World Happiness Report 2023 - EDA

April 23, 2023

0.1 World Happiness Report 2023 - Exploratory Data Analysis

The World Happiness Report is an annual publication that ranks countries based on their happiness levels, as measured by a range of economic, social, and political indicators. The report is produced by the United Nations Sustainable Development Solutions Network. It is designed to provide policymakers, academics, and the general public with insights into the factors that contribute to happiness and well-being around the world.

In this project, an analysis of the World Happiness Report 2023 dataset was conducted, exploring the relationships between happiness scores, regional trends, and key indicators such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption.

The dataset was downloaded from the World Happiness Report website, which provides open access to the data and tools for analysis. By examining these relationships, insights were gained into the complex and multifaceted nature of happiness and well-being, and potential strategies for promoting happiness and improving quality of life around the world were identified. Finally, the happiness scores of all the countries around the world were presented on a map with a timeline from 2005 to 2022.

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
```

World Happiness Report Dataset

[2]:
```
whr_df = pd.read_csv("whr2023.csv")
whr_df.head()
```

[2]:

	Country name	iso alpha	Regional indicator	Happiness score	\
0	Afghanistan	AFG	South Asia	1.859	
1	Albania	ALB	Central and Eastern Europe	5.277	
2	Algeria	DZA	Middle East and North Africa	5.329	
3	Argentina	ARG	Latin America and Caribbean	6.024	
4	Armenia	ARM	Commonwealth of Independent States	5.342	

	Standard error of ladder score	upperwhisker	lowerwhisker	\
0	0.033	1.923	1.795	
1	0.066	5.406	5.148	
2	0.062	5.451	5.207	
3	0.063	6.147	5.900	
4	0.066	5.470	5.213	

	Logged GDP per capita	Social support	Healthy life expectancy	…	\
0	7.324	0.341	54.712	…	
1	9.567	0.718	69.150	…	
2	9.300	0.855	66.549	…	
3	9.959	0.891	67.200	…	
4	9.615	0.790	67.789	…	

	Generosity	Perceptions of corruption	Ladder score in Dystopia	\
0	-0.081	0.847	1.778	
1	-0.007	0.878	1.778	
2	-0.117	0.717	1.778	
3	-0.089	0.814	1.778	
4	-0.155	0.705	1.778	

	Explained by: Log GDP per capita	Explained by: Social support	\
0	0.645	0.000	
1	1.449	0.951	
2	1.353	1.298	
3	1.590	1.388	
4	1.466	1.134	

	Explained by: Healthy life expectancy	\
0	0.087	
1	0.480	
2	0.409	
3	0.427	
4	0.443	

	Explained by: Freedom to make life choices	Explained by: Generosity	\
0	0.000	0.093	
1	0.549	0.133	
2	0.252	0.073	
3	0.587	0.088	
4	0.551	0.053	

	Explained by: Perceptions of corruption	Dystopia + residual
0	0.059	0.976
1	0.037	1.678
2	0.152	1.791
3	0.082	1.861

```
4                                    0.160                 1.534

[5 rows x 21 columns]
```

Filter the columns and copy only the ones that will be used for analysis.

[20]:
```
columns = ['Country name', 'iso alpha', 'Regional indicator', 'Happiness␣
 ↪score', 'Logged GDP per capita', 'Social support', 'Healthy life␣
 ↪expectancy', 'Freedom to make life choices', 'Generosity', 'Perceptions of␣
 ↪corruption']
happy_df = whr_df[columns].copy()
happy_df.head()
```

[20]:
```
  Country name iso alpha                 Regional indicator  Happiness score  \
0  Afghanistan       AFG                         South Asia            1.859
1      Albania       ALB          Central and Eastern Europe            5.277
2      Algeria       DZA        Middle East and North Africa            5.329
3    Argentina       ARG         Latin America and Caribbean            6.024
4      Armenia       ARM  Commonwealth of Independent States            5.342

   Logged GDP per capita  Social support  Healthy life expectancy  \
0                  7.324           0.341                   54.712
1                  9.567           0.718                   69.150
2                  9.300           0.855                   66.549
3                  9.959           0.891                   67.200
4                  9.615           0.790                   67.789

   Freedom to make life choices  Generosity  Perceptions of corruption
0                         0.382      -0.081                      0.847
1                         0.794      -0.007                      0.878
2                         0.571      -0.117                      0.717
3                         0.823      -0.089                      0.814
4                         0.796      -0.155                      0.705
```

Convert all column names to lowercase.

[22]:
```
happy_df.columns = happy_df.columns.str.lower()
happy_df.head()
```

[22]:
```
  country name iso alpha                 regional indicator  happiness score  \
0  Afghanistan       AFG                         South Asia            1.859
1      Albania       ALB          Central and Eastern Europe            5.277
2      Algeria       DZA        Middle East and North Africa            5.329
3    Argentina       ARG         Latin America and Caribbean            6.024
4      Armenia       ARM  Commonwealth of Independent States            5.342

   logged gdp per capita  social support  healthy life expectancy  \
0                  7.324           0.341                   54.712
```

```
1                  9.567           0.718                    69.150
2                  9.300           0.855                    66.549
3                  9.959           0.891                    67.200
4                  9.615           0.790                    67.789

   freedom to make life choices  generosity  perceptions of corruption
0                         0.382      -0.081                      0.847
1                         0.794      -0.007                      0.878
2                         0.571      -0.117                      0.717
3                         0.823      -0.089                      0.814
4                         0.796      -0.155                      0.705
```

Data Info

[5]:
```
happy_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 137 entries, 0 to 136
Data columns (total 10 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   country name                    137 non-null    object
 1   iso alpha                       137 non-null    object
 2   regional indicator              137 non-null    object
 3   happiness score                 137 non-null    float64
 4   logged gdp per capita           137 non-null    float64
 5   social support                  137 non-null    float64
 6   healthy life expectancy         136 non-null    float64
 7   freedom to make life choices    137 non-null    float64
 8   generosity                      137 non-null    float64
 9   perceptions of corruption       137 non-null    float64
dtypes: float64(7), object(3)
memory usage: 10.8+ KB
```

[6]:
```
# Check if there are any missing values
happy_df.isnull().sum()
```

[6]:
```
country name                    0
iso alpha                       0
regional indicator              0
happiness score                 0
logged gdp per capita           0
social support                  0
healthy life expectancy         1
freedom to make life choices    0
generosity                      0
perceptions of corruption       0
dtype: int64
```

[7]:
```
happy_df[happy_df.isnull().any(axis = 1)]
```

[7]:
```
           country name iso alpha            regional indicator  \
116  State of Palestine       PSE  Middle East and North Africa

     happiness score  logged gdp per capita  social support  \
116            4.908                  8.716           0.859

     healthy life expectancy  freedom to make life choices  generosity  \
116                      NaN                         0.694      -0.132

