In [ ]:
```
## Question 1: What is the main driver of happiness?
To find this, I created a correlation heatmap of all the main factors.
###Question 2: Does money really buy happiness?
    To find this I created scatterplot to find outlear for Happiness vs GDP

#Q3.How is happiness distributed?
   To find distribution of HAppiness I have ploted Histogram of count vs Happiness

#(Imp)Q4.What is the biggest driver of happiness in rich countries vs. poor countri
     To find this I have plotted two heatmpa 1st for Happiness vs Low GDP and 2nd
```

In [1]:
```
import pandas as pd
```

In [3]:
```
df=pd.read_csv('WHR_2023.csv')
```

In [4]:
```
df.head()
```

Out[4]:

	country	region	happiness_score	gdp_per_capita	social_support	healthy_life_expec
0	Finland	Western Europe	7.804	1.888	1.585	
1	Denmark	Western Europe	7.586	1.949	1.548	
2	Iceland	Western Europe	7.530	1.926	1.620	
3	Israel	Middle East and North Africa	7.473	1.833	1.521	
4	Netherlands	Western Europe	7.403	1.942	1.488	

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 137 entries, 0 to 136
Data columns (total 9 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   country                         137 non-null    object
 1   region                          137 non-null    object
 2   happiness_score                 137 non-null    float64
 3   gdp_per_capita                  137 non-null    float64
 4   social_support                  137 non-null    float64
 5   healthy_life_expectancy         136 non-null    float64
 6   freedom_to_make_life_choices    137 non-null    float64
 7   generosity                      137 non-null    float64
 8   perceptions_of_corruption       137 non-null    float64
dtypes: float64(7), object(2)
memory usage: 9.8+ KB
```

In [6]:
```
df.describe()
```

Out[6]:

	happiness_score	gdp_per_capita	social_support	healthy_life_expectancy	freedom_to
count	137.000000	137.000000	137.000000	136.000000	
mean	5.539796	1.406985	1.156212	0.366176	
std	1.139929	0.432963	0.326322	0.156691	
min	1.859000	0.000000	0.000000	0.000000	
25%	4.724000	1.099000	0.962000	0.248500	
50%	5.684000	1.449000	1.227000	0.389500	
75%	6.334000	1.798000	1.401000	0.487500	
max	7.804000	2.200000	1.620000	0.702000	

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
country                         0
region                          0
happiness_score                 0
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
```

In [9]:
```
df=df.dropna()
```

In [10]:
```
df=df.rename(columns={
    'happiness_score': 'Happiness',
    'gdp_per_capita  ': 'GDP',
```

```
    'social_support': 'Social',
    'healthy_life_expectancy': 'Life_Expectancy',
    'freedom_to_make_life_choices': 'Freedom',
    'perceptions_of_corruption': 'Corruption'

})
```

In [11]:
```
df.head()
```

Out[11]:

	country	region	Happiness	gdp_per_capita	Social	Life_Expectancy	Freedom	gen
0	Finland	Western Europe	7.804	1.888	1.585	0.535	0.772	
1	Denmark	Western Europe	7.586	1.949	1.548	0.537	0.734	
2	Iceland	Western Europe	7.530	1.926	1.620	0.559	0.738	
3	Israel	Middle East and North Africa	7.473	1.833	1.521	0.577	0.569	
4	Netherlands	Western Europe	7.403	1.942	1.488	0.545	0.672	

