World Happiness Report project

In [53]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [54]:
```
happiness_report =pd.read_csv('happiness_score_dataset.csv')
df.head()
```

Out[54]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.66973
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297

In [55]:
```
#find out the number of columns and rows and the type of the values in each column.
happiness_report.shape
```

Out[55]:
```
(158, 12)
```

In [56]:
```
#check the information of the dataset
happiness_report.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 158 entries, 0 to 157
Data columns (total 12 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Country                        158 non-null    object
 1   Region                         158 non-null    object
 2   Happiness Rank                 158 non-null    int64
 3   Happiness Score                158 non-null    float64
 4   Standard Error                 158 non-null    float64
 5   Economy (GDP per Capita)       158 non-null    float64
 6   Family                         158 non-null    float64
 7   Health (Life Expectancy)       158 non-null    float64
 8   Freedom                        158 non-null    float64
 9   Trust (Government Corruption)  158 non-null    float64
 10  Generosity                     158 non-null    float64
 11  Dystopia Residual              158 non-null    float64
dtypes: float64(9), int64(1), object(2)
memory usage: 14.9+ KB
```

In [57]:
```
#check the columns values
happiness_report.columns
```

Out[57]:
```
Index(['Country', 'Region', 'Happiness Rank', 'Happiness Score',
       'Standard Error', 'Economy (GDP per Capita)', 'Family',
       'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)',
       'Generosity', 'Dystopia Residual'],
      dtype='object')
```

In [58]:
```
#check the missing values of the datset
happiness_report.isnull().sum()
```

Out[58]:
```
Country                          0
Region                           0
Happiness Rank                   0
Happiness Score                  0
Standard Error                   0
Economy (GDP per Capita)         0
Family                           0
Health (Life Expectancy)         0
Freedom                          0
Trust (Government Corruption)    0
Generosity                       0
Dystopia Residual                0
dtype: int64
```

Here no missing values find

In [59]:
```
happiness_report.describe()
```

Out[59]:

	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	(Governr Corrup
count	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.00
mean	79.493671	5.375734	0.047885	0.846137	0.991046	0.630259	0.428615	0.14
std	45.754363	1.145010	0.017146	0.403121	0.272369	0.247078	0.150693	0.12
min	1.000000	2.839000	0.018480	0.000000	0.000000	0.000000	0.000000	0.00
25%	40.250000	4.526000	0.037268	0.545808	0.856823	0.439185	0.328330	0.06
50%	79.500000	5.232500	0.043940	0.910245	1.029510	0.696705	0.435515	0.10
75%	118.750000	6.243750	0.052300	1.158448	1.214405	0.811013	0.549092	0.18
max	158.000000	7.587000	0.136930	1.690420	1.402230	1.025250	0.669730	0.55

In [60]:
```
#Happiness Rank ranges from 1 to 158. Happiness Score ranges from 2.839 to 7.587.
len(happiness_report["Country"].unique())
```

Out[60]:
```
158
```

There are 158 unique countries which means that there are no duplicate countries in the list.

In [61]:
```
len(happiness_report["Region"].unique())
```

Out[61]:
```
10
```

There are 10 unique regions in the list.

In [62]:
```
len(happiness_report["Happiness Rank"].unique())
```

Out[62]:
```
157
```

Out of 158 countries, there are 157 unique Happiness Rank. It means that there are 2 countries with the same rank. Let's find out which countries they are.

In [63]:
```
duplicated_rank = happiness_report[happiness_report["Happiness Rank"].duplicated()].il
happiness_report[happiness_report["Happiness Rank"] == duplicated_rank]
```

Out[63]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedor
81	Jordan	Middle East and Northern Africa	82	5.192	0.04524	0.90198	1.05392	0.69639	0.4066
82	Montenegro	Central and Eastern Europe	82	5.192	0.05235	0.97438	0.90557	0.72521	0.1826

ordan and Montenegro have the same Happiness Rank (#82) with the equal Happiness Score of 5.192.

