HAPPINESS AROUND THE WORLD

By Deepali Lalchandani

CONTENTS

- Context
- Problem Statement
- Methodogy and Technology Used
- Exploring and understanding data
- Data cleaning and transforming data
- Data visualization
- Data Analysis
- Reression Model Building
- Conclusions

CONTEXT

- The World Happiness Report is a publication of the Sustainable Development Solutions Network, powered by the Gallup World Poll data.
- The Gallup World Poll (GWP) has gathered data since 2005, though the first report was published in 2013
- The happiness scores are based on national average response to the main life evaluation questions asked in the poll.

- Data was downloaded from https://worldhappiness.report/

PROBLEM STATEMENTS

- Analyse trends in happiness around the world
- Can we predict happiness of a country if all factors are known?

METHODOLOGY USED

- Exploring and understanding data
- Data cleaning and transformation
- Data visualization
- Data Analysis
- Machine learning to build efficient model

TECHNOLOGY USED

- Power BI
- Jupyter Notebook

EXPLORING AND UNDERSTANDING DATA

- Data available from **2005 to 2020** **+2021**
- Columns in the table
  - ❖ Happiness score
  - ❖ GDP per capita
  - ❖ Health life expectancy
  - ❖ Social support
  - ❖ Freedom to make life choices
  - ❖ Generosity
  - ❖ Corruption Perception
  - ❖ Positive affect
  - ❖ Negative affect

- We don’t know the units but all columns are type **integer** except Country name and year
- There is **missing data** in many columns

Country name	year	Happiness Score	Log GDP per capita	Social support	Healthy life expe	Freedom to make l	Generosity	Perceptions c	Positive affect	Negative affect
Algeria	2014	6.35489845275879	9.33515930175781	0.818189442157745	65.1399993896484				0.625904619693756	0.17686609923839
Algeria	2016	5.34085369110107	9.36202239990234	0.748588263988495	65.5				0.6605104804039	0.37711197137832
Bahrain	2014	6.16513395309448	10.7834672927856		67.4199981689453					
Bahrain	2015	6.00737524032593	10.785270690918	0.852550745010376	67.6999969482422	0.849521160125732	0.11202140152454		0.71554297208786	0.30297210812568
Bahrain	2016	6.16967344284058	10.7808504104614	0.86270010471344	68.0999984741211	0.888691067695618	0.08818747848272		0.787187218666077	0.28346633911132
Bahrain	2017	6.22732067108154	10.7714796066284	0.87574714422226	68.5	0.905858516693115	0.13631758093833		0.813570559024811	0.28975951671600
Bahrain	2019	7.09801244735718	10.7149906158447	0.877929449081421	69.3000030517578	0.906535506248474	0.04786324501037		0.761622965335846	0.31710639595985

