

# DASHBOARD PROJECT

## World Happiness Report Analysis

Pre-processing of datasets through R language and visualisation of the csv file using TABLEAU and drawing conclusions.

**Course** : UCS548, Foundations of Data Science

**Name** : Jasween Kaur Brar

**Roll No** : 102017187

**Sub-Group** : 3CS8

### 1) INTRODUCTION

- The **World Happiness Report** is released by the **UN** Sustainable Development Solutions Network every year.
- It is a landmark survey of the state of global happiness that **ranks** countries by how happy their citizens perceive themselves to be.
- The rankings are based on several factors such as **GDP** per capita, **social support**, **healthy life expectancy**, the **freedom** to make life choices, **generosity** and **trust in government**.

Source of data : UN Sustainable Development Solutions Network  
(<https://worldhappiness.report/>) for years from 2015-2022.

### 2) OBJECTIVE

We will analyse the happiness score of countries and how the happiness score of a country is dependent on following 6 factors :

- **GDP per capita, Social Support, Life Expectancy, Freedom, Trust in Government, Generosity.**

This analysis will be best done through data visualisation and the tool we use is TABLEAU.

Using various graphs and figures we will finally make conclusions.

### 3) DATA TRANSFORMATION & DATA PRE-PROCESSING

~ with help of R Programming

#### Steps we followed:

1. We have data for 8 years (2015 - 2022) with each year of data stored in a separate file. Also the format of the data changes every year. Hence the first step of our analysis would be to read different data files and format the column headers so as to obtain data consistency. Following this step we should be then able to concatenate all the datasets to create a single dataframe.
2. After combining the above datasets, we will apply data transformations to handle missing values.
3. To create a master data file, we will **merge** the world happiness report dataset of the individually processed datasets.

Finally, defining the columns in the final merged & clean dataset :

- 1) **Happiness\_Rank** - Rank 1 means the happiest nation
- 2) **year** - from 8 years (2015 to 2022)
- 3) **Happiness\_Score** - means if you have greater score you can have better possibilities in life
- 4) **GDP\_per\_Capita** - means Gross Domestic Product (GDP) per capita
- 5) **Social\_support** - means that person has a family, friends and other people, who can help them in difficult life situation
- 6) **Life\_Expectancy** - years of healthy life which newborn can expect to have
- 7) **Freedom** - means that people can choose job, family, friends, government and place to live according to their choice
- 8) **Trust\_in\_Government** - means degree of corruption in the country and how much people trust in government
- 9) **Generosity** - means the quality of kindness and generous

