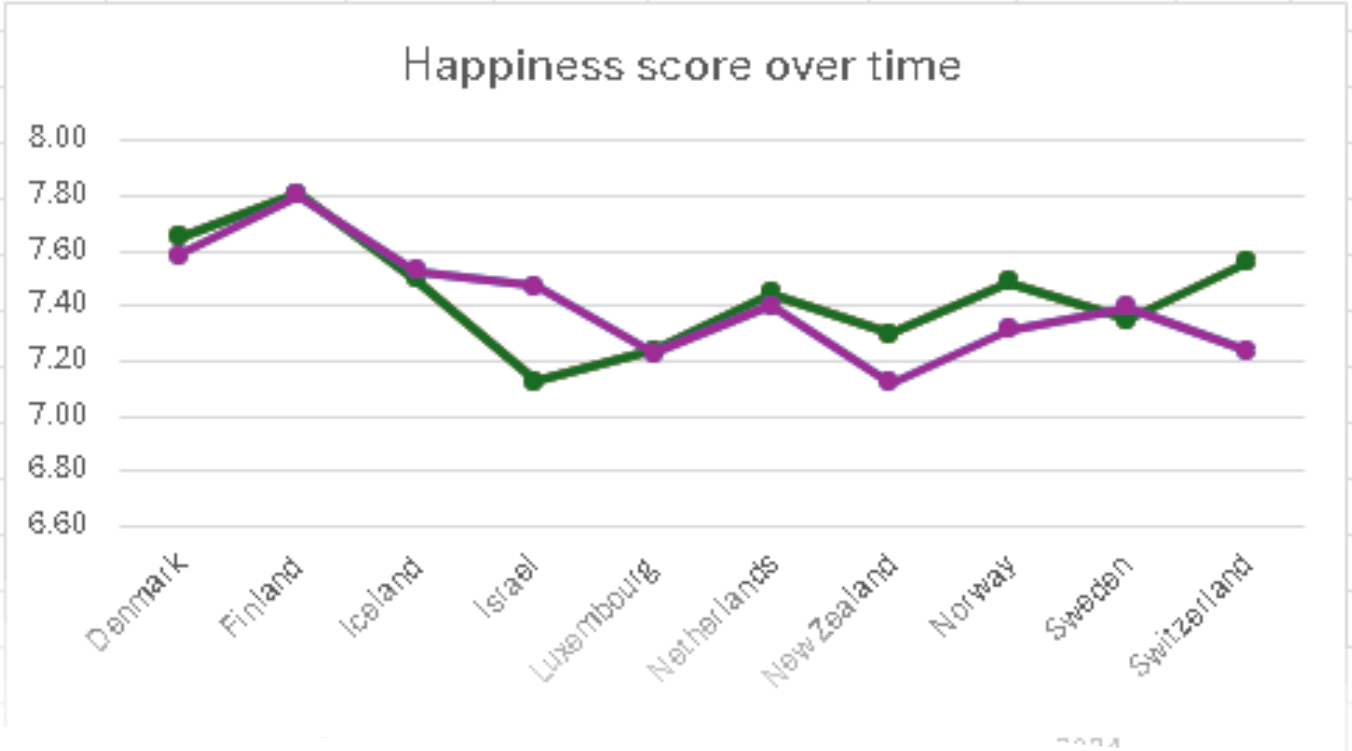


Exploratory Data Analysis using SQL in PostgreSQL and Excel

Used Excel to answer these questions:

- How has the happiness score changed over the years for the top countries?
- Which countries have the highest corruption perception despite being happy?
- Which regions have the highest social support, and does it affect happiness?

