DATASET INFORMATION

The World Happiness Report 2023 is an annual publication by the Sustainable Development Solutions Network, which ranks countries based on their citizens' subjective well-being and happiness. The report draws from global survey data from individuals in over 150 countries.It also reflects the impact of the COVID-19 pandemic on global happiness. While some countries have shown resilience, the pandemic has exacerbated challenges in others, affecting mental health, economic stability, and social connections.

IMPORTANCE AND BENEFITS OF THE DATASET

the World Happiness Report serves as a valuable tool for policymakers, suggesting that investments in social support, health care and community-building can significantly enhance the well-being of citizens. The emphasis on happiness as a measure of societal progress encourages governments to look beyond economic indicators and consider the holistic well-being of their populations.

DATASET SOURCE AND PERIOD

Source: Kaggle - World Happiness Report, Period:2023

In [10]:
```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import numpy as np
import seaborn as sns
```

In [11]:
```
file_path = r'C:\Users\R90FXGJR\Desktop\WHR2023.csv'
df = pd.read_csv(file_path)
df.head()
```

Out[11]:

	Country name	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy
0	Finland	7.804	0.036	7.875	7.733	10.792	0.969	71.150
1	Denmark	7.586	0.041	7.667	7.506	10.962	0.954	71.250
2	Iceland	7.530	0.049	7.625	7.434	10.896	0.983	72.050
3	Israel	7.473	0.032	7.535	7.411	10.639	0.943	72.697
4	Netherlands	7.403	0.029	7.460	7.346	10.942	0.930	71.550

In [12]:
```
df.describe()
```

Out[12]:

	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Health expect
count	137.000000	137.000000	137.000000	137.000000	137.000000	137.000000	137.00
mean	5.539796	0.064715	5.666526	5.412971	9.449796	0.799073	64.97
std	1.139929	0.023031	1.117421	1.163724	1.207302	0.129222	5.72
min	1.859000	0.029000	1.923000	1.795000	5.527000	0.341000	51.53
25%	4.724000	0.047000	4.980000	4.496000	8.591000	0.722000	60.69
50%	5.684000	0.060000	5.797000	5.529000	9.567000	0.827000	65.82
75%	6.334000	0.077000	6.441000	6.243000	10.540000	0.896000	69.35
max	7.804000	0.147000	7.875000	7.733000	11.660000	0.983000	77.28