## CODE :

```
library(dplyr)
library(tibble)

# Read world happiness dataset files
df_2015 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2015.csv")
df_2016 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2016.csv")
df_2017 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2017.csv")
df_2018 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2018.csv")
df_2019 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2019.csv")
df_2020 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2020.csv")
df_2021 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2021.csv")
df_2022 <- read.csv("C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 oddsem\\data science R\\World_Happiness_Report_Analysis\\2022.csv")

# DATA TRANSFORMATION & DATA PRE-PROCESSING

# A quick view of the dataset helps us understand the inconsistencies in the dataset.
# We are going to tackle these and make our dataset consistent:

# Step 1) 2020 & 2021 datasets do not have columns representing "happiness rank" and "happiness score".
# Happiness Score is called "ladder score".
# Adding the Happiness Rank by sorting the Ladder Score column and getting the highest score as number one rank.

adding_index1 <- seq(1,nrow(df_2020),by=1)
arrange(df_2020,'Ladder score')
mutate(df_2020,Happiness_Rank = adding_index1)
df_2020['Happiness_Rank'] = adding_index1

adding_index2 = seq(1,nrow(df_2021),by=1)
arrange(df_2021,'Ladder score')
df_2021['Happiness_Rank'] = adding_index2

# Step 2) Some datasets have "Region" column mapping each country to a specific region.
# We want to add this column across all years.
# Here we will store this mapping from 2015 data and later merge it with our consolidated dataset.

regions_df <- select(df_2015,Country,Region)

# Step 3) We have inconsistent column names across the different years.
```

```

# So, let us make them consistent.

# 2015 world happiness report data
df_2015 = select(df_2015, 'Country', 'Happiness.Rank', 'Happiness.Score', 'Economy..GDP.per.Ca
pita.', 'Family', 'Health..Life.Expectancy.', 'Freedom', 'Trust..Government.Corruption.', 'G
enerosity')
df_2015 <- df_2015 %>% rename('Happiness.Rank'='Happiness_Rank', 'Happiness.Score' = 'Happi
ness_Score', 'Economy..GDP.per.Capita.' = 'GDP_per_Capita', 'Family' ='Social_support', 'Heal
th..Life.Expectancy.' = 'Life_Expectancy', 'Trust..Government.Corruption.' = 'Trust_in_Gove
rnment')
mutate(df_2015, year = '2015')

# 2016 world happiness report data
df_2016 = select(df_2016, 'Country', 'Happiness.Rank', 'Happiness.Score', 'Economy..GDP.per.Ca
pita.', 'Family', 'Health..Life.Expectancy.', 'Freedom', 'Trust..Government.Corruption.', 'G
enerosity')
df_2016 <- df_2016 %>% rename('Happiness.Rank'='Happiness_Rank', 'Happiness.Score' = 'Happi
ness_Score', 'Economy..GDP.per.Capita.' = 'GDP_per_Capita', 'Family' ='Social_support', 'Heal
th..Life.Expectancy.' = 'Life_Expectancy', 'Trust..Government.Corruption.' = 'Trust_in_Gove
rnment')
mutate(df_2016, year = '2016')

# 2017 world happiness report data
df_2017 = select(df_2017, 'Country', 'Happiness.Rank', 'Happiness.Score', 'Economy..GDP.per.Ca
pita.', 'Family', 'Health..Life.Expectancy.', 'Freedom', 'Trust..Government.Corruption.', 'G
enerosity')
df_2017 <- df_2017 %>% rename('Economy..GDP.per.Capita.'='GDP_per_Capita', 'Family'='Social
_support', 'Health..Life.Expectancy.'='Life_Expectancy', 'Trust..Government.Corruption.'='Tr
ust_in_Government', 'Happiness.Rank'='Happiness_Rank', 'Happiness.Score': 'Happiness_Score')
mutate(df_2017, year = '2017')

# 2018 world happiness report data
df_2018 = select(df_2018, 'Country or region', 'Overall rank', 'Score', 'GDP per capita', 'Soci
al support', 'Healthy life expectancy', 'Freedom to make life choices', 'Perceptions of co
rruption', 'Generosity')
df_2018 <- df_2018 %>% rename('GDP per capita'='GDP_per_Capita', 'Social support'='Social_s
upport', 'Healthy life expectancy'='Life_Expectancy', 'Perceptions of corruption'='Trust_in_
Government', 'Overall rank'='Happiness_Rank', 'Score'='Happiness_Score', 'Freedom to make lif
e choices'='Freedom', 'Country or region'='Country')
mutate(df_2018, year = '2018')

# 2019 world happiness report data
df_2019 = select(df_2019, 'Country or region', 'Overall rank', 'Score', 'GDP per capita', 'Soci
al support', 'Healthy life expectancy', 'Freedom to make life choices', 'Perceptions of co
rruption', 'Generosity')
df_2019 <- df_2019 %>% rename('GDP per capita'='GDP_per_Capita', 'Social support'='Social_s
upport', 'Healthy life expectancy'='Life_Expectancy', 'Perceptions of corruption'='Trust_in_
Government', 'Overall rank'='Happiness_Rank', 'Score'='Happiness_Score', 'Freedom to make lif
e choices'='Freedom', 'Country or region'='Country')
mutate(df_2019, year = '2019')

# 2020 world happiness report data
df_2020 = select(df_2020, 'Country name', 'Happiness_Rank', 'Ladder score', 'Explained by: Log
GDP per capita', 'Explained by: Social support', 'Explained by: Healthy life expectancy',