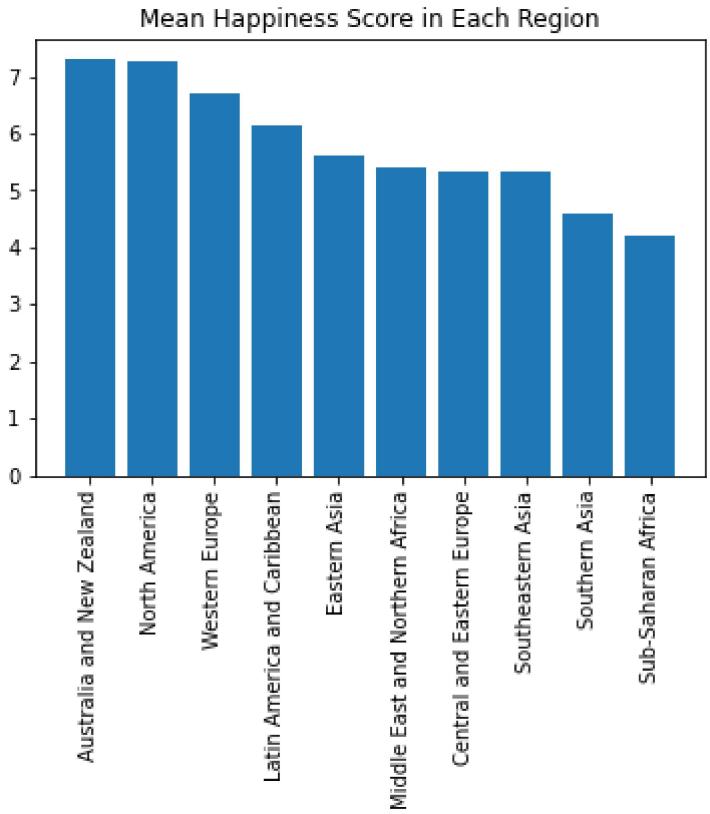
Now, we will load the 2019 dataset and merge the dataframes together based on the year.

Exploratory Analysis and Visualization

In [ ]:

In [64]:
```
region_2015_df = happiness_report.groupby("Region")
happiness_score_region_2015_df = region_2015_df['Happiness Score'].mean().to_frame()
happiness_score_region_2015_df = happiness_score_region_2015_df.sort_values('Happiness
```

In [65]:
```
plt.title("Mean Happiness Score in Each Region");
plt.bar(happiness_score_region_2015_df.index, happiness_score_region_2015_df["Happines
plt.xticks(rotation=90);
```



As shown in the graph above, Australia and New Zealand have the highest mean Happiness Score which means that people who live in this region feel happier than people living in any other region. North America is a close second

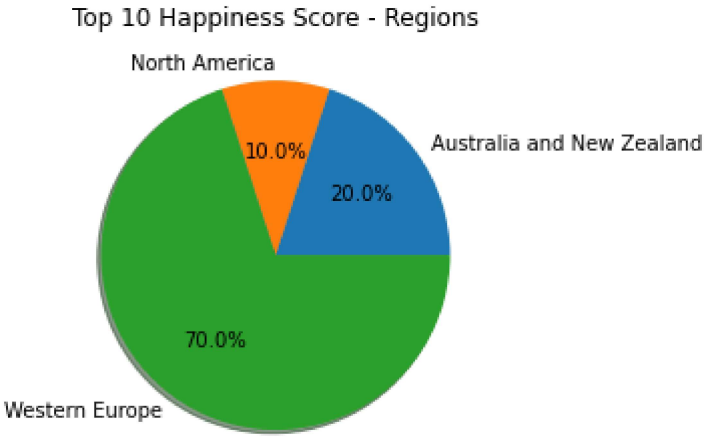
In [66]:
```
#check top 10 happiest countries.
top_10_happiest_countries_df = happiness_report_2015_df.sort_values('Happiness Rank').
top_10_happiest_countries_df
```