In [13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 137 entries, 0 to 136
Data columns (total 19 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   Country name                              137 non-null    object 
 1   Ladder score                              137 non-null    float64
 2   Standard error of ladder score            137 non-null    float64
 3   upperwhisker                              137 non-null    float64
 4   lowerwhisker                              137 non-null    float64
 5   Logged GDP per capita                     137 non-null    float64
 6   Social support                            137 non-null    float64
 7   Healthy life expectancy                   137 non-null    float64
 8   Freedom to make life choices              137 non-null    float64
 9   Generosity                                137 non-null    float64
 10  Perceptions of corruption                 137 non-null    float64
 11  Ladder score in Dystopia                  137 non-null    float64
 12  Explained by: Log GDP per capita          137 non-null    float64
 13  Explained by: Social support              137 non-null    float64
 14  Explained by: Healthy life expectancy     137 non-null    float64
 15  Explained by: Freedom to make life choices 137 non-null   float64
 16  Explained by: Generosity                  137 non-null    float64
 17  Explained by: Perceptions of corruption   137 non-null    float64
 18  Dystopia + residual                       137 non-null    float64
dtypes: float64(18), object(1)
memory usage: 20.5+ KB
```

In [14]:
```
df.columns = df.columns.str.strip()
df
```

Out[14]:

	Country name	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy expectar
0	Finland	7.804	0.036	7.875	7.733	10.792	0.969	71.
1	Denmark	7.586	0.041	7.667	7.506	10.962	0.954	71.2
2	Iceland	7.530	0.049	7.625	7.434	10.896	0.983	72.(
3	Israel	7.473	0.032	7.535	7.411	10.639	0.943	72.(
4	Netherlands	7.403	0.029	7.460	7.346	10.942	0.930	71.5
...	...	...	...	...	...	...	...	
132	Congo (Kinshasa)	3.207	0.095	3.394	3.020	7.007	0.652	55.3
133	Zimbabwe	3.204	0.061	3.323	3.084	7.641	0.690	54.(
134	Sierra Leone	3.138	0.082	3.299	2.976	7.394	0.555	54.9
135	Lebanon	2.392	0.044	2.479	2.305	9.478	0.530	66.
136	Afghanistan	1.859	0.033	1.923	1.795	7.324	0.341	54.

137 rows × 19 columns

In [15]:
```
df.columns=df.columns.str.lower()
df
```

Out[15]:

	country name	ladder score	standard error of ladder score	upperwhisker	lowerwhisker	logged gdp per capita	social support	healthy li expectanc
0	Finland	7.804	0.036	7.875	7.733	10.792	0.969	71.15
1	Denmark	7.586	0.041	7.667	7.506	10.962	0.954	71.25
2	Iceland	7.530	0.049	7.625	7.434	10.896	0.983	72.05
3	Israel	7.473	0.032	7.535	7.411	10.639	0.943	72.69
4	Netherlands	7.403	0.029	7.460	7.346	10.942	0.930	71.55
...	...	...	...	...	...	...	...	
132	Congo (Kinshasa)	3.207	0.095	3.394	3.020	7.007	0.652	55.37
133	Zimbabwe	3.204	0.061	3.323	3.084	7.641	0.690	54.05
134	Sierra Leone	3.138	0.082	3.299	2.976	7.394	0.555	54.90
135	Lebanon	2.392	0.044	2.479	2.305	9.478	0.530	66.14
136	Afghanistan	1.859	0.033	1.923	1.795	7.324	0.341	54.7

137 rows × 19 columns

In [16]:
```
df.shape
```

Out[16]:
```
(137, 19)
```

In [17]:
```
df.drop_duplicates(inplace=True)
df.shape
```

Out[17]:
```
(137, 19)
```

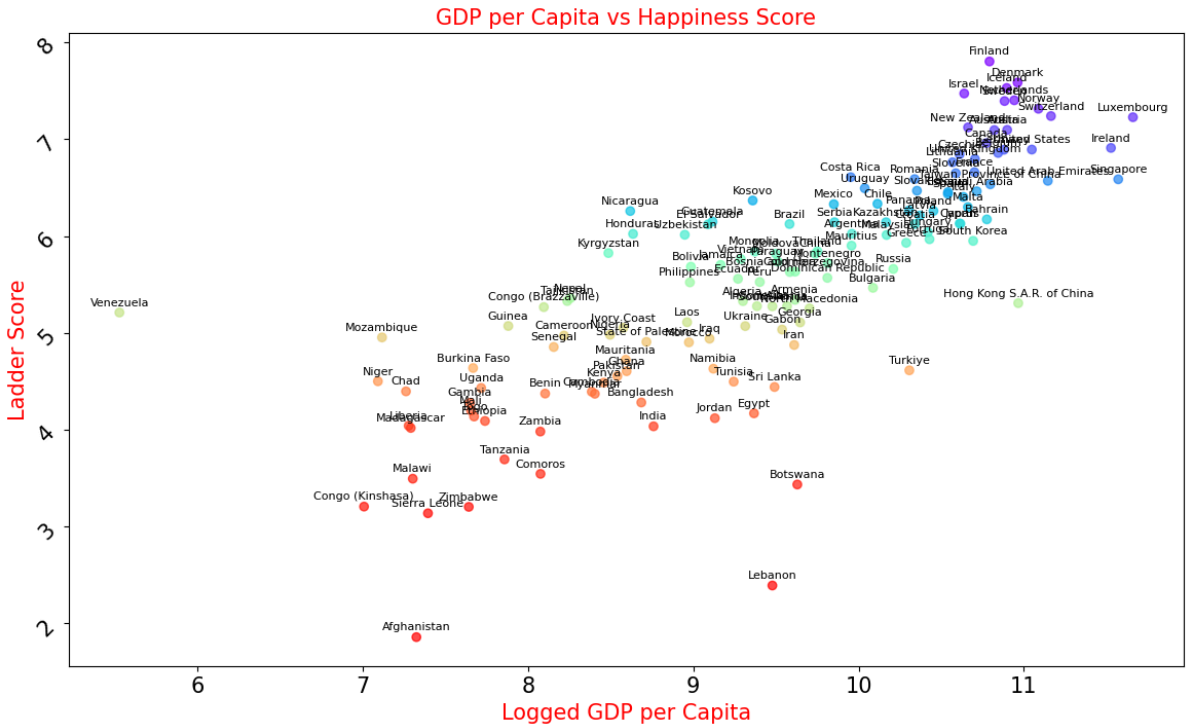
In [18]:
```
df.isna().sum()
```

Out[18]:
```
country name                                0
ladder score                                0
standard error of ladder score              0
upperwhisker                                0
lowerwhisker                                0
logged gdp per capita                       0
social support                              0
healthy life expectancy                     0
freedom to make life choices                0
generosity                                  0
perceptions of corruption                   0
ladder score in dystopia                    0
explained by: log gdp per capita            0
explained by: social support                0
explained by: healthy life expectancy       0
explained by: freedom to make life choices  0
explained by: generosity                    0
explained by: perceptions of corruption     0
dystopia + residual                         0
dtype: int64
```

Q 1: How does the GDP per capita correlate with the overall happiness (ladder score) across these countries?

In [19]:
```
# Extract relevant columns
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import numpy as np
file_path = r'C:\Users\R90FXGJR\Desktop\WHR2023.csv'
df = pd.read_csv(file_path)
df.head()