```

```

'Explained by: Freedom to make life choices', 'Explained by: Perceptions of corruption',
'Explained by: Generosity')
df_2020 <- df_2020 %>% rename('Explained by: Log GDP per capita'='GDP_per_Capita','Explained
by: Social support'='Social_support','Explained by: Healthy life expectancy'='Life_Expectancy',
'Explained by: Perceptions of corruption'='Trust_in_Government','Ladder score'='Happiness_Score',
'Explained by: Freedom to make life choices'='Freedom','Country name'='Country',
'Explained by: Generosity'='Generosity')
mutate(df_2020, year = '2020')

# 2021 world happiness report data
df_2021 = select(df_2021, 'Country name', 'Happiness_Rank', 'Ladder score', 'Explained by: Log
GDP per capita', 'Explained by: Social support', 'Explained by: Healthy life expectancy',
'Explained by: Freedom to make life choices', 'Explained by: Perceptions of corruption',
'Explained by: Generosity')
df_2021 <- df_2021 %>% rename('Explained by: Log GDP per capita'='GDP_per_Capita','Explained
by: Social support'='Social_support','Explained by: Healthy life expectancy'='Life_Expectancy',
'Explained by: Perceptions of corruption'='Trust_in_Government','Ladder score'='Happiness_Score',
'Explained by: Freedom to make life choices'='Freedom','Country name'='Country',
'Explained by: Generosity'='Generosity')
mutate(df_2021, year = '2021')

# 2022 world happiness report data
df_2022 = select(df_2022, 'Country', 'RANK', 'Happiness score', 'Explained by: GDP per capita',
'Explained by: Social support', 'Explained by: Healthy life expectancy', 'Explained by:
Freedom to make life choices', 'Explained by: Perceptions of corruption', 'Explained by: Generosity')
df_2022 <- df_2022 %>% rename('Explained by: GDP per capita'='GDP_per_Capita','Explained by:
Social support'='Social_support','Explained by: Healthy life expectancy'='Life_Expectancy',
'Explained by: Perceptions of corruption'='Trust_in_Government','Happiness score'='Happiness_Score',
'Explained by: Freedom to make life choices'='Freedom','Country name'='Country',
'Explained by: Generosity'='Generosity', 'RANK'='Happiness_Rank')
mutate(df_2022, year = '2022')

# Step 4) Merging the datasets of all years together into a dataframe

happiness_index_df_final <- append(df_2015, df_2016)
happiness_index_df_final <- append(happiness_index_df_final, df_2017)
happiness_index_df_final <- append(happiness_index_df_final, df_2018)
happiness_index_df_final <- append(happiness_index_df_final, df_2019)
happiness_index_df_final <- append(happiness_index_df_final, df_2020)
happiness_index_df_final <- append(happiness_index_df_final, df_2021)
happiness_index_df_final <- append(happiness_index_df_final, df_2022)

# Step 5) Keep the records of countries having data for all 8 years and remove other countries
# Number of countries in report for each year is different.
# We want to create a unified set of countries for which we have data across 8 years to perform analysis.
# Hence we would not perform analysis on countries with missing/incomplete information.

arrange(happiness_index_df_final, year)
happiness_index_df_final <- group_by(happiness_index_df_final, country)
happiness_index_df_final = happiness_index_df_final[happiness_index_df_final.Country.isin(
country_list)==True]

```

```
# Step 6) Further cleaning of data
# There are null values in the "Trust_in_Government" column specifically for UAE.
# We will impute the missing values for this country with average "Trust_in_Government" value for UAE calculated using available data.
meanTrust<-mean(happiness_index_df_final$'Trust_in_Government', na.rm=TRUE)
happiness_index_df_final$'Trust_in_Government'[is.na(happiness_index_df_final$'Trust_in_Government')]  
<- meanTrust

# Finally, we write our compiled dataset to a csv file.
write.csv(happiness_index_df_final, "C:\\Users\\91922\\Desktop\\Fifth Sem Material 2022 od dsem\\data science R\\World_Happiness_Report_Analysis\\HappinessIndex.csv", row.names=FALSE)
```

The final merged and processed dataset is available in the file “**HappinessIndex.csv**”

