Nicolas Braverman EDA Project

The World Happiness Report examines data from different countries to analyze what makes certain countries happier than others. This project explores and investigates correlations between the data through different forms of visualizations in order to determine what goes into making a country happy. The variables for the data include the region of the country, overall happiness score, and the extent to which GDP per capita, social support, life expectancy, freedom, generosity, and perceptions of corruption contributed to the overall happiness score. When collecting the data, I discovered that the respective regions for the countries was not included in the data. After collecting and preparing this missing information, it was ready to be analyzed for trends.

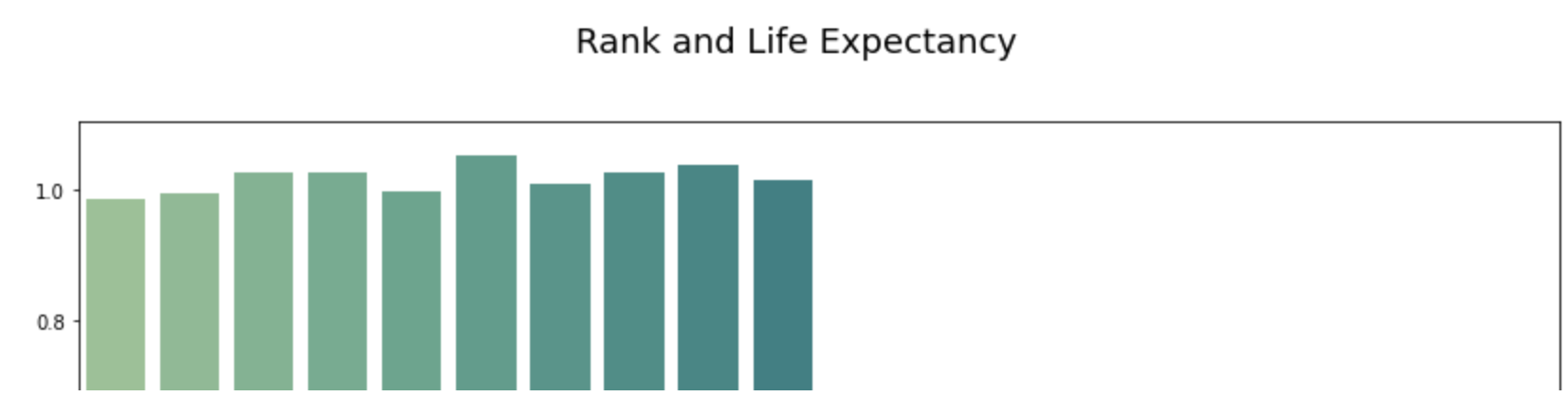
Some questions that arose after exploring the data include:

- does life expectancy contribute to overall rank of happiness?
- does GDP per capita contribute to overall rank of happiness?
- does social support contribute to overall rank of happiness?
- does perception of corruption impact the overall rank of happiness?
- does generosity impact the overall rank of happiness?
- which region has the majority of happy countries?
- which region has the majority of least happy countries?
- is there any correlation between the variables?

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns


frame = pd.read_csv("dataeda.csv") # read file


frameTop = frame.head(10) # top ten happiest countries
frameBottom = frame.tail(10) # lowest ten happiest countries
frameConcat = pd.concat([frameTop, frameBottom]) # puts separate data frames together