Out[66]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.66973
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297
5	Finland	Western Europe	6	7.406	0.03140	1.29025	1.31826	0.88911	0.64169
6	Netherlands	Western Europe	7	7.378	0.02799	1.32944	1.28017	0.89284	0.61576
7	Sweden	Western Europe	8	7.364	0.03157	1.33171	1.28907	0.91087	0.65980
8	New Zealand	Australia and New Zealand	9	7.286	0.03371	1.25018	1.31967	0.90837	0.63938
9	Australia	Australia and New Zealand	10	7.284	0.04083	1.33358	1.30923	0.93156	0.65124

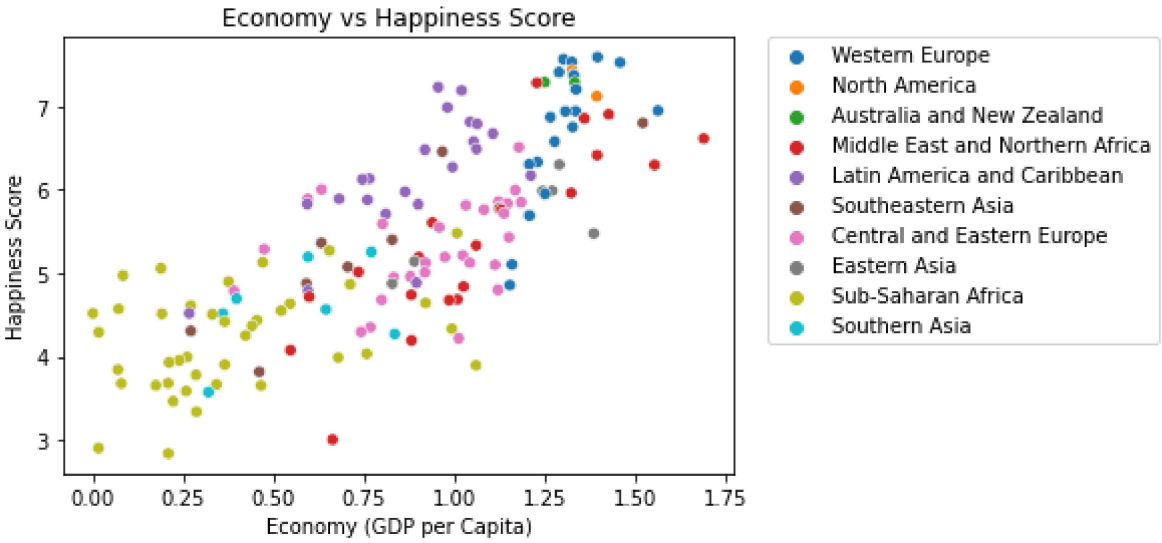
In [67]:

```
top_10_happiest_countries_in_regions_df = top_10_happiest_countries_df.groupby(["Regio
top_10_happiest_countries_in_regions_df
plt.title("Top 10 Happiness Score - Regions");
plt.pie(top_10_happiest_countries_in_regions_df["Country"], labels=top_10_happiest_cou
        autopct='%1.1f%%', shadow=True);
```