Sum of Happiness score	year				
Country name	2020	2021	2022	2023	2024
Denmark	7.65	7.65	7.65	7.59	7.59
Finland	7.81	7.81	7.81	7.80	7.80
Iceland	7.50	7.50	7.50	7.53	7.53
Israel	7.13	7.13	7.13	7.47	7.47
Luxembourg	7.24	7.24	7.24	7.23	7.23
Netherlands	7.45	7.45	7.45	7.40	7.40
New Zealand	7.30	7.30	7.30	7.12	7.12
Norway	7.49	7.49	7.49	7.32	7.32
Sweden	7.35	7.35	7.35	7.40	7.40
Switzerland	7.56	7.56	7.56	7.24	7.24
Grand Total					



Exploratory Data Analysis using Excel

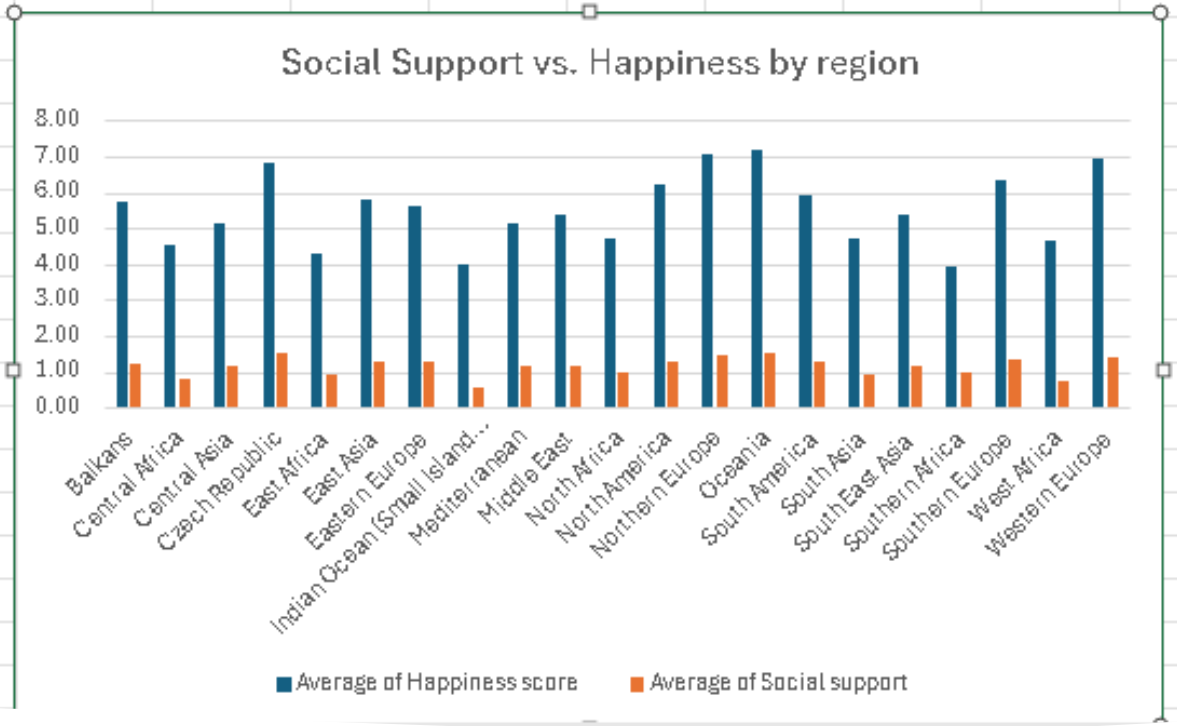
How has the happiness score changed over the years for the top countries?

	A	B	C	D	E	F
3	Country name	Average of Happiness score	Average of Perceptions of corruption			
4	Finland	7.81	0.50			
5	Denmark	7.62	0.51			
6	Iceland	7.51	0.16			
7	Switzerland	7.43	0.44			
8	Netherlands	7.43	0.38			
9	Norway	7.42	0.44			
10	Sweden	7.37	0.47			
11	Israel	7.27	0.12			
12	Luxembourg	7.24	0.39			
13	New Zealand	7.23	0.46			
14	Austria					

Exploratory Data Analysis using Excel

Which countries have the highest corruption perception despite being happy?

Region	Average of Happiness score	Average of Social support
Balkans	5.74	1.25
Central Africa	4.56	0.77
Central Asia	5.13	1.18
Czech Republic	6.85	1.54
East Africa	4.26	0.90
East Asia	5.80	1.29
Eastern Europe	5.64	1.29
Indian Ocean (Small Island Nations	3.99	0.56
Mediterranean	5.15	1.18
Middle East	5.38	1.17
North Africa	4.74	0.96
North America	6.21	1.27
Northern Europe	7.05	1.49
Oceania	7.20	1.50
South America	5.90	1.30



Exploratory Data Analysis using Excel

Which regions have the highest social support, and does it affect happiness?