CONTD…

- **Statistics** of all columns

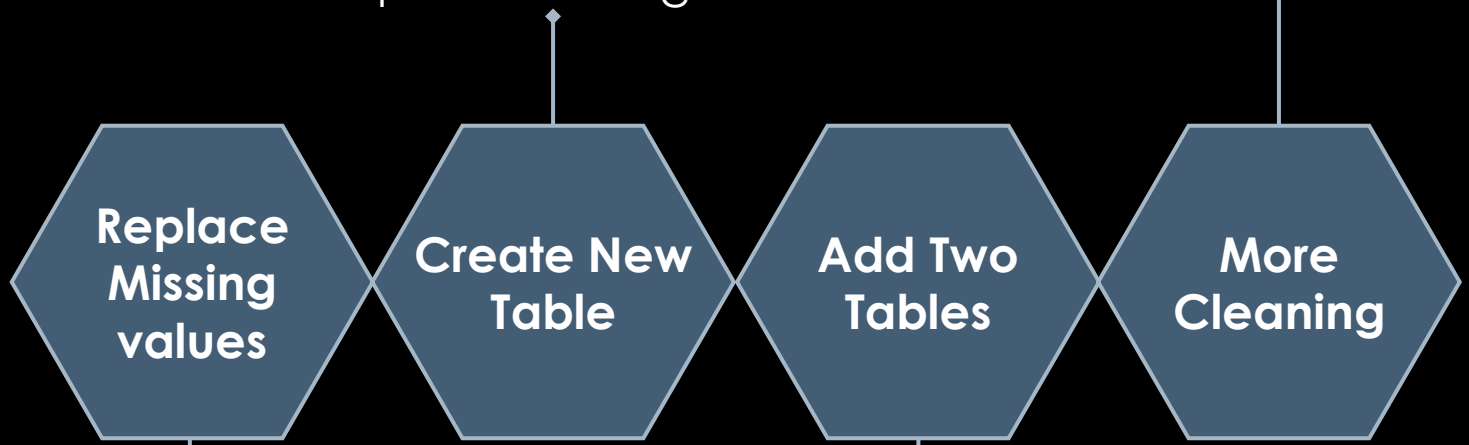
	year	Happiness Score	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect
count	1949.000000	1949.000000	1913.000000	1936.000000	1894.000000	1917.000000	1860.000000	1839.000000	1927.000000	1933.000000
mean	2013.216008	5.466707	9.368459	0.812553	63.359375	0.742567	0.000108	0.747111	0.709998	0.268552
std	4.166828	1.115717	1.154091	0.118480	7.510244	0.142104	0.162221	0.186793	0.107106	0.085176
min	2005.000000	2.375092	6.635322	0.290184	32.299999	0.257534	-0.335040	0.035198	0.321690	0.082737
25%	2010.000000	4.640079	8.463744	0.749390	58.685000	0.647048	-0.112973	0.690305	0.625373	0.206403
50%	2013.000000	5.386025	9.460323	0.835167	65.199997	0.763476	-0.025393	0.802428	0.722391	0.258117
75%	2017.000000	6.283498	10.352778	0.905291	68.589998	0.856030	0.090967	0.871942	0.799276	0.319716
max	2020.000000	8.018934	11.648169	0.987343	77.099998	0.985178	0.698099	0.983276	0.943621	0.704590

- **Missing Data**

DATA CLEANING AND TRANSFORMATION

- Only Happiness scores through years
- Splitting
- Merging
- Replace missing values