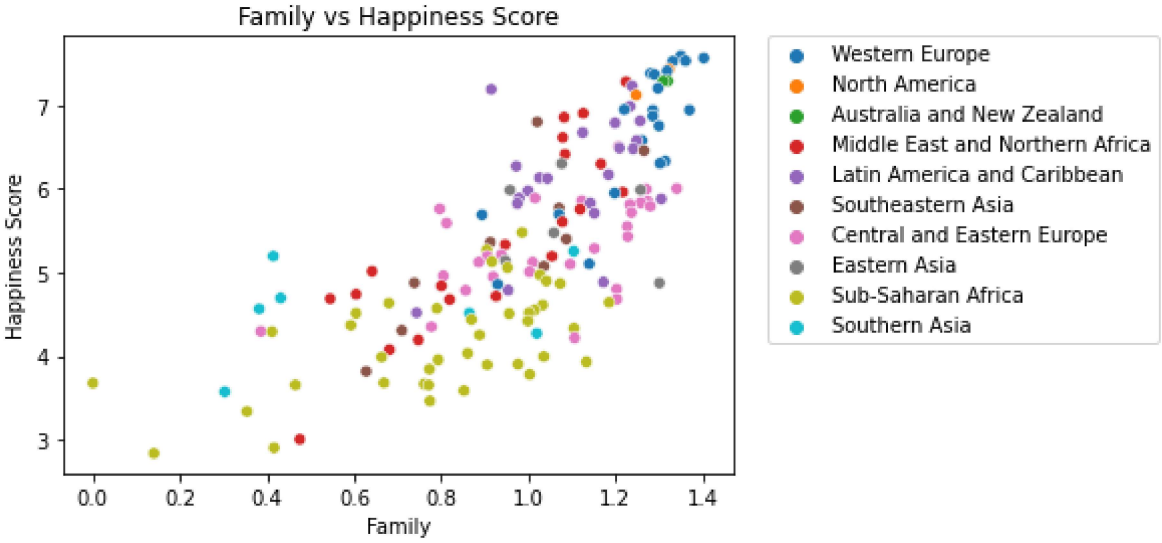
As shown in the pie chart above, 70% of the top 10 happiest countries are from Western Europe. 20% are from Australia and New Zealand and 10% are from North America.

In [68]:
```
#Check how economy is related to the happiness score.
plt.title("Economy vs Happiness Score");
sns.scatterplot(x=happiness_report_2015_df["Economy (GDP per Capita)"], y=happiness_re

# Put the legend out of the figure
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.);
```



As the economy (GDP per Capita) increases, the happiness score also increases fairly linearly. This is as I expected. People say more money more problem but I guess that is not true.

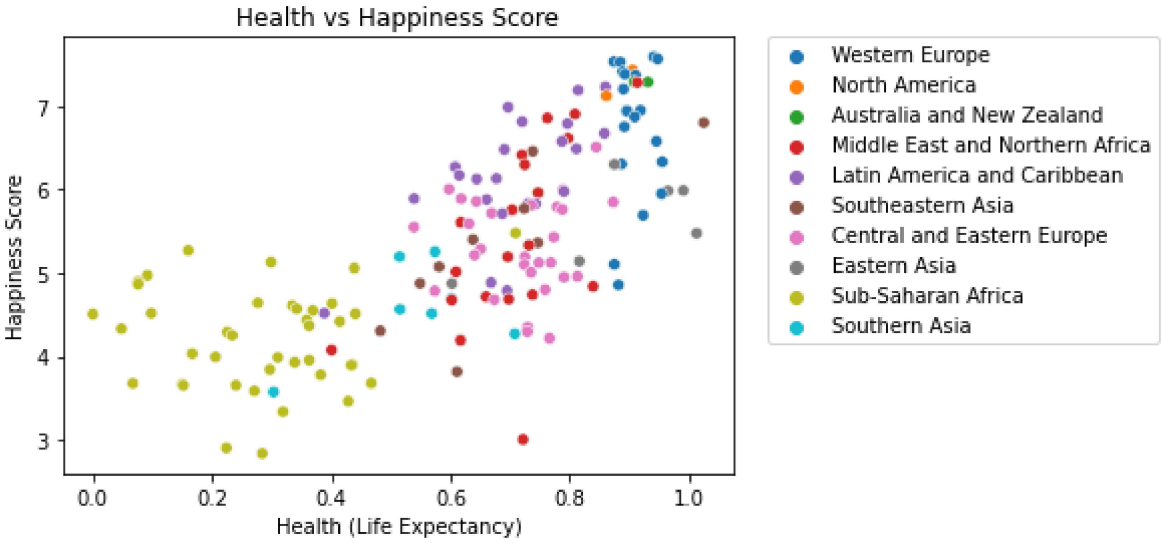
In [69]:
```
#Check family aspect affect the happiness score
plt.title("Family vs Happiness Score");
sns.scatterplot(x=happiness_report_2015_df["Family"], y=happiness_report_2015_df["Happ

# Put the legend out of the figure
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.);
```



The plot represents the relationship between the family contribution and the happiness score.