     perceptions of corruption
116                      0.836
```

- A missing value has been identified in the healthy life expectancy column for the country State of Palestine. It has been decided to leave the value as is, as assumptions could bias the information and the value is not provided in the report. It has been noted for future reference. The decision to leave this missing value as is will not affect the analysis.

[8]:
```
# Choose the numerical columns for further analysis
numerical_columns = ['happiness score', 'logged gdp per capita', 'social␣
 ↪support', 'healthy life expectancy', 'freedom to make life choices',␣
 ↪'generosity', 'perceptions of corruption']
happy_df[numerical_columns].describe()
```

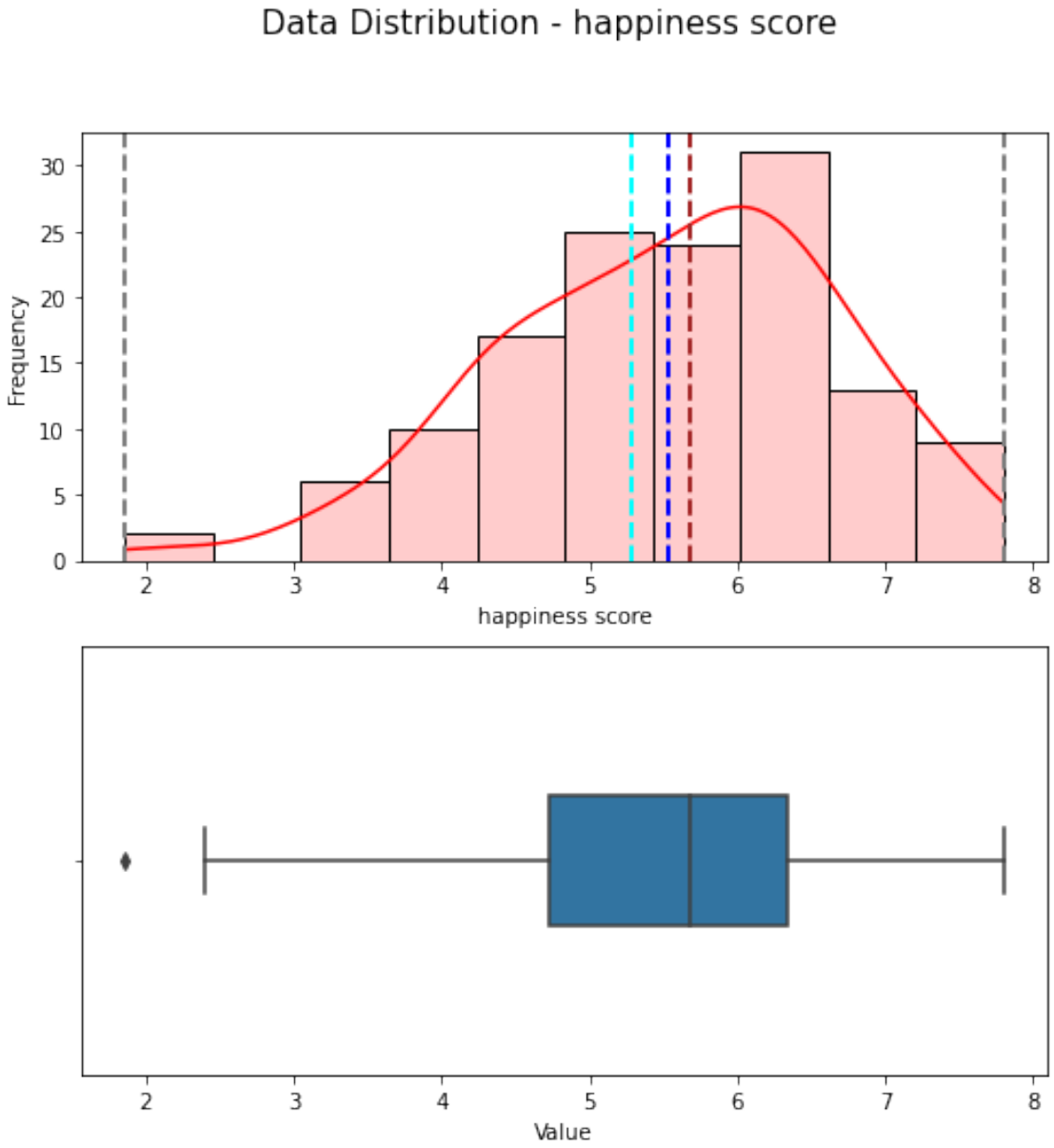
[8]:
```
       happiness score  logged gdp per capita  social support  \
count       137.000000             137.000000      137.000000
mean          5.539796               9.449796        0.799073
std           1.139929               1.207302        0.129222
min           1.859000               5.527000        0.341000
25%           4.724000               8.591000        0.722000
50%           5.684000               9.567000        0.827000
75%           6.334000              10.540000        0.896000
max           7.804000              11.660000        0.983000

       healthy life expectancy  freedom to make life choices  generosity  \
count               136.000000                    137.000000  137.000000
mean                 64.967632                      0.787394    0.022431
std                   5.750390                      0.112371    0.141707
min                  51.530000                      0.382000   -0.254000
25%                  60.648500                      0.724000   -0.074000
50%                  65.837500                      0.801000    0.001000
75%                  69.412500                      0.874000    0.117000
max                  77.280000                      0.961000    0.531000