print(frameConcat[['Rank', 'Country', 'Region', 'Score']]) # scores of top and bottom countries
```

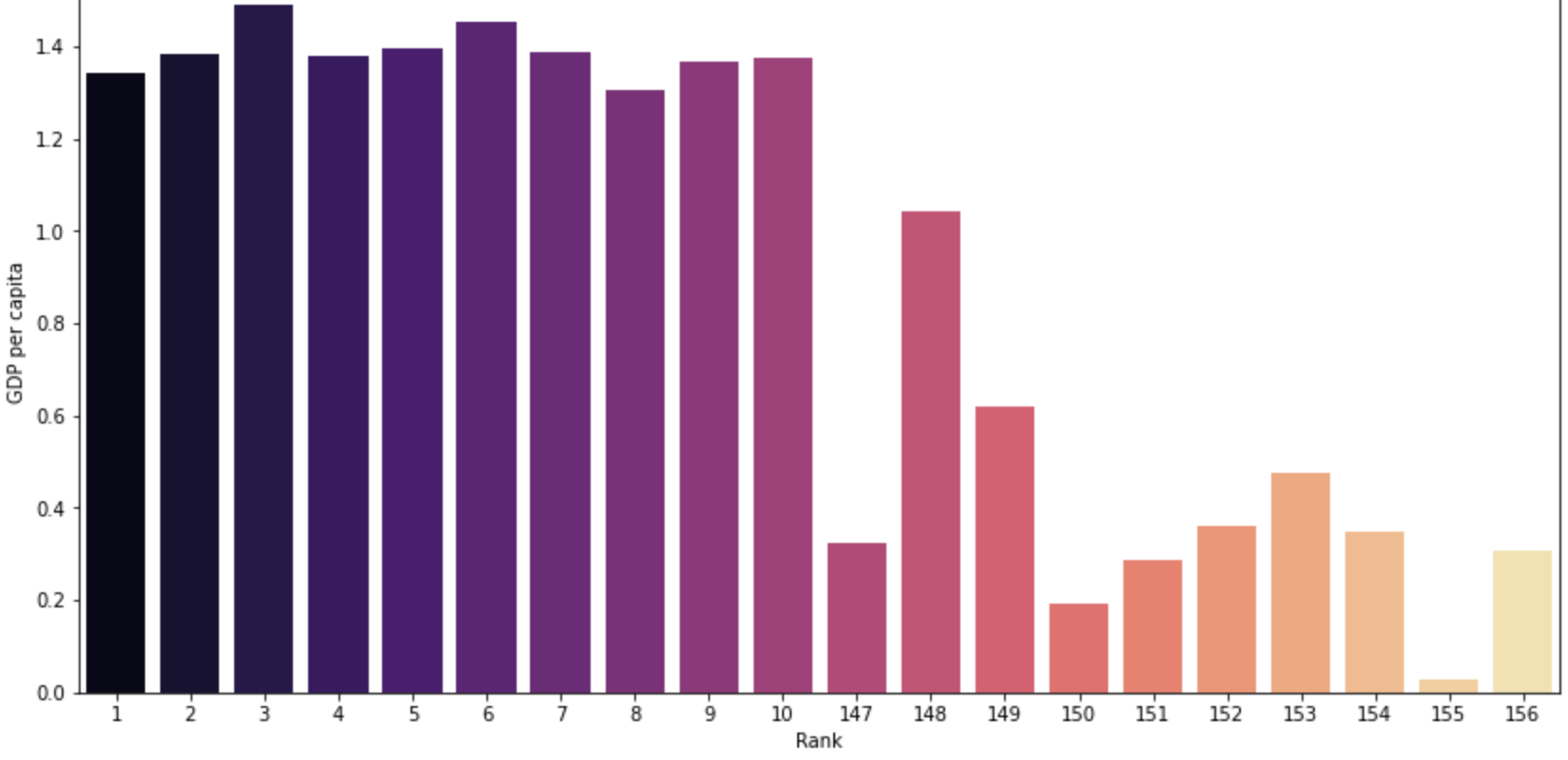
	Rank	Country	Region	Score
0	1	Finland	Western Europe	7.769
1	2	Denmark	Western Europe	7.600
2	3	Norway	Western Europe	7.554
3	4	Iceland	Western Europe	7.494
4	5	Netherlands	Western Europe	7.488
5	6	Switzerland	Western Europe	7.480
6	7	Sweden	Western Europe	7.343
7	8	New Zealand	Oceania	7.307
8	9	Canada	North America	7.278
9	10	Austria	Western Europe	7.246
146	147	Haiti	Latin America and Caribbean	3.597
147	148	Botswana	Sub-Saharan Africa	3.488
148	149	Syria	Middle East and Northern Africa	3.462
149	150	Malawi	Sub-Saharan Africa	3.410
150	151	Yemen	Middle East and Northern Africa	3.380
151	152	Rwanda	Sub-Saharan Africa	3.334
152	153	Tanzania	Sub-Saharan Africa	3.231
153	154	Afghanistan	Southern Asia	3.203
154	155	Central African Republic	Sub-Saharan Africa	3.083
155	156	South Sudan	Sub-Saharan Africa	2.853

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(14,7))
fig.suptitle('Rank and Life Expectancy', fontsize='18')
ax = sns.barplot(x="Rank", y="Life Expectancy", data=frameConcat, palette = "crest")
```