logged_gdp_per_capita = df['Logged GDP per capita']
ladder_score = df['Ladder score']
country_name = df['Country name']

# Create a color map
colors = cm.rainbow(np.linspace(0, 1, len(country_name)))

# Plot scatter plot
plt.figure(figsize=(14, 8))
scatter = plt.scatter(logged_gdp_per_capita, ladder_score, marker='o', colo

# Add labels and title
plt.title('GDP per Capita vs Happiness Score',fontsize=15, color='red')
plt.xlabel('Logged GDP per Capita',fontsize=15, color='red')
```

```
plt.ylabel('Ladder Score',fontsize=15, color='red')
plt.yticks(fontsize=15,rotation=45)
plt.xticks(fontsize=15)

# Annotate points with country names
for i, country in enumerate(country_name):
    plt.annotate(country, (logged_gdp_per_capita[i], ladder_score[i]), texto

# Show the plot
plt.show()
```



In [ ]:

Observation

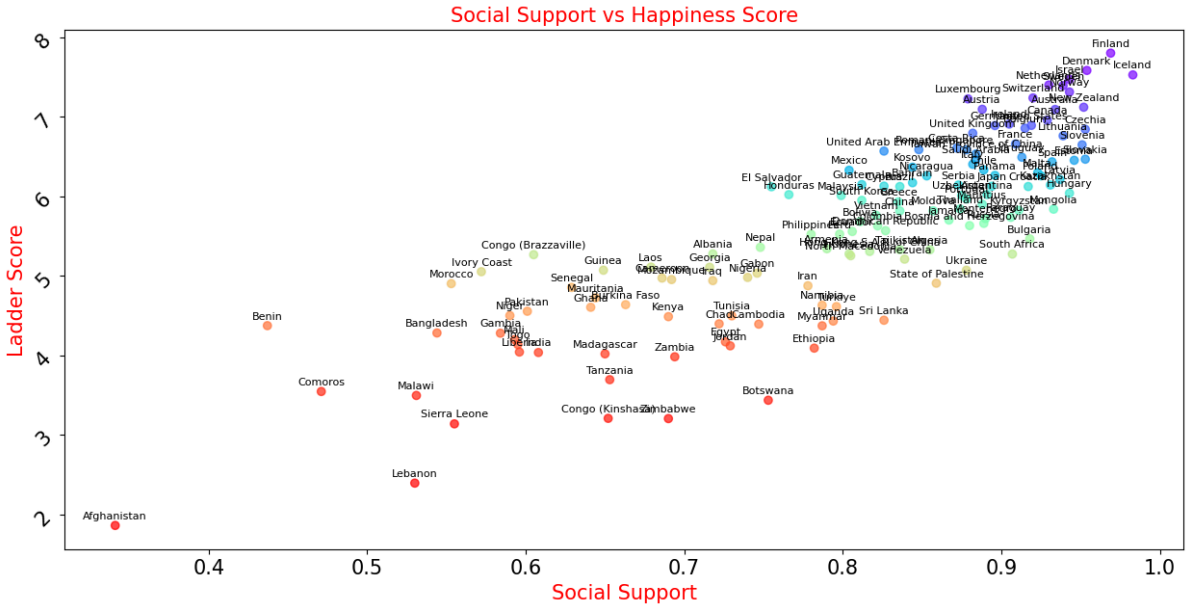
Based on the scatter plot it's evident that there is a general positive trend between higher GDP per capita and higher happiness scores. However, there are notable exceptions where countries with relatively lower GDP per capita still achieve high happiness scores, indicating that factors beyond economic wealth, such as social support, health, and personal freedom, also significantly influence overall happiness levels.

Q 2. What is the impact of social support on happiness scores across countries?

In [20]:

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import numpy as np
file_path = r'C:\Users\R90FXGJR\Desktop\WHR2023.csv'
df = pd.read_csv(file_path)
df.head()
```

```
# Extract relevant columns
social_support = df['Social support']
ladder_score = df['Ladder score']
country_name = df['Country name']

# Create a color map
colors = cm.rainbow(np.linspace(0, 1, len(country_name)))

# Plot scatter plot
plt.figure(figsize=(15, 7))
scatter = plt.scatter(social_support, ladder_score, marker='o', color=colors

# Add labels and title
plt.title('Social Support vs Happiness Score', fontsize=15, color='red')
plt.xlabel('Social Support', fontsize=15, color='red')
plt.ylabel('Ladder Score', fontsize=15, color='red')
plt.yticks(fontsize=15, rotation=45)
plt.xticks(fontsize=15)

# Annotate points with country names
for i, country in enumerate(country_name):
    plt.annotate(country, (social_support[i], ladder_score[i]), textcoords='

# Show the plot
plt.show()
```



