The World Happiness Report dataset

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.
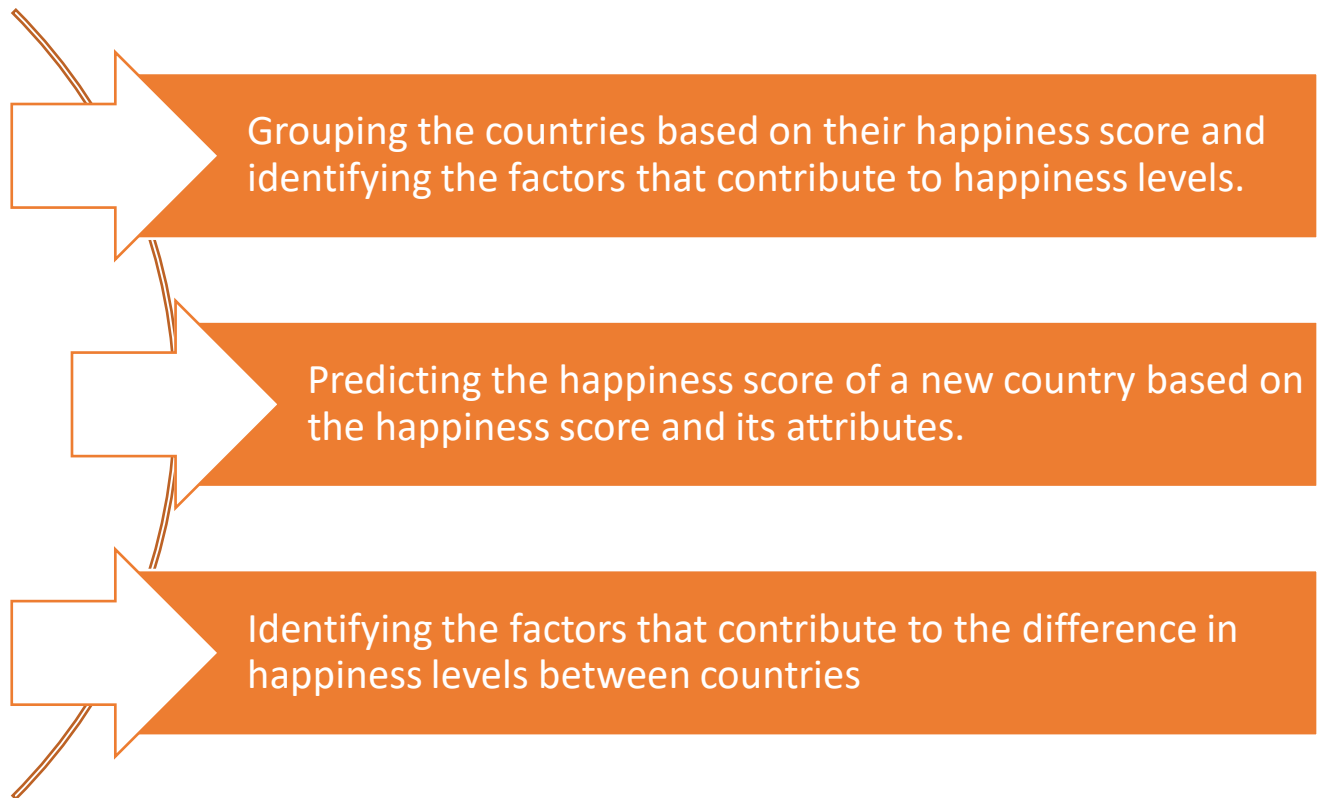
6. Perception of corruption. [1]

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.

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.