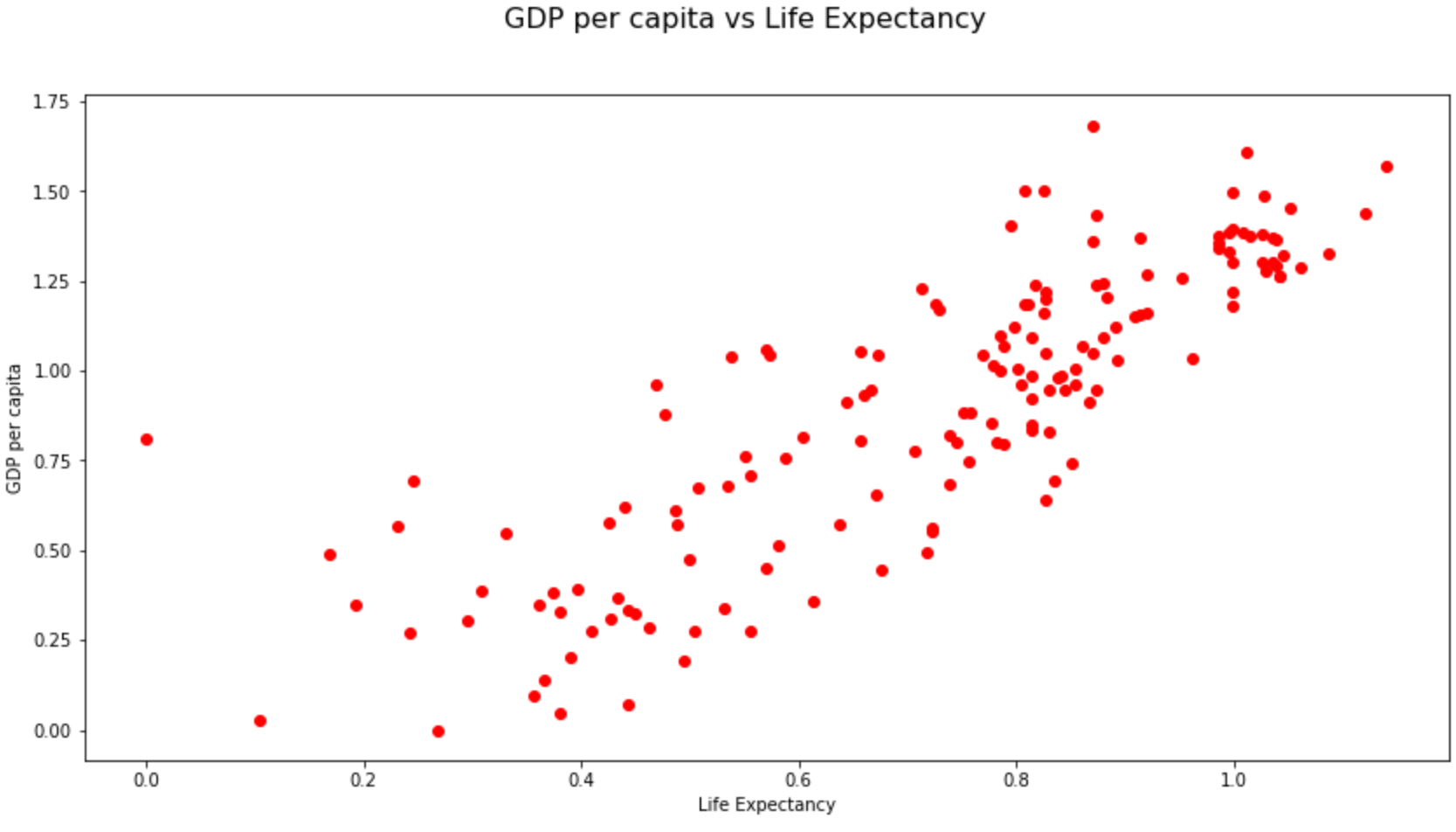
This visualization demonstrates that life expectancy is correlated with having a greater happiness score. The top ten happiest countries all consistently have higher life expectancy than the lowest ten happiest countries.

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(14,7))
fig.suptitle('Rank and GDP per capita', fontsize='18')
ax = sns.barplot(x="Rank", y="GDP per capita", data=frameConcat, palette = "magma")
```



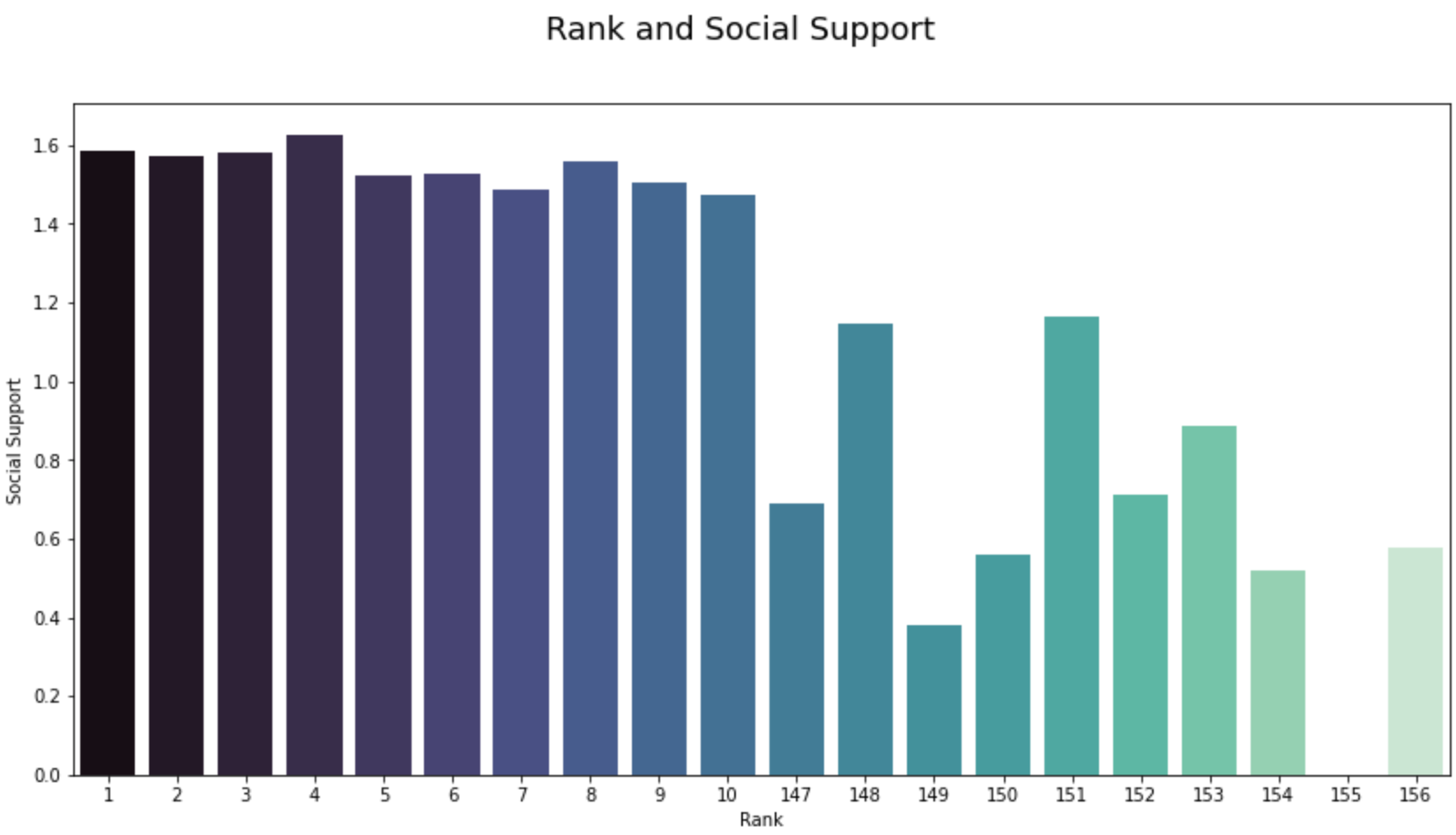
Similar to life expectancy, this visualization shows that the greater the GDP per capita of a country is, the higher its happiness score.