       perceptions of corruption
count                 137.000000
mean                    0.725401
```

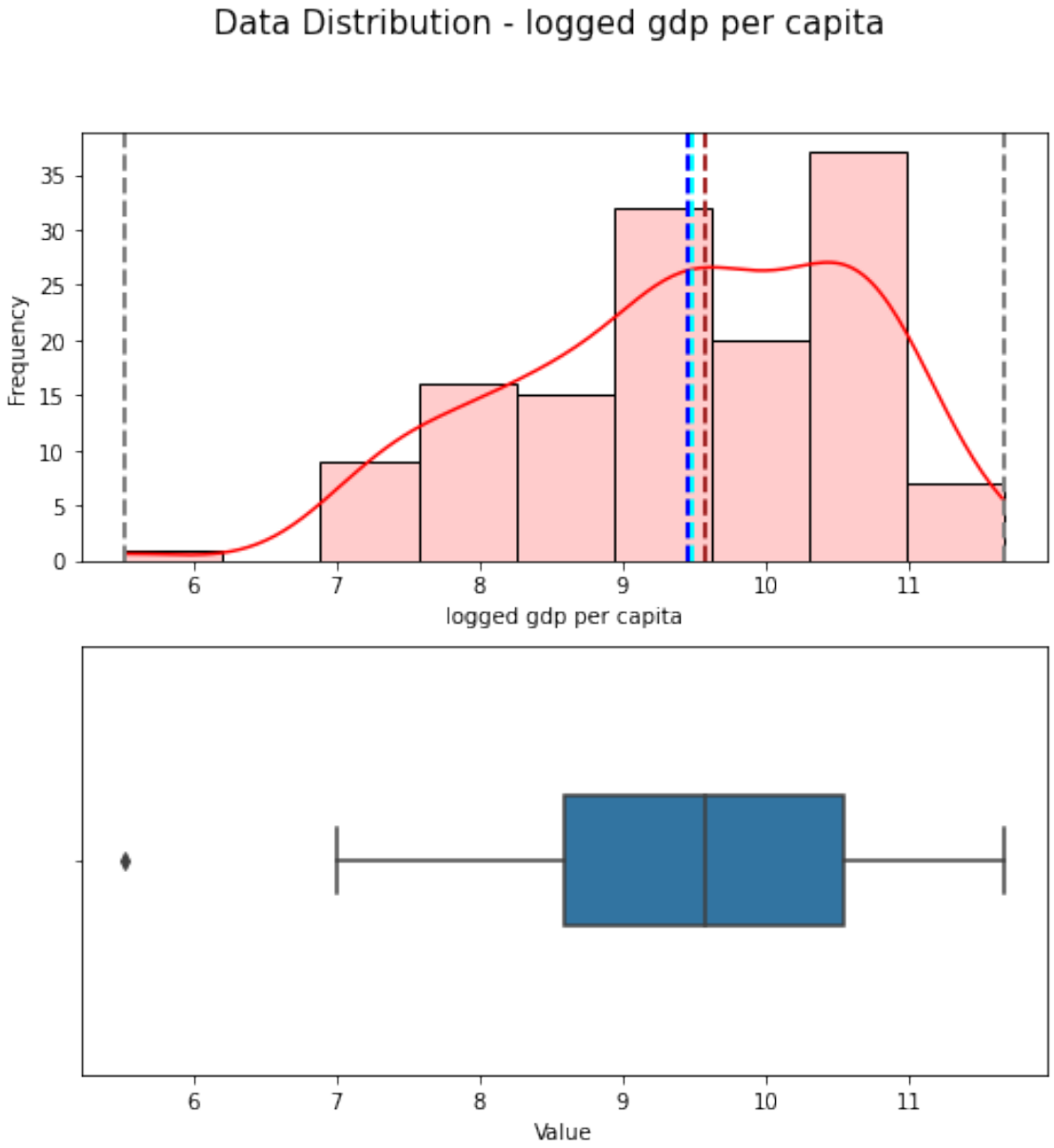
```
std                   0.176956
min                   0.146000
25%                   0.668000
50%                   0.774000
75%                   0.846000
max                   0.929000
```

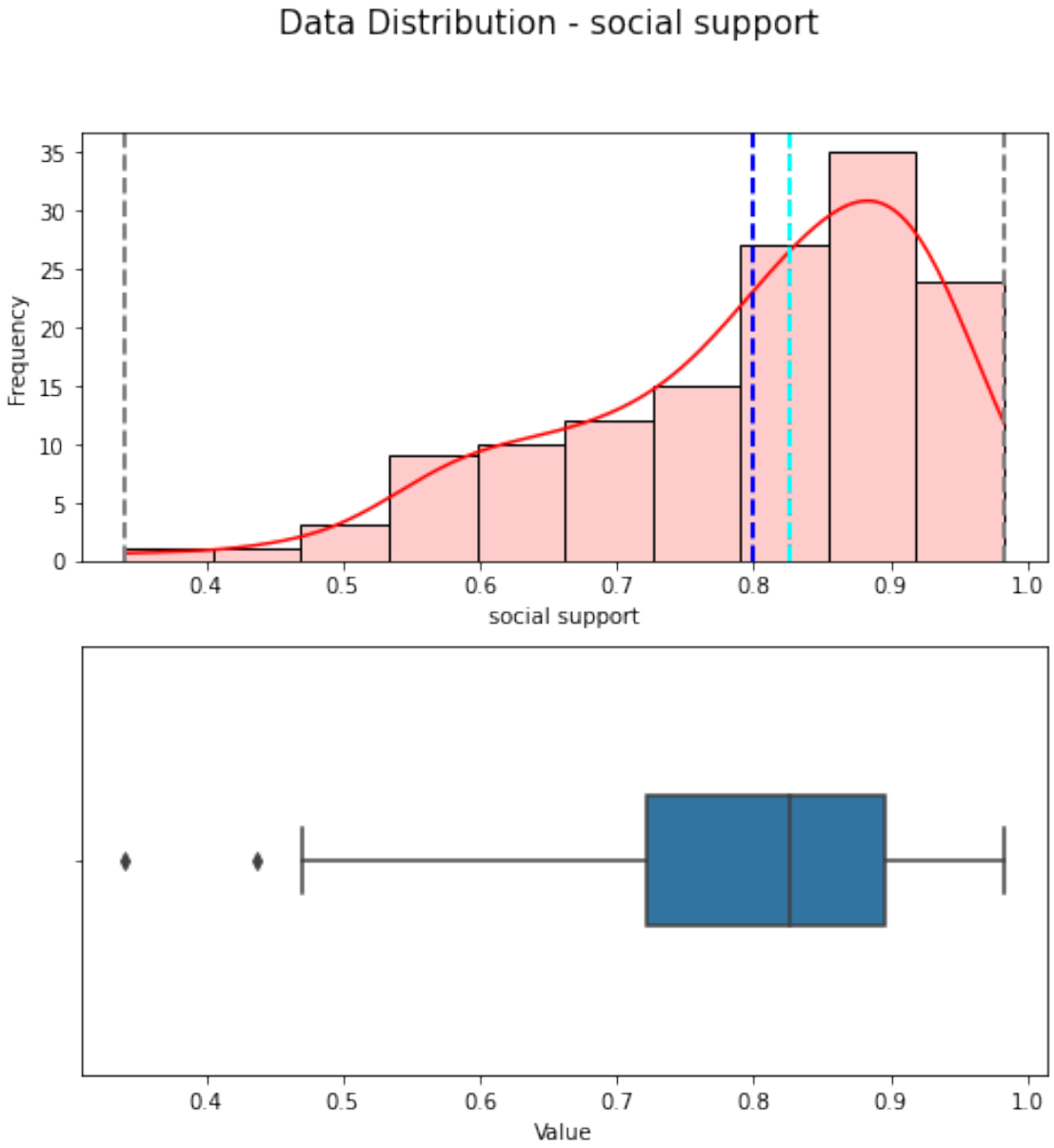
The distribution of numerical data using both histograms and boxplots.

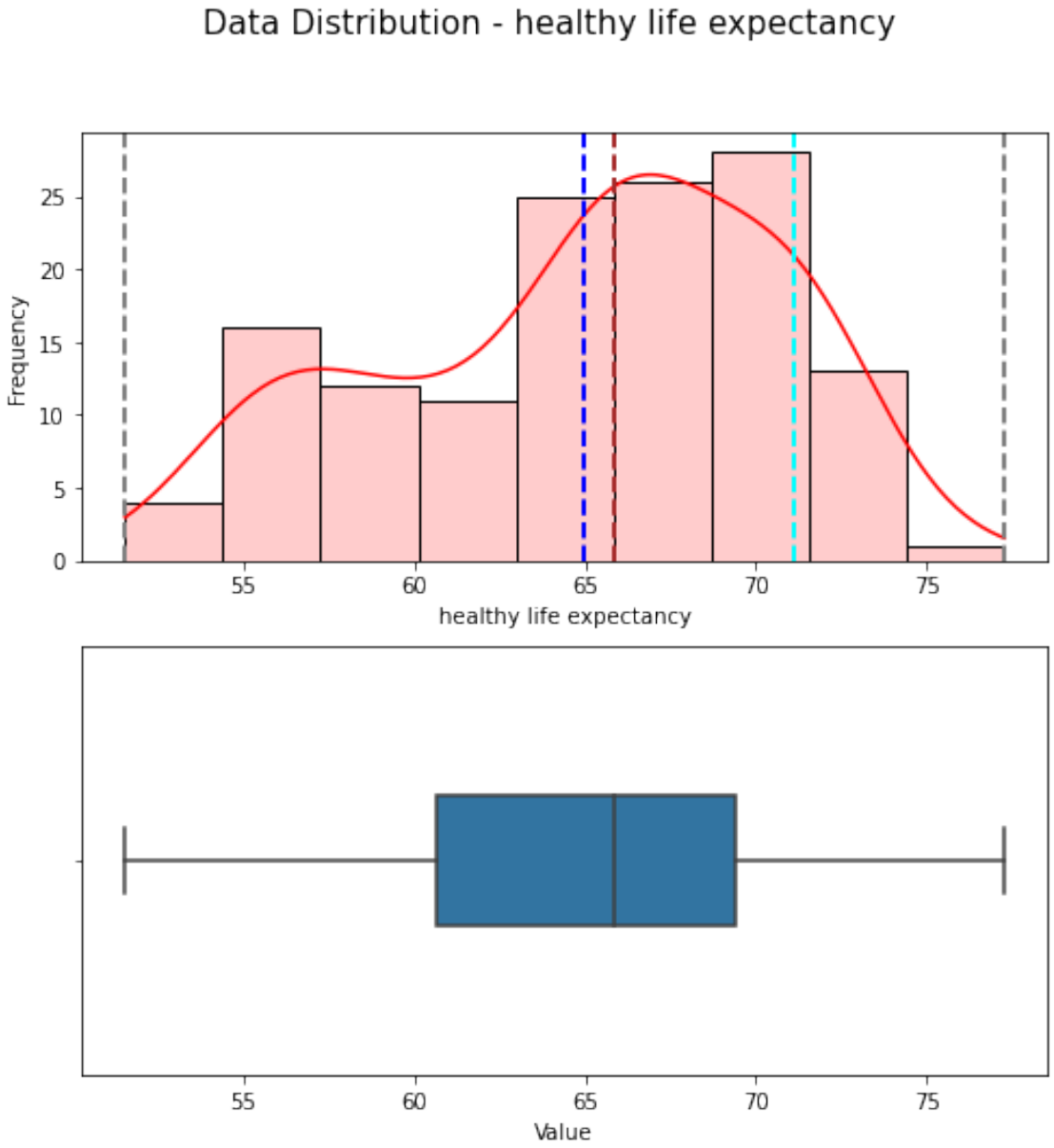
[9]:
```
# Create a function that displays the distribution of numerical data using both␣
 ↪histograms and boxplots.
def show_distribution(var):
    var_min = happy_df[var].min()
    var_mean = happy_df[var].mean()
    var_median = happy_df[var].median()
    var_mode = happy_df[var].mode()[0]
    var_max = happy_df[var].max()

    # Histogram
    fig, ax = plt.subplots(2, 1, figsize = (8,8))
    sns.histplot(happy_df, x = var, kde = True, color = 'red', alpha = 0.2, ax␣
 ↪= ax[0])
    ax[0].set_ylabel("Frequency")

    ax[0].axvline(x = var_min, color = 'gray', linewidth = 2, linestyle = "--")
    ax[0].axvline(x = var_mean, color = 'blue', linewidth = 2, linestyle = "--")
    ax[0].axvline(x = var_median, color = 'brown', linewidth = 2, linestyle =␣
 ↪"--")
    ax[0].axvline(x = var_mode, color = 'cyan', linewidth = 2, linestyle = "--")
    ax[0].axvline(x = var_max, color = 'gray', linewidth = 2, linestyle = "--")