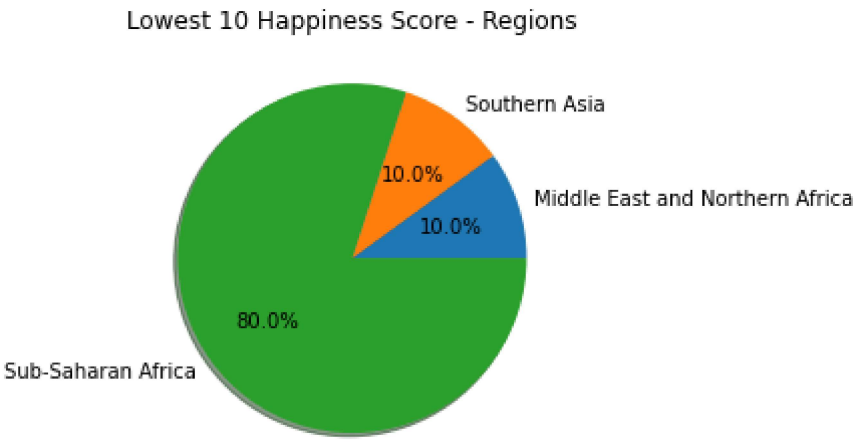
Anything below 0.6 is inconclusive. However, family contribution that is larger than 0.6 shows a linear relationship with the happiness score.

In [70]:
```
#Look at health vs happiness score.
plt.title("Health vs Happiness Score");
sns.scatterplot(x=happiness_report_2015_df["Health (Life Expectancy)"], y=happiness_re

# Put the legend out of the figure
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.);
```



As it shows on the plot, the healthier you are, the more happier you would feel. However, there are many outliers. Many countries in Sub-Saharan Africa scored very low in health or life expectancy, but the happiness score varied from 3-5. On the other hand, one of the countires in Middle East and Norther Africa has relatively high health score but scored around 3. We can conclude that the health do affect the happiness score, but not necessarily as much as family or economy.