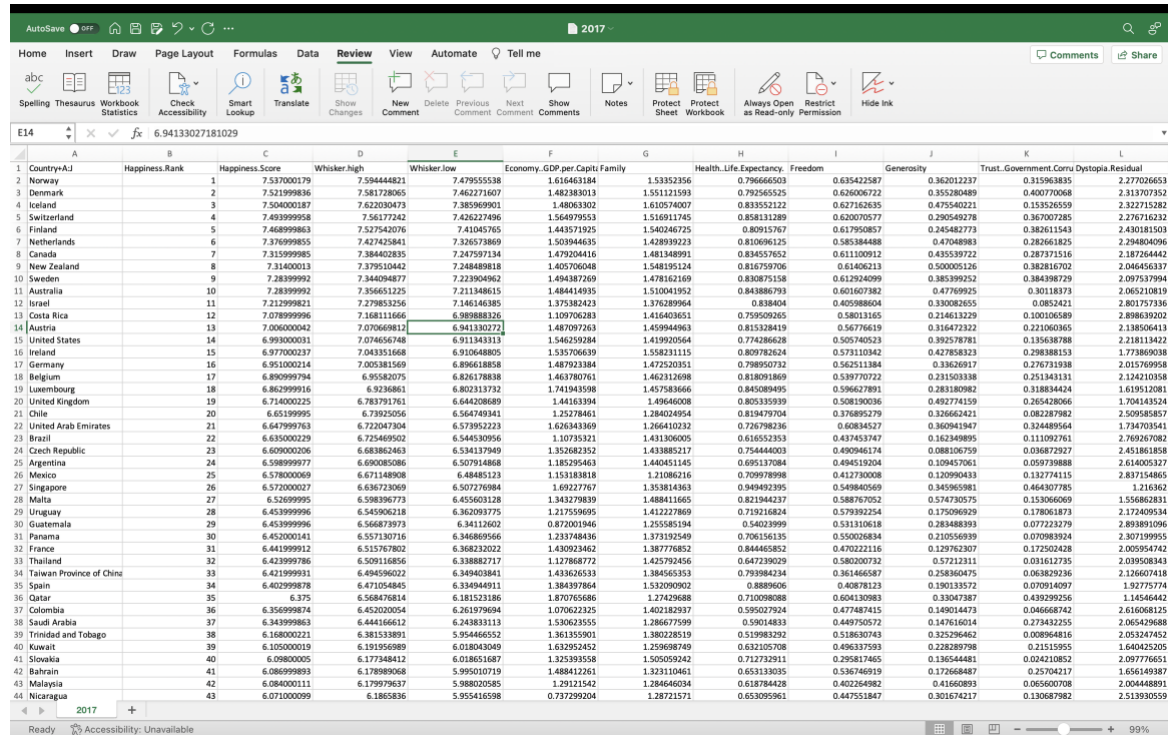Identifying the factors that contribute to the difference in happiness levels between countries

Preprocessing:

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:

After:

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.
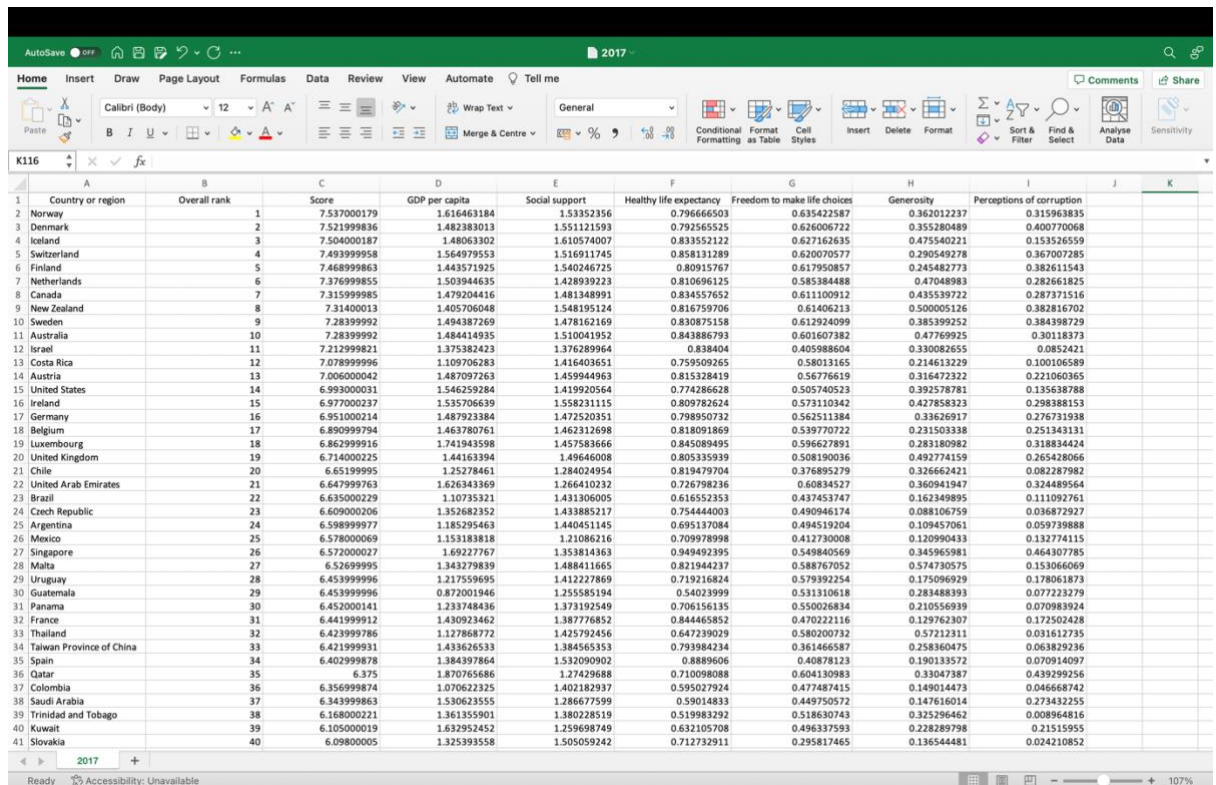
Before:

Country+A:J	Happiness.Rank	Happiness.Score	Economy..GDP.per.Capita	Family	Health..Life.Expectancy.	Freedom	Generosity	Trust..Government.Corruption.
Norway	1	7.537000179	1.616463184	1.53352356	0.796666503	0.635422587	0.362012237	0.315963835
Denmark	2	7.521999836	1.482383013	1.551121593	0.792565525	0.626006722	0.355280489	0.400770068
Iceland	3	7.504000187	1.48063302	1.610574007	0.833552122	0.627162635	0.475540221	0.153526559
Switzerland	4	7.493999958	1.564979553	1.516911745	0.858131289	0.620070577	0.290549278	0.367007285
Finland	5	7.468999863	1.443571925	1.540246725	0.80915767	0.617950857	0.245482773	0.382611543
Netherlands	6	7.376999855	1.503944635	1.428939223	0.810696125	0.585384488	0.47048983	0.282661825
Canada	7	7.315999985	1.479204416	1.481348991	0.834557652	0.611100912	0.435539722	0.287371516
New Zealand	8	7.31400013	1.405706048	1.548195124	0.816759706	0.61406213	0.500005126	0.382816702
Sweden	9	7.28399992	1.494387269	1.478162169	0.830875158	0.612924099	0.385399252	0.384398729
Australia	10	7.28399992	1.484414935	1.510041952	0.843886793	0.601607382	0.47769925	0.30118373
Israel	11	7.212999821	1.375382423	1.376289964	0.838404	0.405988604	0.330082655	0.0852421
Costa Rica	12	7.078999996	1.109706283	1.416403651	0.759509265	0.58013165	0.214613229	0.100106589
Austria	13	7.006000042	1.487097263	1.459944963	0.815328419	0.56776619	0.316472322	0.221060365
United States	14	6.993000031	1.546259284	1.419920564	0.774286628	0.505740523	0.392578781	0.135638788
Ireland	15	6.977000237	1.535706639	1.558231115	0.809782624	0.573110342	0.427858323	0.298388153
Germany	16	6.951000214	1.487923384	1.472520351	0.798950732	0.562511384	0.33626917	0.276731938
Belgium	17	6.890999794	1.463780761	1.462312698	0.818091869	0.539770722	0.231503338	0.251343131
Luxembourg	18	6.862999916	1.741943598	1.457583666	0.845089495	0.596627891	0.283180982	0.318834424
United Kingdom	19	6.714000225	1.44163394	1.49646008	0.805335939	0.508190036	0.492774159	0.265428066
Chile	20	6.65199995	1.25278461	1.284024954	0.819479704	0.376895279	0.326662421	0.082287982
United Arab Emirates	21	6.647999763	1.626343369	1.266410232	0.726798236	0.60834527	0.360941947	0.324489564
Brazil	22	6.635000229	1.10735321	1.431306005	0.616552353	0.437453747	0.162349895	0.111092761
Czech Republic	23	6.609000206	1.352682352	1.433885217	0.754444003	0.490946174	0.088106759	0.036872927
Argentina	24	6.598999977	1.185295463	1.440451145	0.695137084	0.494519204	0.109457061	0.059739888
Mexico	25	6.578000069	1.153183818	1.21086216	0.709978998	0.412730008	0.120990433	0.132774115
Singapore	26	6.572000027	1.69227767	1.353814363	0.949492395	0.549840569	0.345965981	0.464307785
Malta	27	6.52699995	1.343279839	1.488411665	0.821944237	0.588767052	0.574730575	0.153066069
Uruguay	28	6.453999996	1.217559695	1.412227869	0.719216824	0.579392254	0.175096929	0.178061873
Guatemala	29	6.453999996	0.872001946	1.255585194	0.54023999	0.531310618	0.283488393	0.077223279
Panama	30	6.452000141	1.233748436	1.373192549	0.706156135	0.550026834	0.210556939	0.070983924
France	31	6.441999912	1.430923462	1.387776852	0.844465852	0.470222116	0.129762307	0.172502428
Thailand	32	6.423999786	1.127868772	1.425792456	0.647239029	0.580200732	0.57212311	0.031612735
Taiwan Province of China	33	6.421999931	1.433626533	1.384565353	0.793984234	0.361466587	0.258360475	0.063829236
Spain	34	6.402999878	1.384397864	1.532090902	0.8889606	0.40878123	0.190133572	0.070914097
Qatar	35	6.375	1.870765686	1.27429688	0.710098088	0.604130983	0.33047387	0.439299256
Colombia	36	6.356999874	1.070622325	1.402182937	0.595027924	0.477487415	0.149014473	0.046668742
Saudi Arabia	37	6.343999863	1.530623555	1.286677599	0.59014833	0.449750572	0.147616014	0.273432255
Trinidad and Tobago	38	6.168000221	1.361355901	1.380228519	0.519983292	0.518630743	0.325296462	0.008964816
Kuwait	39	6.105000019	1.632952452	1.259698749	0.632105708	0.496337593	0.228289798	0.21515955
Slovakia	40	6.09800005	1.325393558	1.505059242	0.712732911	0.295817465	0.136544481	0.024210852
Bahrain	41	6.086999893	1.488412261	1.323110461	0.653133035	0.536746919	0.172668487	0.25704217
Malaysia	42	6.084000111	1.29121542	1.284646034	0.618784428	0.402264982	0.41660893	0.065600708
Nicaragua	43	6.071000099	0.737299204	1.28721571	0.653095961	0.447551847	0.301674217	0.130687982

After:

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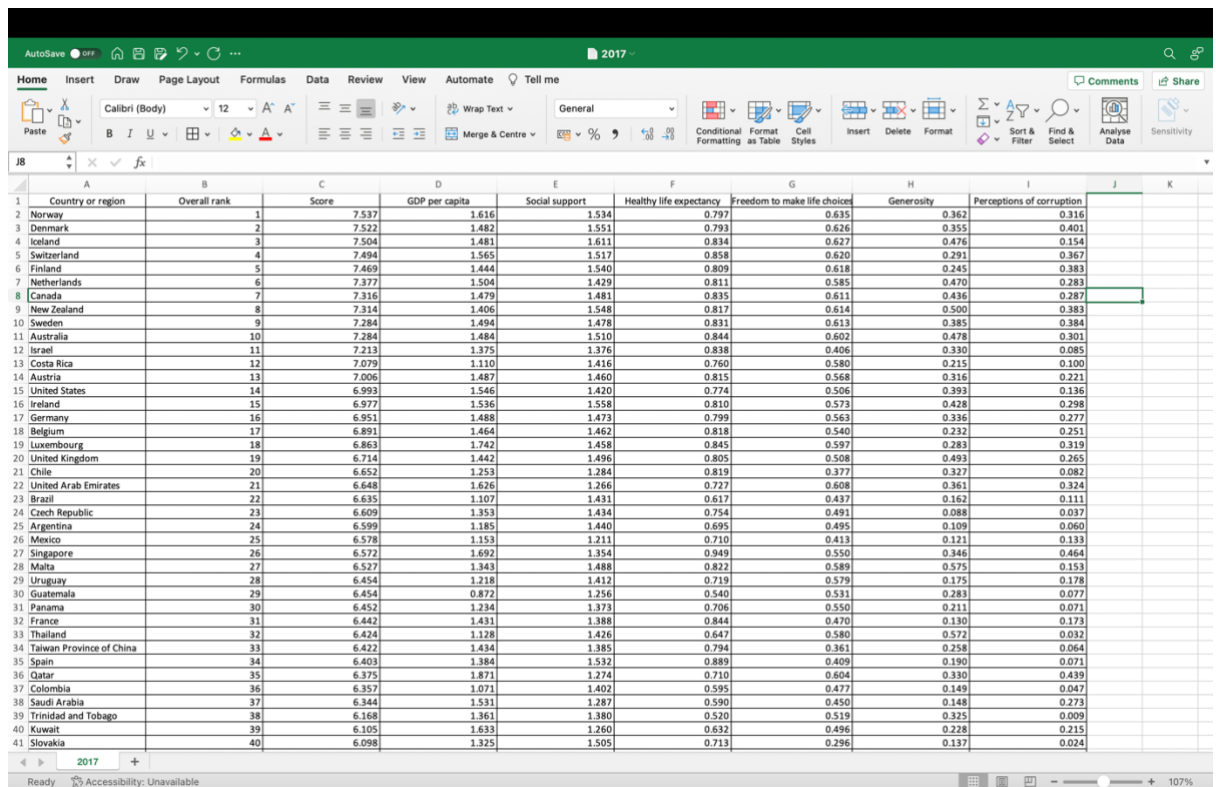
Before:

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

After:

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

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.
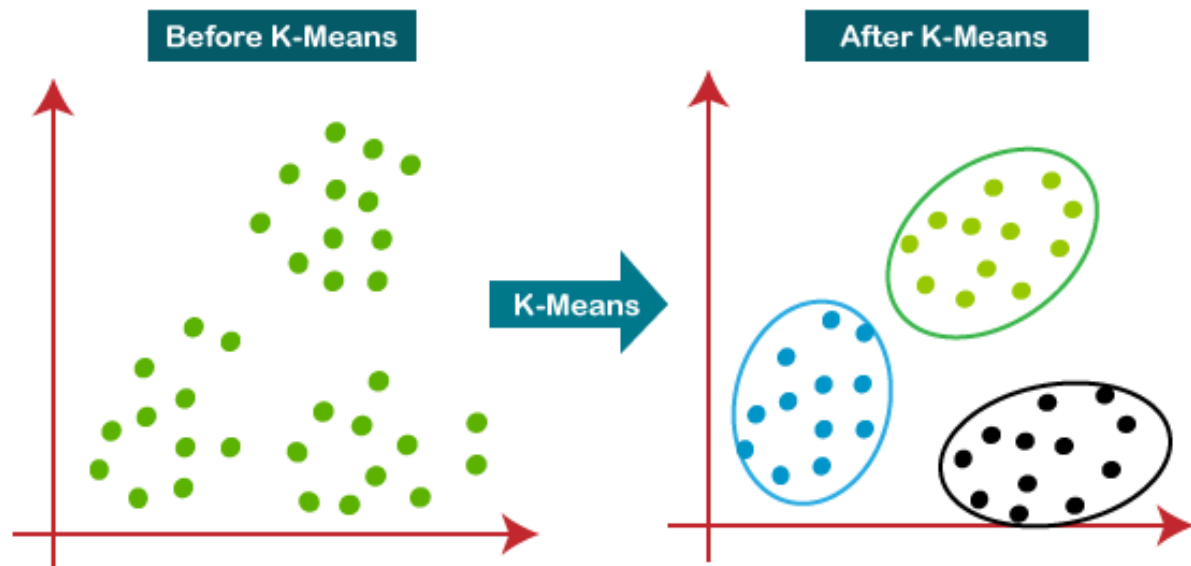
Hypothesis 1:
Grouping the countries based on their happiness score and identifying the factors that contribute to happiness levels.

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. This information can be useful for the improvement of the well-being of the citizens. We can perform K-Means Clustering on the dataset using the happiness score as the target variable and other attributes such as GDP, social support, life expectancy, etc. as features.

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. This method is useful when we have a limited amount of data for the new country and would want to predict its happiness score based on the similarities with existing countries.

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. This information can be useful for people that are considering moving or doing business in these countries, or for countries who want to improve the well-being of their citizens. We can use K-Means Clustering with happiness score as the target variable and other attributes as features to identify it.

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).