    # Box plot
    sns.boxplot(happy_df, x = var, width = 0.3, ax = ax[1],)
    ax[1].set_xlabel("Value")

    fig.suptitle("Data Distribution - " + var, size = 15)
    fig.show()
```

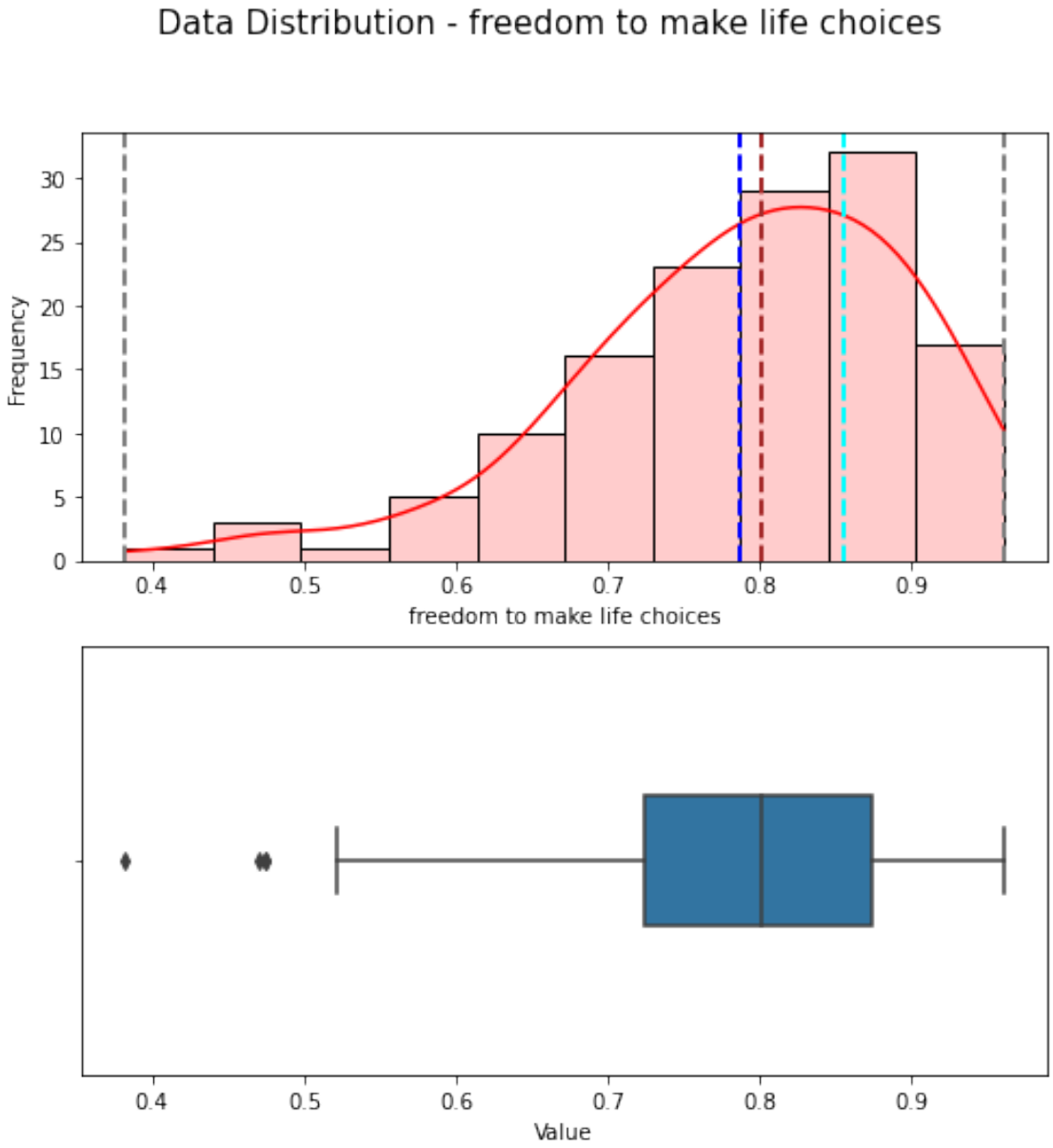
[10]:
```
# Call show_distribution function
for column in numerical_columns:
    show_distribution(column)
```

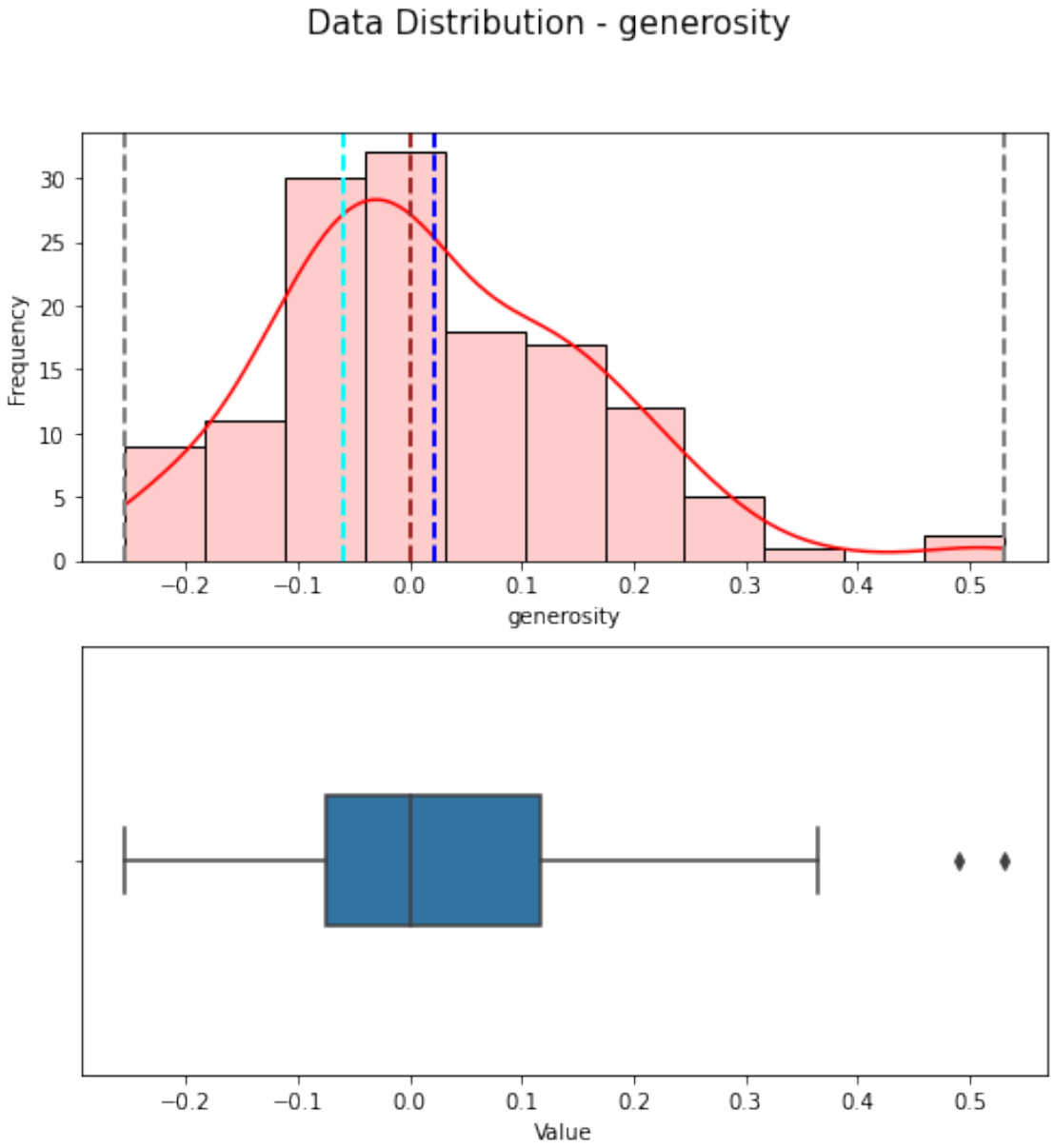

Data Distribution - happiness score
Frequency
happiness score
Value
0
5
10
15
20
25
30
2
3
4
5
6
7
8

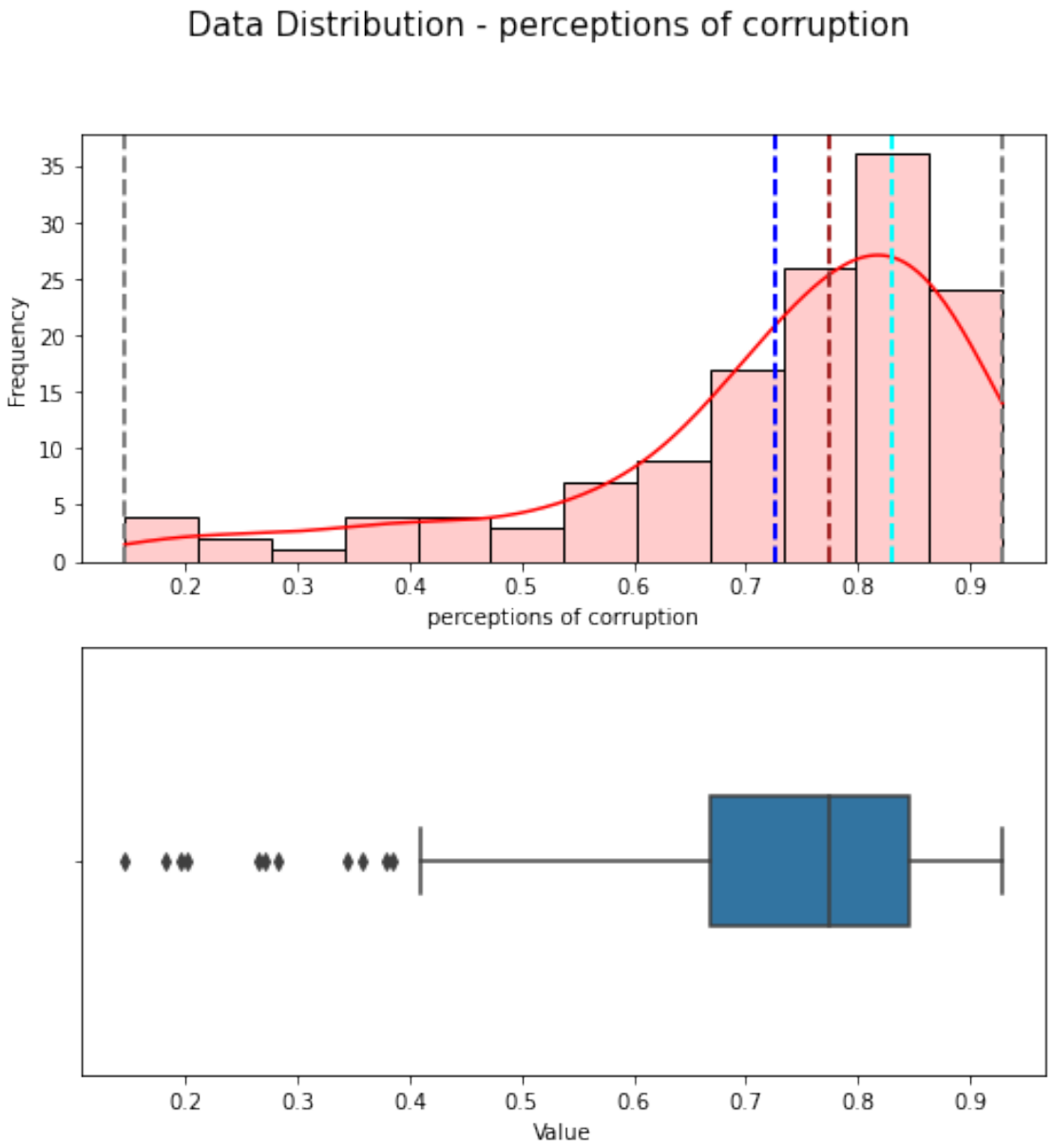

Data Distribution - logged gdp per capita
Frequency
0
5
10
15
20
25
30
35
6
7
8
9
10
11
logged gdp per capita
6
7
8
9
10
11
Value


Data Distribution - social support
Frequency
social support
Value
0
5
10
15
20
25
30
35
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Data Distribution - healthy life expectancy
Frequency
0
5
10
15
20
25
55
60
65
70
75
healthy life expectancy
55
60
65
70
75
Value


Data Distribution - freedom to make life choices
Frequency
0
5