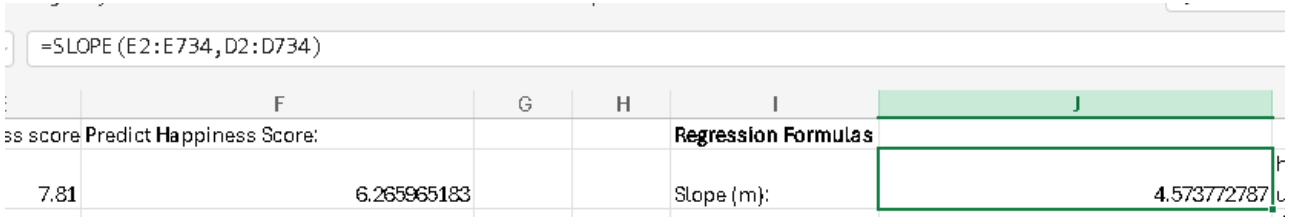
Statistical Analysis

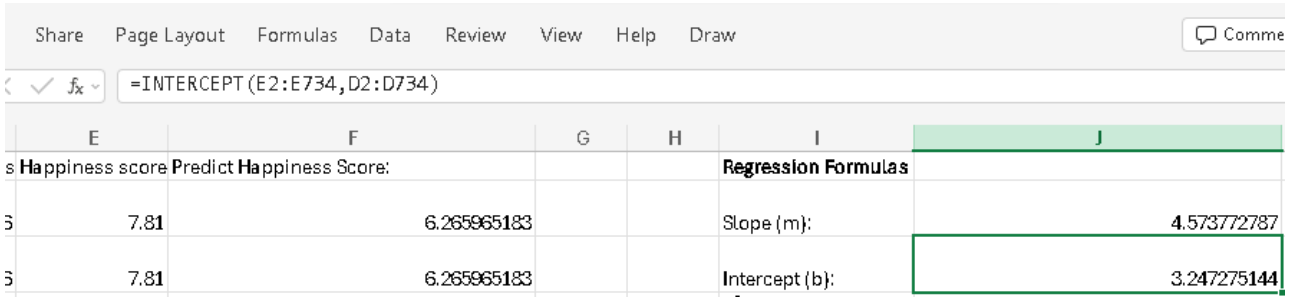
We want to **predict Happiness Score based on Freedom to Make Life Choices** using **Simple Linear Regression**:

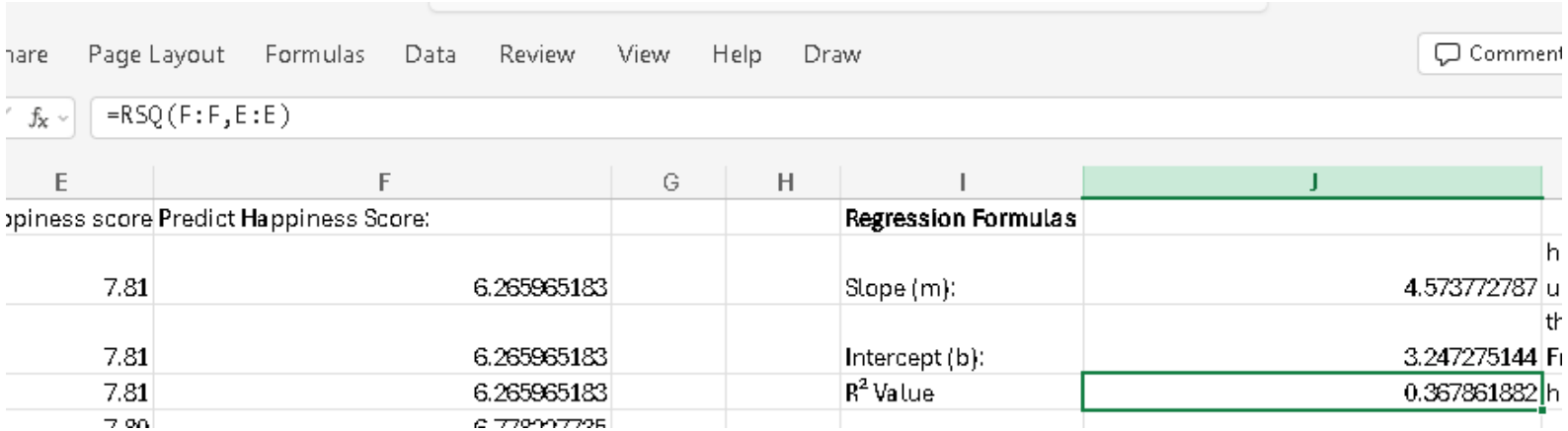
$Y=mX+bY = mX + bY=mX+b$

Where:

- **Y (Dependent Variable):** Happiness Score
- **X (Independent Variable):** Freedom to Make Life Choices
- **m (Slope)** and **b (Intercept)** are regression coefficients

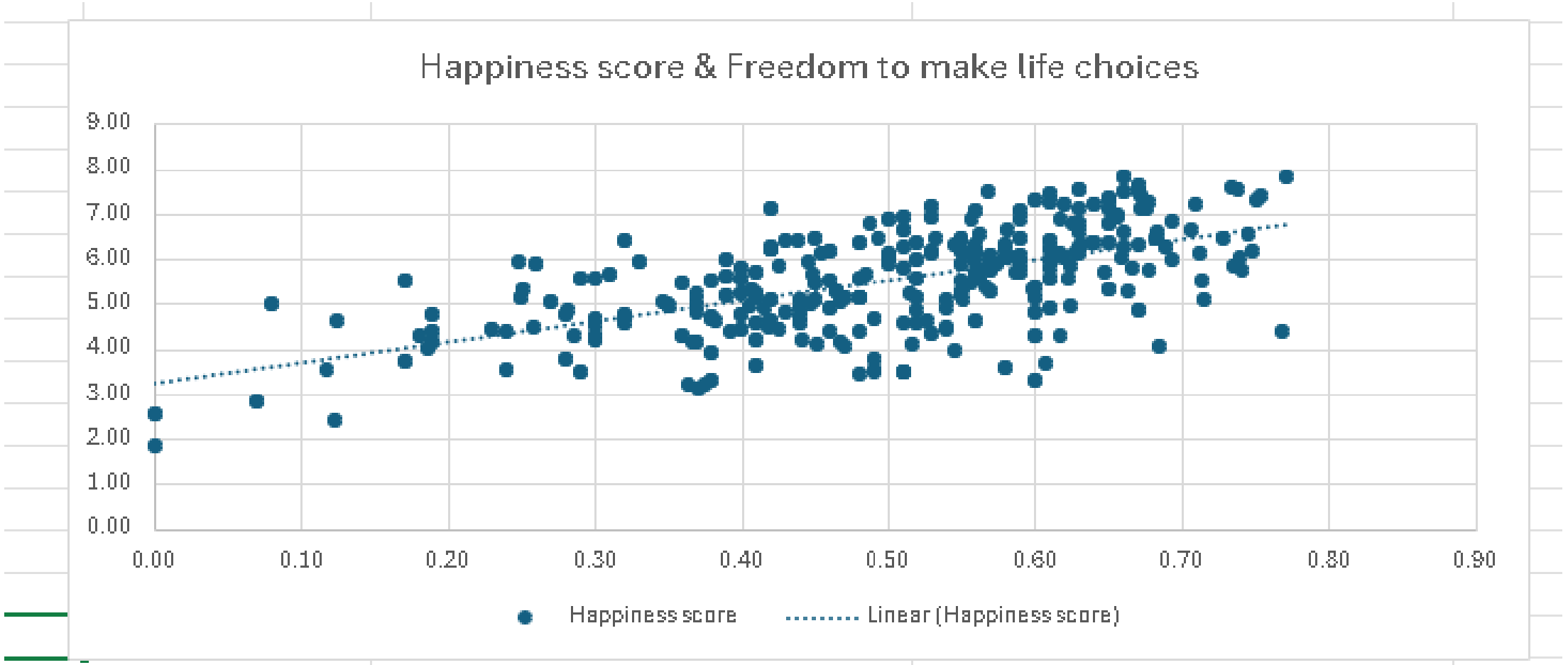






Regression Coefficients and R^2

- Slope (m), to understand the change in Happiness Score for every 1-unit change in Freedom to Make Life Choices.
- Intercept (b), to see the predicted Happiness Score when Freedom = 0
- R^2, to evaluate how strong the relationship between "happiness score" and "Freedom to Make Life Choices" is