```
fig = plt.figure(figsize=(14,7))
fig.suptitle('GDP per capita vs Life Expectancy', fontsize='16')
plt.scatter(frame["Life Expectancy"],frame["GDP per capita"],color="red")
plt.xlabel("Life Expectancy")
plt.ylabel("GDP per capita")
plt.show()
```



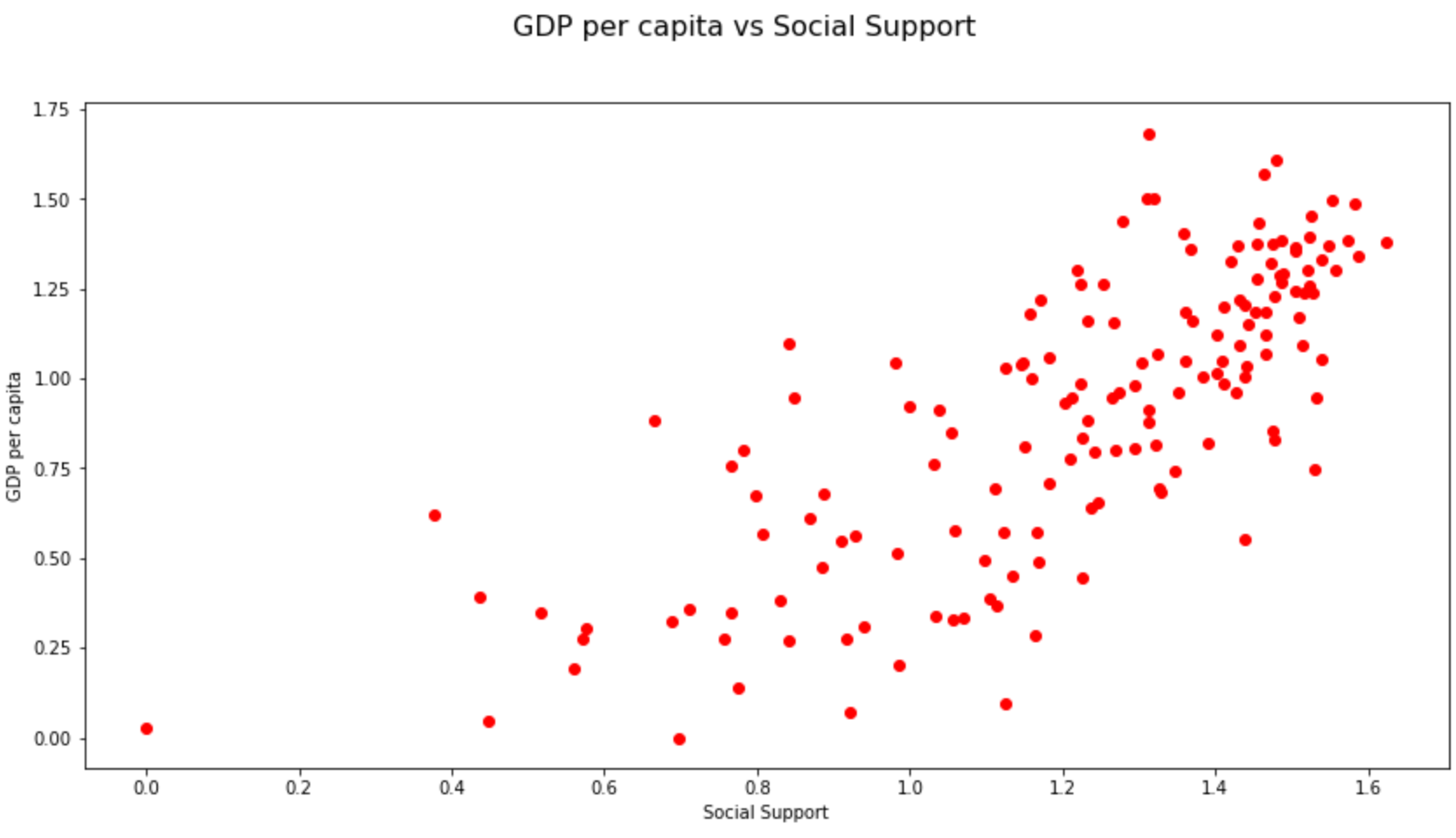
Due to both higher life expectancy and higher GDP per capita resulting in higher happiness scores, I wanted to see if these two variables had positive correlations. After analyzing this visualization, it shows that a country having a higher life expectancy is positvely correlated with a higher GDP per capita

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(14,7))
fig.suptitle('Rank and Social Support', fontsize='18')
ax = sns.barplot(x="Rank", y="Social Support", data=frameConcat, palette = "mako")
```