In [12]:
```
df.isnull().sum()
```

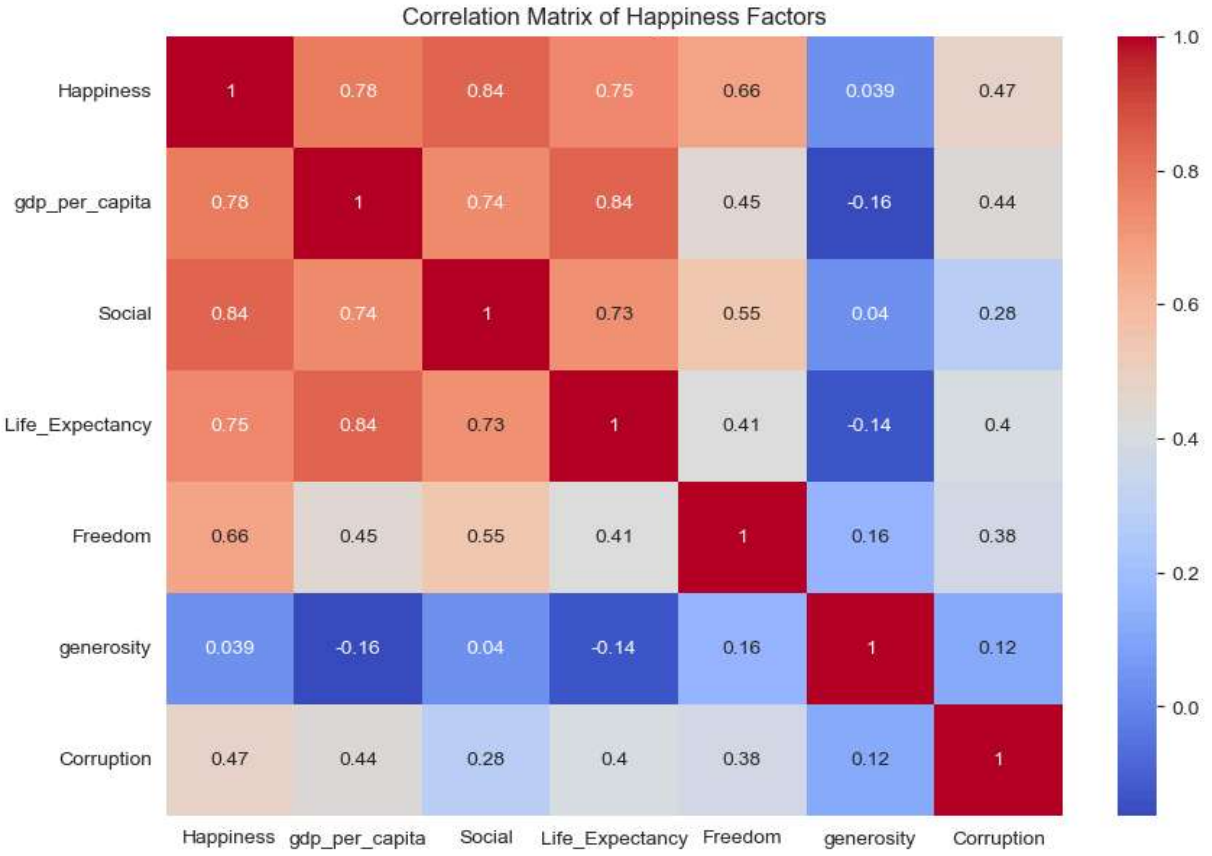
Out[12]:
```
country            0
region             0
Happiness          0
gdp_per_capita     0
Social             0
Life_Expectancy    0
Freedom            0
generosity         0
Corruption         0
dtype: int64
```

In [13]:
```
##Visualization
import seaborn as sns
import matplotlib.pyplot as plt
```