10
15
20
25
30
0.4
0.5
0.6
0.7
0.8
0.9
freedom to make life choices
0.4
0.5
0.6
0.7
0.8
0.9
Value


Data Distribution - generosity
Frequency
generosity
Value
0
5
10
15
20
25
30
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
0.5

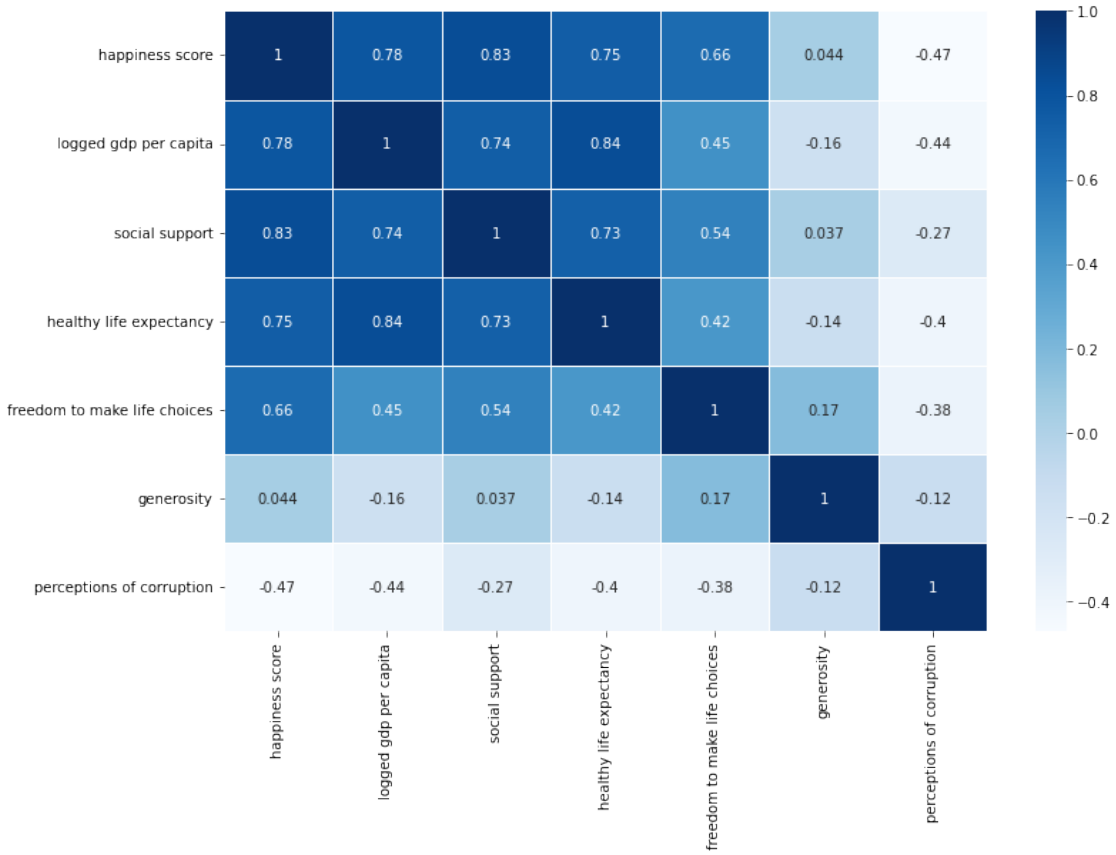


- Outliers have been identified in the distribution of perceptions of corruption. Despite this, it has been decided to keep them in the analysis as they may indicate the existence of extraordinary situations in the country.

Generate a heatmap to identify the correlation

[11]:
```
# Generate a heatmap to identify the correlation
corr = happy_df.corr()
fig = plt.figure(figsize = (12,8))
sns.heatmap(corr, cmap = 'Blues', linewidth = 0.5, annot = True)
```

[11]: <Axes: >

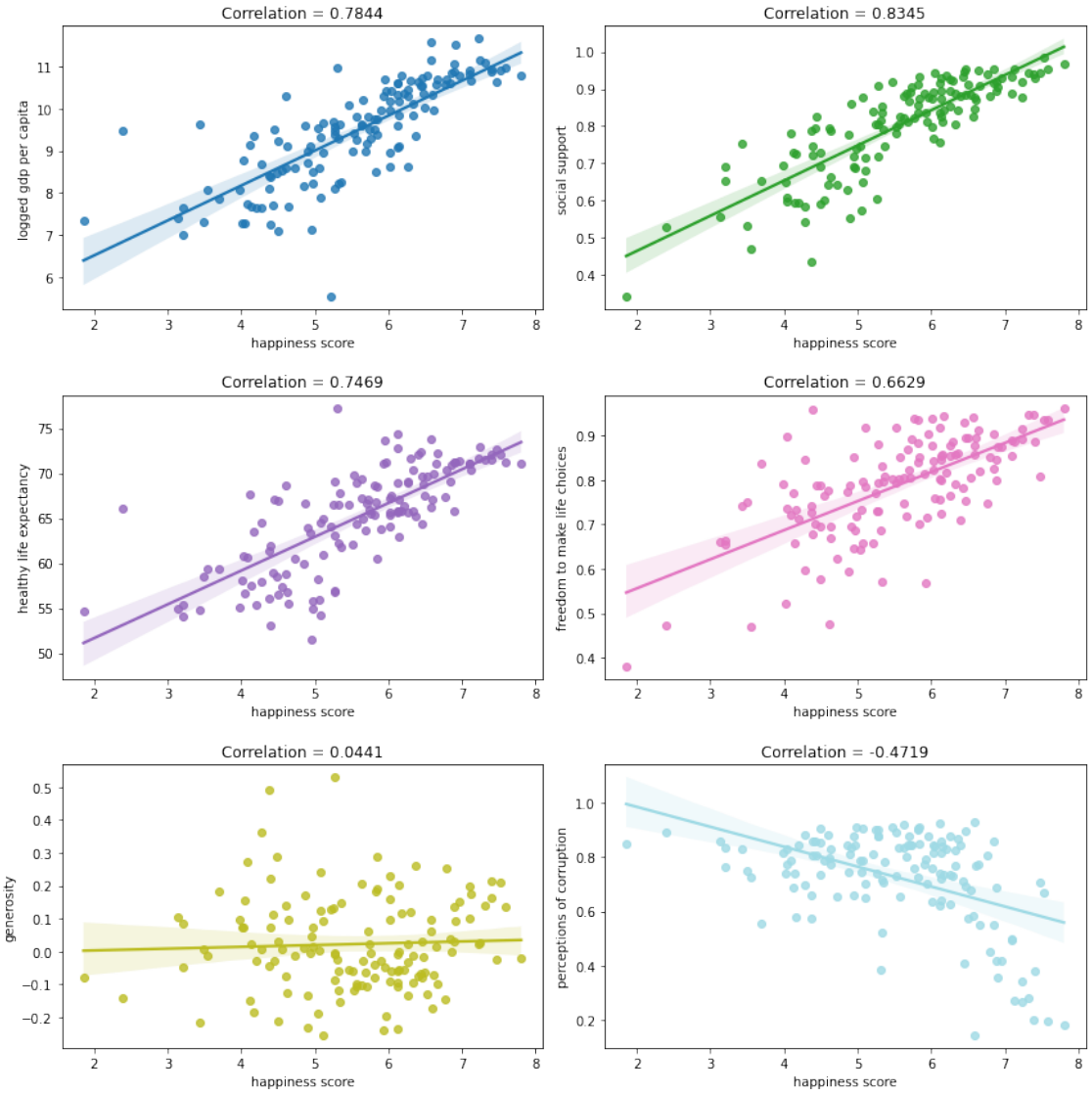


- This heat map reveals several strong correlations between factors and happiness score. It has been found that countries with higher GDP per capita, healthier life expectancies, greater freedom