Dataset pic :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Index	Country	year	Happiness_Rank	Happiness_Score	GDP_per_Capita	Social_support	Life_Expectancy	Freedom	Trust_in_Government	Generosity	Region_x	Region_y
2	0	0	Switzerland	2015	1	7.587	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678	Western Europe	Western Europe
3	1	1	Switzerland	2016	2	7.509	1.52733	1.14524	0.86903	0.58557	0.41203	0.28083	Western Europe	Western Europe
4	2	2	Switzerland	2017	4	7.493999958	1.564979553	1.516911745	0.858131289	0.620070577	0.367007285	0.290549278	Western Europe	Western Europe
5	3	3	Switzerland	2018	5	7.487	1.42	1.549	0.927	0.66	0.357	0.256	Western Europe	Western Europe
6	4	4	Switzerland	2019	6	7.48	1.452	1.526	1.052	0.572	0.343	0.263	Western Europe	Western Europe
7	5	5	Switzerland	2020	3	7.559899807	1.39077425	1.472403407	1.040533185	0.62895447	0.407945901	0.269055754	Western Europe	Western Europe
8	6	6	Switzerland	2021	3	7.571	1.566	1.079	0.816	0.653	0.413	0.204	Western Europe	Western Europe
9	7	7	Switzerland	2022	4	7.512	2.026	1.226	0.822	0.677	0.461	0.147	Western Europe	Western Europe
10	8	8	Iceland	2015	2	7.561	1.30232	1.40223	0.94784	0.62877	0.14145	0.4363	Western Europe	Western Europe
11	9	9	Iceland	2016	3	7.501	1.42666	1.18326	0.86733	0.56624	0.14975	0.47678	Western Europe	Western Europe
12	10	10	Iceland	2017	3	7.504000187	1.48063302	1.610574007	0.833552122	0.627162635	0.153526559	0.475540221	Western Europe	Western Europe
13	11	11	Iceland	2018	4	7.495	1.343	1.644	0.914	0.677	0.138	0.353	Western Europe	Western Europe
14	12	12	Iceland	2019	4	7.494	1.38	1.624	1.026	0.591	0.118	0.354	Western Europe	Western Europe
15	13	13	Iceland	2020	4	7.504499912	1.326501608	1.547567487	1.000843406	0.661980748	0.144540772	0.362330228	Western Europe	Western Europe
16	14	14	Iceland	2021	4	7.554	1.482	1.172	0.702	0.698	0.17	0.293	Western Europe	Western Europe
17	15	15	Iceland	2022	3	7.557	1.936	1.32	0.873	0.718	0.191	0.27	Western Europe	Western Europe
18	16	16	Denmark	2015	3	7.527	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	Western Europe	Western Europe
19	17	17	Denmark	2016	1	7.526	1.44178	1.16374	0.79504	0.57941	0.44453	0.36171	Western Europe	Western Europe
20	18	18	Denmark	2017	2	7.521999836	1.482383013	1.551121593	0.792565525	0.626006722	0.400770068	0.355280489	Western Europe	Western Europe
21	19	19	Denmark	2018	3	7.555	1.351	1.59	0.868	0.683	0.408	0.284	Western Europe	Western Europe
22	20	20	Denmark	2019	2	7.6	1.383	1.573	0.996	0.592	0.41	0.252	Western Europe	Western Europe
23	21	21	Denmark	2020	2	7.645599842	1.326948524	1.503449202	0.979332566	0.665039897	0.495260328	0.242793396	Western Europe	Western Europe
24	22	22	Denmark	2021	2	7.62	1.502	1.108	0.763	0.686	0.485	0.208	Western Europe	Western Europe
25	23	23	Denmark	2022	2	7.636	1.953	1.243	0.777	0.719	0.532	0.188	Western Europe	Western Europe
26	24	24	Norway	2015	4	7.522	1.459	1.33095	0.88521	0.66973	0.36503	0.34699	Western Europe	Western Europe
27	25	25	Norway	2016	4	7.498	1.57744	1.1269	0.79579	0.59609	0.35776	0.37895	Western Europe	Western Europe
28	26	26	Norway	2017	1	7.537000179	1.616463184	1.53352356	0.796666503	0.635422587	0.315963835	0.362012237	Western Europe	Western Europe
29	27	27	Norway	2018	2	7.594	1.456	1.582	0.861	0.686	0.34	0.286	Western Europe	Western Europe
30	28	28	Norway	2019	3	7.554	1.488	1.582	1.028	0.603	0.341	0.271	Western Europe	Western Europe
31	29	29	Norway	2020	5	7.487999916	1.42420733	1.495172501	1.008071899	0.670200884	0.434100568	0.287985086	Western Europe	Western Europe
32	30	30	Norway	2021	6	7.392	1.543	1.108	0.782	0.703	0.427	0.249	Western Europe	Western Europe
33	31	31	Norway	2022	8	7.365	1.997	1.239	0.786	0.728	0.474	0.217	Western Europe	Western Europe