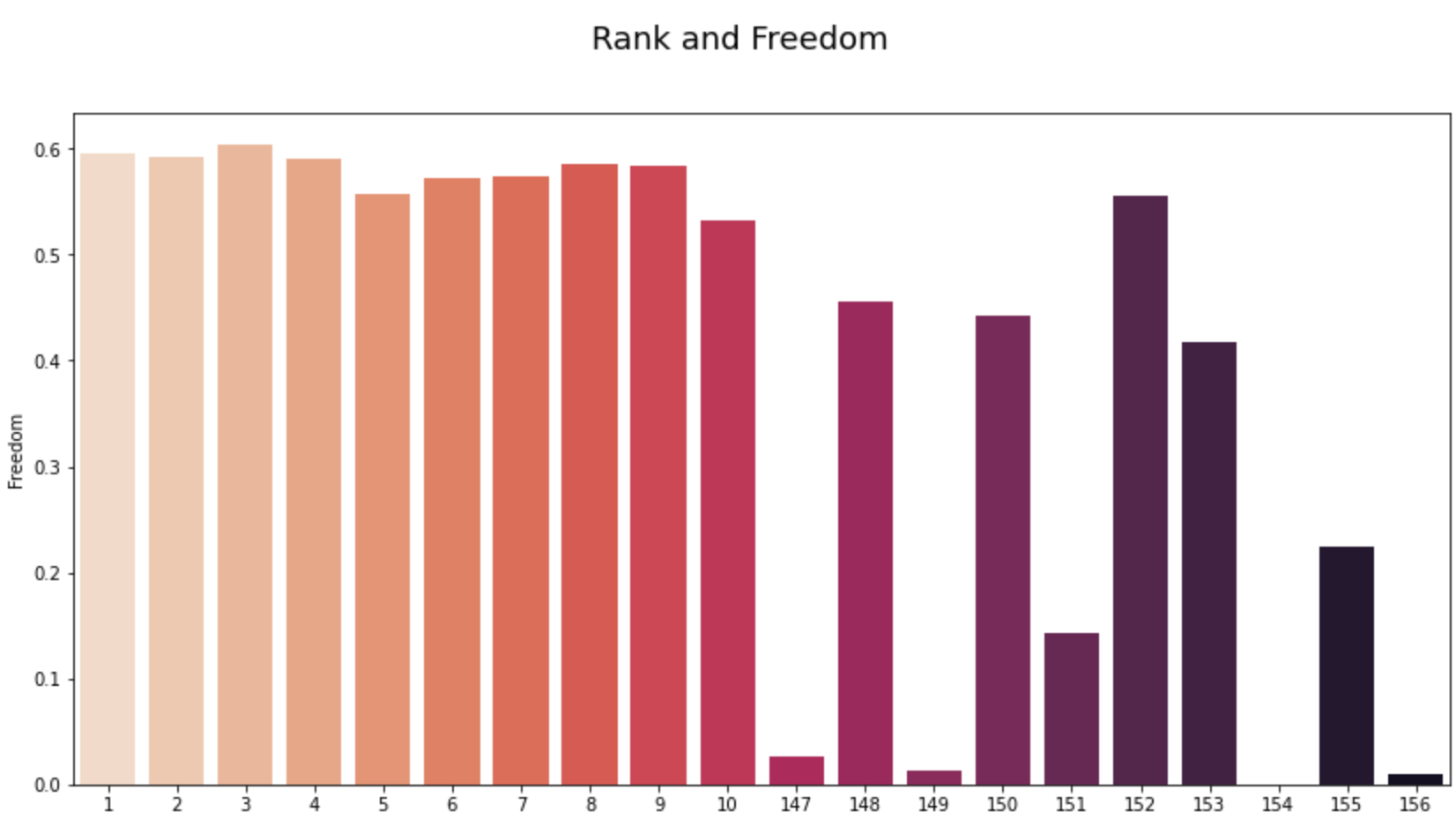
This visualization shows that the greater the social support of a country is, the higher its happiness score. However, there are a couple of outliers, demonstrating that it is not as highly correlated as the previous two variables.

```
fig = plt.figure(figsize=(14,7))
fig.suptitle('GDP per capita vs Social Support', fontsize='16')
plt.scatter(frame["Social Support"],frame["GDP per capita"],color="red")
plt.xlabel("Social Support")
plt.ylabel("GDP per capita")
plt.show()
```