to make choices, and stronger social support tend to have higher happiness scores. Furthermore, a strong positive correlation has been found between GDP per capita, healthy life expectancy, and social support.
- In addition, a negative correlation has been found between happiness score and perceptions of corruption. This suggests that maintaining high levels of happiness among citizens may be challenging for countries with higher levels of corruption.
- Overall, these findings highlight the importance of economic, social, and political factors in determining happiness levels across countries.

A scatter plot has been created to further explore the distribution of the factors, depicting the relationship between the happiness score and other factors.

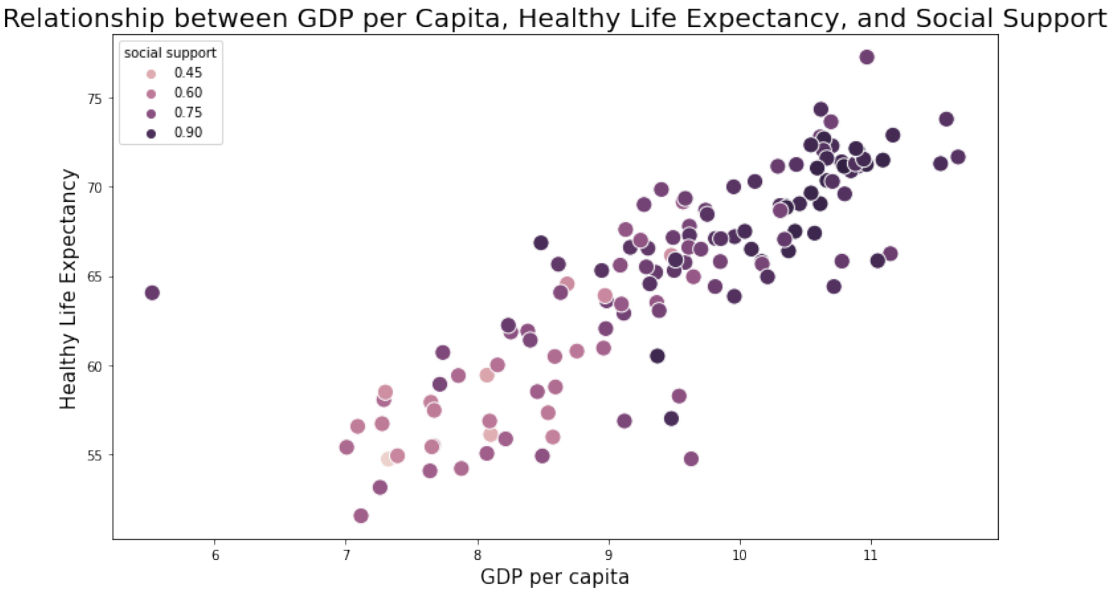
[12]:
```
#Happiness score vs other factors
tab_20_colors = ["#1f77b4", "#2ca02c", "#9467bd", "#e377c2", "#bcbd22",␣
 ↪"#9edae5"]
fig, ax = plt.subplots(3, 2, figsize = (12,12))
i = 0
```