- Column names
- Data type
- Data Format
- Hide Columns

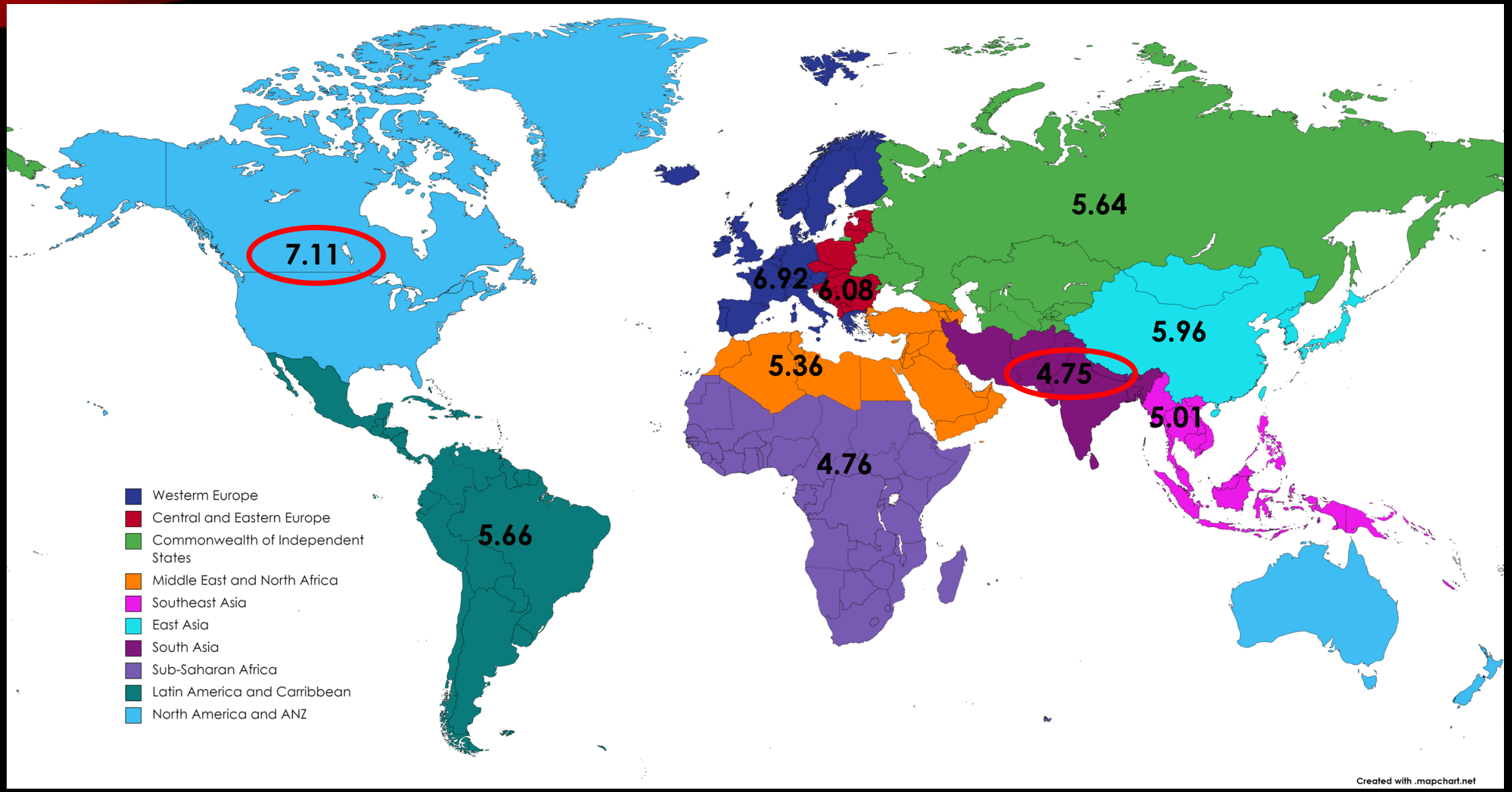


- Calculated Columns
- New measure
- Conditional column

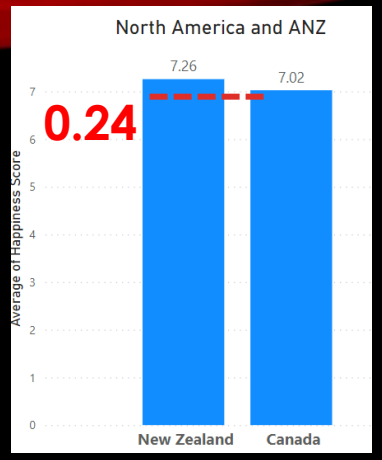
- Regions
- Data for 2021