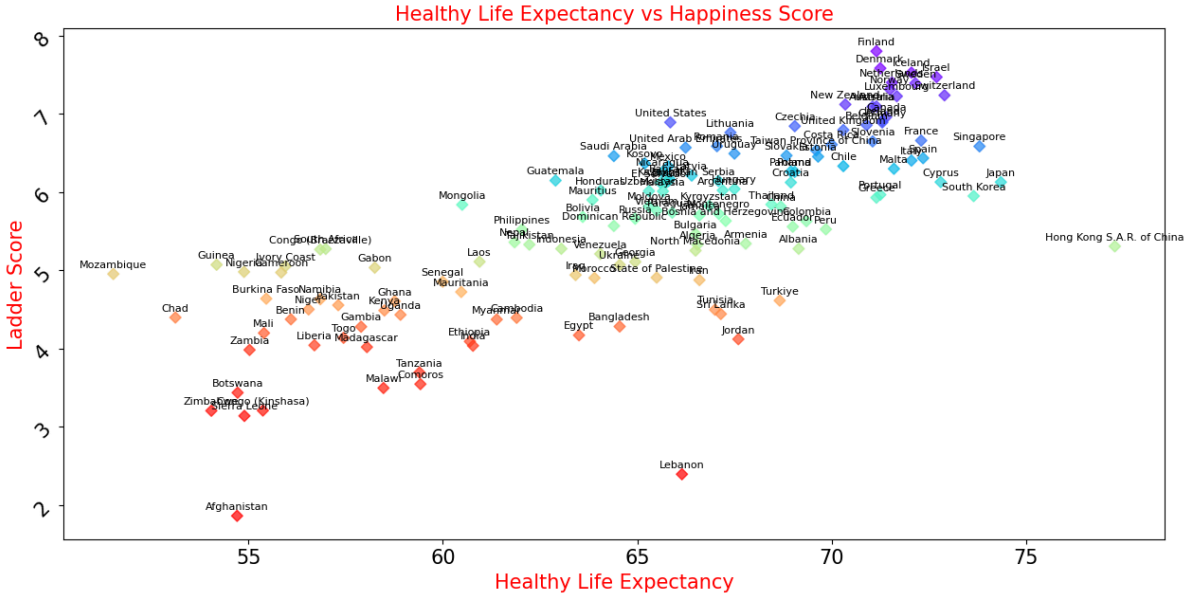
Observation:

It's visually confirms the positive correlation between higher levels of social support and higher happiness scores observed in many countries

