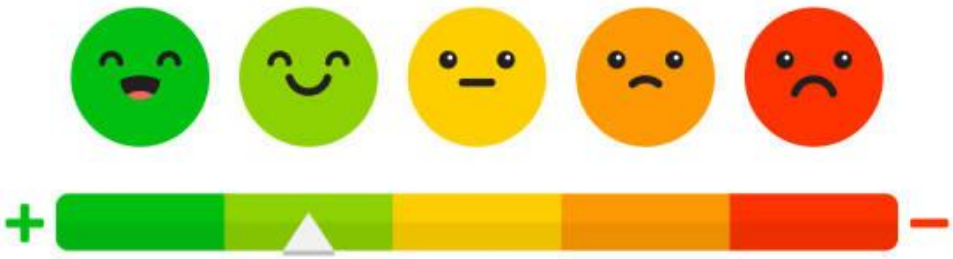
The World Happiness Report dataset

Data Science



MAY 11, 2023

NOUHA LAMRANI
LAM21490091

Table of Contents

Introduction:

One of the most common life objectives of humans all over the world is happiness. The study of happiness started to be done from a social science perspective since the 21st century, people for example who believe they have enough social support report being happier than those who don't, and they also experience less psychological issues, such as eating disorders and mental disease. Various analyses of happiness throughout the years have connected it to several factors, such as:

1. GPD per capita.

2. Social support.

3. Healthy life expectancy.

4. Freedom to make life choices.

5. Generosity.

6. Perception of corruption. [1]

The dataset that I have chosen is the "World Happiness Report dataset", which can be found on the website called "Kaggle" at the following URL: (https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2017.csv). Collaborated by Sustainable Development Solutions Network (Owner), and last updated 3 years ago. The dataset includes data on the happiness levels of different countries around the world, in addition to various factors that contributes to it.

	Characters	Description
Categorical Data	Country or Region	156 countries and region's names all over the world.
Numerical Data	Overall Rank	Rank of the country based on the Happiness Score.
	Score	A metric measured in 2017 by asking the sampled people the question: "How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest". [2]
	GDP per capita	The extent to which GDP (Gross Domestic Product) contributes to the calculation of the Happiness Score.
	Social support	Determines if family or friends would support the individual in case of facing problems.
	Healthy life expectancy	Data from the World Health Organisation (WHO) are used to calculate the healthy life expectancy at birth.
	Freedom to make life choices	The level of satisfaction of a person with having the freedom to select what to do in life.
	Generosity	The residual after regressing the national mean of answers to the question "Have you donated money to a charity in the last month?" .
	Perceptions of corruption	an index that ranks nations based on their perceived levels of corruption in the public sector.

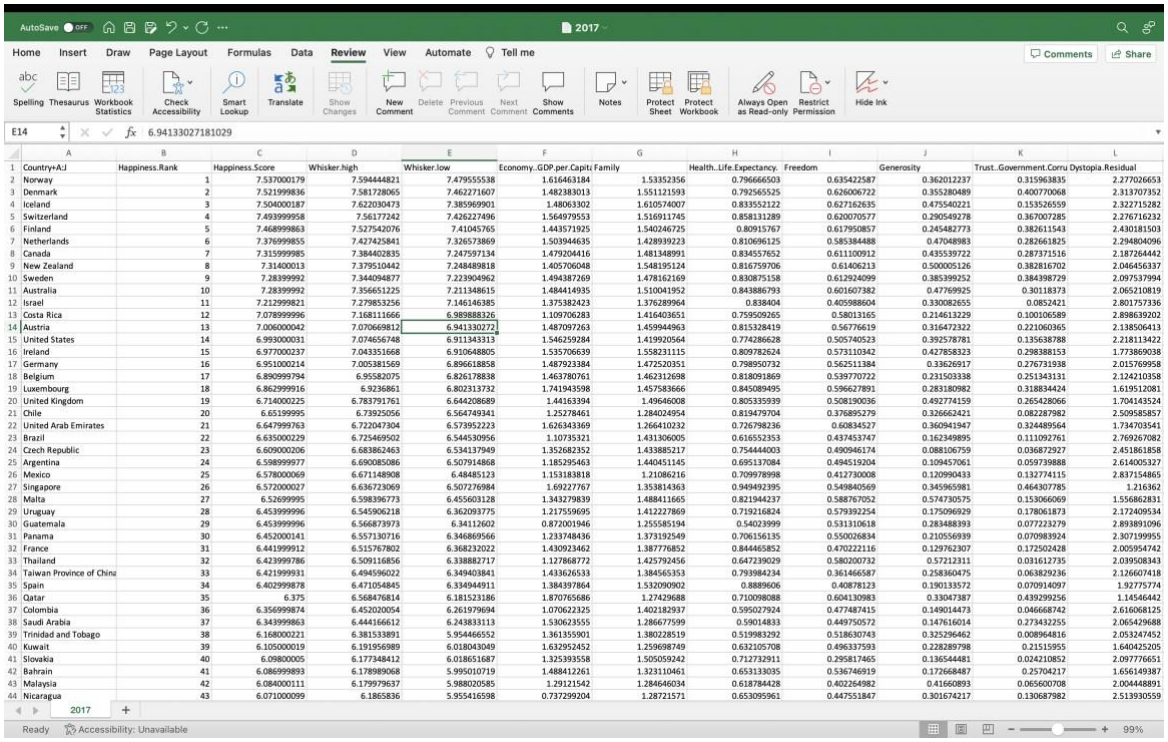
Preprocessing:

Manually:

The preprocessing steps that I have applied for making the data fittingly utilisable for my purpose are as follows:

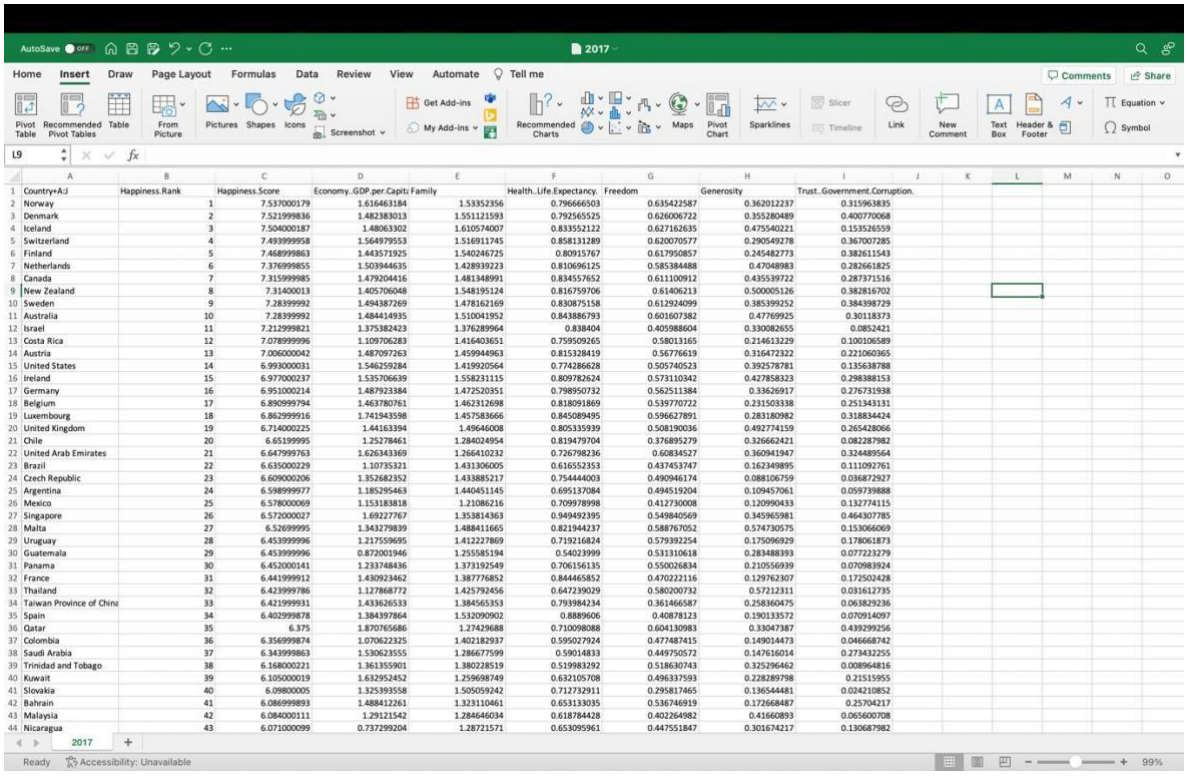
1. Removed columns that are not applicable for my analysis: I removed the columns on Whisker.high, Whisker.low, and Dystopia.Residual, as I will not be using these in my analysis.