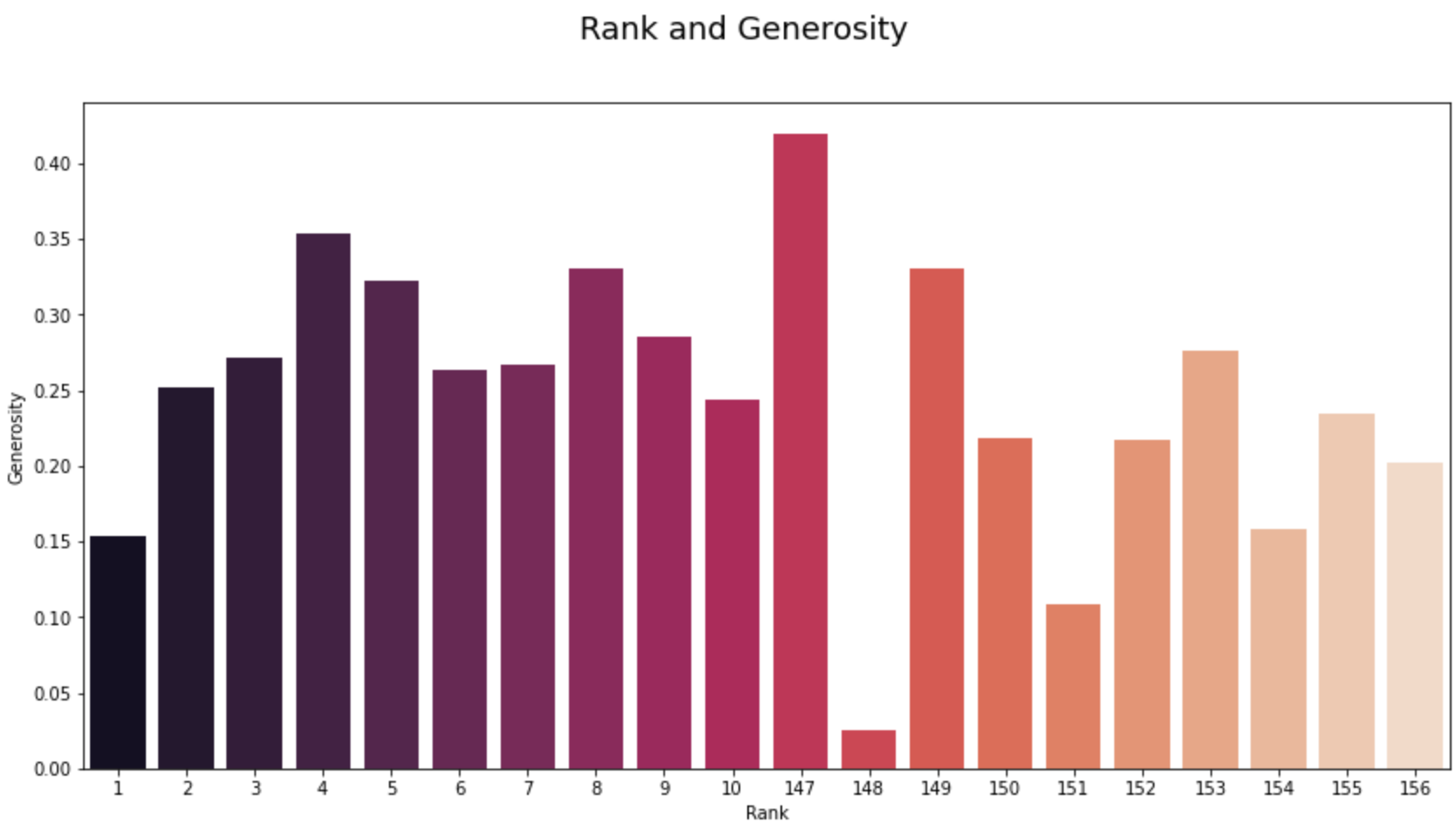
Due to both higher social support and higher GDP per capita resulting in higher happiness scores, I wanted to see if these two variables had positive correlations. After analyzing this visualization, it shows that a country having higher social support is positvely correlated with a higher GDP per capita.