Q 3. How does healthy life expectancy correlate with the average happiness score

In [21]:
```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import numpy as np
```

```
file_path = r'C:\Users\R90FXGJR\Desktop\WHR2023.csv'
df = pd.read_csv(file_path)
df.head()

# Extract relevant columns
healthy_life_expectancy = df['Healthy life expectancy']
ladder_score = df['Ladder score']
country_name = df['Country name']

# Create a color map
colors = cm.rainbow(np.linspace(0, 1, len(country_name)))

# Plot scatter plot
plt.figure(figsize=(15, 7))
scatter = plt.scatter(healthy_life_expectancy, ladder_score, marker='D', co

# Add labels and title
plt.title('Healthy Life Expectancy vs Happiness Score', fontsize=15, color=
plt.xlabel('Healthy Life Expectancy', fontsize=15, color='red')
plt.ylabel('Ladder Score', fontsize=15, color='red')
plt.yticks(fontsize=15, rotation=45)
plt.xticks(fontsize=15)

# Annotate points with country names
for i, country in enumerate(country_name):
    plt.annotate(country, (healthy_life_expectancy[i], ladder_score[i]), te

# Show the plot
plt.show()
```



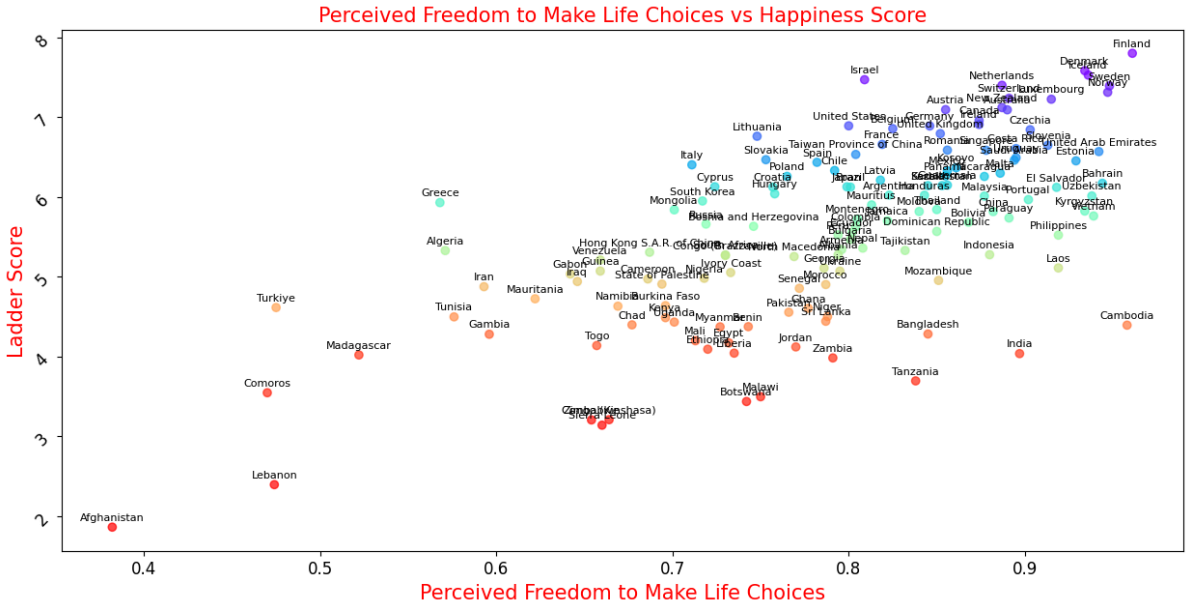
Observation:

The graph shows a positive correlation between healthy life expectancy and happiness score. Countries with higher healthy life expectancy have higher happiness scores.Again countries are relatively close to each other, indicating a similar range of healthy life expectancy and happiness scores among these countries.

Q 4

Which countries have the highest and lowest levels of perceived freedom to make life choices, and how does this impact their happiness scores

In [27]:
```
colors = cm.rainbow(np.linspace(0, 1, len(country_name)))

plt.figure(figsize=(15, 7))
scatter = plt.scatter(freedom_to_make_life_choices, ladder_score, marker='o
plt.title('Perceived Freedom to Make Life Choices vs Happiness Score', fonts
plt.xlabel('Perceived Freedom to Make Life Choices', fontsize=15, color='red
plt.ylabel('Ladder Score', fontsize=15, color='red')
plt.yticks(fontsize=12, rotation=45)
plt.xticks(fontsize=12)

# Annotate points with country names
for i, country in enumerate(country_name):
    plt.annotate(country, (freedom_to_make_life_choices[i], ladder_score[i])

plt.show()
```



