

1. Data

1.1.Data source

The information related to the World Happiness Report rank has been acquired in the excel format from Kaggle website. However, the main source of data is Gallup and different reports have been generated by different research teams who were partnered with this institution. There are six separated excel files which contain information regarding countries' score and rank as well as their score in each different variable from 2015 to 2020.

1.2.Data pre-processing

At first sight, it could be observed that the number of countries under study vary over different years. Therefore, I conducted a VLOOKUP function in excel to find what countries are missing in each file. The total number of countries and missing countries are shown in the table below:

year	Total number of countries	Compared with	Number of missing countries	Name of missing countries
2015	158	2016	5	Namibia, South Sudan, Belize, Somalia, Purto Rico
		2017	6	Belize, Taiwan, Hong Kong, Somalia, Namibia, South Sudan
		2018	6	Trinidad and Tobago, Belize, North Cyprus, Somalia, Namibia, South Sudan
		2019	7	Trinidad and Tobago, North Cyprus, North Macedonia, Somalia, Namibia, Gambia, South Sudan
		2020	6	Taiwan, Hong Kong, Maldives, Gambia, Namibia, South Sudan
2016	157	2015	3	Oman, Djibouti, Central Africa
		2017	5	Taiwan, Hong Kong, Mozambique, Lesotho, Central African Republic
		2018	5	Trinidad and Tobago, North Cyprus, Mozambique, Lesotho, Central African Republic
		2019	8	Trinidad and Tobago, North Cyprus, North Macedonia, Gambia, Mozambique, Swaziland, Lesotho, Central African Republic
		2020	8	Taiwan, Congo, Hong Kong, Gambia, Mozambique, Swaziland, Lesotho, Central African Republic
2017	155	2015	9	Oman, Djibouti, Swaziland, Taiwan, Suriname, Hong kong, Somaliland region, Laos, Comoros
		2016	7	Purto Rico, Taiwan, Suriname, Hong kong, Somaliland region, Laos, Comoros

		2018	5	Taiwan, Trinidad and Tobago, North Cyprus, Hong Kong, Laos
		2019	9	Taiwan, Trinidad and Tobago, North Cyprus, Hong Kong, Laos, North Macedonia, Gambia, Swaziland, Comoros
		2020	6	Hong Kong, Maldives, Laos, Gambia, Swaziland, Comoros
2018	156	2015	8	Oman, Suriname, Trinidad and Tobago, North Cyprus, Swaziland, Somaliland region, Comoros, Djibouti
		2016	6	Puerto Rico, Suriname, Trinidad and Tobago, North Cyprus, Somaliland region, Comoros
		2017	4	Taiwan, Trinidad and Tobago, North Cyprus, Hong Kong
		2019	4	North Macedonia, Gambia, Swaziland, Comoros
		2020	8	Taiwan, Trinidad and Tobago, North Cyprus, Hong Kong, Maldives, Gambia, Swaziland, Comoros
2019	156	2015	9	Oman, Suriname, Trinidad and Tobago, North Cyprus, Somaliland region, Macedonia, Sudan, Djibouti, Angola
		2016	9	Puerto Rico, Suriname, Trinidad and Tobago, Belize, North Cyprus, Macedonia, Somaliland region, Sudan, Angola
		2017	8	Taiwan, Trinidad and Tobago, Belize, North Cyprus, Macedonia, Hong Kong, Sudan, Angola
		2018		Belize, Macedonia, Sudan, Angola
		2020	10	Taiwan, Trinidad and Tobago, North Cyprus, Hong Kong, Maldives, Macedonia
2020	153	2015	11	Oman, Qatar, Taiwan, Suriname, Hong Kong, Bhutan, Somaliland region, Sudan, Djibouti, Angola, Syria
		2016	11	Puerto Rico, Suriname, Qatar, Taiwan, Belize, Hong Kong, Somalia, Somaliland region, Sudan, Angola, Syria
		2017	8	Qatar, Belize, Hong Kong, Somalia, Bhutan, Sudan, Angola, Syria

		2018	11	Taiwan, Qatar, Trinidad and Tobago, Belize, Hong Kong, North Cyprus, Bhutan, Somalia, Sudan, Angola, Syria
		2019	9	Taiwan, Qatar, Trinidad and Tobago, North Cyprus, Hong Kong, Bhutan, Somalia, North Macedonia, Syria

Since I do not have the information for the missing countries to complete the report of 2020, I will drop the rows related to these countries while data cleaning.

Moreover, there are some features that include redundant information and should be dropped. All Data should be normalized. Then Exploratory data analysis including correlations, Descriptive analysis, etc. will be done to prepare data for further analysis.