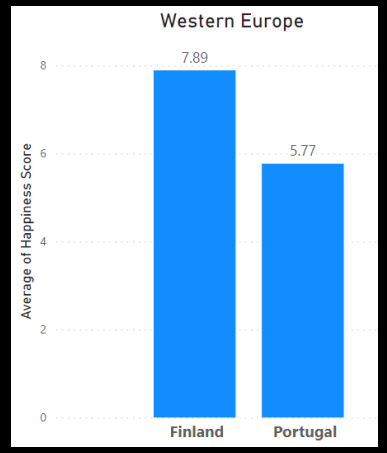
DATA VISUALIZATION

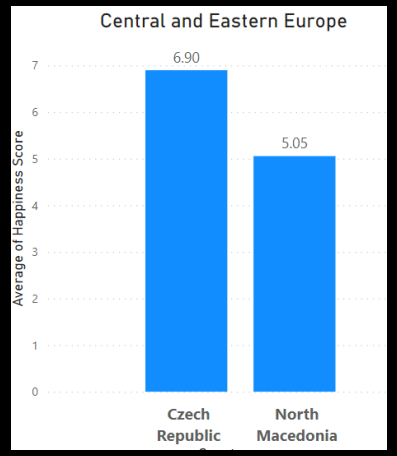
HAPPINESS ACROSS THE WORLD

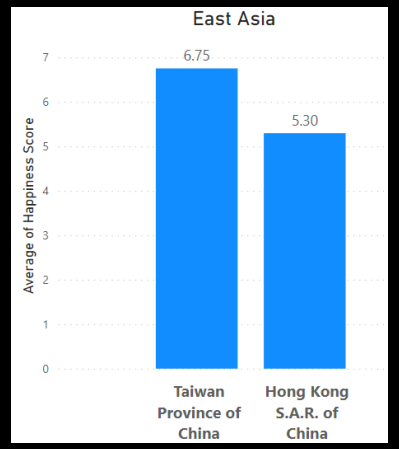


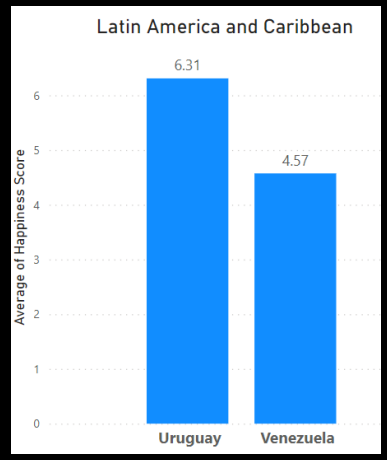
COMPARISON OF HAPPIEST AND UNHAPPIEST COUNTRY IN EACH REGION