Before:



	A	B	C	D	E	F	G	H	I	J	K	L
1	Country+A:J	Happiness.Rank	Happiness.Score	Whisker.high	Whisker.low	Economy..GDP.per.Capita	Family	Health..Life.Expectancy.	Freedom	Generosity	Trust..Government.Corru	Dystopia.Residual
2	Norway	1	7.537000179	7.594444821	7.479555538	1.616463184	1.53352356	0.796666503	0.635422587	0.362012237	0.315963835	2.277026653
3	Denmark	2	7.521999836	7.581728065	7.462271607	1.482383013	1.551121593	0.792565525	0.626006722	0.355280489	0.400770068	2.313707352
4	Iceland	3	7.504000187	7.622030473	7.385969901	1.48063302	1.610574007	0.833552122	0.627162635	0.475540221	0.153526559	2.322715282
5	Switzerland	4	7.493999958	7.56177242	7.426227496	1.564979553	1.516911745	0.858131289	0.620070577	0.290549278	0.367007285	2.276716232
6	Finland	5	7.468999863	7.527542076	7.41045765	1.443571925	1.540246725	0.80915767	0.617950857	0.245482773	0.382611543	2.430181503
7	Netherlands	6	7.376999855	7.427425841	7.326573869	1.503944635	1.428939223	0.810696125	0.585384488	0.47048983	0.282661825	2.294804096
8	Canada	7	7.315999985	7.384402835	7.247597134	1.479204416	1.481348991	0.834557652	0.611100912	0.435539722	0.287371516	2.187264442
9	New Zealand	8	7.31400013	7.379510442	7.248489818	1.405706048	1.548195124	0.816759706	0.61406213	0.500005126	0.382816702	2.046456337
10	Sweden	9	7.28399992	7.344094877	7.223904962	1.494387269	1.478162169	0.830875158	0.612924099	0.385399252	0.384398729	2.097537994
11	Australia	10	7.28399992	7.356651225	7.211348615	1.484414935	1.510041952	0.843886793	0.601607382	0.47769925	0.30118373	2.065210819
12	Israel	11	7.212999821	7.279853256	7.146146385	1.375382423	1.376289964	0.838404	0.405988604	0.330082655	0.0852421	2.801757336
13	Costa Rica	12	7.078999996	7.168111666	6.989888326	1.109706283	1.416403651	0.759509265	0.58013165	0.214613229	0.100106589	2.898639202
14	Austria	13	7.006000042	7.070669812	6.941330272	1.487097263	1.459944963	0.815328419	0.56776619	0.316472322	0.221060365	2.138506413
15	United States	14	6.993000031	7.074656748	6.911343313	1.546259284	1.419920564	0.774286628	0.505740523	0.392578781	0.135638788	2.218113422
16	Ireland	15	6.977000237	7.043351668	6.910648805	1.535706639	1.558231115	0.809782624	0.573110342	0.427858323	0.298388153	1.773869038
17	Germany	16	6.951000214	7.005381569	6.896618858	1.487923384	1.472520351	0.798950732	0.562511384	0.33626917	0.276731938	2.015769958
18	Belgium	17	6.890999794	6.95582075	6.826178838	1.463780761	1.462312698	0.818091869	0.539770722	0.231503338	0.251343131	2.124210358
19	Luxembourg	18	6.862999916	6.9236861	6.802313732	1.741943598	1.457583666	0.845089495	0.596627891	0.283180982	0.318834424	1.619512081
20	United Kingdom	19	6.714000225	6.783791761	6.644208689	1.44163394	1.49646008	0.805335939	0.508190036	0.492774159	0.265428066	1.704143524
21	Chile	20	6.65199995	6.73925056	6.564749341	1.25278461	1.284024954	0.819479704	0.376895279	0.326662421	0.082287982	2.509585857
22	United Arab Emirates	21	6.647999763	6.722047304	6.573952223	1.626343369	1.266410232	0.726798236	0.60834527	0.360941947	0.324489564	1.734703541
23	Brazil	22	6.635000229	6.725469502	6.544530956	1.10735321	1.431306005	0.616552353	0.437453747	0.162349895	0.111092761	2.769267082
24	Czech Republic	23	6.609000206	6.683862463	6.534137949	1.352682352	1.433885217	0.754444003	0.490946174	0.088106759	0.036872927	2.451861858
25	Argentina	24	6.598999977	6.690085086	6.507914868	1.185295463	1.440451145	0.695137084	0.494519204	0.109457061	0.059739888	2.614005327
26	Mexico	25	6.578000069	6.671148908	6.48485123	1.153183818	1.21086216	0.709978998	0.412730008	0.120990433	0.132774115	2.837154865
27	Singapore	26	6.572000027	6.636723069	6.507276984	1.69227767	1.353814363	0.949492395	0.549840569	0.345965981	0.464307785	1.216362
28	Malta	27	6.52699995	6.598396773	6.455603128	1.343279839	1.488411665	0.821944237	0.588767052	0.574730575	0.153066069	1.556862831
29	Uruguay	28	6.453999996	6.545906218	6.362093775	1.217559695	1.412227869	0.719216824	0.579392254	0.175096929	0.178061873	2.172409534
30	Guatemala	29	6.453999996	6.566873973	6.34112602	0.872001946	1.255585194	0.54023999	0.531310618	0.283488393	0.077232379	2.893891096
31	Panama	30	6.452000141	6.557130716	6.346869566	1.233748436	1.373192549	0.706156135	0.550026834	0.210556939	0.070983924	2.307199955
32	France	31	6.441999912	6.515767802	6.368232022	1.430923462	1.387776852	0.844465852	0.470222116	0.129762307	0.172502428	2.005954742
33	Thailand	32	6.423999786	6.509116856	6.338882717	1.127868772	1.425792456	0.647239029	0.580200732	0.57212311	0.031612735	2.039508343
34	Taiwan Province of China	33	6.421999931	6.494596022	6.349403841	1.433626533	1.384565353	0.793984234	0.361466587	0.258360475	0.063829236	2.126607418
35	Spain	34	6.402999878	6.471054845	6.334944911	1.384397864	1.532090902	0.8889606	0.40878123	0.190133572	0.070914097	1.92775774
36	Qatar	35	6.375	6.568476814	6.181523186	1.870765686	1.27429688	0.710098088	0.604130983	0.33047387	0.439299256	1.14546442
37	Colombia	36	6.356999874	6.452020054	6.261979694	1.070622325	1.402182937	0.595027924	0.477487415	0.149014473	0.046668742	2.616068125
38	Saudi Arabia	37	6.343999863	6.444166612	6.243833113	1.530623555	1.286677599	0.59014833	0.449750572	0.147616014	0.273432255	2.065429688
39	Trinidad and Tobago	38	6.168000221	6.381533891	5.954466552	1.361355901	1.380228519	0.519983292	0.518630743	0.325296462	0.008964816	2.053247452
40	Kuwait	39	6.105000019	6.191956989	6.018043049	1.632952452	1.259698749	0.632105708	0.496337593	0.228289798	0.21515955	1.640425205
41	Slovakia	40	6.09800005	6.177348412	6.018651687	1.325393558	1.505059242	0.712732911	0.295817465	0.136544481	0.024210852	2.097776651
42	Bahrain	41	6.086999893	6.178989068	5.995010719	1.488412261	1.323110461	0.653133035	0.536746919	0.172668487	0.25704217	1.656149387
43	Malaysia	42	6.084000111	6.179979637	5.988020585	1.29121542	1.284646034	0.618784428	0.402264982	0.41660893	0.065600708	2.004448891
44	Nicaragua	43	6.071000099	6.1865836	5.955416598	0.737299204	1.28721571	0.653095961	0.447551847	0.301674217	0.130687982	2.513930559