In [14]:
```
sns.set_style('darkgrid')
```

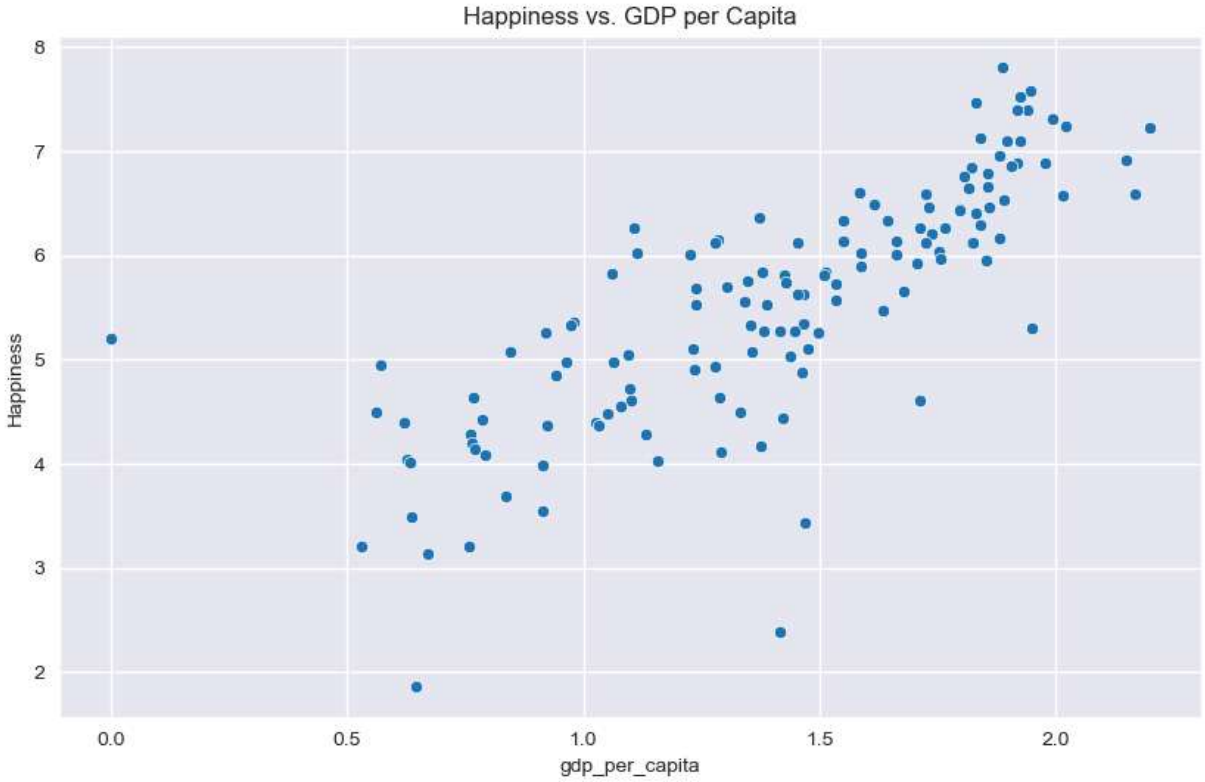
In [15]:
```
columns_to_correlate=['Happiness','gdp_per_capita','Social','Life_Expectancy','Free
correlation_matrix=df[columns_to_correlate].corr()

#draw heatmap
plt.figure(figsize=(10,7))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
```

```
plt.title('Correlation Matrix of Happiness Factors')
plt.show()
```