## 4) MySQL QUERIES WHICH CAN BE USED INSTEAD OF R CODE

```
-- The UNION operator is used to combine the result-set of two or more SELECT statements.
-- Hence we use UNION operator to combine all the 8 year data tables into a new table with all unique rows
```

```
CREATE TABLE HappinessIndex
(SELECT * FROM df_2015)
```

```

UNION
(SELECT * FROM df_2016)
UNION
(SELECT * FROM df_2017)
UNION
(SELECT * FROM df_2018)
UNION
(SELECT * FROM df_2019)
UNION
(SELECT * FROM df_2020)
UNION
(SELECT * FROM df_2021)
UNION
(SELECT * FROM df_2022);

```

## 5) DATA VISUALISATION IN TABLEAU & ANALYSIS

~ Using our final file “HappinessIndex.csv”

a) **DROPDOWN LIST**, it can be used to select a country and a particular year to dynamically find out statistics about the country. It displays the global happiness rank, happiness score of that year, GDP Per Capita and Average Health Life Expectancy.

<b>Year</b>	Global Happiness Rank	Happiness Score
2022	136	3.777
<b>Country</b>	GDP Per Capita	Health Life Expectancy
India	1.167	0.4710

As shown in this **example**, it displays data for India for the year 2022.

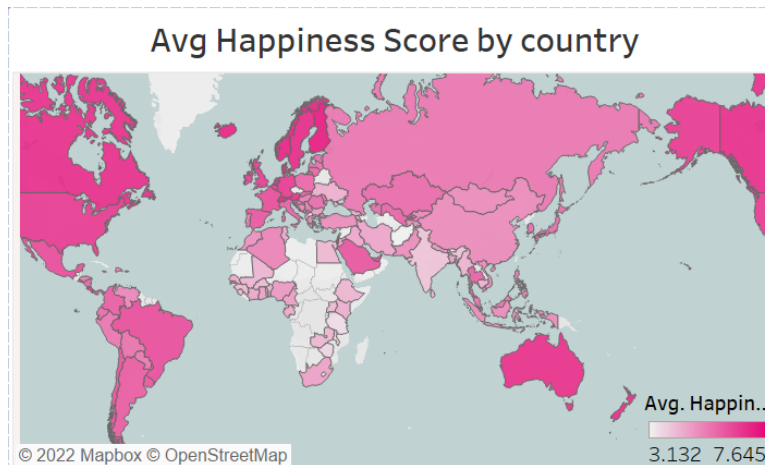
b) **SPATIAL CHART** depicts precise locations and geographical patterns.

Tableau tooltip used :

```

Country: <Country>
Avg. Happiness Score: <AVG(Happiness Score)>