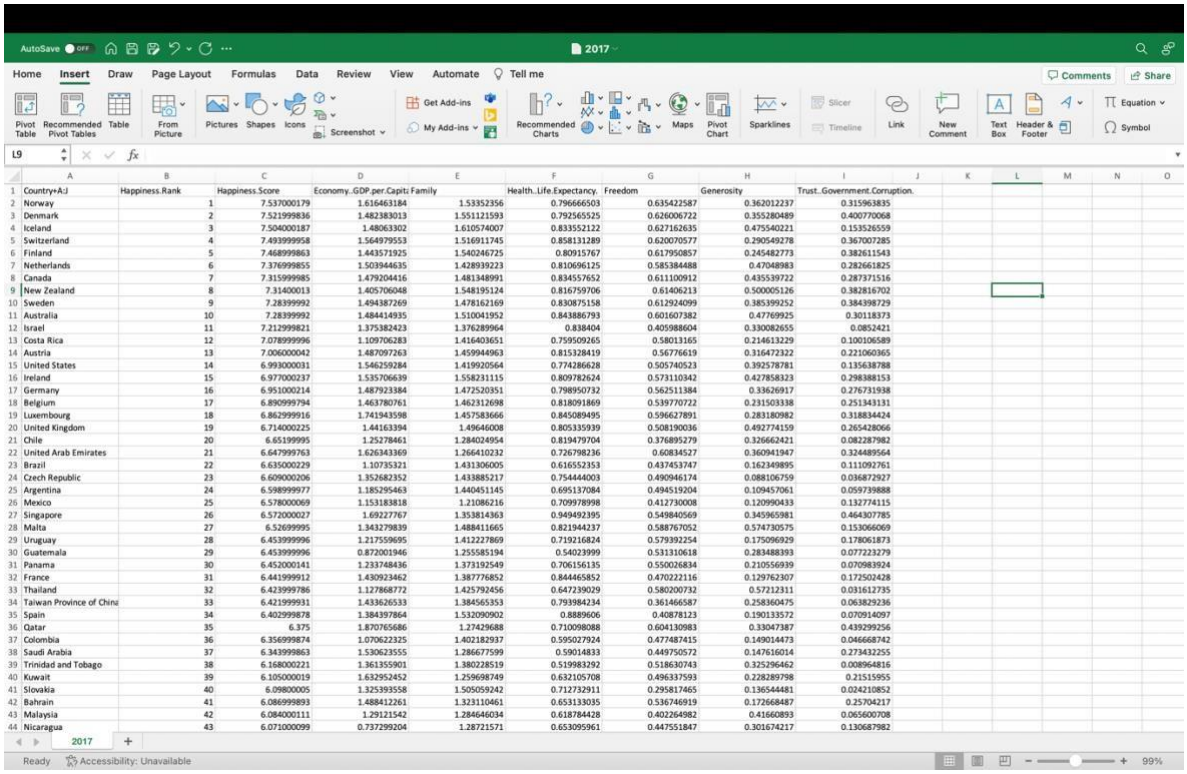
After:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Country+A:J	Happiness.Rank	Happiness.Score	Economy..GDP.per.Capita	Family	Health..Life.Expectancy.	Freedom	Generosity	Trust..Government.Corruption.						
2	Norway	1	7.537000179	1.616463184	1.53352356	0.796666503	0.635422587	0.362012237	0.315963835						
3	Denmark	2	7.521999836	1.482383013	1.551121593	0.792565525	0.626006722	0.355280489	0.400770068						
4	Iceland	3	7.504000187	1.48063302	1.610574007	0.833552122	0.627162635	0.475540221	0.153526559						
5	Switzerland	4	7.493999958	1.564979553	1.516911745	0.858131289	0.620070577	0.290549278	0.367007285						
6	Finland	5	7.468999863	1.443571925	1.540246725	0.80915767	0.617950857	0.245482773	0.382611543						
7	Netherlands	6	7.376999855	1.503944635	1.428939223	0.810696125	0.585384488	0.47048983	0.282661825						
8	Canada	7	7.315999985	1.479204416	1.481348991	0.834557652	0.611100912	0.435539722	0.287371516						
9	New Zealand	8	7.31400013	1.405706048	1.548195124	0.816759706	0.61406213	0.500005126	0.382816702						
10	Sweden	9	7.28399992	1.494387269	1.478162169	0.830875158	0.612924099	0.385399252	0.384398729						
11	Australia	10	7.28399992	1.484414935	1.510041952	0.843886793	0.601607382	0.47769925	0.30118373						
12	Israel	11	7.212999821	1.375382423	1.376289964	0.838404	0.405988604	0.330082655	0.0852421						
13	Costa Rica	12	7.078999996	1.109706283	1.416403651	0.759509265	0.58013165	0.214613229	0.100106589						
14	Austria	13	7.006000042	1.487097263	1.459944963	0.815328419	0.56776619	0.316472322	0.221060365						
15	United States	14	6.993000031	1.546259284	1.419920564	0.774286628	0.505740523	0.392578781	0.135638788						
16	Ireland	15	6.977000237	1.535706639	1.558231115	0.809782624	0.573110342	0.427858323	0.298388153						
17	Germany	16	6.951000214	1.487923384	1.472520351	0.798950732	0.562511384	0.33626917	0.276731938						
18	Belgium	17	6.890999794	1.463780761	1.462312698	0.818091869	0.539770722	0.231503338	0.251343131						
19	Luxembourg	18	6.862999916	1.741943598	1.457583666	0.845089495	0.596627891	0.283180982	0.318834424						
20	United Kingdom	19	6.714000225	1.44163394	1.49646008	0.805335939	0.508190036	0.492774159	0.265428066						
21	Chile	20	6.65199995	1.25278461	1.284024954	0.819479704	0.376895279	0.326662421	0.082287982						
22	United Arab Emirates	21	6.647999763	1.626343369	1.266410232	0.726798236	0.60834527	0.360941947	0.324489564						
23	Brazil	22	6.635000229	1.10735321	1.431306005	0.616552353	0.437453747	0.162349895	0.111092761						
24	Czech Republic	23	6.609000206	1.352682352	1.433885217	0.754444003	0.490946174	0.088106759	0.036872927						
25	Argentina	24	6.598999977	1.185295463	1.440451145	0.695137084	0.494519204	0.109457061	0.059739888						
26	Mexico	25	6.578000069	1.153183818	1.21086216	0.709978998	0.412730008	0.120990433	0.132774115						
27	Singapore	26	6.572000027	1.69227767	1.353814363	0.949492395	0.549840569	0.345965981	0.464307785						
28	Malta	27	6.52699995	1.343279839	1.488411665	0.821944237	0.588767052	0.574730575	0.153066069						
29	Uruguay	28	6.453999996	1.212559695	1.412227869	0.719216824	0.579392254	0.175096929	0.178061873						
30	Guatemala	29	6.453999996	0.872001946	1.255585194	0.54023999	0.531310618	0.283488393	0.077232379						
31	Panama	30	6.452000141	1.233748436	1.373192549	0.706156135	0.550026834	0.210556939	0.070983924						
32	France	31	6.441999912	1.430923462	1.387776852	0.844465852	0.470222116	0.129762307	0.172502428						
33	Thailand	32	6.423999786	1.127868772	1.425792456	0.647239029	0.580200732	0.57212311	0.031612735						
34	Taiwan Province of China	33	6.421999931	1.433626533	1.384565353	0.793984234	0.361466587	0.258360475	0.063829236						
35	Spain	34	6.402999878	1.384397864	1.532090902	0.8889606	0.40878123	0.190133572	0.070914097						
36	Qatar	35	6.375	1.870765686	1.27429688	0.710098088	0.604130983	0.33047387	0.439299256						
37	Colombia	36	6.356999874	1.070622325	1.402182937	0.595027924	0.477487415	0.149014473	0.046668742						
38	Saudi Arabia	37	6.343999863	1.530623555	1.286677599	0.59014833	0.449750572	0.147616014	0.273432255						
39	Trinidad and Tobago	38	6.168000221	1.361355901	1.380228519	0.519983292	0.518630743	0.325296462	0.008964816						
40	Kuwait	39	6.105000019	1.632952452	1.259698749	0.632105708	0.496337593	0.228289798	0.21515955						
41	Slovakia	40	6.09800005	1.325393558	1.505059242	0.712732911	0.295817465	0.136544481	0.024210852						
42	Bahrain	41	6.086999893	1.488412261	1.323110461	0.653133035	0.536746919	0.172668487	0.25704217						
43	Malaysia	42	6.084000111	1.29121542	1.284646034	0.618784428	0.402264982	0.41660893	0.065600708						
44	Nicaragua	43	6.071000099	0.737299204	1.28721571	0.653095961	0.447551847	0.301674217	0.130687982						

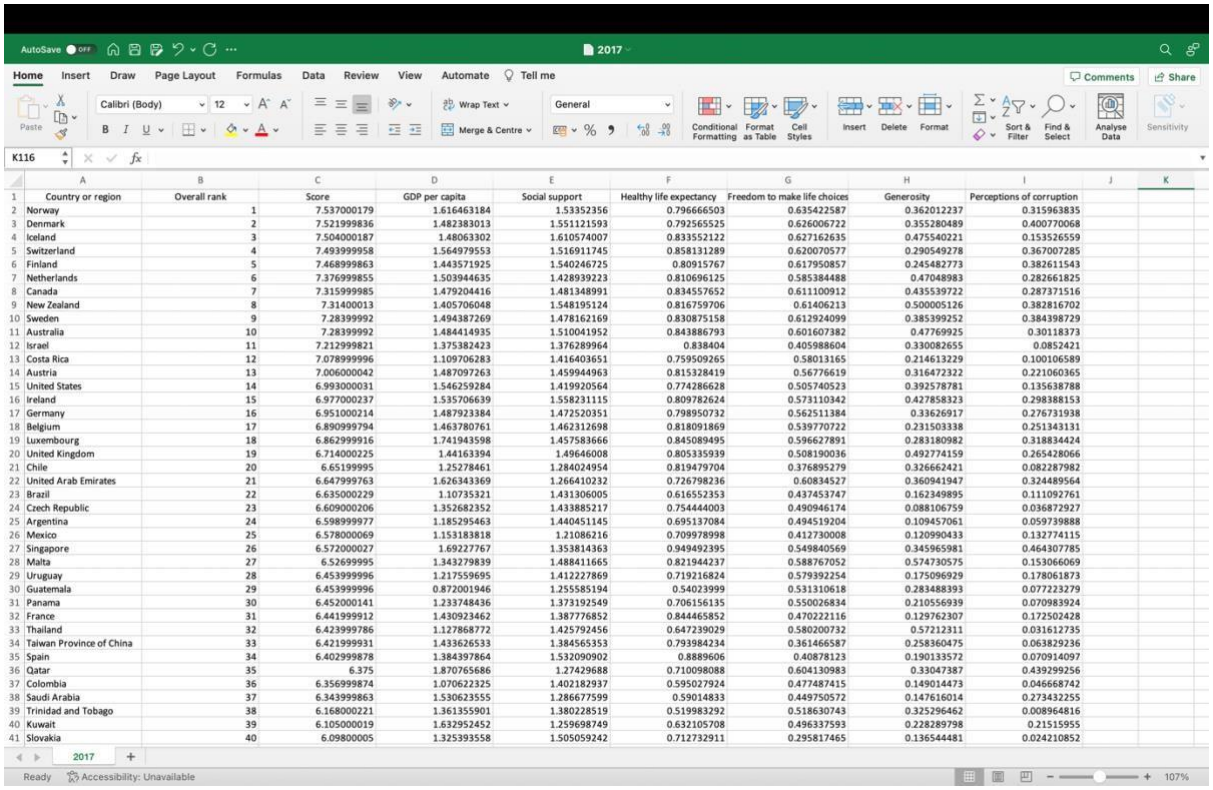
2. Renamed the columns for better legibility: I renamed the columns as some of them are difficult to understand and not very clear. This way I made them more descriptive and easier to read.