Statistical Analysis

Scatter Plot with Trendline

Data Visualization with Tableau

1. Line Chart – "Happiness Score Trend Over the Years"

Question: *How has the Happiness Score changed over time for the top countries?*

2. Scatter Plot – "Which Countries Have High Happiness Despite High Corruption?"

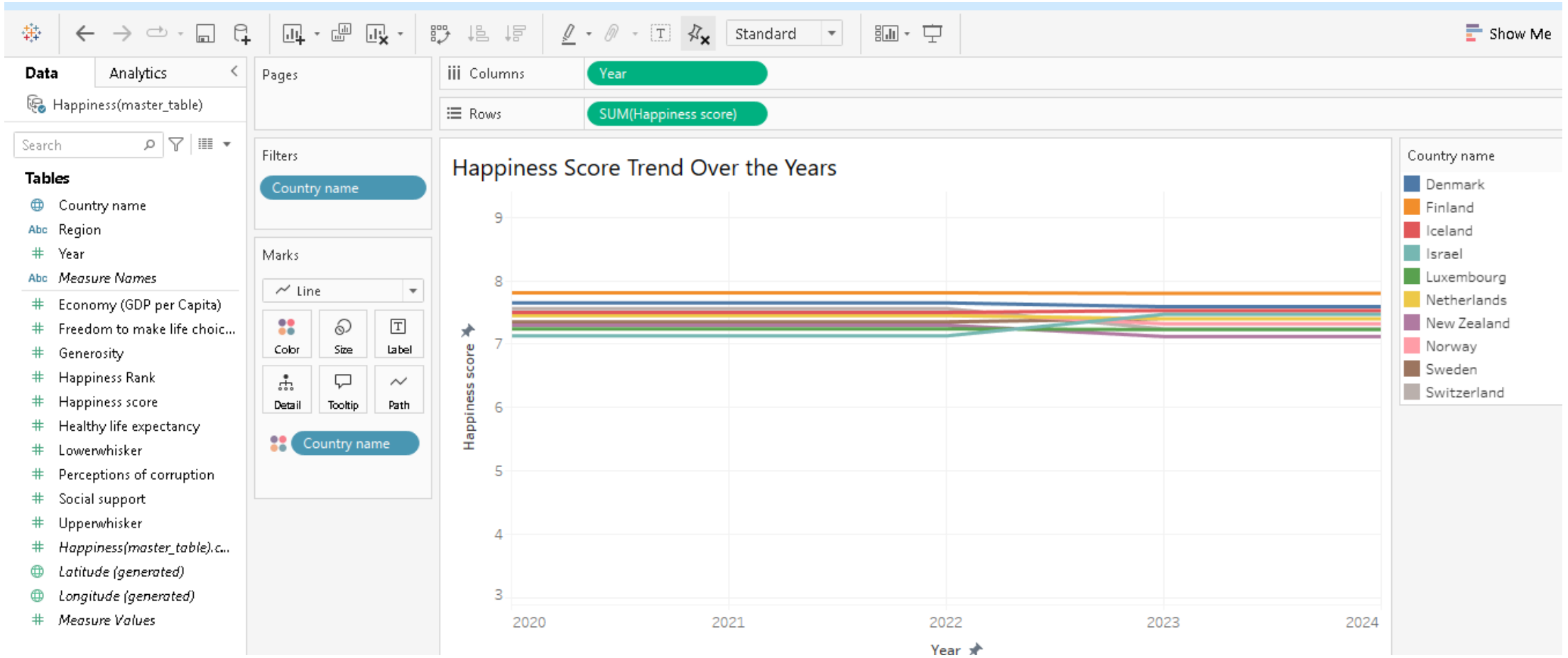
Question: *Do some countries remain happy despite high corruption perception?*