```

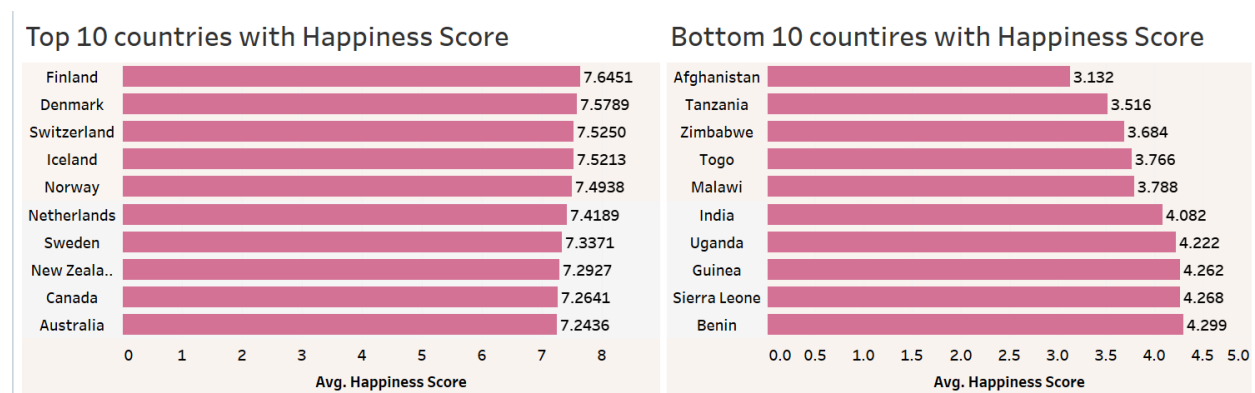


**Analysis :** This gives us a clear picture of the happiness distribution across the world. North America, Scandinavia(Northern Europe) and Australia lead the world in happiness score, whereas Africa & Asia are least happy.

**c) BAR GRAPH (Rankings)** show the top-most and bottom-most countries according to their average happiness score.

Tableau tooltip used :

Country: <Country>  
Avg. Happiness Score: <AVG(Happiness Score)>



**Analysis :** Scandinavian countries such as Finland, Denmark, Switzerland have the most 'happy' citizens in the world based on the data, whereas Afghanistan, Tanzania

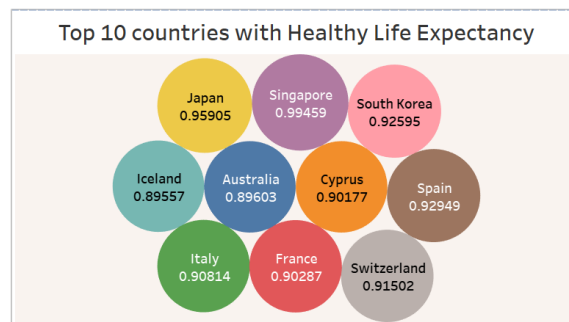


and other countries from Africa rank very low on the happiness index.

d) **PACKED BUBBLE CHART** shows that top 10 countries with highest average life expectancy.

Tableau tooltip used :

```
Country: <Country>  
Avg. Life Expectancy: <AVG(Life Expectancy)>
```

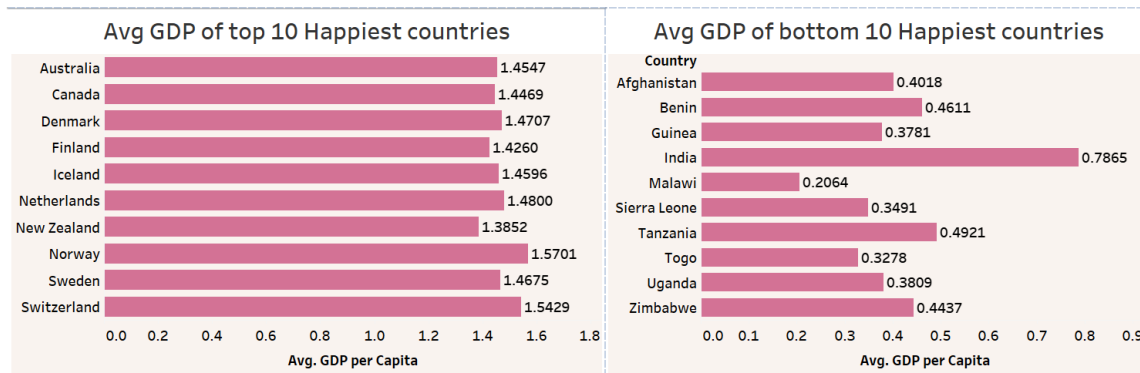


**Analysis :** Developed countries in South east Asia such as Singapore, Japan, South Korea and European countries such as Spain, Italy, Switzerland have excellent healthcare infrastucture which lead to higher life expectancy of their citizens. Hence, these countries are among the happiest countries.

e) **HORIZONTAL BAR GRAPH** show the average GDP of top 10 and bottom 10 happiest countries.