```
for j, column in enumerate(numerical_columns[1:]):
    corr = happy_df['happiness score'].corr(happy_df[column])
    sns.regplot(happy_df, x = "happiness score", y = column, color =␣
 ↪tab_20_colors[i], ax = ax[j//2, j%2])
    ax[j//2, j%2].set_title("Correlation = {:.4f}".format(corr))
    i += 1
fig.tight_layout()
fig.subplots_adjust(hspace = 0.3)
fig.show()
```



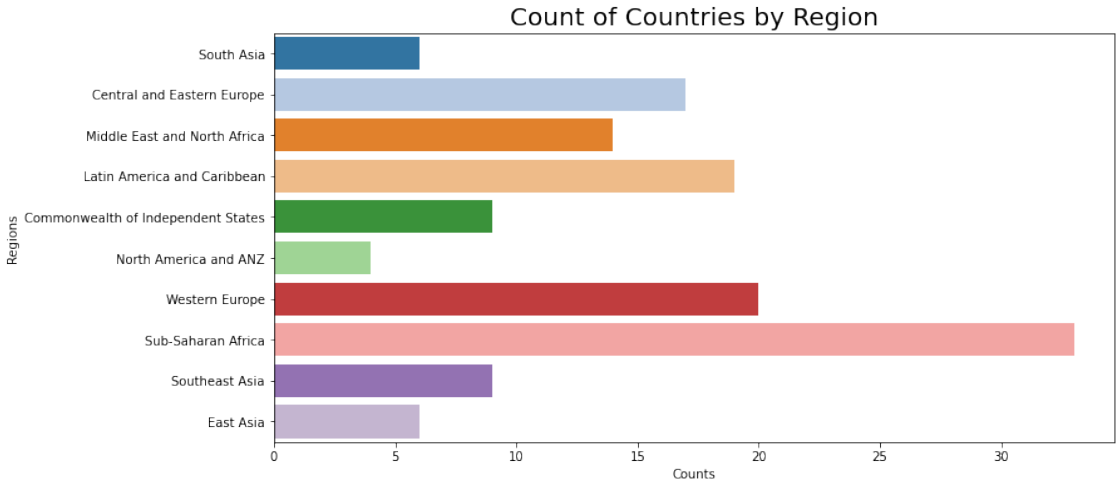
The distribution of the strong positive correlation between GDP per capita, healthy life expectancy, and social support is depicted in a scatter plot below.

[13]:
```
fig = plt.figure(figsize = (12,7))
sns.scatterplot(happy_df, x = "logged gdp per capita", y = "healthy life␣
 ↪expectancy", hue = "social support", s = 150)
plt.xlabel("GDP per capita", size = 15)
plt.ylabel("Healthy Life Expectancy", size = 15)
plt.title("Relationship between GDP per Capita, Healthy Life Expectancy, and␣
 ↪Social Support", size = 20)
plt.show()
```



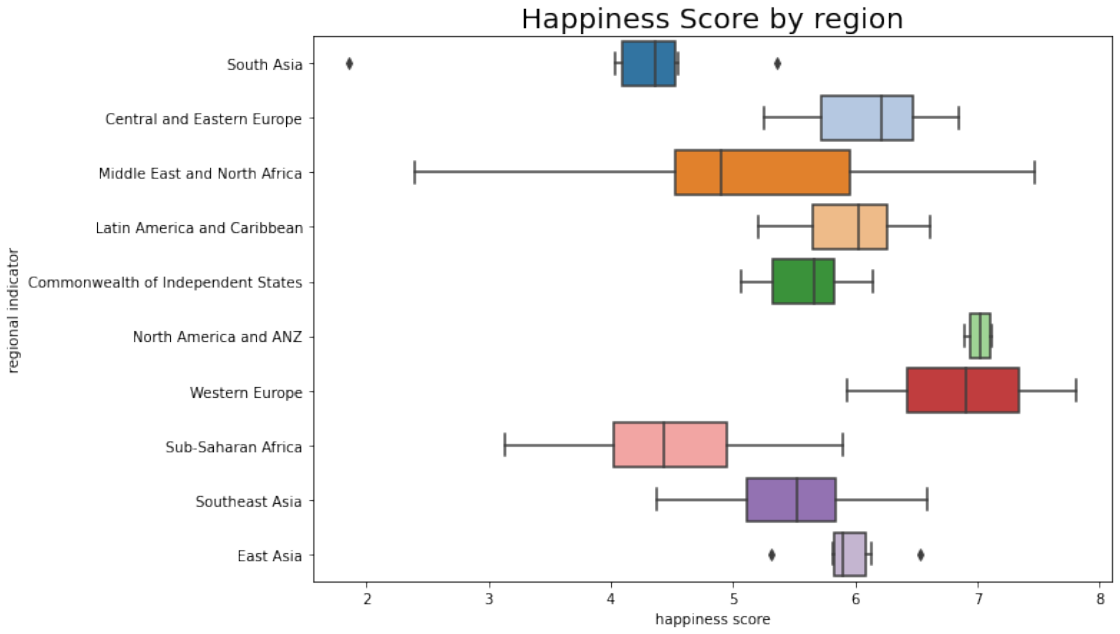
Analysis of happiness score and other factors across different region.

[14]:
```
#Count of Countries by Region
fig = plt.figure(figsize = (12,6))
sns.countplot(happy_df, y = 'regional indicator', palette = 'tab20')
plt.ylabel("Regions")
plt.xlabel("Counts")
plt.title("Count of Countries by Region", size = 20)
plt.show()
```



[15]:
```
#Boxplot
fig = plt.figure(figsize = (10,7))
sns.boxplot(happy_df, y = 'regional indicator', x = "happiness score", palette
 ↪= 'tab20')
plt.title("Happiness Score by region", size = 20)
```

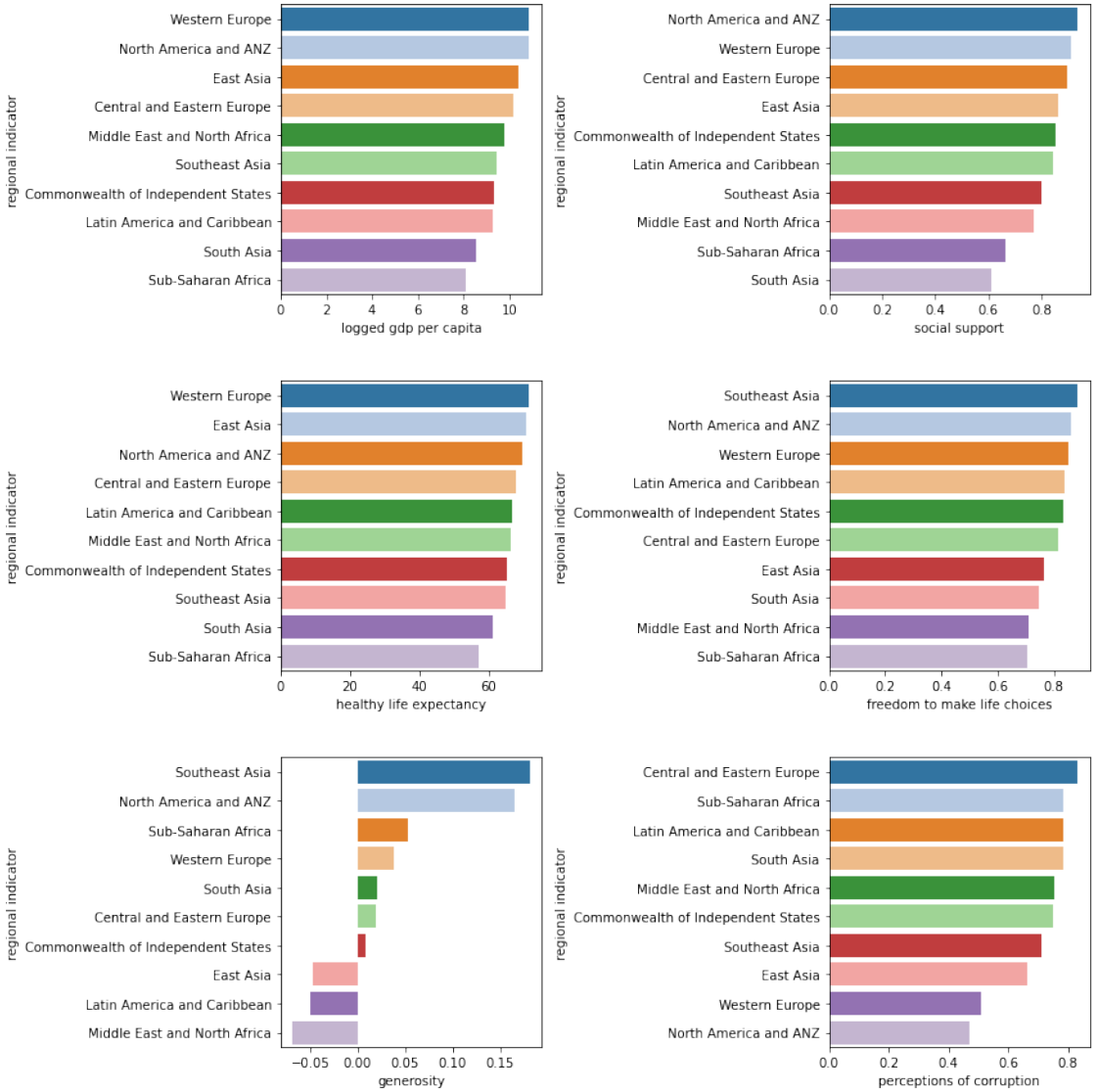
[15]:
```
Text(0.5, 1.0, 'Happiness Score by region')
```



- The box plot analysis indicates that higher median happiness scores are observed in the Western Europe, North America, and ANZ regions compared to other regions.
- An outlier is observed in the South Asia region, where Afghanistan has a notably lower happiness score. This may be attributed to political instability and a lower healthy life expectancy in Afghanistan.

An analysis was conducted to further explore the factors that may contribute to higher or lower happiness scores across different regions. A bar chart was created to depict the average values of various factors, such as GDP per capita, social support, healthy life expectancy, etc., across different regions.

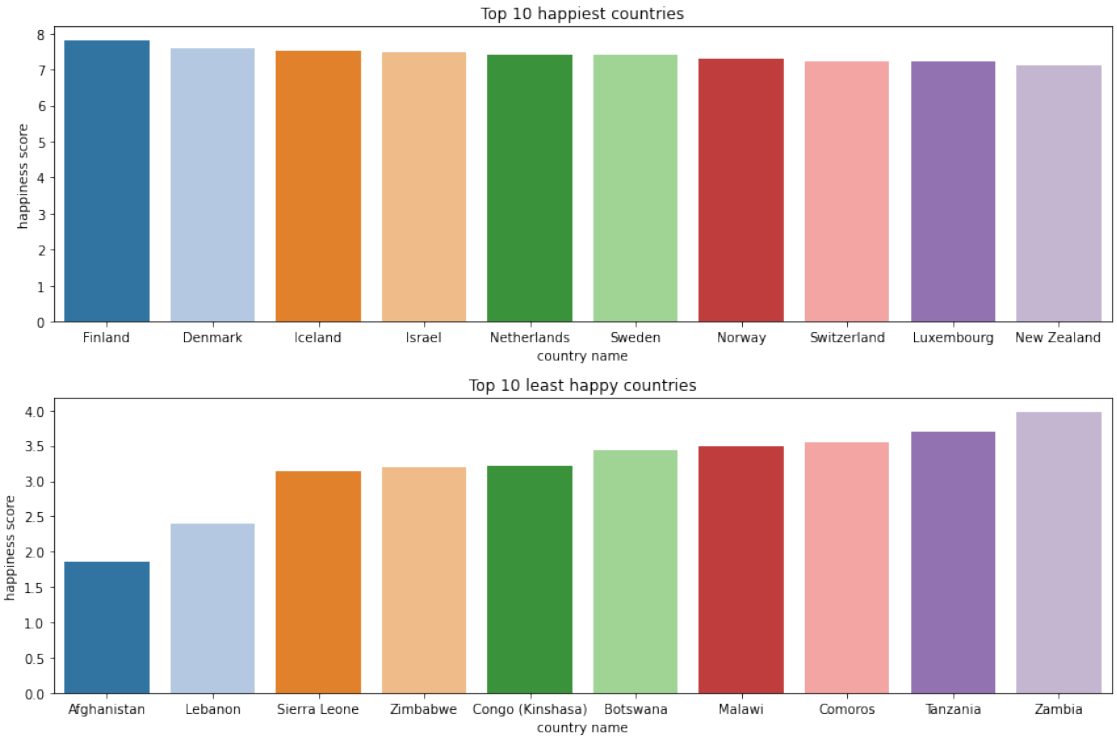
[16]:
```
region_wise = happy_df.groupby("regional indicator")[numerical_columns].mean().
 ↪reset_index()
fig, ax = plt.subplots(3, 2, figsize = (12,12))
for i, column in enumerate(numerical_columns[1:]):
    region_wise.sort_values(column, ascending = False, inplace = True)
    sns.barplot(region_wise, x = column, y = 'regional indicator', palette =␣
 ↪'tab20', ax = ax[i//2, i%2])
fig.tight_layout()
fig.subplots_adjust(hspace = 0.3)
fig.show()
```



- It is suggested by this analysis that regions with higher median happiness scores, such as Western Europe, North America, and ANZ, tend to have higher average values in various key factors, including GDP per capita, social support, healthy life expectancy, and freedom to make life choices, compared to regions with lower median happiness scores.
- Moreover, these regions exhibit lower average values in factors that are negatively associated with happiness levels, such as perceptions of corruption.
- These findings underscore the significance of economic prosperity, social support networks, access to healthcare, and freedom to make life choices in promoting happiness and well-being across different regions.

Top 10 happiest and least happy countries.

[17]:
```
#A column chart was created to analyze the top 10 happiest and least happy␣
 ↪countries.
top_10_happy_country = happy_df.sort_values('happiness score', ascending =␣
 ↪False)
bottom_10_happy_country = happy_df.sort_values('happiness score', ascending =␣
 ↪True)