3. Bar Chart – "Regions with Low GDP but High Happiness"

Question: *Which regions are happy despite low economic performance?*

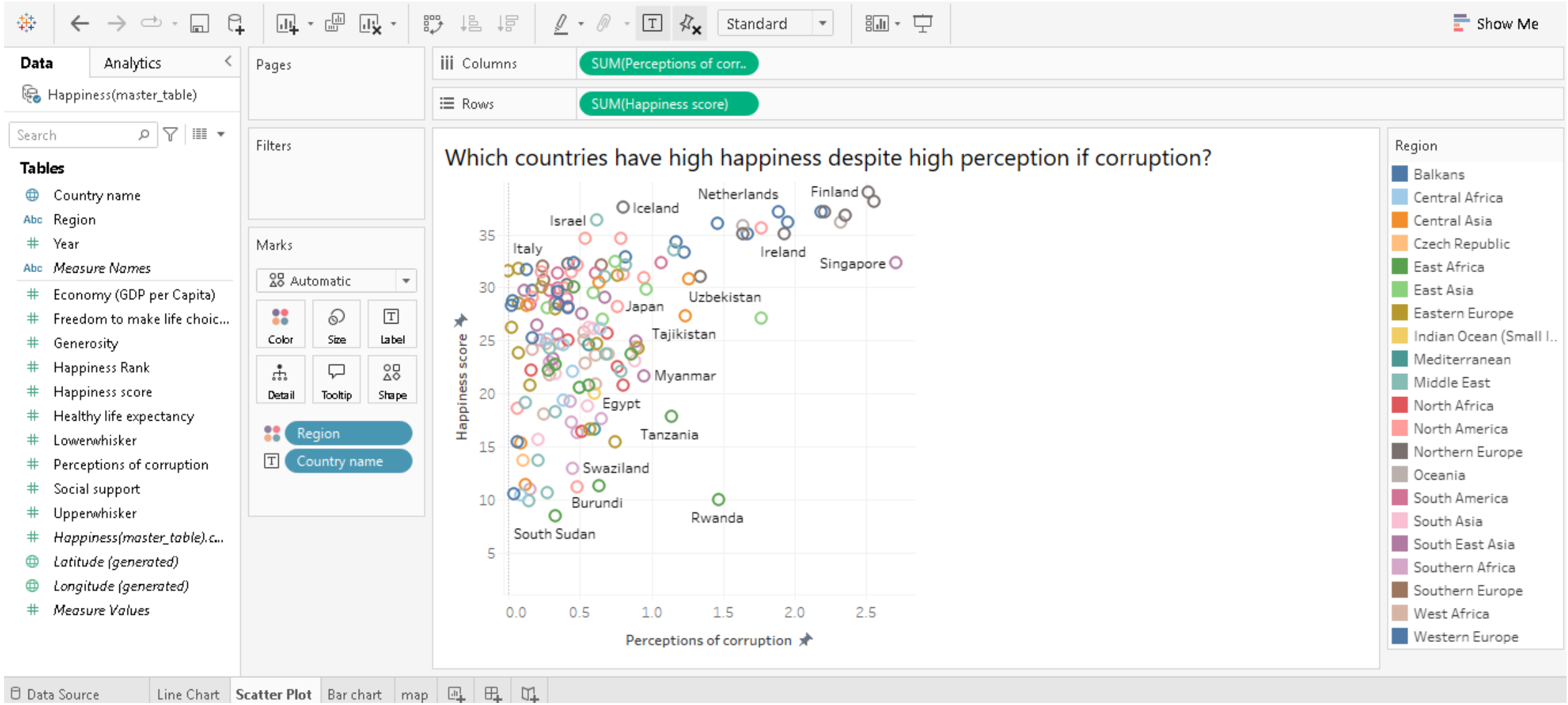
4. Map Visualization – "Global Happiness Distribution"

Question: *Which countries have the highest and lowest happiness?*



Data Visualization with Tableau

Line Chart – "Happiness Score Trend Over the Years"