In [31]:
```
# Extract relevant columns
freedom_to_make_life_choices = df['Freedom to make life choices']
ladder_score = df['Ladder score']
country_name = df['Country name']
# Find the countries with the highest and lowest levels of perceived freedom
highest_freedom_country = df.loc[df['Freedom to make life choices'].idxmax()
lowest_freedom_country = df.loc[df['Freedom to make life choices'].idxmin()]
# Print the results
print("Country with the highest perceived freedom to make life choices:")
print(highest_freedom_country)
print("\nCountry with the lowest perceived freedom to make life choices:")
print(lowest_freedom_country)
```

```
Country with the highest perceived freedom to make life choices:
Country name                                  Finland
Ladder score                                    7.804
Standard error of ladder score                  0.036
upperwhisker                                    7.875
lowerwhisker                                    7.733
Logged GDP per capita                          10.792
Social support                                  0.969
Healthy life expectancy                         71.15
Freedom to make life choices                    0.961
Generosity                                     -0.019
Perceptions of corruption                       0.182
Ladder score in Dystopia                        1.778
Explained by: Log GDP per capita                1.888
Explained by: Social support                    1.585
Explained by: Healthy life expectancy           0.535
Explained by: Freedom to make life choices      0.772
Explained by: Generosity                        0.126
Explained by: Perceptions of corruption         0.535
Dystopia + residual                             2.363
Name: 0, dtype: object

Country with the lowest perceived freedom to make life choices:
Country name                                 Afghanistan
Ladder score                                       1.859
Standard error of ladder score                     0.033
upperwhisker                                       1.923
lowerwhisker                                       1.795
Logged GDP per capita                              7.324
Social support                                     0.341
Healthy life expectancy                           54.712
Freedom to make life choices                       0.382
Generosity                                        -0.081
Perceptions of corruption                          0.847
Ladder score in Dystopia                           1.778
Explained by: Log GDP per capita                   0.645
Explained by: Social support                         0.0
Explained by: Healthy life expectancy              0.087
Explained by: Freedom to make life choices           0.0
Explained by: Generosity                           0.093
Explained by: Perceptions of corruption            0.059
Dystopia + residual                                0.976
Name: 136, dtype: object
```

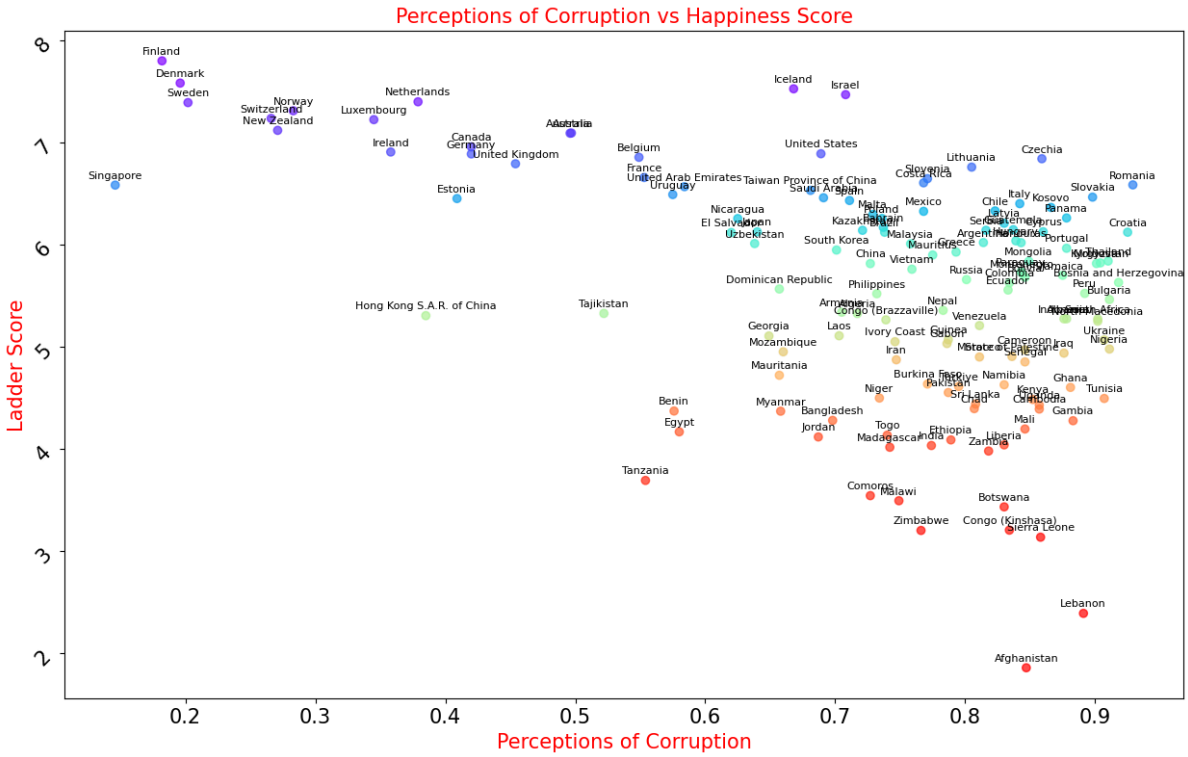
Q 5

How do perceptions of corruption affect the happiness scores?