fig, ax = plt.subplots(2, 1, figsize = (12,8))

sns.barplot(top_10_happy_country.iloc[:10], y = 'happiness score', x = 'country␣
 ↪name', palette = 'tab20', ax = ax[0])
ax[0].set_title("Top 10 happiest countries")



sns.barplot(bottom_10_happy_country.iloc[:10], y = 'happiness score', x =␣
 ↪'country name', palette = 'tab20', ax = ax[1])
ax[1].set_title("Top 10 least happy countries")

fig.tight_layout()
fig.show()
```



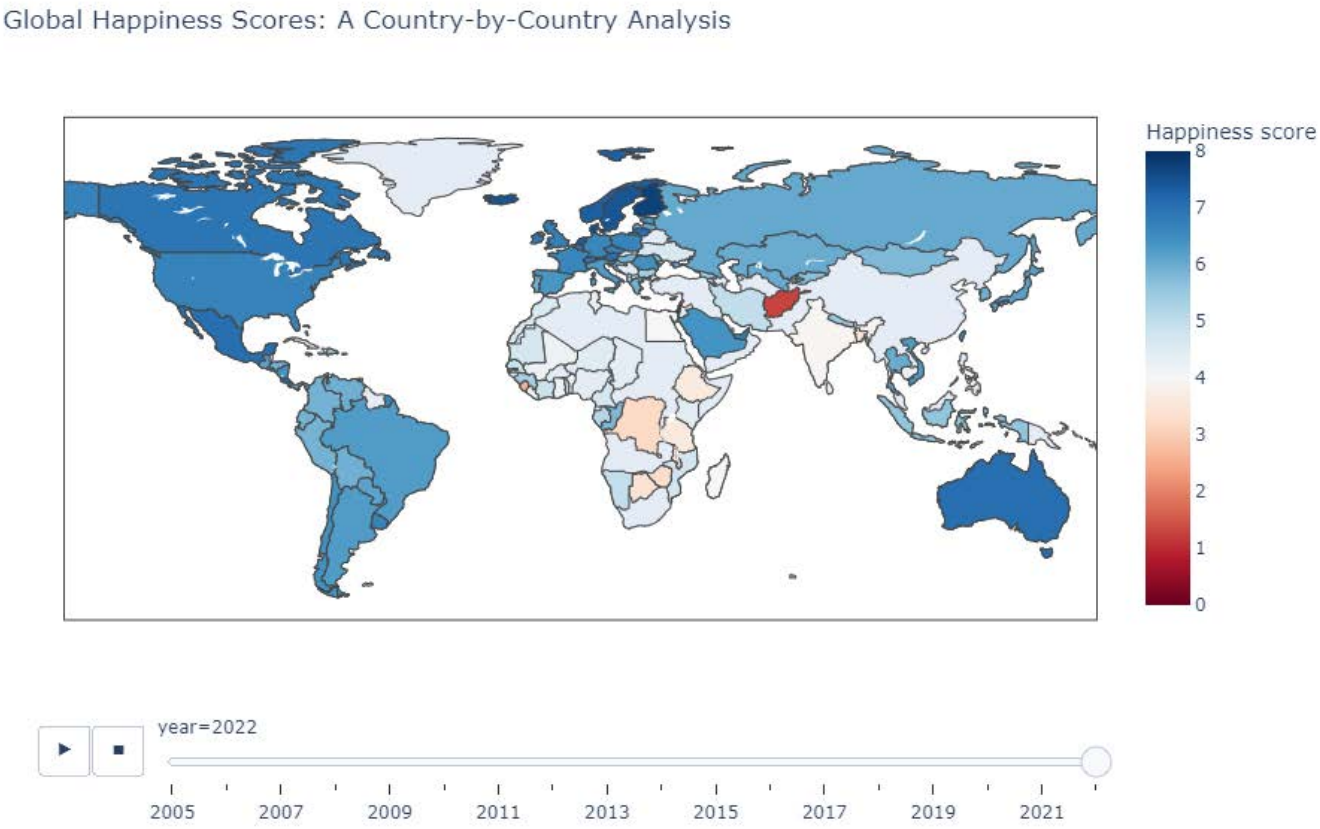
- The analysis revealed that eight out of the top 10 happiest countries are from the Western Europe region.

- Similarly, the analysis of the top 10 least happy countries showed that eight out of 10 countries are from the Sub-Saharan Africa region.

A country-by-country analysis

[18]:
```
happy_df_time_line = pd.read_csv("whr_200522.csv")
happy_df_time_line.sort_values('year', inplace = True)
```

[19]:
```
px.choropleth(happy_df_time_line, locations="Iso alpha", color="Happiness␣
↪score", animation_frame="year" ,hover_name="Country name",␣
↪range_color=[0,8], color_continuous_scale=px.colors.diverging.RdBu, width =␣
↪950, height = 600, title = "Global Happiness Scores: A Country-by-Country␣
↪Analysis")
```



- The map shows that happiness scores have generally remained stable or increased slightly over time, with a few notable dips during times of economic and political instability.

0.2 Conclusion

In conclusion, this project has provided a comprehensive analysis of happiness scores around the world, highlighting the importance of economic, social, and health factors in determining individual and societal well-being. The findings suggest that promoting economic prosperity, social support networks, and access to healthcare can contribute to higher levels of happiness, while reducing corruption can help ensure that these gains are sustainable over time. By prioritizing happiness as a key goal for individuals, communities, and policymakers, we can work towards creating a world that is more just, equitable, and fulfilling for all.