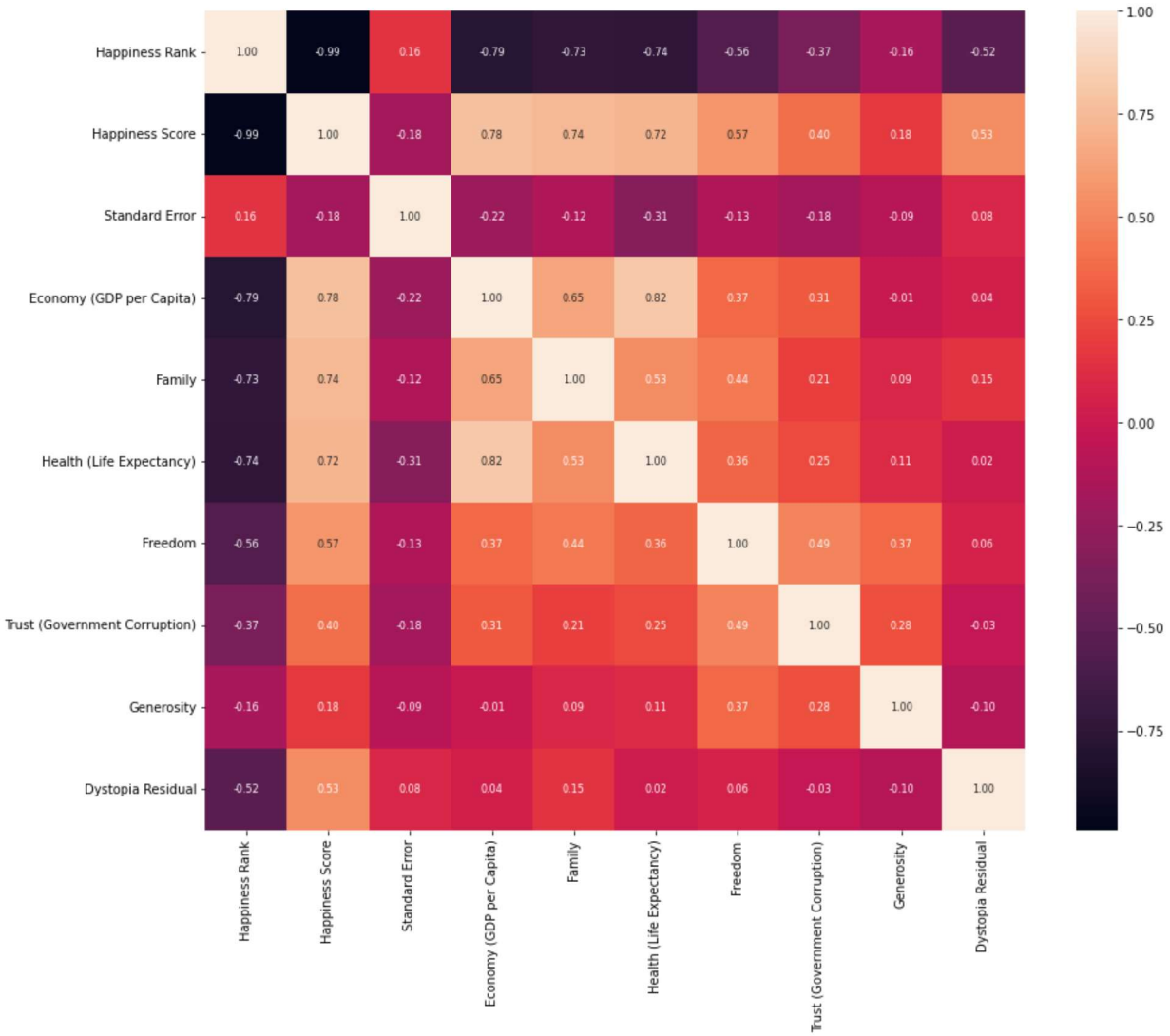
In [71]:
```
#Which region do the least happiest 10 countries belong to?
lowest_happiness_rank_2015_df = happiness_report_2015_df[happiness_report_2015_df["Hap
lowest_happiness_rank_region_df = lowest_happiness_rank_2015_df.groupby(["Region"])["C

plt.title("Lowest 10 Happiness Score - Regions");
plt.pie(lowest_happiness_rank_region_df["Country"], labels=lowest_happiness_rank_regic
        autopct='%1.1f%%', shadow=True);
```



Out of 10 countries, 8 of them belong to Sub-Saharan Africa region and 1 from Southern Asia and 1 from Middle East and Norther Africa.

In [72]:
```
plt.figure(figsize=(15,12))
sns.heatmap(happiness_report_2015_df.corr(),cbar=True,square=True,fmt='.2f',annot=True
```

Out[72]:
```
<AxesSubplot:>
```



As shown in the heatmap, the economy score seems to be the most strongly correlated with the happiness score. Along with the economy score, the family and health is almost as strongly correlated with the happiness score as the economy score. This concludes that in order to be happy, we need to be economically well, need to have good relationship with the family, and need to be health.

Conclusion

We have looked into the world happiness report. We have learned that majority of the top 10 happiest country belongs to Western Europe and the majority of the bottom 10 happiest country belongs to Sub-Saharan Africa region. We should remember to work hard, spend good amount of time with our family and keep having an apple a day to stay healthy! This is a key to become a happier person.

In [ ]: