DataVisualization

October 21, 2022

[1]:
```
import pandas as pd
import numpy as np
import missingno as msno
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
import math
warnings.filterwarnings('ignore')
sns.set_theme()
```

1 GDP per capita and Life Expectancy

[2]:
```
#reading the dataframe
df = pd.read_csv('Data/life-expectancy-vs-gdp-per-capita.csv')
df.head()
```

[2]:

	Entity	Code	Year	Life expectancy	GDP per capita	\
0	Abkhazia	OWID_ABK	2015	NaN	NaN	
1	Afghanistan	AFG	1950	27.638	1156.0	
2	Afghanistan	AFG	1951	27.878	1170.0	
3	Afghanistan	AFG	1952	28.361	1189.0	
4	Afghanistan	AFG	1953	28.852	1240.0	

	145446-annotations	Population (historical estimates)	Continent
0	NaN	NaN	Asia
1	NaN	7752117.0	NaN
2	NaN	7840151.0	NaN
3	NaN	7935996.0	NaN
4	NaN	8039684.0	NaN

[3]:
```
#generating descriptive statistics
df.describe()
```

[3]:

	Year	Life expectancy	GDP per capita	\
count	60066.000000	19028.000000	19876.000000	
mean	1606.366297	61.751767	6707.679440	

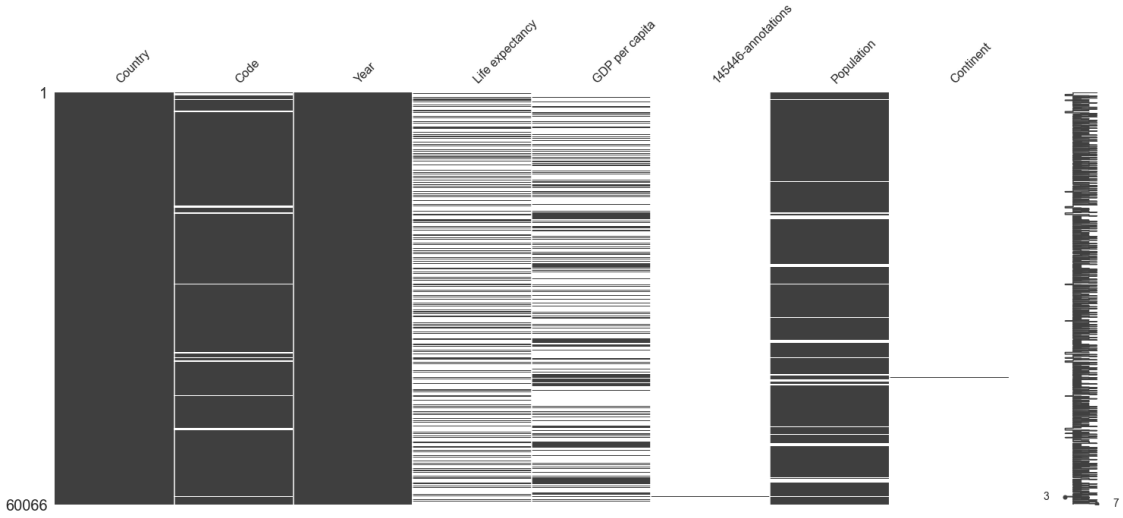
```
std     1364.912223        13.091632    10120.349224
min   -10000.000000        17.760000      295.000000
25%     1819.000000        52.314750     1553.000000
50%     1892.000000        64.713000     2798.000000
75%     1962.000000        71.984250     7130.298500
max     2021.000000        86.751000   156299.000000

       Population (historical estimates)
count                       5.565600e+04
mean                        3.246352e+07
std                         2.503028e+08
min                         1.000000e+00
25%                         1.338740e+05
50%                         1.218570e+06
75%                         5.396250e+06
max                         7.874966e+09
```

[4]:
```
#Renaming the columns
df.rename(columns={'Entity': 'Country'}, inplace=True)
df.rename(columns={'Population (historical estimates)': 'Population'},␣
↪inplace=True)
```

[5]:
```
#Showing data sparsity across all dataframe columns.
msno.matrix(df)
```

[5]: <AxesSubplot:>



From the plot above you can see the amount of missing data per column in the dataset. The columns '145446-annotations' and 'continent' have the most number of missing

data.

[6]:
```
# creating a new dataframe with selected columns
dfa = df[['Country','Year','Life expectancy','GDP per capita', 'Population']]
dfa = dfa[dfa['Year'] == 2018] #Latest year include 'GDP per capita'

dfa.head()
```

[6]:

	Country	Year	Life expectancy	GDP per capita	Population
69	Afghanistan	2018	64.486	1934.5550	3.717192e+07
330	Africa	2018	62.839	NaN	1.275921e+09
588	Albania	2018	78.458	11104.1665	2.882735e+06
850	Algeria	2018	76.693	14228.0250	4.222842e+07
1106	American Samoa	2018	73.679	NaN	5.546100e+04

[7]:
```
#checking for Na values
dfa.isna().any()
```

[7]:
```
Country            False
Year               False
Life expectancy     True
GDP per capita      True
Population          True
dtype: bool
```

[8]:
```
#removing Na values
dfa = dfa.dropna()
```

[9]:
```
#Top 10 ountries ranked by life expectancy in 2018
dfa_sorted = dfa.sort_values(by =['Life expectancy'], ascending=False)
dfa_sorted.head(10)
```

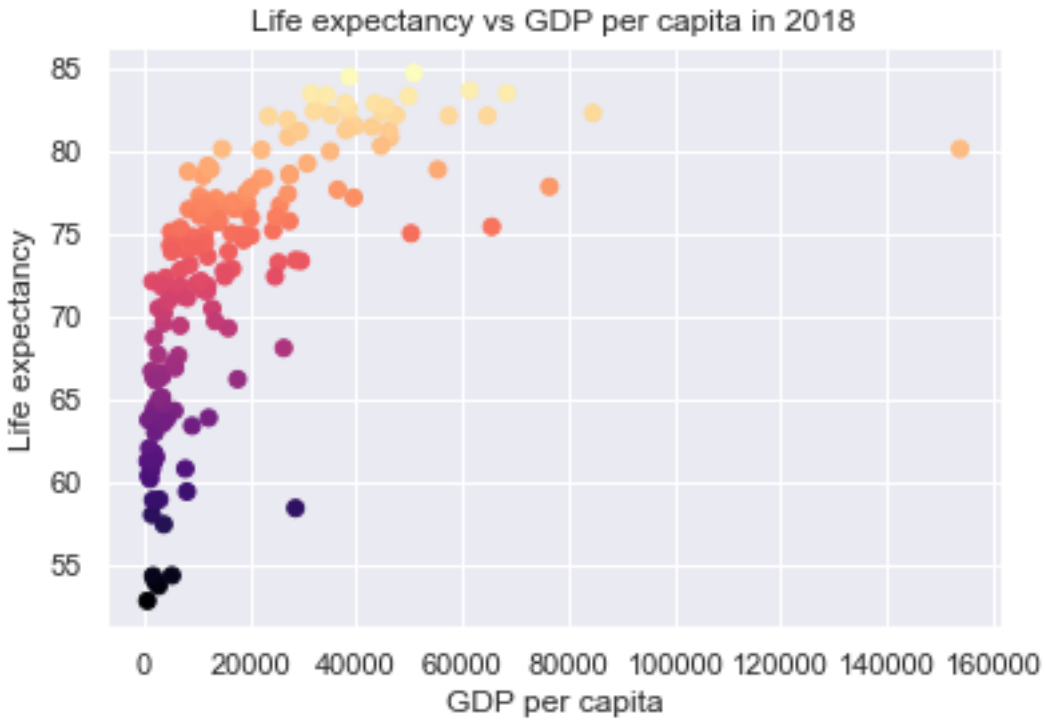
[9]:

	Country	Year	Life expectancy	GDP per capita	Population
22575	Hong Kong	2018	84.687	50839.3714	7371728.0
25991	Japan	2018	84.470	38673.8081	127202190.0
51740	Switzerland	2018	83.630	61372.7301	8525614.0
47306	Singapore	2018	83.458	68402.3451	5757503.0
49808	Spain	2018	83.433	31496.5200	46692863.0
25062	Italy	2018	83.352	34364.1682	60627291.0
3110	Australia	2018	83.281	49830.7993	24898153.0
23205	Iceland	2018	82.855	43438.5412	336712.0
49270	South Korea	2018	82.846	37927.6095	51171700.0
24724	Israel	2018	82.819	32954.7701	8381507.0

Scatter plot of GDP per capita vs life expectancy

[10]:
```
#Simple scatter plot
x = dfa["GDP per capita"]
```

```
y = dfa["Life expectancy"]
plt.scatter(x, y, c = dfa["Life expectancy"], cmap = 'magma')
plt.title('Life expectancy vs GDP per capita in 2018')
plt.xlabel('GDP per capita')
plt.ylabel('Life expectancy')
plt.rcParams['figure.figsize'] = [15, 15]
plt.show()
```



1. We chose 2018 as the most recent year for which we could find data on the 'GDP per capita' column.
2. Because we feel that population has an impact on a country's GDP, we decided to include population as one of the elements in our research.

From the scatter plot we can interpret that the increase in life expectancy is accompanied with the increase in Gross Domestic Product per capita income. We also discussed that the inclusion of population growth rate would be an important factor contributing towards GDP and if this is also included our interpretation could be more precise.

In the beggining I used msno.matrix in the dataframe to create a axessubplot to check missing values in all the columns where we could see that the columns 145446-annotations' and 'continent' had the most number of missing data. Since we didn't have to use those columns for the other questions we created another dataframe and selected the columns I needed to work with(dfa). I also chose 2018 as the most recent year for which we could find data on the 'GDP per capita' column.

Which countries have a life expectancy higher than one standard deviation above the mean?

[11]:
```
import statistics
```

[12]:
```
# calculating the mean of Life expectancy
meanLife_Exp = statistics.mean(dfa['Life expectancy'])
round(meanLife_Exp,2)
```

[12]: 72.66

[13]:
```
# calculating the standard deviation of Life expectancy
std_dev_life = statistics.stdev(dfa['Life expectancy'])
round(std_dev_life,2)
```

[13]: 7.72

[14]:
```
# One standard deviation above the mean of Life expectancy
std_dev_above =  meanLife_Exp + std_dev_life
round(std_dev_above,2)
```

[14]: 80.39

[15]:
```
# Countries having a life expectancy higher than one standard deviation above␣
 ↪the mean
result_above = dfa[dfa['Life expectancy'] > std_dev_above]
result_above
```

[15]:

	Country	Year	Life expectancy	GDP per capita	Population
3110	Australia	2018	83.281	49830.7993	24898153.0
3347	Austria	2018	81.434	42988.0709	8891383.0
5248	Belgium	2018	81.468	39756.2031	11482180.0
9029	Canada	2018	82.315	44868.7435	37074558.0
12639	Cyprus	2018	80.828	27184.4166	1189262.0
13622	Denmark	2018	80.784	46312.3443	5752131.0
17473	Finland	2018	81.736	38896.7005	5522585.0
18013	France	2018	82.541	38515.9193	64990512.0
19617	Germany	2018	81.180	46177.6187	83124413.0
20214	Greece	2018	82.072	23450.7658	10522244.0
22575	Hong Kong	2018	84.687	50839.3714	7371728.0
23205	Iceland	2018	82.855	43438.5412	336712.0
24399	Ireland	2018	82.103	64684.3020	4818694.0
24724	Israel	2018	82.819	32954.7701	8381507.0
25062	Italy	2018	83.352	34364.1682	60627291.0
25991	Japan	2018	84.470	38673.8081	127202190.0
29967	Luxembourg	2018	82.102	57427.5003	604244.0
31507	Malta	2018	82.376	32028.9124	439255.0
36073	Netherlands	2018	82.143	47474.1095	17059560.0

36857	New Zealand	2018	82.145	35336.1363	4743131.0
38966	Norway	2018	82.271	84580.1362	5337960.0
42481	Portugal	2018	81.857	27035.6002	10256192.0
47306	Singapore	2018	83.458	68402.3451	5757503.0
47887	Slovenia	2018	81.172	29244.9198	2077835.0
49270	South Korea	2018	82.846	37927.6095	51171700.0
49808	Spain	2018	83.433	31496.5200	46692863.0
51120	Sweden	2018	82.654	45541.8921	9971630.0
51740	Switzerland	2018	83.630	61372.7301	8525614.0
56358	United Kingdom	2018	81.236	38058.0856	67141678.0

Which countries have high life expectancy but have low GDP?

[16]:
```
# Calculating the mean of GDP
mean_GDP = statistics.mean(dfa['GDP per capita'])
round(mean_GDP,2)
```

[16]:
```
18936.93
```

[17]:
```
# Selecting the countries which have the GDP lower than the GDP mean but have␣
 ↪Life expectancy higher than the mean of Life expectancy
result_1e = dfa[(dfa['GDP per capita'] < mean_GDP) & (dfa['Life expectancy'] >␣
 ↪meanLife_Exp)]
result_1e
```

[17]:

	Country	Year	Life expectancy	GDP per capita	\
588	Albania	2018	78.458	11104.1665	
850	Algeria	2018	76.693	14228.0250	
2210	Argentina	2018	76.520	18556.3831	
2454	Armenia	2018	74.945	11454.4251	
3597	Azerbaijan	2018	72.864	16628.0553	
4633	Barbados	2018	79.081	11995.1868	
4886	Belarus	2018	74.590	18727.3176	
6586	Bosnia and Herzegovina	2018	77.262	10460.5201	
7111	Brazil	2018	75.672	14033.5656	
7706	Bulgaria	2018	74.928	18444.2602	
9250	Cape Verde	2018	72.782	6831.2160	
10409	China	2018	76.704	13101.7064	
10700	Colombia	2018	77.109	13545.0495	
11528	Costa Rica	2018	80.095	14686.2539	
12310	Cuba	2018	78.726	8325.6313	
14025	Dominica	2018	74.806	9021.1737	
14282	Dominican Republic	2018	73.892	15912.3995	
14579	Ecuador	2018	76.800	10638.8251	
15104	El Salvador	2018	73.096	8598.1982	
19349	Georgia	2018	73.600	11984.9049	
21024	Guatemala	2018	74.063	7402.1146	

22319	Honduras	2018	75.088	5041.6354
23871	Iran	2018	76.479	17011.3042
25790	Jamaica	2018	74.368	7272.9805
26237	Jordan	2018	74.405	11506.3383
28364	Lebanon	2018	78.875	12558.9669
29141	Libya	2018	72.724	15013.3124
32737	Mexico	2018	74.992	16494.0790
34455	Morocco	2018	76.453	8451.1355
37095	Nicaragua	2018	74.275	4952.4772
38460	North Macedonia	2018	75.688	13074.2313
40120	Palestine	2018	73.895	5207.7569
40866	Paraguay	2018	74.131	9338.9484
41121	Peru	2018	76.516	12310.0847
44633	Saint Lucia	2018	76.057	10475.3689
46564	Serbia	2018	75.849	14124.1177
50083	Sri Lanka	2018	76.812	11662.9064
52966	Thailand	2018	76.931	16648.6237
54305	Tunisia	2018	76.505	11353.8865
58603	Vietnam	2018	75.317	6814.1423

	Population
588	2.882735e+06
850	4.222842e+07
2210	4.436115e+07
2454	2.951741e+06
3597	9.949537e+06
4633	2.866400e+05
4886	9.452615e+06
6586	3.323929e+06
7111	2.094693e+08
7706	7.051610e+06
9250	5.437640e+05
10409	1.427648e+09
10700	4.966106e+07
11528	4.999443e+06
12310	1.133815e+07
14025	7.162600e+04
14282	1.062715e+07
14579	1.708436e+07
15104	6.420740e+06
19349	4.002946e+06
21024	1.724786e+07
22319	9.587523e+06
23871	8.180020e+07
25790	2.934853e+06
26237	9.965322e+06
28364	6.859408e+06

```
29141  6.678565e+06
32737  1.261908e+08
34455  3.602909e+07
37095  6.465502e+06
38460  2.082956e+06
40120  4.862978e+06
40866  6.956069e+06
41121  3.198926e+07
44633  1.818900e+05
46564  8.802741e+06
50083  2.122876e+07
52966  6.942845e+07
54305  1.156520e+07
58603  9.554596e+07
```

1.f Does every strong economy (normally indicated by GDP) have high life expectancy?
We assume that a strong economy is equal to countries with higher GDP per capita than one standard deviation above the mean.

[18]:
```
# calculating the standard deviation of GDP per capita
std_dev_GDP = statistics.stdev(dfa['GDP per capita'])
round(std_dev_GDP,2)
```

[18]:
```
20261.81
```

[19]:
```
# One standard deviation above the mean of GDP per capita
stdmGDP =  mean_GDP + std_dev_GDP
round(stdmGDP,2)
```

[19]:
```
39198.74
```

[20]:
```
result_1f = dfa[dfa['GDP per capita'] > stdmGDP ]
result_1f = result_1f.sort_values(by =['GDP per capita'], ascending=False)
result_1f
```

[20]:

	Country	Year	Life expectancy	GDP per capita
43107	Qatar	2018	80.100	153764.1643
38966	Norway	2018	82.271	84580.1362
55931	United Arab Emirates	2018	77.814	76397.8181
47306	Singapore	2018	83.458	68402.3451
27239	Kuwait	2018	75.398	65520.7367
24399	Ireland	2018	82.103	64684.3020
51740	Switzerland	2018	83.630	61372.7301
29967	Luxembourg	2018	82.102	57427.5003
57035	United States	2018	78.851	55334.7394
22575	Hong Kong	2018	84.687	50839.3714
46045	Saudi Arabia	2018	74.998	50304.7502

```
3110              Australia  2018          83.281      49830.7993
36073           Netherlands  2018          82.143      47474.1095
13622               Denmark  2018          80.784      46312.3443
19617               Germany  2018          81.180      46177.6187
51120                Sweden  2018          82.654      45541.8921
9029                 Canada  2018          82.315      44868.7435
52185                Taiwan  2018          80.283      44663.8642
23205               Iceland  2018          82.855      43438.5412
3347                Austria  2018          81.434      42988.0709
5248                Belgium  2018          81.468      39756.2031
4115                Bahrain  2018          77.163      39498.7672
```

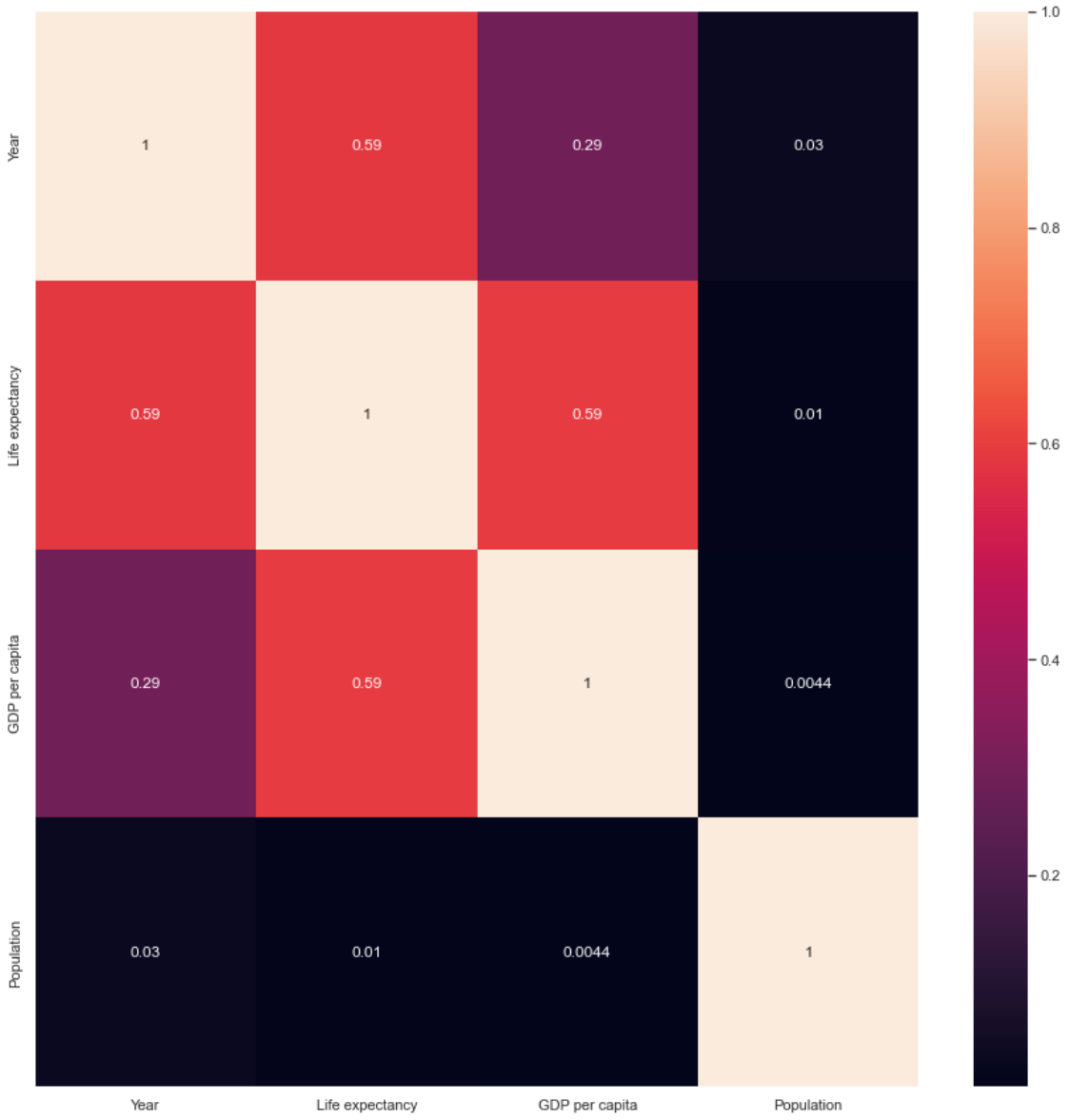
```
        Population
43107    2781682.0
38966    5337960.0
55931    9630966.0
47306    5757503.0
27239    4137314.0
24399    4818694.0
51740    8525614.0
29967     604244.0
57035  327096263.0
22575    7371728.0
46045   33702757.0
3110    24898153.0
36073   17059560.0
13622    5752131.0
19617   83124413.0
51120    9971630.0
9029    37074558.0
52185   23726460.0
23205     336712.0
3347     8891383.0
5248    11482180.0
4115     1569440.0
```

We took the highest and the second last GDP per capita to compare their life expectancy; Qatar had 153764.1643(USD) GDP per capita in 2018 but only had 80% Life expectancy, while Belgium had 39756.2031(USD) GDP per capita but have higher life expectancy than Qatar (81.468%). #### We can assume that not every strong country have a high life expectancy.

1.g Related to question f, what would happen if you use GDP per capita as an indicator of strong economy? Explain the results you obtained, and discuss any insights you get from comparing the results of g and f.

[21]:
```
#plotting a correlation matrix for comparison
corrMatrix = df.corr()
sns.heatmap(corrMatrix, annot=True)
```

```
plt.show()
```



There is a positive linear correlation between life expectancy and GDP per capita from the correlation matrix. Although, as we can see from e and f, some countries have low GDP but have high Life expectancy, and some countries have the highest GDP but do not have the highest Life expectancy. Therefore we can conclude that life expectancy depends not only on the GDP per capita.

2 Happiness and life satisfaction, trust, corruption.

[22]:
```
# dataframe Life satisfaction
satisfaction_df = pd.read_csv('Data/happiness-cantril-ladder.csv')
satisfaction_df.rename(columns={'Life satisfaction in Cantril Ladder (World␣
↪Happiness Report 2021)': 'Life satisfaction'}, inplace=True)
satisfaction_df.rename(columns={'Entity': 'Country'}, inplace=True)
satisfaction_df.head(10)
```

[22]:

	Country	Code	Year	Life satisfaction
0	Afghanistan	AFG	2008	3.724
1	Afghanistan	AFG	2009	4.402
2	Afghanistan	AFG	2010	4.758
3	Afghanistan	AFG	2011	3.832
4	Afghanistan	AFG	2012	3.783
5	Afghanistan	AFG	2013	3.572
6	Afghanistan	AFG	2014	3.131
7	Afghanistan	AFG	2015	3.983
8	Afghanistan	AFG	2016	4.220
9	Afghanistan	AFG	2017	2.662

Link of the dataset: https://ourworldindata.org/happiness-and-life-satisfaction

[23]:
```
# human development index dataframe
HDI_df = pd.read_csv('Data/human-development-index.csv')
HDI_df.rename(columns={'Entity': 'Country'}, inplace=True)
HDI_df.head()
```

[23]:

	Country	Code	Year	Human Development Index (UNDP)
0	Afghanistan	AFG	1980	0.228
1	Afghanistan	AFG	1985	0.273
2	Afghanistan	AFG	2002	0.373
3	Afghanistan	AFG	2003	0.383
4	Afghanistan	AFG	2004	0.398

Data set description: The Human Development Index (HDI) is an index that measures key dimensions of human development. The three key dimensions are:1

– A long and healthy life – measured by life expectancy.

– Access to education – measured by expected years of schooling of children at school-entry age and mean years of schooling of the adult population.

– And a decent standard of living – measured by Gross National Income per capita adjusted for the price level of the country.

This entry provides a basic overview of the Human Development Index over the last decades using the standard HDI methodology of the UNDP.

In addition we are looking at long-term development by relying on the Historical Index of Human

Development (HIHD), developed by historian Leandro Prados de la Escosura.

The metrics of the HDI and HIHD are similar, but differ slightly in how they are used to derive the development index – details on these measures can be found in the Data Quality & Definitions section below.

Link of the data set: https://ourworldindata.org/human-development-index

[24]:
```
#Corruption dataframe
corruption_df = pd.read_csv('Data/average-rating-of-corruption-perception.csv')
corruption_df.rename(columns={'Entity': 'Country'}, inplace=True)
corruption_df.head()
```

[24]:

	Country	Code	Year	Corruption Perception Rating
0	Afghanistan	AFG	2013	4.1
1	Albania	ALB	2013	4.2
2	Algeria	DZA	2013	4.6
3	Argentina	ARG	2013	4.5
4	Armenia	ARM	2013	4.4

Average rating of perceived corruption in public sector, 2013

CORRUPTION PERCEPTION RATING

Variable description: Average of all individuals' perception ratings on a scale from 1 (corruption is not a problem) to 5 (corruption is a very serious problem).

Variable time span: 2013 – 2013

Data published by: Transparency International - Global Corruption Barometer

Data publisher's source: Population surveys

Link of the data set: https://ourworldindata.org/corruption

Meaningful questions that can be answered with these data, make several informative visualisations to answer those questions. Questions that can be answered with these data :

1. What is the correlation between corruption and economic development?

2. How satisfied are people with their lives in different continent? How life satisfaction effect on life expectancy?

3. What is the relationship between the HDI and life expectancy?

4. What is the relationship between Gross Domestic Product (GDP) and the Human Development Index (HDI)?

2.0.1 Question 1: What is the correlation between corruption and economic development?

[25]:
```
#merging
corr_eco = pd.merge(df, corruption_df, on=['Country','Year'])
corr_eco = corr_eco[['Country','GDP per capita','Corruption Perception Rating']]
corr_eco = corr_eco.sort_values(by =['Corruption Perception Rating'],␣
↪ascending=False)
corr_eco.head()
```

[25]:

	Country	GDP per capita	Corruption Perception Rating
59	Mongolia	11545.0	4.8
49	Liberia	900.0	4.8
105	Zimbabwe	1604.0	4.7
75	Russia	24224.0	4.7
57	Mexico	15680.0	4.7

[26]:
```
corr_eco.tail()
```

[26]:

	Country	GDP per capita	Corruption Perception Rating
29	Finland	37246.0	2.9
89	Switzerland	59036.0	2.7
88	Sudan	3451.0	2.6
23	Denmark	43733.0	2.2
76	Rwanda	1554.0	2.0

[27]:
```
#plotting a corruption perception rating graph for the year 2013
import plotly.graph_objs as go
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
init_notebook_mode(connected=True)

data = dict(type = 'choropleth',
           locations = corr_eco['Country'],
           locationmode = 'country names',
           z = corr_eco['Corruption Perception Rating'],
           text = corr_eco['Country'],
           colorbar = {'title':'Corruption Perception Rating'})
layout = dict(title = 'Corruption Perception Rating in 2013',
             geo = dict(showframe = False))
choromap3 = go.Figure(data = [data], layout=layout)
iplot(choromap3)
```

It can be concluded from the graph above there was a high score in Eastern Europe and Russia, and Latin America. Indeed, the country with the most increased corruption Perception Rating is Mongolia, and the lowest corruption Perception Rating is Rwanda.

[28]:
```
#GDP per capita to corruption percentage rating graph
import plotly.express as px
```

```
fig = px.scatter(corr_eco, x='Corruption Perception Rating', y='GDP per␣
 ↪capita', text='Country',title ='Corruption and Economic Growth(GDP) in 2013␣
 ↪over the world')
fig.update_traces(textposition='top center')
fig.update_layout(
    height=1000,
)
fig.show()
```

It is clear to see that there is a curve in the plot of the corruption and GDP per capita. The group of the group of low to middle income countries (GDP per capita below 30k USD per year) is tend to have a high corruption perception rating. While the top 5 highest GDP countries have 3 Europe countries (Norway, Switzerland, Luxemburg) have the corruption less than 3.5. We can assume that for most of low to middle income class countries the corruption scale does affect with the GDP of the country.

2.0.2 Question 2: How satisfied are people with their lives in different continent? How life satisfaction effect on life expectancy?

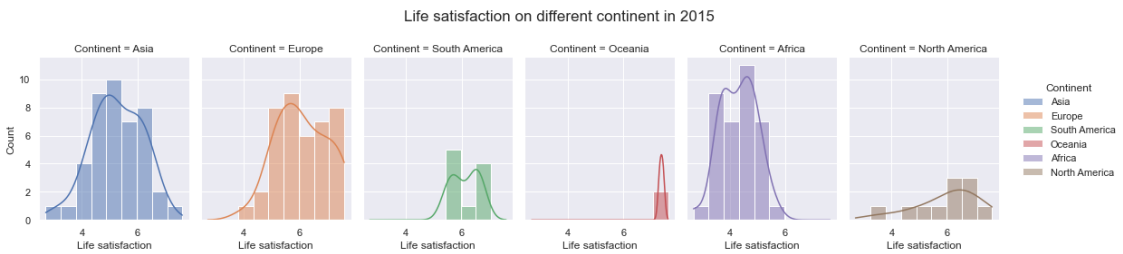
[29]:
```
sat_expe = pd.merge(df, satisfaction_df, on=['Country','Year'])
sat_expe = sat_expe[['Country','Year','Life expectancy', 'Life␣
 ↪satisfaction','Continent']]
sat_expe.head()
```

[29]:

	Country	Year	Life expectancy	Life satisfaction	Continent
0	Afghanistan	2008	59.930	3.724	NaN
1	Afghanistan	2009	60.484	4.402	NaN
2	Afghanistan	2010	61.028	4.758	NaN
3	Afghanistan	2011	61.553	3.832	NaN
4	Afghanistan	2012	62.054	3.783	NaN

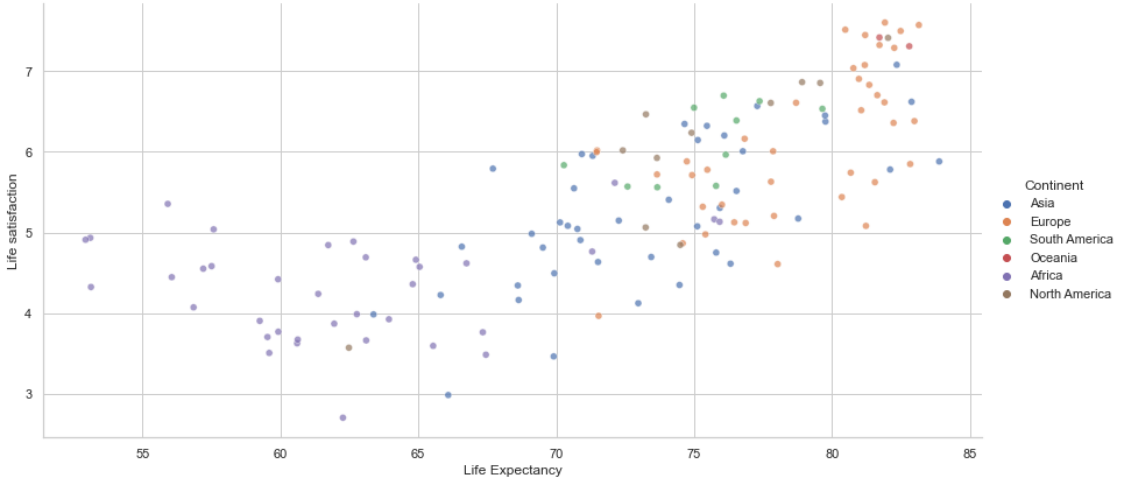
[30]:
```
#taking only the year 2015 because that is the only year where continent is␣
 ↪available
sat_expe_2015 = sat_expe[sat_expe['Year'] == 2015]
```

[31]:
```
plot = sns.displot(data= sat_expe_2015, x='Life satisfaction',kde=True,␣
 ↪hue='Continent', col='Continent', height=3.5, aspect=.75)
plot.fig.suptitle('Life satisfaction on different continent in 2015',y=1.
 ↪1,fontsize = 17)
sns.despine()
```



As we can see from the figure above, Asia had on average higher levels of satisfaction allover the year of 2015, while Oceania and North America had much lower levels of satisfaction compared to Asia.

[32]:
```
sns.set_style('whitegrid')
plot = sns.relplot(data= sat_expe_2015, x='Life expectancy', y='Life␣
↪satisfaction', hue='Continent', alpha=0.7, edgecolors="grey", linewidth=0.5,␣
↪sizes=(30, 250), height=6, aspect=2)
plot.set_axis_labels("Life Expectancy","Life satisfaction");
```



The presented graph ilustrates the correlation between Life satisfaction and Life expectancy on different continents. As can be seen, most of the Africa and Asia countries have lower life expectancy as well as the life satisfaction in 2015. On the other hand, the countries of Europe and South America have higher life expectancy and life satisfaction. In conclusion, there was a strong relationship between Life expectancy and Life satisfaction.

2.0.3 Question 3: What is the relationship between the HDI and life expectancy?

[33]:
```
#Merging HDI and Life expectancy dataframes
HDI_Life = pd.merge(df, HDI_df, on=['Country','Year'])
HDI_Life = HDI_Life[HDI_Life['Year'] == 2016]
HDI_Life.head()
```

[33]:

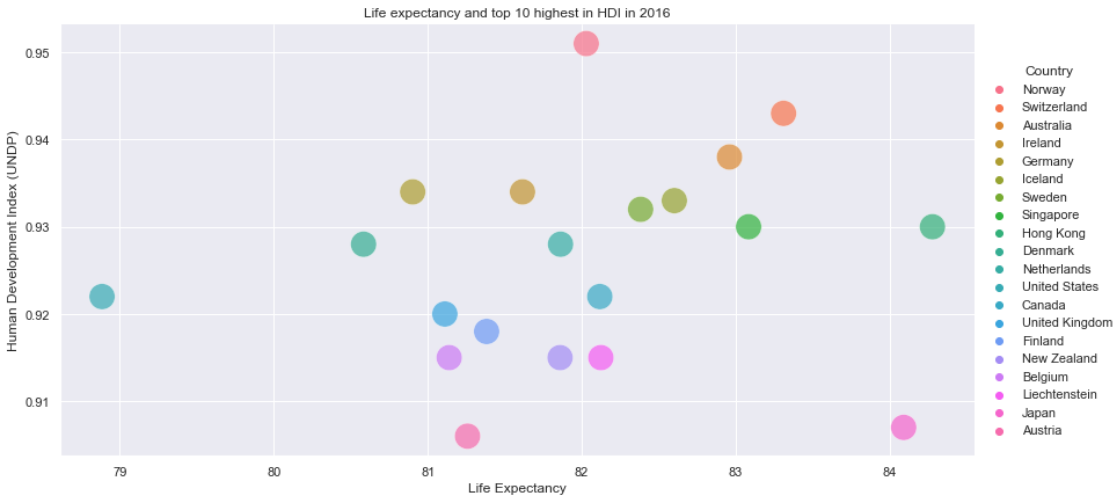
	Country	Code_x	Year	Life expectancy	GDP per capita	\
16	Afghanistan	AFG	2016	63.763	1929.0	
46	Albania	ALB	2016	78.194	10342.0	
74	Algeria	DZA	2016	76.298	14331.0	
92	Andorra	AND	2016	83.274	NaN	
111	Angola	AGO	2016	59.925	8453.0	

	145446-annotations	Population	Continent	Code_y	\
16	NaN	35383028.0	NaN	AFG	
46	NaN	2886427.0	NaN	ALB	
74	NaN	40551398.0	NaN	DZA	
92	NaN	77295.0	NaN	AND	
111	NaN	28842482.0	NaN	AGO	

	Human Development Index (UNDP)
16	0.494
46	0.782
74	0.753
92	0.856
111	0.577

[34]:
```
#Sorting values on HDI
HDI_Life_Temp = HDI_Life.sort_values(by =['Human Development Index (UNDP)'],␣
 ↪ascending=False)
HDI_Life_Temp = HDI_Life_Temp.head(20)
```

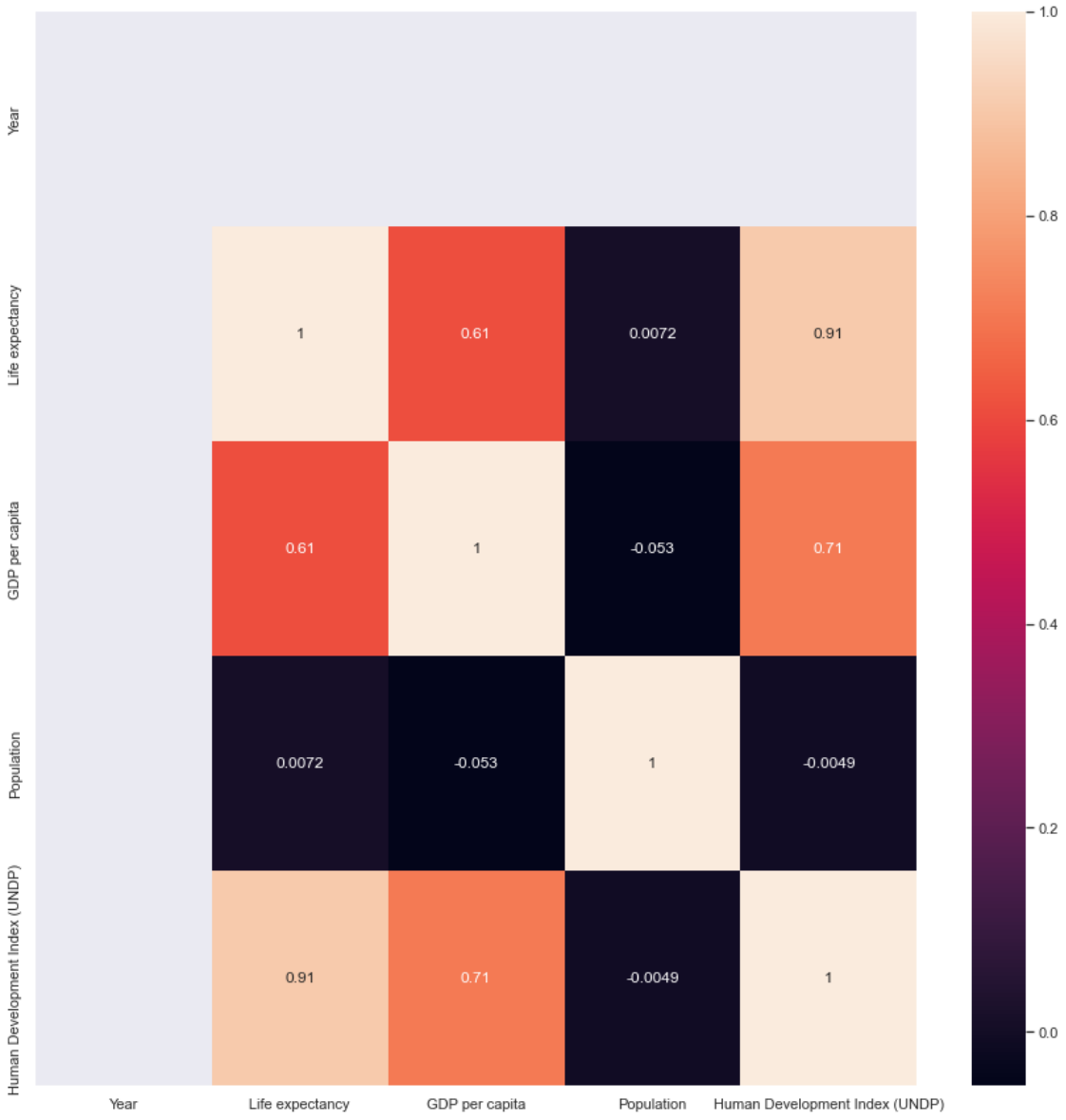
[35]:
```
#Plotting a graph on Life Expectancy against HDI
sns.set_style('darkgrid')
plot = sns.relplot(data = HDI_Life_Temp, x='Life expectancy', y='Human␣
 ↪Development Index (UNDP)', hue='Country', alpha=0.7, edgecolors="grey",␣
 ↪linewidth=0.5, s = 500, height=6, aspect=2).set(title='Life expectancy and␣
 ↪top 10 highest in HDI in 2016')
plot.set_axis_labels("Life Expectancy","Human Development Index (UNDP)");
```



It can be concleded from the graph above that, even Norway have the highest HDI score but the Life expectancy is 82 lower than Janpan which have lower HDI score. Because the sample of the data set we used to plot was very few (top 20 highest Human Development Index Countries over the world in 2016) lead to we hardly see the relationship between HDI and Life expectancy.

We decided to make an corr heat map to check the relationship between those two variables

[36]:
```
#plotting a correlation matrix for comparison
corrMatrixHDI_Life = HDI_Life.corr()
sns.heatmap(corrMatrixHDI_Life, annot=True)
plt.show()
```

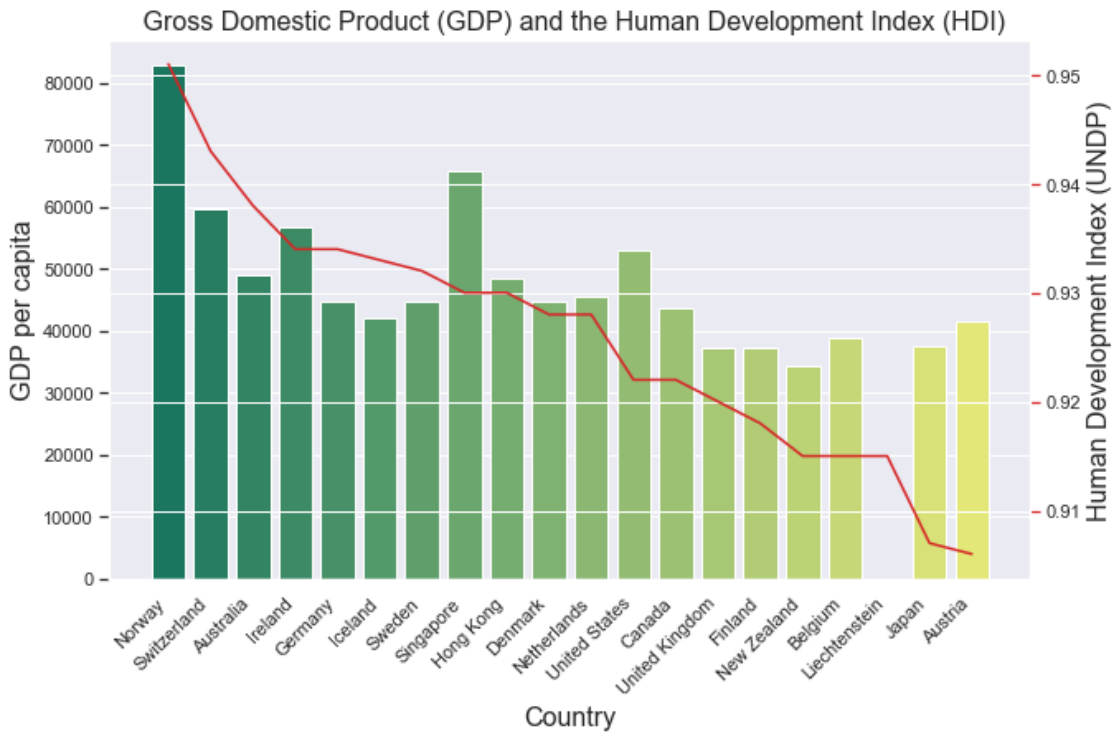


Apparently from the heat map above, the correlation between HDI and Life expectancy is 0.91 which is very high positve linear correlation. To summarise, there is a strong relationship between HDI and Life expectancy

2.0.4 Question 4. What is the relationship between Gross Domestic Product (GDP) and the Human Development Index (HDI)

[37]:
```
#Create combo chart
fig, ax1 = plt.subplots(figsize=(10,6))
color = 'tab:green'
#bar plot creation
```

```
ax1.set_title(' Gross Domestic Product (GDP) and the Human Development Index␣
 ↪(HDI)', fontsize=16)
ax1.set_xlabel('Month', fontsize=16)
ax1.set_ylabel('Avg Temp', fontsize=16)
ax1 = sns.barplot(x='Country', y='GDP per capita', data = HDI_Life_Temp,␣
 ↪palette='summer')
ax1.tick_params(axis='y')
#specify we want to share the same x-axis
ax2 = ax1.twinx()
color = 'tab:red'
#line plot creation
ax2.set_ylabel('Human Development Index (UNDP)', fontsize=16)
ax2 = sns.lineplot(x='Country', y='Human Development Index (UNDP)', data =␣
 ↪HDI_Life_Temp, sort=False, color=color)
ax2.tick_params(axis='y', color=color)
ax1.set_xticklabels(
    ax1.get_xticklabels(),
    rotation=45,
    horizontalalignment='right');
#show plot
plt.show()
```



Here you can see that Norway has the highest GDP and a higher HDI whilst Denmark has a quite average GDP but a lower HDI. So we can say thay HDI is one of the variable that influences the

GDP.

[38]:
```
#Bubble plot on GDP vs HDI(2016)
from plotly.offline import init_notebook_mode, iplot
init_notebook_mode()
from bubbly.bubbly import bubbleplot

figure = bubbleplot(dataset = HDI_Life,x_column='GDP per capita',␣
↪y_column='Human Development Index (UNDP)',
    bubble_column='Country', time_column='Year', size_column='Population',␣
↪color_column='Country',
    x_title="GDP per Capita", y_title="Human Development Index (UNDP)",␣
↪title='GDP per capita vs Human Development Index (UNDP) in 2016',
    x_logscale=True, scale_bubble=3, height=650)

iplot(figure, config={'scrollzoom': True})
```

It can be easily seen that the level of HDI can affect GDP per capita and also on the other way around Economic growth can lead to increase the Human develop index.