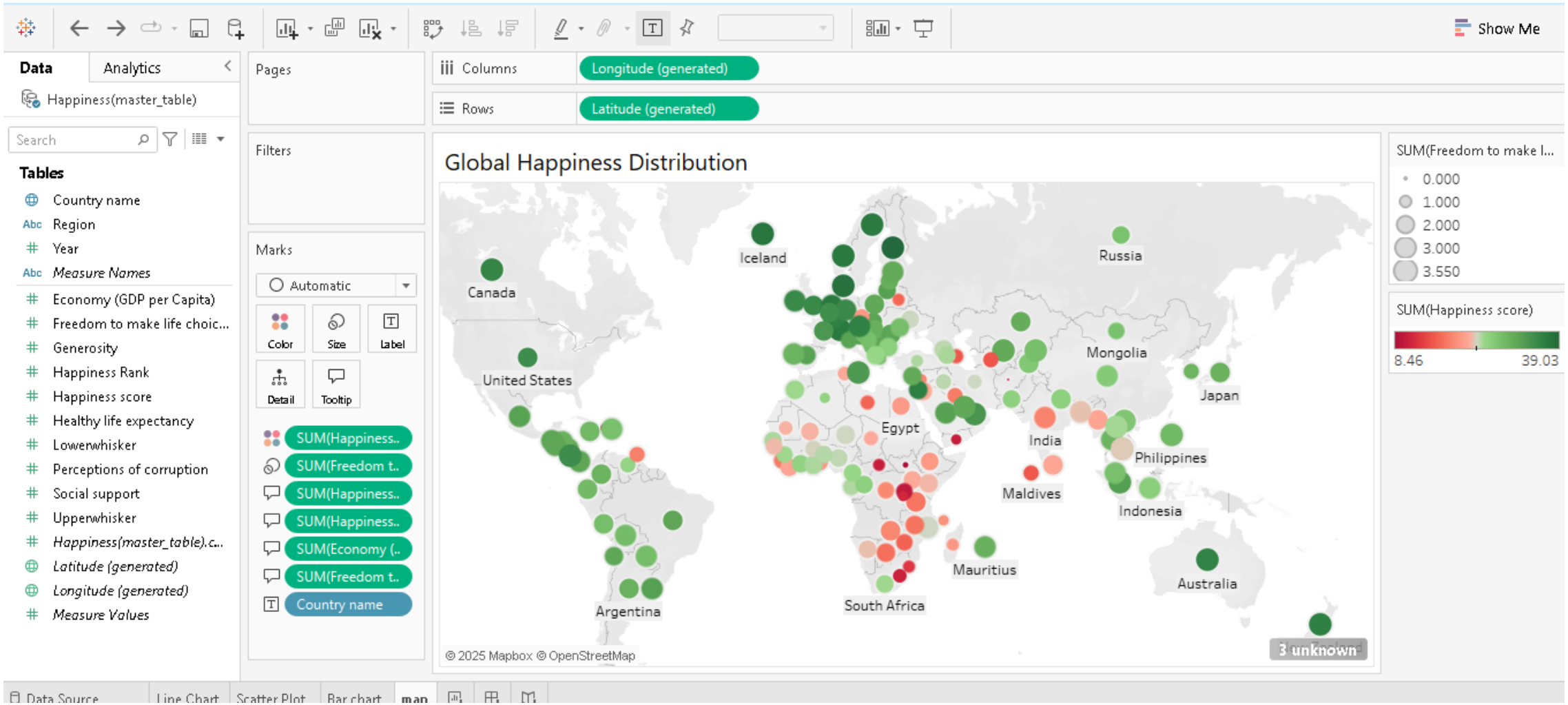


Data Visualization with Tableau

Scatter Plot – "Which Countries Have High Happiness Despite High Corruption?"