Tableau tooltip used :

```
Country: <Country>  
Avg. GDP per Capita: <AVG(GDP per Capita)>
```

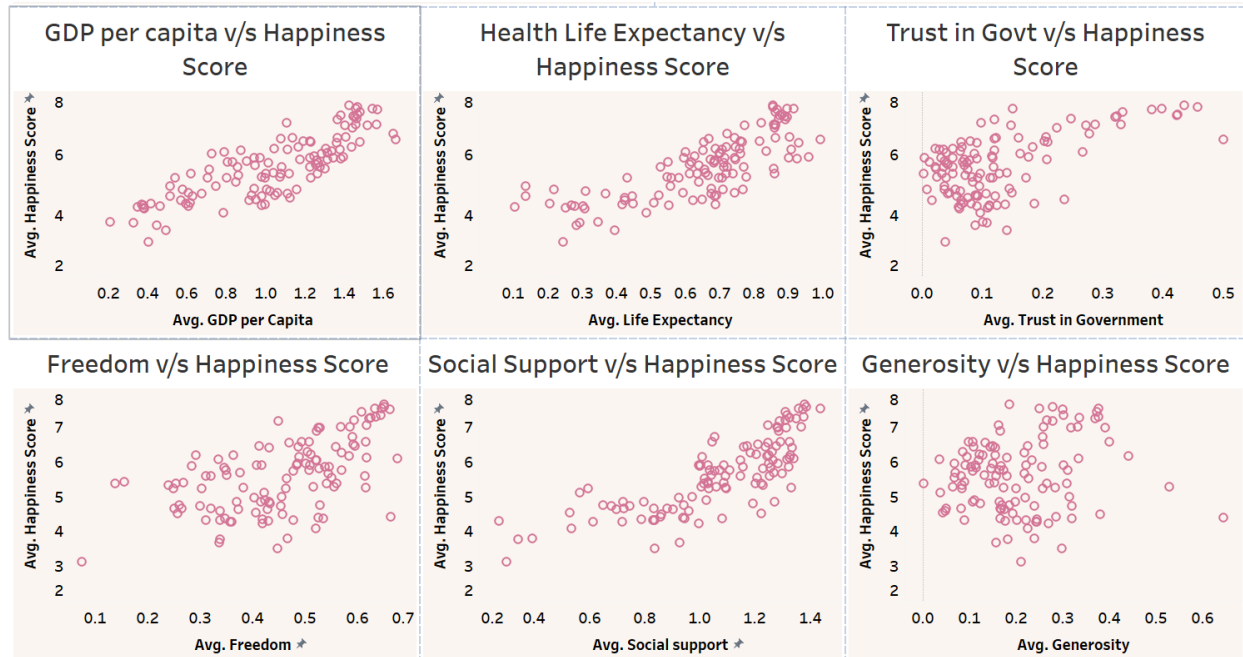


**Analysis :** “GDP per capita” variable which represents the strength of economy of every country. Higher the Gross Domestic Product (GDP) per person, higher the affluence of the country. Average GDP per capita score of top 10 happy countries in world = 1.47 .

**f) SCATTER PLOTS** to understand the **trends** between Happiness score & other features.

Tableau tooltip used :

```
Country: <Country>
Avg. GDP per Capita: <AVG(GDP per Capita)>
Avg. Happiness Score: <AVG(Happiness Score)>
```