Before:



	A	B	C	D	E	F	G	H	I
1	Country+A:J	Happiness.Rank	Happiness.Score	Economy..GDP.per.Capita	Family	Health..Life.Expectancy.	Freedom	Generosity	Trust..Government.Corruption.
2	Norway	1	7.537000179	1.616463184	1.53352356	0.796666503	0.635422587	0.362012237	0.315963835
3	Denmark	2	7.521999836	1.482383013	1.551121593	0.792565525	0.626006722	0.355280489	0.400770068
4	Iceland	3	7.504000187	1.48063302	1.610574007	0.833552122	0.627162635	0.475540221	0.153526559
5	Switzerland	4	7.493999958	1.564979553	1.516911745	0.858131289	0.620070577	0.290549278	0.367007285
6	Finland	5	7.468999863	1.443571925	1.540246725	0.80915767	0.617950857	0.245482773	0.382611543
7	Netherlands	6	7.376999855	1.503944635	1.428939223	0.810696125	0.585384488	0.47048983	0.282661825
8	Canada	7	7.315999985	1.479204416	1.481348991	0.834557652	0.611100912	0.435539722	0.287371516
9	New Zealand	8	7.31400013	1.405706048	1.548195124	0.816759706	0.61406213	0.500005126	0.382816702
10	Sweden	9	7.28399992	1.494387269	1.478162169	0.830875158	0.612924099	0.385399252	0.384398729
11	Australia	10	7.28399992	1.484414935	1.510041952	0.843886793	0.601607382	0.47769925	0.30118373
12	Israel	11	7.212999821	1.375382423	1.376289964	0.838404	0.405988604	0.330082655	0.0852421
13	Costa Rica	12	7.078999996	1.109706283	1.416403651	0.759509265	0.58013165	0.214613229	0.100106589
14	Austria	13	7.006000042	1.487097263	1.459944963	0.815328419	0.56776619	0.316472322	0.221060365
15	United States	14	6.993000031	1.546259284	1.419920564	0.774286628	0.505740523	0.392578781	0.135638788
16	Ireland	15	6.977000237	1.535706639	1.558231115	0.809782624	0.573110342	0.427858323	0.296388153
17	Germany	16	6.951000214	1.487923384	1.472520351	0.798950732	0.562511384	0.33626917	0.276731938
18	Belgium	17	6.890999794	1.463780761	1.462312698	0.818091869	0.539770722	0.231503338	0.251343131
19	Luxembourg	18	6.862999916	1.741943598	1.457583666	0.845089495	0.596627891	0.283180982	0.318834424
20	United Kingdom	19	6.714000225	1.44163394	1.49646008	0.805335939	0.508190036	0.492774159	0.265428066
21	Chile	20	6.65199995	1.25278461	1.284024954	0.819479704	0.376895279	0.326662421	0.082287982
22	United Arab Emirates	21	6.647999763	1.626343369	1.266410232	0.726798236	0.60834527	0.360941947	0.324489564
23	Brazil	22	6.635000229	1.10735321	1.431306005	0.616552353	0.437453747	0.162349895	0.111092761
24	Czech Republic	23	6.609000206	1.352682352	1.433885217	0.754444003	0.490946174	0.088106759	0.036872927
25	Argentina	24	6.598999977	1.185295463	1.440451145	0.695137084	0.494519204	0.109457061	0.059739888
26	Mexico	25	6.578000069	1.153183818	1.21086216	0.709978998	0.412730008	0.120990433	0.132774115
27	Singapore	26	6.572000027	1.69227767	1.353814363	0.949492395	0.549840569	0.345965981	0.464307785
28	Malta	27	6.52699995	1.343279839	1.488411665	0.821944237	0.588767052	0.574730575	0.153066069
29	Uruguay	28	6.453999996	1.217559695	1.412227869	0.719216824	0.579392254	0.175096929	0.178061873
30	Guatemala	29	6.453999996	0.872001946	1.255585194	0.54023999	0.531310618	0.283488393	0.077223279
31	Panama	30	6.452000141	1.233748436	1.373192549	0.706156135	0.550026834	0.210556939	0.070983924
32	France	31	6.441999912	1.430923462	1.387776852	0.844465852	0.470222116	0.129762307	0.172502428
33	Thailand	32	6.423999786	1.127868772	1.425792456	0.647239029	0.580200732	0.57212311	0.031612735
34	Taiwan Province of China	33	6.421999931	1.433626533	1.384565353	0.793984234	0.361466587	0.258360475	0.063829236
35	Spain	34	6.402999878	1.384397864	1.532090902	0.8889606	0.40878123	0.190133572	0.070914097
36	Qatar	35	6.375	1.870765686	1.27429688	0.710098088	0.604130983	0.33047387	0.439299256
37	Colombia	36	6.356999874	1.070622325	1.402182937	0.595027924	0.477487415	0.149014473	0.046668742
38	Saudi Arabia	37	6.343999863	1.530623555	1.286677599	0.59014833	0.449750572	0.147616014	0.273432255
39	Trinidad and Tobago	38	6.168000221	1.361355901	1.380228519	0.519983292	0.518630743	0.325296462	0.008964816
40	Kuwait	39	6.105000019	1.632952452	1.259698749	0.632105708	0.496337593	0.228289798	0.21515955
41	Slovakia	40	6.09800005	1.325393558	1.505059242	0.712732911	0.295817465	0.136544481	0.024210852
42	Bahrain	41	6.086999893	1.488412261	1.323110461	0.653133035	0.536746919	0.172668487	0.25704217
43	Malaysia	42	6.084000111	1.29121542	1.284646034	0.618784428	0.402264982	0.41660893	0.065600708
44	Nicaragua	43	6.071000099	0.737299204	1.28721571	0.653095961	0.447551847	0.301674217	0.130687982

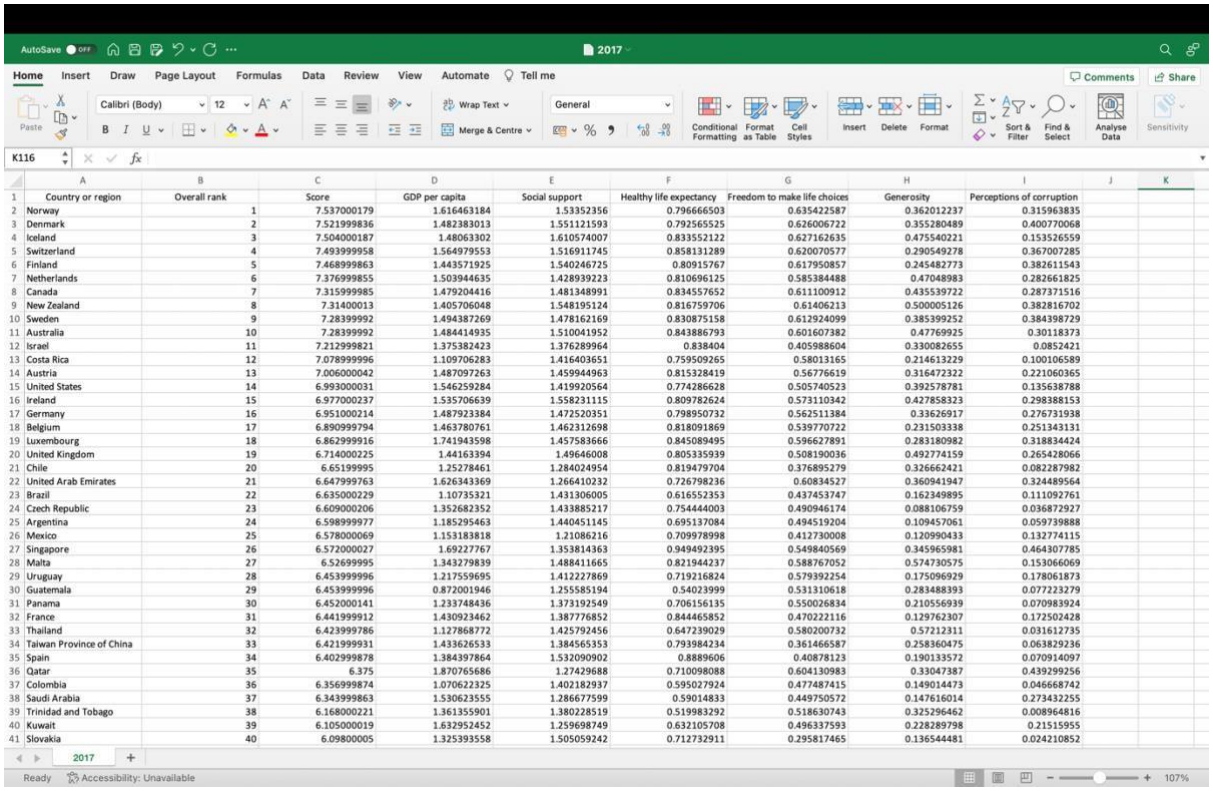
After:



	A	B	C	D	E	F	G	H	I	J	K
1	Country or region	Overall rank	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption		
2	Norway	1	7.537000179	1.616463184	1.53352356	0.796666503	0.635422587	0.362012237	0.315963835		
3	Denmark	2	7.521999836	1.482383013	1.551121593	0.792565525	0.626006722	0.355280489	0.400770068		
4	Iceland	3	7.504000187	1.48063302	1.610574007	0.833552122	0.627162635	0.475540221	0.153526559		
5	Switzerland	4	7.493999958	1.564979553	1.516911745	0.858131289	0.620070577	0.290549278	0.367007285		
6	Finland	5	7.468999863	1.443571925	1.540246725	0.80915767	0.617950857	0.245482773	0.382611543		
7	Netherlands	6	7.376999855	1.503944635	1.428939223	0.810696125	0.585384488	0.47048983	0.282661825		
8	Canada	7	7.315999985	1.479204416	1.481348991	0.834557652	0.611100912	0.435539722	0.287371516		
9	New Zealand	8	7.31400013	1.405706048	1.548195124	0.816759706	0.61406213	0.500005126	0.382816702		
10	Sweden	9	7.28399992	1.494387269	1.478162169	0.830875158	0.612924099	0.385399252	0.384398729		
11	Australia	10	7.28399992	1.484414935	1.510041952	0.843886793	0.601607382	0.47769925	0.30118373		
12	Israel	11	7.212999821	1.375382423	1.376289964	0.838404	0.405988604	0.330082655	0.0852421		
13	Costa Rica	12	7.078999996	1.109706283	1.416403651	0.759509265	0.58013165	0.214613229	0.100106589		
14	Austria	13	7.006000042	1.487097263	1.459944963	0.815328419	0.56776619	0.316472322	0.221060365		
15	United States	14	6.993000031	1.546259284	1.419920564	0.774286628	0.505740523	0.392578781	0.135638788		
16	Ireland	15	6.977000237	1.535706639	1.558231115	0.809782624	0.573110342	0.427858323	0.296388153		
17	Germany	16	6.951000214	1.487923384	1.472520351	0.798950732	0.562511384	0.33626917	0.276731938		
18	Belgium	17	6.890999794	1.463780761	1.462312698	0.818091869	0.539770722	0.231503338	0.251343131		
19	Luxembourg	18	6.862999916	1.741943598	1.457583666	0.845089495	0.596627891	0.283180982	0.318834424		
20	United Kingdom	19	6.714000225	1.44163394	1.49646008	0.805335939	0.508190036	0.492774159	0.265428066		
21	Chile	20	6.65199995	1.25278461	1.284024954	0.819479704	0.376895279	0.326662421	0.082287982		
22	United Arab Emirates	21	6.647999763	1.626343369	1.266410232	0.726798236	0.60834527	0.360941947	0.324489564		
23	Brazil	22	6.635000229	1.10735321	1.431306005	0.616552353	0.437453747	0.162349895	0.111092761		
24	Czech Republic	23	6.609000206	1.352682352	1.433885217	0.754444003	0.490946174	0.088106759	0.036872927		
25	Argentina	24	6.598999977	1.185295463	1.440451145	0.695137084	0.494519204	0.109457061	0.059739888		
26	Mexico	25	6.578000069	1.153183818	1.21086216	0.709978998	0.412730008	0.120990433	0.132774115		
27	Singapore	26	6.572000027	1.69227767	1.353814363	0.949492395	0.549840569	0.345965981	0.464307785		
28	Malta	27	6.52699995	1.343279839	1.488411665	0.821944237	0.588767052	0.574730575	0.153066069		
29	Uruguay	28	6.453999996	1.217559695	1.412227869	0.719216824	0.579392254	0.175096929	0.178061873		
30	Guatemala	29	6.453999996	0.872001946	1.255585194	0.54023999	0.531310618	0.283488393	0.077223279		
31	Panama	30	6.452000141	1.233748436	1.373192549	0.706156135	0.550026834	0.210556939	0.070983924		
32	France	31	6.441999912	1.430923462	1.387776852	0.844465852	0.470222116	0.129762307	0.172502428		
33	Thailand	32	6.423999786	1.127868772	1.425792456	0.647239029	0.580200732	0.57212311	0.031612735		
34	Taiwan Province of China	33	6.421999931	1.433626533	1.384565353	0.793984234	0.361466587	0.258360475	0.063829236		
35	Spain	34	6.402999878	1.384397864	1.532090902	0.8889606	0.40878123	0.190133572	0.070914097		
36	Qatar	35	6.375	1.870765686	1.27429688	0.710098088	0.604130983	0.33047387	0.439299256		
37	Colombia	36	6.356999874	1.070622325	1.402182937	0.595027924	0.477487415	0.149014473	0.046668742		
38	Saudi Arabia	37	6.343999863	1.530623555	1.286677599	0.59014833	0.449750572	0.147616014	0.273432255		
39	Trinidad and Tobago	38	6.168000221	1.361355901	1.380228519	0.519983292	0.518630743	0.325296462	0.008964816		
40	Kuwait	39	6.105000019	1.632952452	1.259698749	0.632105708	0.496337593	0.228289798	0.21515955		
41	Slovakia	40	6.09800005	1.325393558	1.505059242	0.712732911	0.295817465	0.136544481	0.024210852		

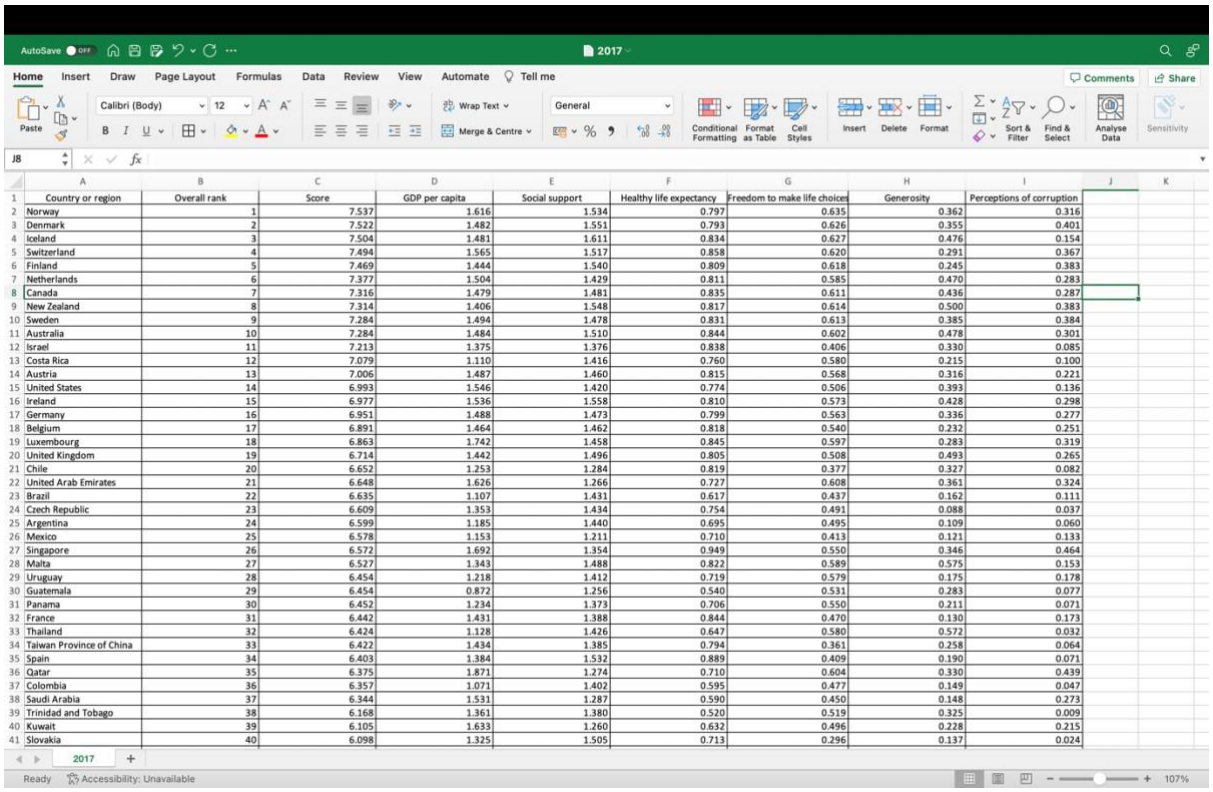
3. Change the values format and handle the missing values: I checked for any missing values and made of the values only 3 number after the dot.

Before:



	A	B	C	D	E	F	G	H	I	J	K
1	Country or region	Overall rank	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption		
2	Norway	1	7.537000179	1.616463184	1.53352356	0.796666503	0.635422587	0.362012237	0.315963835		
3	Denmark	2	7.521999836	1.482383013	1.551121593	0.792565525	0.626006722	0.355280489	0.400770068		
4	Iceland	3	7.504000187	1.48063302	1.610574007	0.833552122	0.627162635	0.475540221	0.153526559		
5	Switzerland	4	7.493999958	1.564979553	1.516911745	0.858131289	0.620070577	0.290549278	0.367007285		
6	Finland	5	7.468999863	1.443571925	1.540246725	0.80915767	0.617950857	0.245482773	0.382611543		
7	Netherlands	6	7.376999855	1.503944635	1.428939223	0.810696125	0.585384488	0.47048983	0.282661825		
8	Canada	7	7.315999985	1.479204416	1.481348991	0.834557652	0.611100912	0.435539722	0.287371516		
9	New Zealand	8	7.31400013	1.405706048	1.548195124	0.816759706	0.61406213	0.500005126	0.382816702		
10	Sweden	9	7.28399992	1.494387269	1.478162169	0.830875158	0.612924099	0.385399252	0.384398729		
11	Australia	10	7.28399992	1.484414935	1.510041952	0.843886793	0.601607382	0.47769925	0.30118373		
12	Israel	11	7.212999821	1.375382423	1.376289964	0.838404	0.405988604	0.330082655	0.0852421		
13	Costa Rica	12	7.078999996	1.109706283	1.416403651	0.759509265	0.58013165	0.214613229	0.100106589		
14	Austria	13	7.006000042	1.487097263	1.459944963	0.815328419	0.56776619	0.316472322	0.221060365		
15	United States	14	6.993000031	1.546259284	1.419920564	0.774286628	0.505740523	0.392578781	0.135638788		
16	Ireland	15	6.977000237	1.535706639	1.558231115	0.809782624	0.573110342	0.427858323	0.298388153		
17	Germany	16	6.951000214	1.487923384	1.472520351	0.798950732	0.562511384	0.33626917	0.276731938		
18	Belgium	17	6.890999794	1.463780761	1.462312698	0.818091869	0.539770722	0.231503338	0.251343131		
19	Luxembourg	18	6.862999916	1.741943598	1.457583666	0.845089495	0.596627891	0.283180982	0.318834424		
20	United Kingdom	19	6.714000225	1.44163394	1.49646008	0.805335939	0.508190036	0.492774159	0.265428066		
21	Chile	20	6.65199995	1.25278461	1.284024954	0.819479704	0.376895279	0.326662421	0.082287982		
22	United Arab Emirates	21	6.647999763	1.626343369	1.266410232	0.726798236	0.60834527	0.360941947	0.324489564		
23	Brazil	22	6.635000229	1.10735321	1.431306005	0.616552353	0.437453747	0.162349895	0.111092761		
24	Czech Republic	23	6.609000206	1.352682352	1.433885217	0.754444003	0.490946174	0.088106759	0.036872927		
25	Argentina	24	6.598999977	1.185295463	1.440451145	0.695137084	0.494519204	0.109457061	0.059739888		
26	Mexico	25	6.578000069	1.153183818	1.21086216	0.709978998	0.412730008	0.120990433	0.132774115		
27	Singapore	26	6.572000027	1.69227767	1.353814363	0.949492395	0.549840569	0.345965981	0.464307785		
28	Malta	27	6.52699995	1.343279839	1.488411665	0.821944237	0.588767052	0.574730575	0.153066069		
29	Uruguay	28	6.453999996	1.217559695	1.412227869	0.719216824	0.579392254	0.175096929	0.178061873		
30	Guatemala	29	6.453999996	0.872001946	1.255585194	0.54023999	0.531310618	0.283488393	0.077223279		
31	Panama	30	6.452000141	1.233748436	1.373192549	0.706156135	0.550026834	0.210556939	0.070983924		
32	France	31	6.441999912	1.430923462	1.387776852	0.844465852	0.470222116	0.129762307	0.172502428		
33	Thailand	32	6.423999786	1.127868772	1.425792456	0.647239029	0.580200732	0.57212311	0.031612735		
34	Taiwan Province of China	33	6.421999931	1.433626533	1.384565353	0.793984234	0.361466587	0.258360475	0.063829236		
35	Spain	34	6.402999878	1.384397864	1.532090902	0.8889606	0.40878123	0.190133572	0.070914097		
36	Qatar	35	6.375	1.870765686	1.27429688	0.710098088	0.604130983	0.33047387	0.439299256		
37	Colombia	36	6.356999874	1.070622325	1.402182937	0.595027924	0.477487415	0.149014473	0.046668742		
38	Saudi Arabia	37	6.343999863	1.530623555	1.286677599	0.59014833	0.449750572	0.147616014	0.273432255		
39	Trinidad and Tobago	38	6.168000221	1.361355901	1.380228519	0.519983292	0.518630743	0.325296462	0.008964816		
40	Kuwait	39	6.105000019	1.632952452	1.259698749	0.632105708	0.496337593	0.228289798	0.21515955		
41	Slovakia	40	6.09800005	1.325393558	1.505059242	0.712732911	0.295817465	0.136544481	0.024210852		

After:

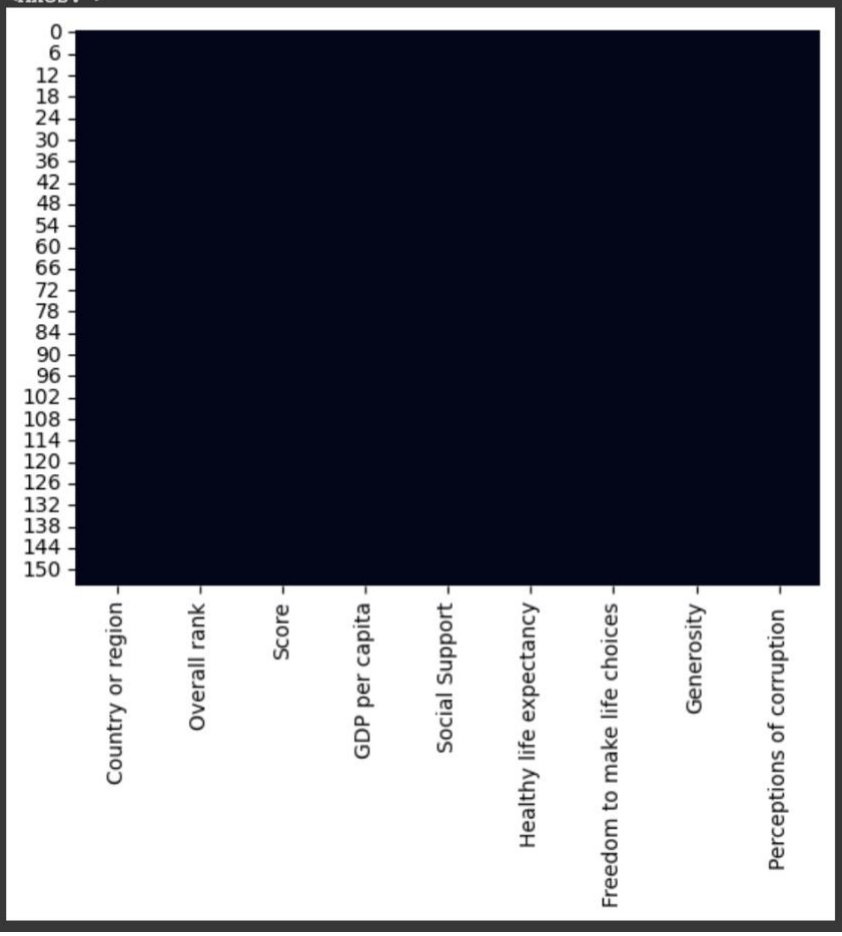


	A	B	C	D	E	F	G	H	I	J	K
1	Country or region	Overall rank	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption		
2	Norway	1	7.537	1.616	1.534	0.797	0.635	0.362	0.316		
3	Denmark	2	7.522	1.482	1.551	0.793	0.626	0.355	0.401		
4	Iceland	3	7.504	1.481	1.611	0.834	0.627	0.476	0.154		
5	Switzerland	4	7.494	1.565	1.517	0.858	0.620	0.291	0.367		
6	Finland	5	7.469	1.444	1.540	0.809	0.618	0.245	0.383		
7	Netherlands	6	7.377	1.504	1.429	0.811	0.585	0.470	0.283		
8	Canada	7	7.316	1.479	1.481	0.835	0.611	0.436	0.287		
9	New Zealand	8	7.314	1.406	1.548	0.817	0.614	0.500	0.383		
10	Sweden	9	7.284	1.494	1.478	0.831	0.613	0.385	0.384		
11	Australia	10	7.284	1.484	1.510	0.844	0.602	0.478	0.301		
12	Israel	11	7.213	1.375	1.376	0.838	0.406	0.330	0.085		
13	Costa Rica	12	7.079	1.110	1.416	0.760	0.580	0.215	0.100		
14	Austria	13	7.006	1.487	1.460	0.815	0.568	0.316	0.221		
15	United States	14	6.993	1.546	1.420	0.774	0.506	0.393	0.136		
16	Ireland	15	6.977	1.536	1.558	0.810	0.573	0.428	0.298		
17	Germany	16	6.951	1.488	1.473	0.799	0.563	0.336	0.277		
18	Belgium	17	6.891	1.464	1.462	0.818	0.540	0.232	0.251		
19	Luxembourg	18	6.863	1.742	1.458	0.845	0.597	0.283	0.319		
20	United Kingdom	19	6.714	1.442	1.496	0.805	0.508	0.493	0.265		
21	Chile	20	6.652	1.253	1.284	0.819	0.377	0.327	0.082		
22	United Arab Emirates	21	6.648	1.626	1.266	0.727	0.608	0.361	0.324		
23	Brazil	22	6.635	1.107	1.431	0.617	0.437	0.162	0.111		
24	Czech Republic	23	6.609	1.353	1.434	0.754	0.491	0.088	0.037		
25	Argentina	24	6.599	1.185	1.440	0.695	0.495	0.109	0.060		
26	Mexico	25	6.578	1.153	1.211	0.710	0.413	0.121	0.133		
27	Singapore	26	6.572	1.692	1.354	0.949	0.550	0.346	0.464		
28	Malta	27	6.527	1.343	1.488	0.822	0.589	0.575	0.153		
29	Uruguay	28	6.454	1.218	1.412	0.719	0.579	0.175	0.178		
30	Guatemala	29	6.454	0.872	1.256	0.540	0.531	0.283	0.077		
31	Panama	30	6.452	1.234	1.373	0.706	0.550	0.211	0.071		
32	France	31	6.442	1.431	1.388	0.844	0.470	0.130	0.173		
33	Thailand	32	6.424	1.128	1.426	0.647	0.580	0.572	0.032		
34	Taiwan Province of China	33	6.422	1.434	1.385	0.794	0.361	0.258	0.064		
35	Spain	34	6.403	1.384	1.532	0.889	0.409	0.190	0.071		
36	Qatar	35	6.375	1.871	1.274	0.710	0.604	0.330	0.439		
37	Colombia	36	6.357	1.071	1.402	0.595	0.477	0.149	0.047		
38	Saudi Arabia	37	6.344	1.531	1.287	0.590	0.450	0.148	0.273		
39	Trinidad and Tobago	38	6.168	1.361	1.380	0.520	0.519	0.325	0.009		
40	Kuwait	39	6.105	1.633	1.260	0.632	0.496	0.228	0.215		
41	Slovakia	40	6.098	1.325	1.505	0.713	0.296	0.137	0.024		

Using pyhton Commands:

1. Check that there is no Null Values in the dataset:

```
Country or region               0
Overall rank                    0
Score                           0
GDP per capita                  0
Social Support                  0
Healthy life expectancy         0
Freedom to make life choices    0
Generosity                      0
Perceptions of corruption       0
dtype: int64
<Axes: >
```



2. Split the data into two catgories: Categorical data and numerical data:

```
Categorical Columns:
['Country or region']

Numerical Columns:
['Overall rank', 'Score', 'GDP per capita', 'Social Support', 'Healthy life expectancy', 'Freedom to make life choices', 'Generosity', 'Perceptions of corruption ']
```

Hypothesis:

Grouping the countries based on their happiness score and identifying the factors that contribute to happiness levels.

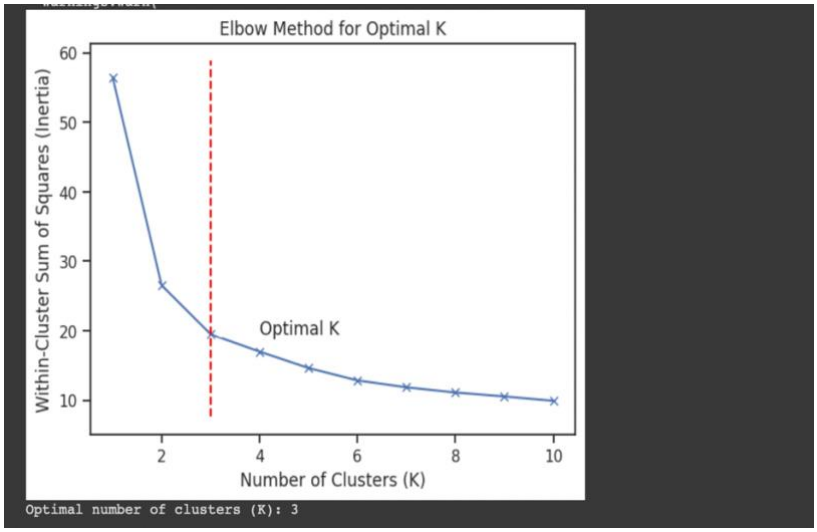
Predicting the happiness score of a new country based on the happiness score and its attributes.

Identifying the factors that contribute to the difference in happiness levels between countries

Proposed solutions:

As a solution for these problems, I picked a technique called “K-Means Clustering”. Clustering is an example of unsupervised learning and a method of grouping a set of objects in a way that ensures the objects that are like each other remain in the same group.

We first need to import the dataset, according to the description, the dataset consists of samples described by 8 numeric predictor attributes (Overall rank, Score, GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity and Perceptions of corruption) plus a label (Country or region).
We will only work with the last seven attributes (Score, GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity and Perceptions of corruption). The number of clusters we wish to construct is decided before we run the K-means algorithm. Scikit-Learn's KMeans will be imported. Then, using the dataset to train the K-means model, we will determine the dispersion of the clusters for each value of K, ranging from 1 to 11. We can observe from the graphic below that as K increases, the Sum of Overall Distances between the centroids and the datapoints in the clusters decreases. However, greater values of K also result in a slower rate of decline for this number. In other words, we see the largest decrease of that quantity between K=1 and K=2 and then the second largest drop is between K=2 and K=3. The Sum of Overall Distances is thus only slightly affected by the subsequent drop in K. As a result, we decide that K=3 is the elbow point.
We may also accomplish this programmatically by identifying the elbow point "k" value by installing the "Kneed" package.



In this code, you fit the K-means clustering algorithm with the selected K value after figuring out the ideal value of K using the elbow approach. The DataFrame (data['Cluster']) is then updated with the cluster labels. Using data['Cluster'], you may display the total number of data points in each cluster. To gain a general idea of how the data points are distributed within clusters, use value_counts() .Printing the 'Country' and 'Cluster' columns from the DataFrame will allow you to view the clustered output in its final form. This will display the cluster that was given to each nation in the dataset based on the ideal K value.

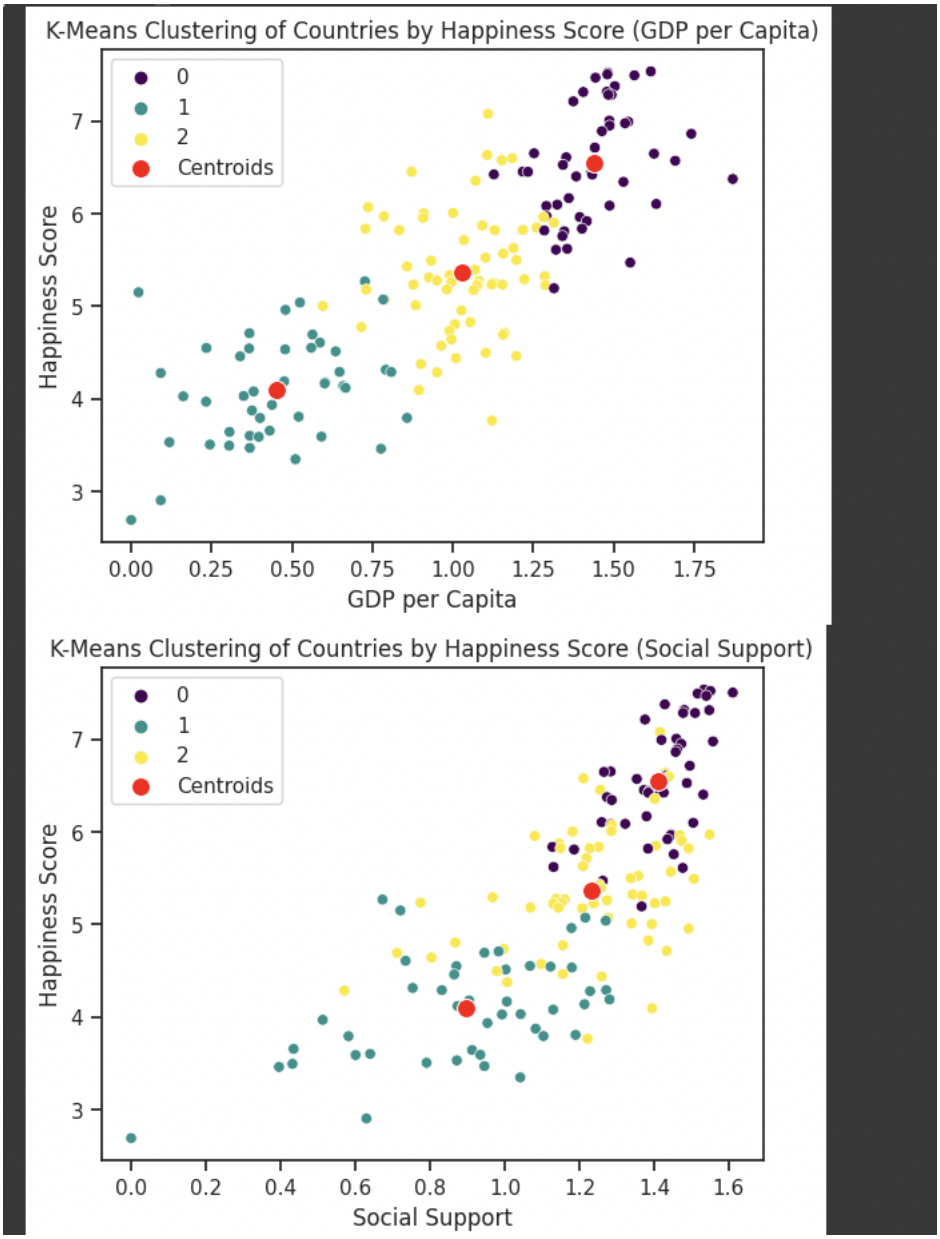
```
2    63
0    47
1    45
Name: Cluster, dtype: int64
            Country or region  Cluster
0                      Norway        0
1                     Denmark        0
2                     Iceland        0
3                 Switzerland        0
4                     Finland        0
..                          ...      ...
150                    Rwanda        1
151                     Syria        1
152                  Tanzania        1
153                   Burundi        1
154  Central African Republic        1
```

Hypothesis 1:

Grouping the countries based on their happiness score and identifying the factors that contribute to happiness levels.

Solution:

K-Means Clustering is a useful method for grouping countries with similar happiness scores. By analysing the characteristics of each group, we can identify the factors that contribute to happiness in different regions of the world. The code that I implemented starts by importing the necessary libraries for data manipulation, K-means clustering, plotting, and visualization. It reads the dataset from the CSV file into a pandas DataFrame. The relevant columns for clustering are selected from the DataFrame. The number of desired clusters is set, and K-means clustering is performed on the selected features. The cluster labels obtained from the clustering algorithm are added as a new column in the DataFrame. The data is then grouped by the cluster column, and the mean values of each feature for each cluster are calculated. A function is defined to plot the clusters and centroids, where scatter plots are created to visualize the data points and centroids, with the specified feature as the x-axis and the 'Score' column as the y-axis. The plot includes labels, a title, and a legend. The function is called for each feature of interest. Therefore, we are able to analyse the results and identify the factors that contributes to happiness levels. [3]

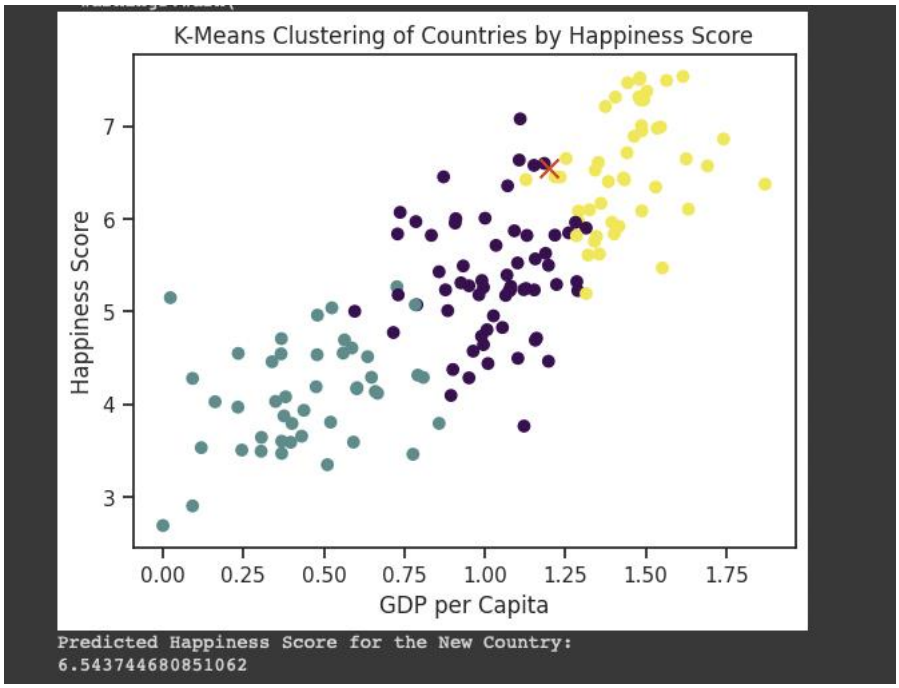


Hypothesis 2:

Predicting the happiness score of a new country based on the happiness score and its features.

Solution:

K-Means Clustering can be used to predict the happiness score of a new country. We can first group the existing countries based on their happiness score and features, and then assign the new country to the most similar group. The average happiness score of that cluster can be used as a prediction for the happiness score of the new country. The code that I provided performs K-means clustering on the dataset using the features 'GDP per capita', 'Social Support', 'Healthy life expectancy', 'Freedom to make life choices', 'Generosity', and 'Perceptions of corruption'. It starts by reading the dataset into a DataFrame and selecting the relevant columns. Then, K-means clustering is performed with a specified number of clusters (k=3), and the cluster labels are added to the DataFrame. The code includes an example of predicting the cluster for a new country based on its features and calculating the average happiness score for that predicted cluster. The clusters and the new country are visualized using a scatter plot, with the data points coloured according to their cluster, and the predicted new country shown as a red 'x' marker. We received results that the predicted happiness score of the new country added is around 6.54 therefore, our hypotheses is verified.

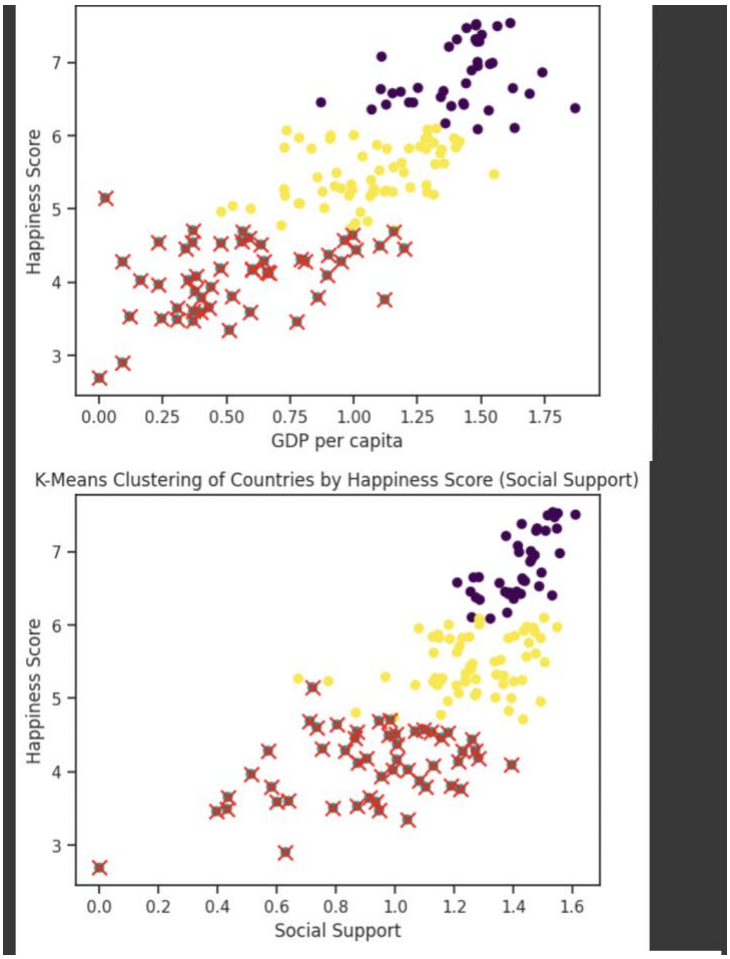


Hypothesis 3:

Identifying the factors that contribute to the difference in happiness levels between countries.

Solution:

K-Means Clustering can be used to identify countries that have significantly different happiness scores and attributes than the rest of the dataset. By analysing the attributes of these countries, we can identify factors that contribute to happiness or its absence in these countries. My next code performs K-means clustering, It starts by importing the necessary libraries. The dataset is read from the CSV file and stored in a pandas DataFrame. Relevant columns for clustering, including 'GDP per capita', 'Social Support', 'Healthy life expectancy', 'Freedom to make life choices', 'Generosity', 'Perceptions of corruption', and 'Score', are selected. K-means clustering is then performed with k=3 clusters. The cluster labels are obtained and added as a new column in the DataFrame. The code identifies the cluster with the lowest average happiness score. It filters the data to extract the countries belonging to that cluster. A function is defined to plot the clusters and the countries with the lowest happiness scores for each feature. The function is called for each feature, creating scatter plots. Finally, the countries with significantly different happiness scores and attributes are printed. As you can see below all the countries with low happiness score were crossed in each diagram which verifies our hypothesis in identifying the lowest happiness score between countries.



	Country or region	GDP per capita	Social Support	\
92	Somalia	0.023	0.721	
105	Sierra Leone	0.368	0.984	
106	Cameroon	0.564	0.946	
107	Iran	1.157	0.712	
108	Albania	0.996	0.804	
109	Bangladesh	0.587	0.735	
110	Namibia	0.964	1.098	
111	Kenya	0.560	1.068	
112	Mozambique	0.234	0.871	
113	Myanmar	0.367	1.123	
114	Senegal	0.479	1.180	
115	Zambia	0.636	1.003	
116	Iraq	1.103	0.979	
117	Gabon	1.198	1.156	
118	Ethiopia	0.339	0.865	
119	Sri Lanka	1.010	1.260	
120	Armenia	0.901	1.007	
121	India	0.792	0.754	
122	Mauritania	0.648	1.272	
123	Congo (Brazzaville)	0.809	0.832	
124	Georgia	0.951	0.571	
125	Congo (Kinshasa)	0.092	1.229	
126	Mali	0.476	1.281	
127	Ivory Coast	0.603	0.905	
128	Cambodia	0.602	1.006	
129	Sudan	0.660	1.214	
130	Ghana	0.667	0.874	
131	Ukraine	0.895	1.395	
132	Uganda	0.381	1.130	
133	Burkina Faso	0.350	1.043	
134	Niger	0.162	0.993	
135	Malawi	0.233	0.513	
136	Chad	0.438	0.954	
137	Zimbabwe	0.376	1.083	
138	Lesotho	0.521	1.190	
139	Angola	0.858	1.104	
140	Afghanistan	0.401	0.582	
141	Botswana	1.122	1.222	
142	Benin	0.431	0.435	
143	Madagascar	0.306	0.913	
144	Haiti	0.369	0.640	
145	Yemen	0.592	0.935	
146	South Sudan	0.397	0.601	
147	Liberia	0.119	0.872	
148	Guinea	0.245	0.791	
149	Togo	0.305	0.432	
150	Rwanda	0.369	0.946	
151	Syria	0.777	0.396	
152	Tanzania	0.511	1.042	
153	Burundi	0.092	0.630	
154	Central African Republic	0.000	0.000	

Reflection:

This coursework is the implementation of the previous one, we explored the application of K-means clustering to a dataset of country happiness scores. The goal was to verify hypothesis from coursework 1 to identify clusters of countries with similar attributes and analyse their happiness scores. We started by importing the necessary libraries and reading the dataset into a DataFrame. We then selected the relevant columns for clustering, which included features such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, perceptions of corruption, and the happiness score itself. Next, we performed K-means clustering with a chosen number of clusters, k. We used the KMeans class from scikit-learn, fitted the model to the selected features, and obtained the cluster labels. We added the cluster labels to the original DataFrame. To analyse the characteristics of each cluster, we calculated the mean values of the features for each cluster. These mean values represented the centroids of the clusters. We plotted the clusters and centroids for each feature against the happiness score using scatter plots. The clusters were visualized with different colours, while the centroids were marked with red

dots. Additionally, we identified the cluster with the lowest average happiness score and extracted the countries belonging to that cluster. We plotted the clusters and highlighted the countries with the lowest happiness scores using red 'x' markers. Finally, we printed the countries with significantly different happiness scores and attributes, focusing on the countries in the cluster with the lowest happiness score.
Overall, this coursework helped demonstrate the use of K-means clustering to identify clusters and analyse their characteristics based on specific features. It also showcased the visualization of clusters and centroids to gain insights into the distribution of data points. The analysis provided valuable information about the countries with the lowest happiness scores, allowing for further investigation and examination of potential factors contributing to their lower happiness levels.

References:

[1] *Laccei* (no date). Available at: https://laccei.org/LACCEI2021-VirtualEdition/full_papers/FP303.pdf (Accessed: April 18, 2023).

[2] Network, S.D.S. (2019) *World happiness report*, *Kaggle*. Available at: https://www.kaggle.com/datasets/unsdsn/world-happiness/code?datasetId=894&sortBy=voteCount&select=2017.csv (Accessed: April 18, 2023).

[3] “StatQuest: K-means clustering,” YouTube, https://www.youtube.com/watch?v=4b5d3muPQmA (accessed May 11, 2023).