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(14,7))
fig.suptitle('Rank and Freedom', fontsize='18')
ax = sns.barplot(x="Rank", y="Freedom", data=frameConcat, palette = "rocket_r")
```



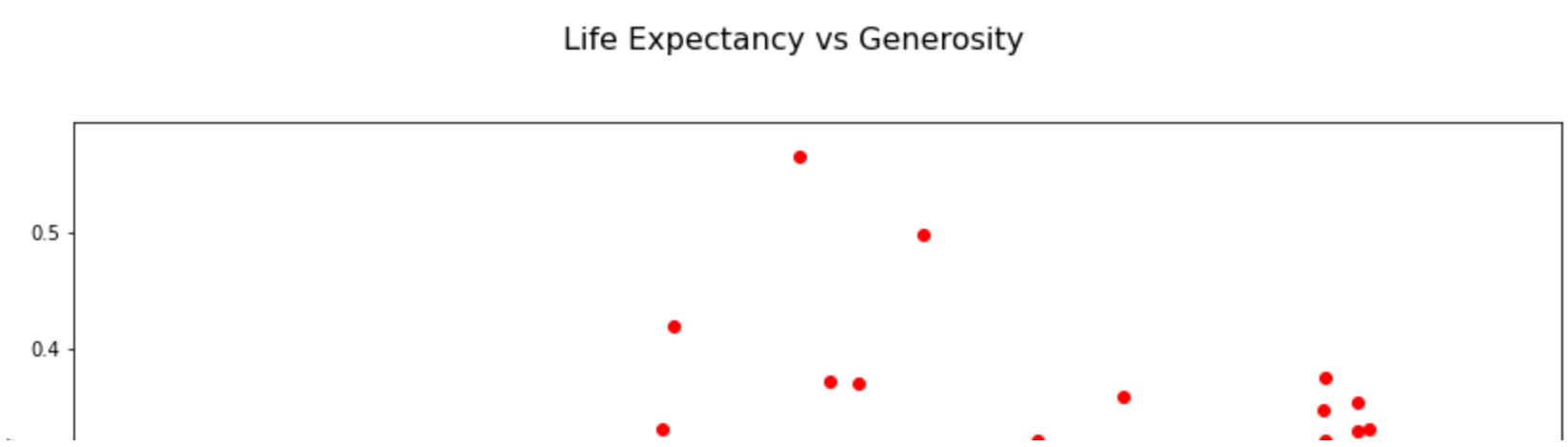
This visualization surprised me due to the number of outliers in the lowest top ten happiest countries. This showed that although freedom to make choices is a component in making a country happy, it is not the most influential.

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(14,7))
fig.suptitle('Rank and Generosity', fontsize='18')
ax = sns.barplot(x="Rank", y="Generosity", data=frameConcat, palette = "rocket")
```



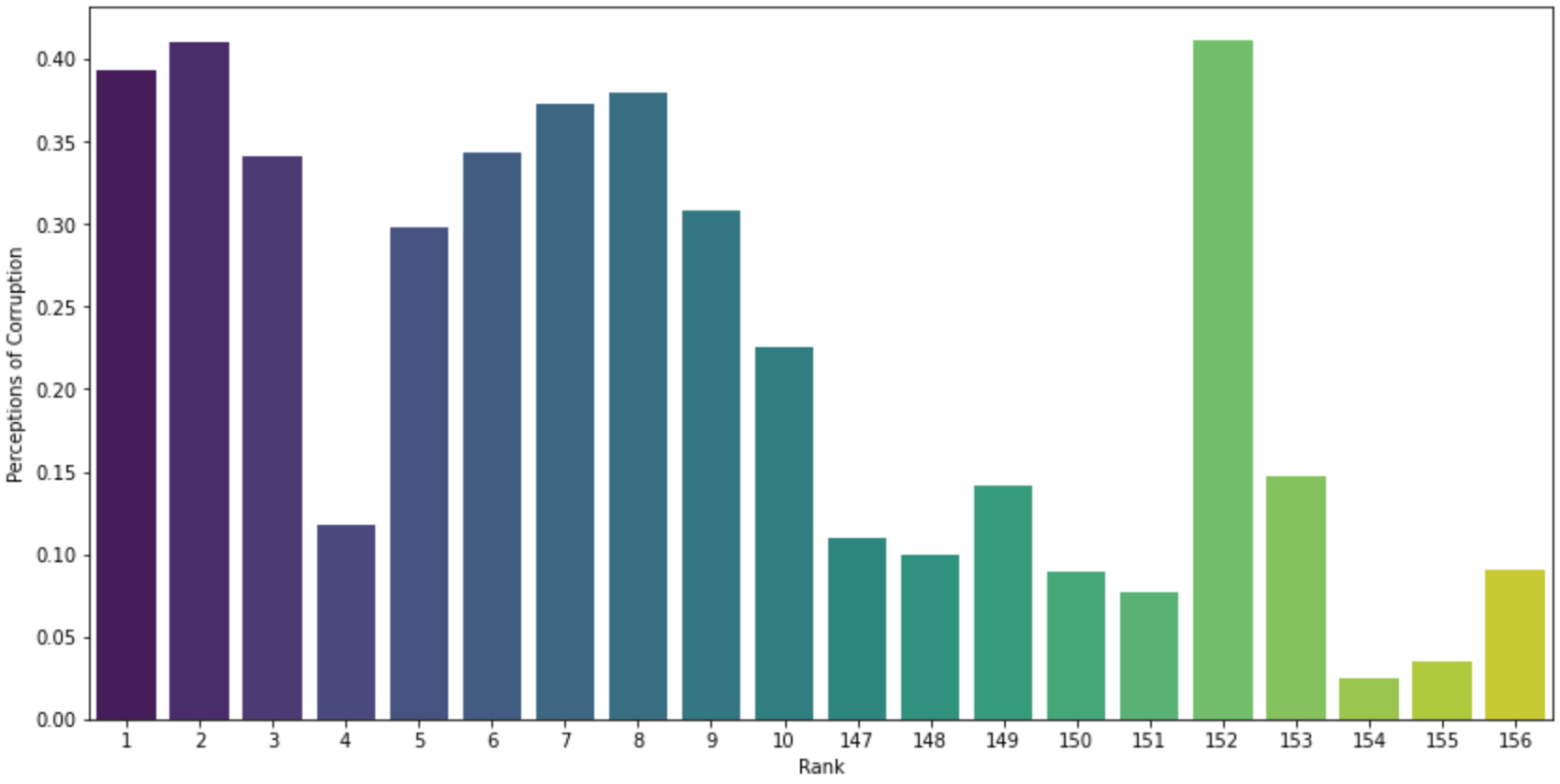
This graph demonstrates very little correlation between happiness and generosity as the valuables are spread out pretty evenly across the graph.

```
fig = plt.figure(figsize=(14,7))
fig.suptitle('Life Expectancy vs Generosity', fontsize='16')
plt.scatter(frame["Life Expectancy"],frame["Generosity"],color="red")
plt.xlabel("Life Expectancy")
plt.ylabel("Genorosity")
plt.show()
```