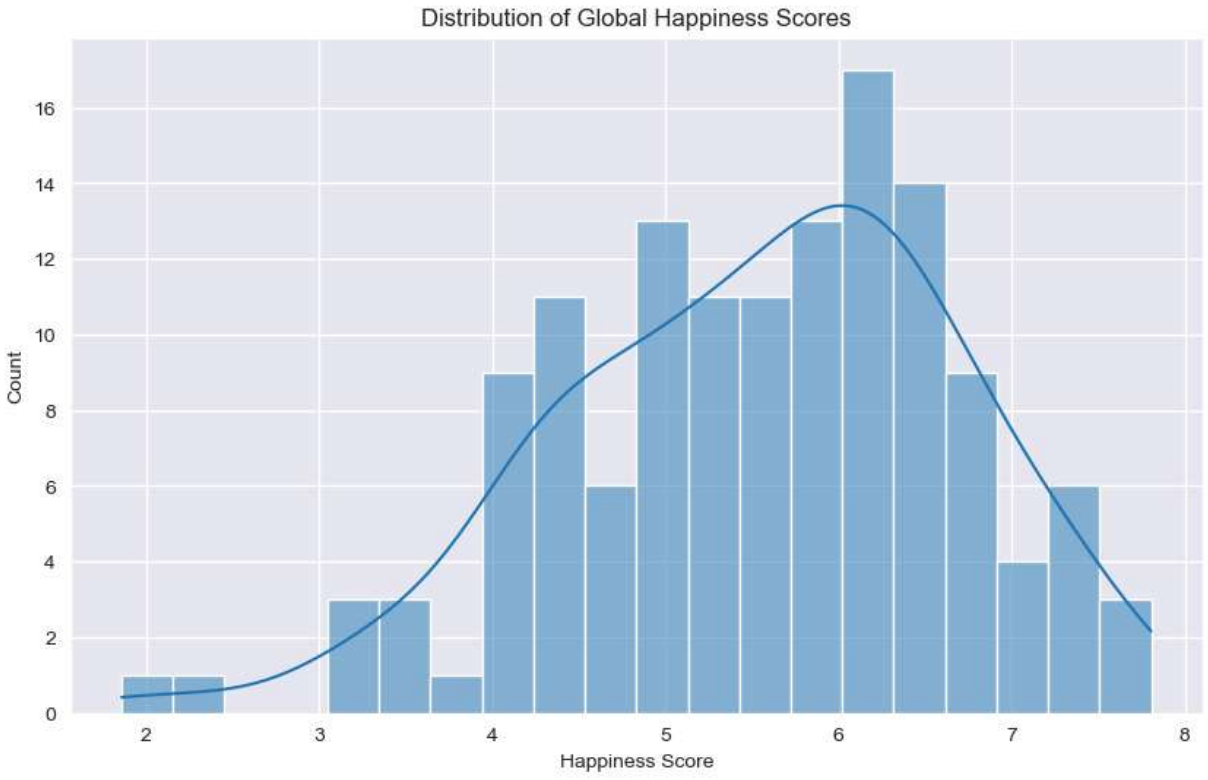
In [16]:
```
#Question 1: What is the single biggest driver of happiness?
     #Look at the "Happiness" row. The number closest to +1.0 (bright red)
     #is the strongest positive driver. The number closest to -1.0 (bright blue) is
```

In [20]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='gdp_per_capita', y='Happiness')
plt.title('Happiness vs. GDP per Capita')
plt.show()
```



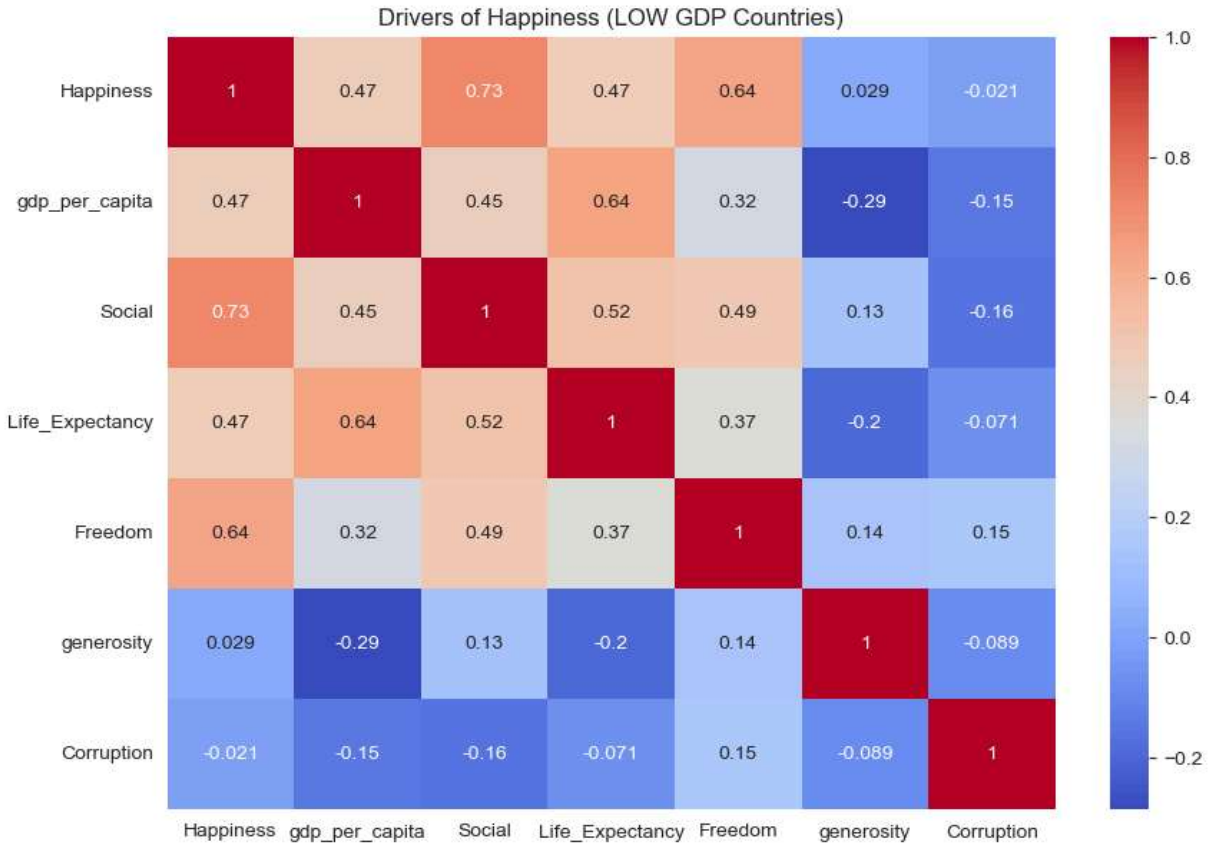
In [21]:
```
#Question 2: Does money really buy happiness?
  #Does the line go up and up? (More money = more happiness).

#Does it go up and then flatten out? (This is a huge insight: "Money buys happiness

#Are there outliers? (e.g., A rich, sad country or a poor, happy country).
```

In [25]:
```
plt.figure(figsize=(10, 6))
sns.histplot(df['Happiness'], kde=True, bins=20)
plt.title('Distribution of Global Happiness Scores')
plt.xlabel('Happiness Score')
plt.show()