Data Visualization with Tableau

Bar Chart – "Regions with Low GDP but High Happiness



Data Visualization with Tableau

Map Visualization – "Global Happiness Distribution"

Results

Statistical Analysis

Based on these results:

- **Slope (m) = 4.57**
- **Intercept (b) = 3.25**
- **R^2 = 0.37**

The regression equation is:Happiness Score=(4.57×Freedom)+3.25Happiness\ Score = (4.57 \times Freedom) + 3.25Happiness Score=(4.57×Freedom)+3.25

- This means for every **1-unit increase in Freedom to Make Life Choices**, the **Happiness Score increases by 4.57 points** (on average).

- If a country had **zero** freedom (Freedom = 0), its predicted **Happiness Score would be 3.25** (Intercept).How Strong is the Relationship? (R^2 Interpretation)**R^2 value = 0.37**

- **37% of the variation** in Happiness Score can be explained by Freedom to Make Life Choices.
- The remaining **63%** is influenced by other factors (e.g., GDP, social support, health, etc.).

This indicates a **moderate** correlation but not a very strong one.

Looking at the **scatter plot** comparing **Freedom to Make Life Choices (X-axis)** and **Happiness Score (Y-axis),** the **dots form an upward pattern**, this shows a **positive relationship** between Freedom andHappiness Score.

The regression equation shows a positive slope (4.57), meaning more freedom generally leads tohigher happiness.

Since the R^2 is 0.37, confirms what is shown on the chart: some scatter but an overall positive trend.

Discussion

Insights

- **Freedom matters, but isn’t everything**
  A 1-point increase in "Freedom to make life choices" correlates with a 4.57-point increase in happiness score. However, $R^2 = 0.37$ indicates only moderate explanatory power—other factors also play large roles.
- **GDP is correlated with happiness**, but not equally across regions
  Some low-GDP countries (especially in Latin America) rank surprisingly high in happiness, suggesting cultural, environmental, or community factors at play.
- **Corruption perception has mixed effects**
  Several countries with high corruption perception still report high happiness, showing that trust in local social structures or cultural norms may offset institutional flaws.
- **Social support appears consistently important**
  Regions with strong social support networks tend to have higher happiness scores, even in the absence of strong economic indicators.

Conclusion

Summary

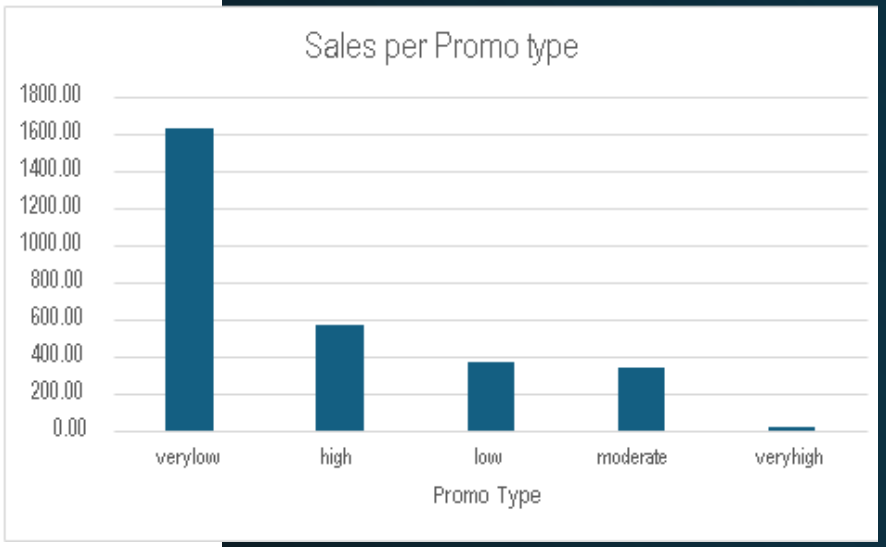
This analysis highlights that happiness is driven by a mix of economic, social, and personal factors—no single metric can fully explain national well-being.

While freedom and GDP show moderate correlations, exceptions point to the importance of contextual factors like culture, social capital, and public trust.

Future work could include multivariate regression to better isolate the contribution of each factor or clustering analysis to group countries by happiness profiles.

Appendix

Appendix


Sales per Promo type
1800.00
1600.00
1400.00
1200.00
1000.00
800.00
600.00
400.00
200.00
0.00
verylow
high
low
moderate
veryhigh
Promo Type