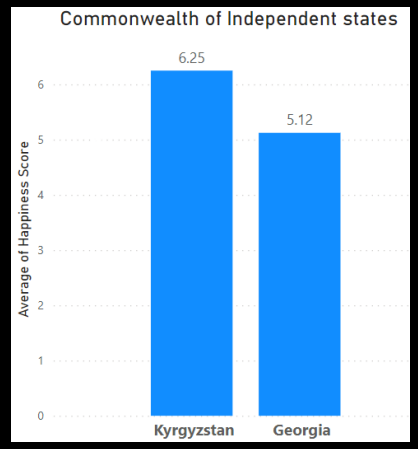


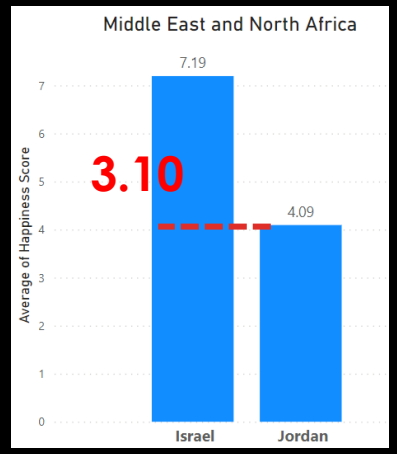


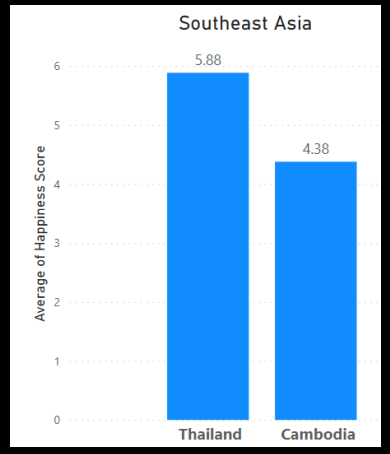


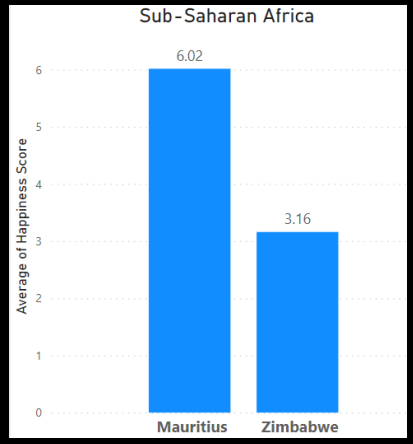


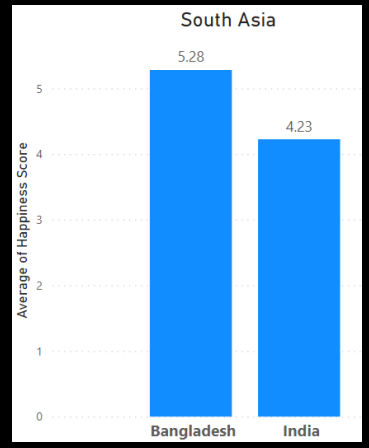