#Q3.How is happiness distributed?
```

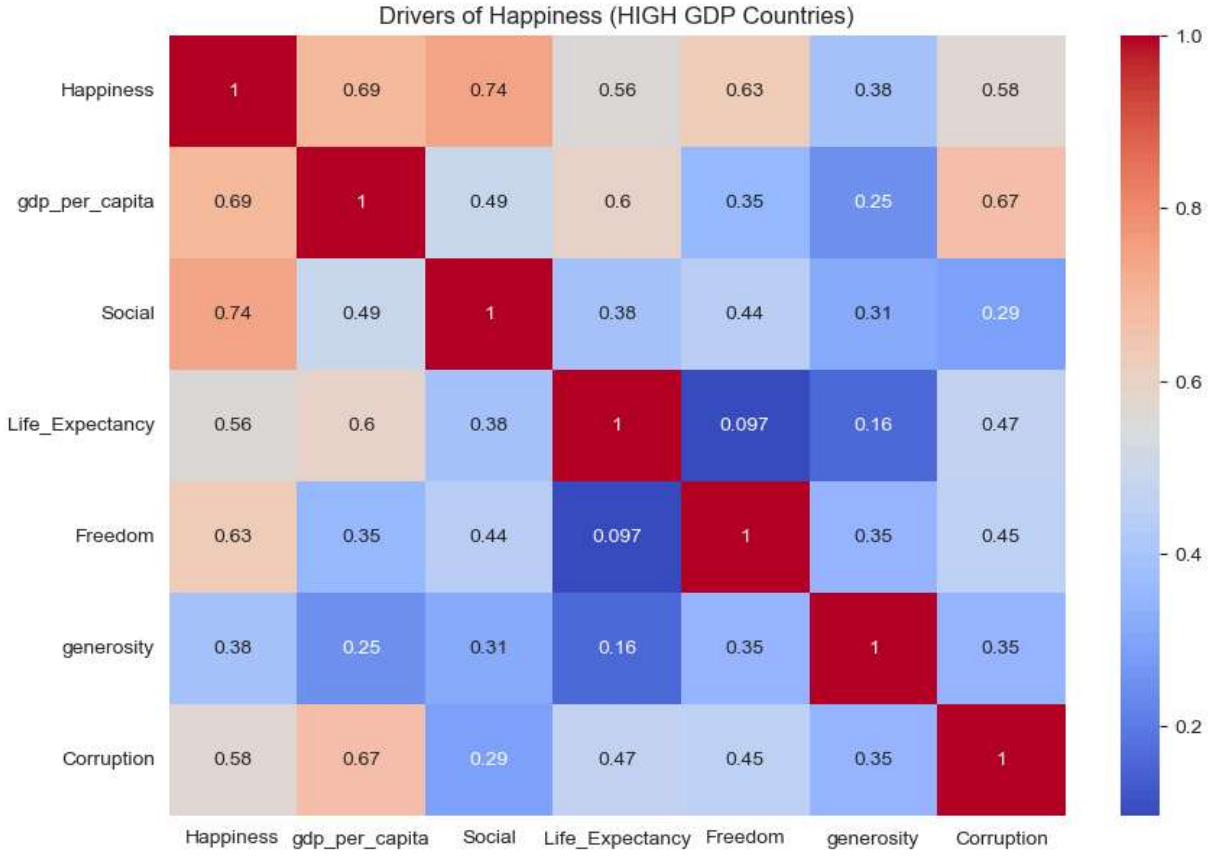


In [26]:
```
#Question 4: The "Difficult Question" (Your Unique Insight)
  #This is what makes your project impressive. "What is the biggest driver of happi

# 1. Find the middle-point for GDP
median_gdp = df['gdp_per_capita'].median()
print(f"The median GDP is: {median_gdp}")

# 2. Create two NEW dataframes: one for "low" GDP, one for "high"
low_gdp_df = df[df['gdp_per_capita'] < median_gdp]
high_gdp_df = df[df['gdp_per_capita'] >= median_gdp]

# 3. Now, just re-run the heatmap code from Question 1 on the first new dataframe
print("\n--- Correlation for LOW GDP Countries ---")
low_gdp_corr = low_gdp_df[columns_to_correlate].corr()
plt.figure(figsize=(10, 7))
sns.heatmap(low_gdp_corr, annot=True, cmap='coolwarm')
plt.title('Drivers of Happiness (LOW GDP Countries)')
plt.show()

# 4. ...and run it again on the second new dataframe
print("\n--- Correlation for HIGH GDP Countries ---")
high_gdp_corr = high_gdp_df[columns_to_correlate].corr()
plt.figure(figsize=(10, 7))
sns.heatmap(high_gdp_corr, annot=True, cmap='coolwarm')
plt.title('Drivers of Happiness (HIGH GDP Countries)')
plt.show()
```

```
The median GDP is: 1.4515

--- Correlation for LOW GDP Countries ---
```




```
--- Correlation for HIGH GDP Countries ---
```



In [ ]:
```
# **Finding:** The heatmap clearly shows that 'GDP' and 'Social' have the highest c
#(around +0.7 or +0.8). This means that, in general, richer and more supportive soc
```