In [23]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import matplotlib.cm as cm

# Extract relevant columns
perceptions_of_corruption = df['Perceptions of corruption']
ladder_score = df['Ladder score']
country_name = df['Country name']
colors = cm.rainbow(np.linspace(0, 1, len(country_name)))
plt.figure(figsize=(15, 9))
scatter = plt.scatter(perceptions_of_corruption, ladder_score, marker='o', c
plt.title('Perceptions of Corruption vs Happiness Score', fontsize=15, colo
```

```
plt.xlabel('Perceptions of Corruption', fontsize=15, color='red')
plt.ylabel('Ladder Score', fontsize=15, color='red')
plt.yticks(fontsize=15, rotation=45)
plt.xticks(fontsize=15)

# Annotate points with country names
for i, country in enumerate(country_name):
    plt.annotate(country, (perceptions_of_corruption[i], ladder_score[i]), t

plt.show()
```



In [ ]:
```
# Calculate the correlation coefficient
correlation = data['Perceptions of corruption'].corr(data['Ladder score'])
print(f"Correlation between perceptions of corruption and happiness scores:
```

Observation

From the graph countries with lower perceptions of corruption have higher happiness scores. There is a visible trend where increased corruption correlates with decreased happiness.

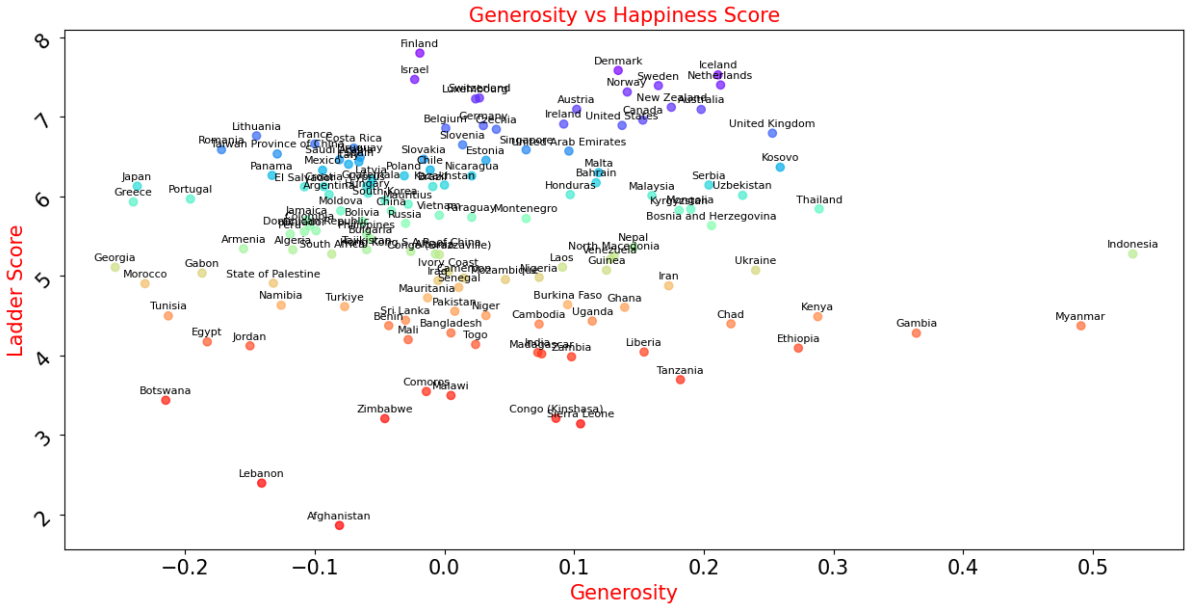
Correlation Coefficient: The calculated correlation coefficient is negative, indicating an inverse relationship between perceptions of corruption and happiness scores. This means that as perceived corruption increases, happiness scores tend to decrease.

Q 6

What is the relationship between generosity and happiness scores?

In [24]:
```
# Extract relevant columns
generosity = df['Generosity']
ladder_score = df['Ladder score']
```