### Analysis :

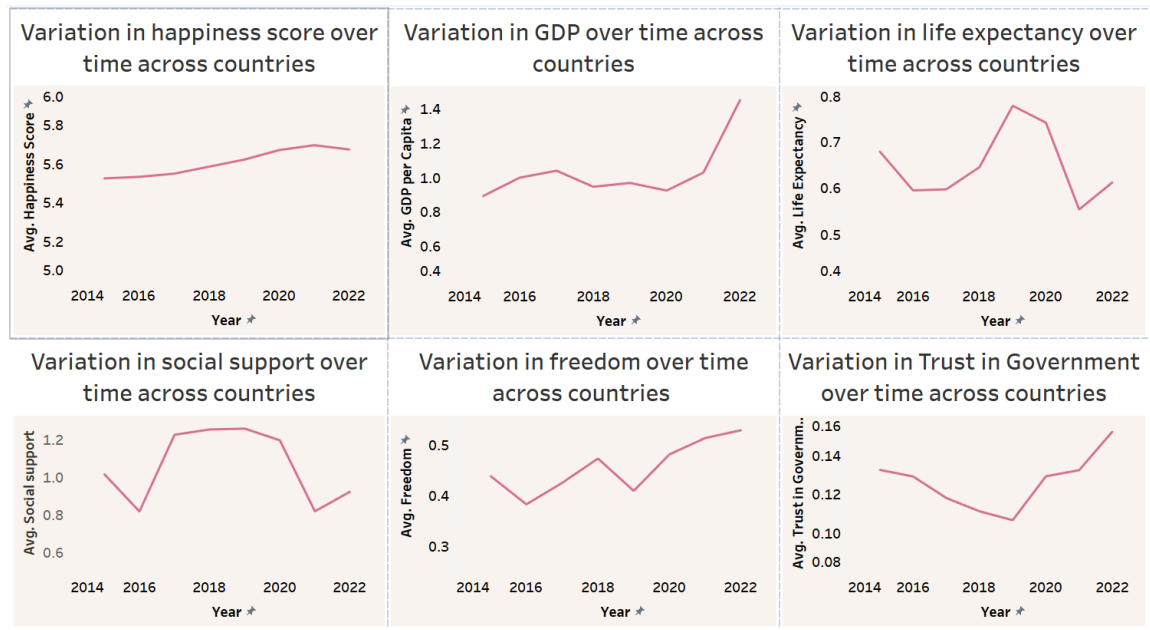
- (i) GDP, Social Support (Family & Friends) and Life Expectancy are strongly correlated with the Happiness score.
- (ii) Freedom correlates quite well with the Happiness score.
- (iii) Government Trust still has a mediocre correlation with the Happiness score.
- (iv) As common wisdom dictates - money makes you happy up to a certain threshold (about 70,000 in the US). However, having a good social network is important and family + friends tend to provide that. High life expectancy and health make you worry less about how you'll survive thus making you more happy.

**g) LINES** to observe change in happiness metrics over **time**.

Tableau tooltip used :

---

Year: <Year>  
 Avg. Happiness Score: <AVG(Happiness Score)>



### Analysis :

- (i) On average worldwide, the happiness score has been rising over the years.
- (ii) The most interesting plot by far is that of "Average Life expectancy" which had been steadily rising until 2019 and we see a significant drop between 2019 to 2021. This clearly demonstrates the effect of COVID-19 pandemic which has taken a serious toll on human life.
- (iii) All of us have been severely affected by the effects of the COVID-19 pandemic and many people lost their loved ones. This explains the drop in social support (family & friends) score in 2021.
- (iv) Trust in Governments across the world has been on a steady decline since 2015 until 2019. This pattern can be explained by the numerous civil unrests across the world and rise of autocratic leadership in many countries. However, the trust has started growing since 2020 and could be attributed to effective government leadership in tackling the pandemic.

## 6) CONCLUSION

### What makes us happy ?

From the data, it can be confirmed that good financial condition can be a leading factor to happiness but up to an extent. After achieving the threshold everyone needs

freedom to enjoy and friends to celebrate with. Good life style with better health care leads to happiness and better happiness score.

### **Impact of Covid-19 on happiness of people.**

The data clearly states that "Average Life expectancy" which had been steadily rising until 2019, had a significant drop between 2019 to 2021. This clearly demonstrates the effect of COVID-19 pandemic which has taken a serious toll on human life. This also impacted the normal life style and had a great impact on 'Trust in Government' factor.

### **Biasness of data towards particular variables.**

Happiness score is strongly correlated with GDP, Social Support (Family & Friends) and Life Expectancy. Most happy countries in the world are much more affluent (Average GDP score = 1.47)