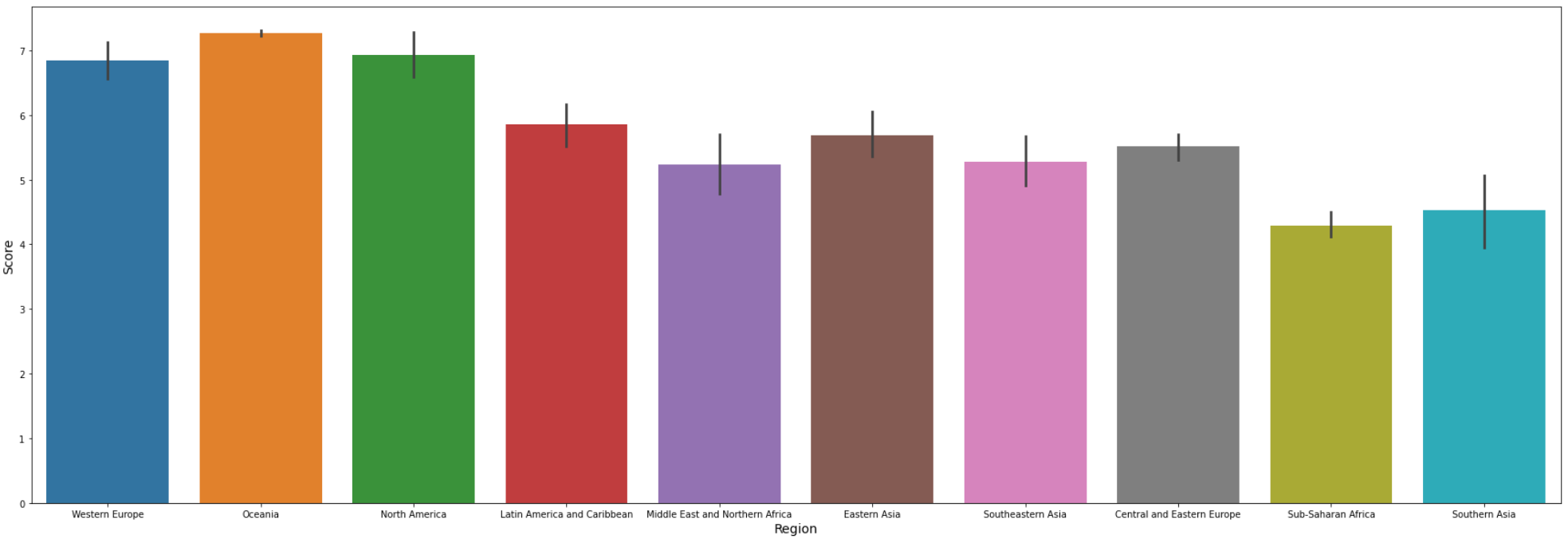
Since the genorosity graph demonstrated little correlation to higher happiness scores, I wanted to see if life expectancy and generosity were positvely correlated. The scatter plot shows that life expectancy and generosity have little correlation. This makes sense since generosity did not show to have much impact in the overall happiness scores.

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(14,7))
fig.suptitle('Rank and Perceptions of Corruption', fontsize='18')
ax = sns.barplot(x="Rank", y="Perceptions of Corruption", data=frameConcat, palette = "viridis")
```



The top ten countries have consistently higher trust in their government than the bottom ten countries. However, this visualization contains outliers and in my opinion does not best represent what makes a country a happy.

```
frame = pd.read_csv("dataeda.csv")
fig = plt.figure(figsize=(30,10))
fig.suptitle('Average Score and Region', fontsize='20')
plt.rcParams["font.size"] = 10
ax = sns.barplot(x="Region", y="Score", data=frame)
```



This plot displays the average scores of happiness of each region. It is interesting to see how the top three regions consist of the most developed countries in the world, the middle consists of countries with emerging economies, and the bottom consists of third-world and underdeveloped countries.