```
country_name = df['Country name']
colors = cm.rainbow(np.linspace(0, 1, len(country_name)))
plt.figure(figsize=(15, 7))
scatter = plt.scatter(generosity, ladder_score, marker='o', color=colors, a
plt.title('Generosity vs Happiness Score', fontsize=15, color='red')
plt.xlabel('Generosity', fontsize=15, color='red')
plt.ylabel('Ladder Score', fontsize=15, color='red')
plt.yticks(fontsize=15, rotation=45)
plt.xticks(fontsize=15)

# Annotate points with country names
for i, country in enumerate(country_name):
    plt.annotate(country, (generosity[i], ladder_score[i]), textcoords="offs

# Show the plot
plt.show()
```



Observation

There is a positive relationship between generosity and happiness scores. Countries that exhibit higher levels of generosity tend to have higher happiness scores. This suggests that fostering a culture of generosity might contribute to greater overall happiness in a country.

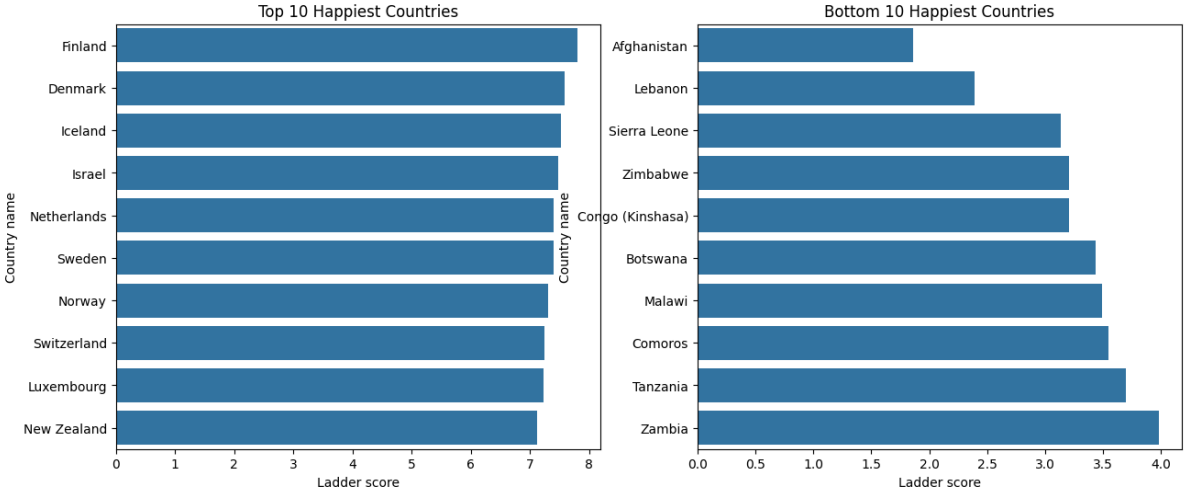
Q 7

How do the ladder score (Happiness sscore) differ between the top 10 happiest countries and the bottom 10 happiest countries, and what insights can be drawn from these differences?"

In [35]:

```
# Highest and Lowest Ladder Scores (Bar Plot)
top_10_happy = df.nlargest(10, 'Ladder score')[['Country name', 'Ladder sco
bottom_10_happy = df.nsmallest(10, 'Ladder score')[['Country name', 'Ladder

fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15, 6))

# Top 10 Happiest Countries
```

```
sns.barplot(x='Ladder score', y='Country name', data=top_10_happy, ax=axes[0
axes[0].set_title('Top 10 Happiest Countries')

# Bottom 10 Happiest Countries
sns.barplot(x='Ladder score', y='Country name', data=bottom_10_happy, ax=axe
axes[1].set_title('Bottom 10 Happiest Countries')

plt.show()
```



Observations

Top 10 Countries: Top 10 happiest countries include Finland, Denmark, Switzerland, Iceland, and the Netherlands typically score high on factors like social support, freedom, and low corruption.

Bottom 10 Countries: Bottom 10 happiest countries include countries from regions experiencing conflict, economic hardship, or poor governance. These countries have lower scores in GDP per capita, social support, and life expectancy.