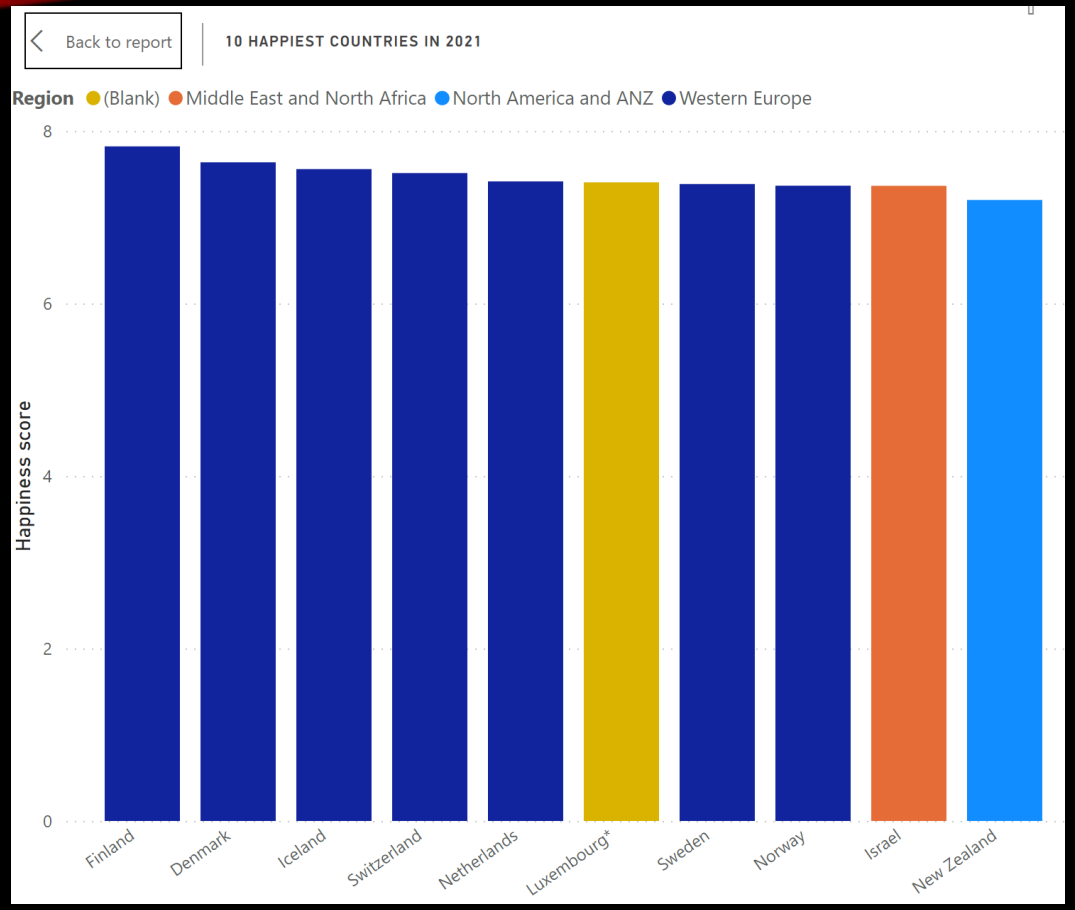


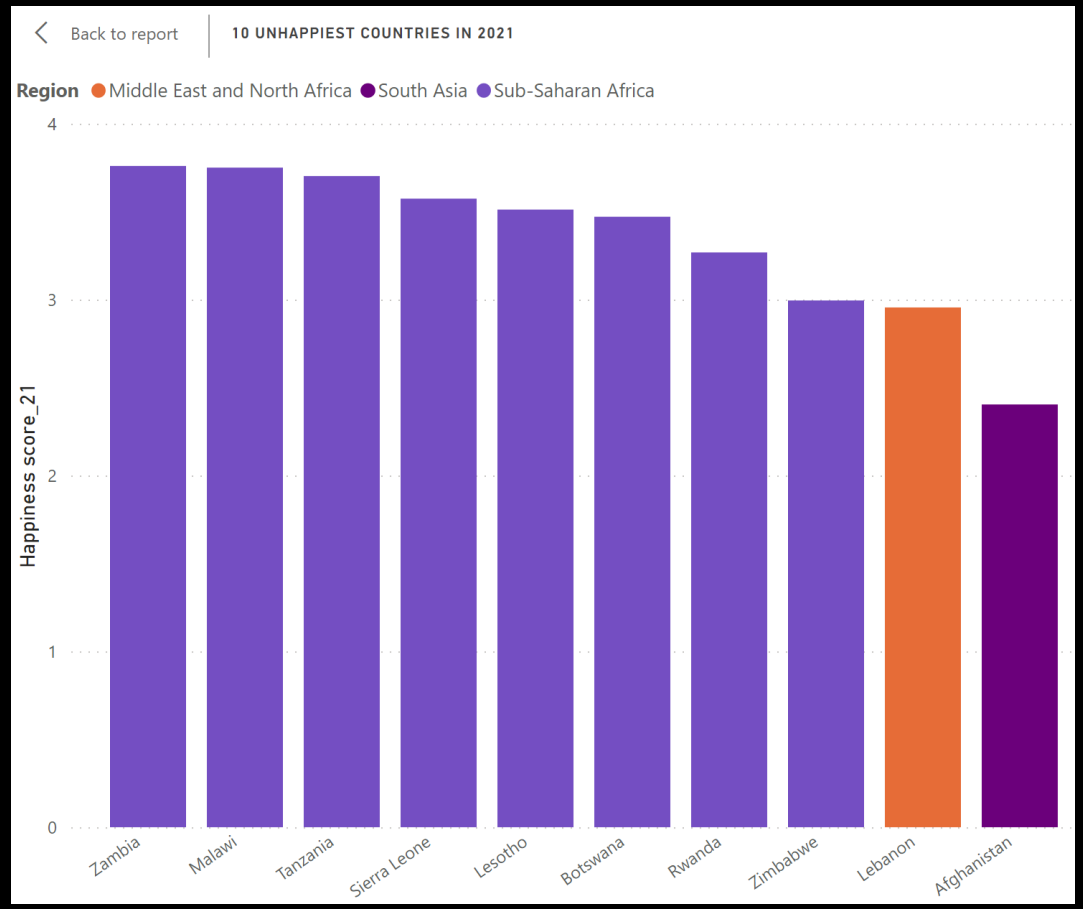




- Difference in Happiness score of the happiest and unhappiest countries in most in Middle East and North Africa and least in North America and ANZ

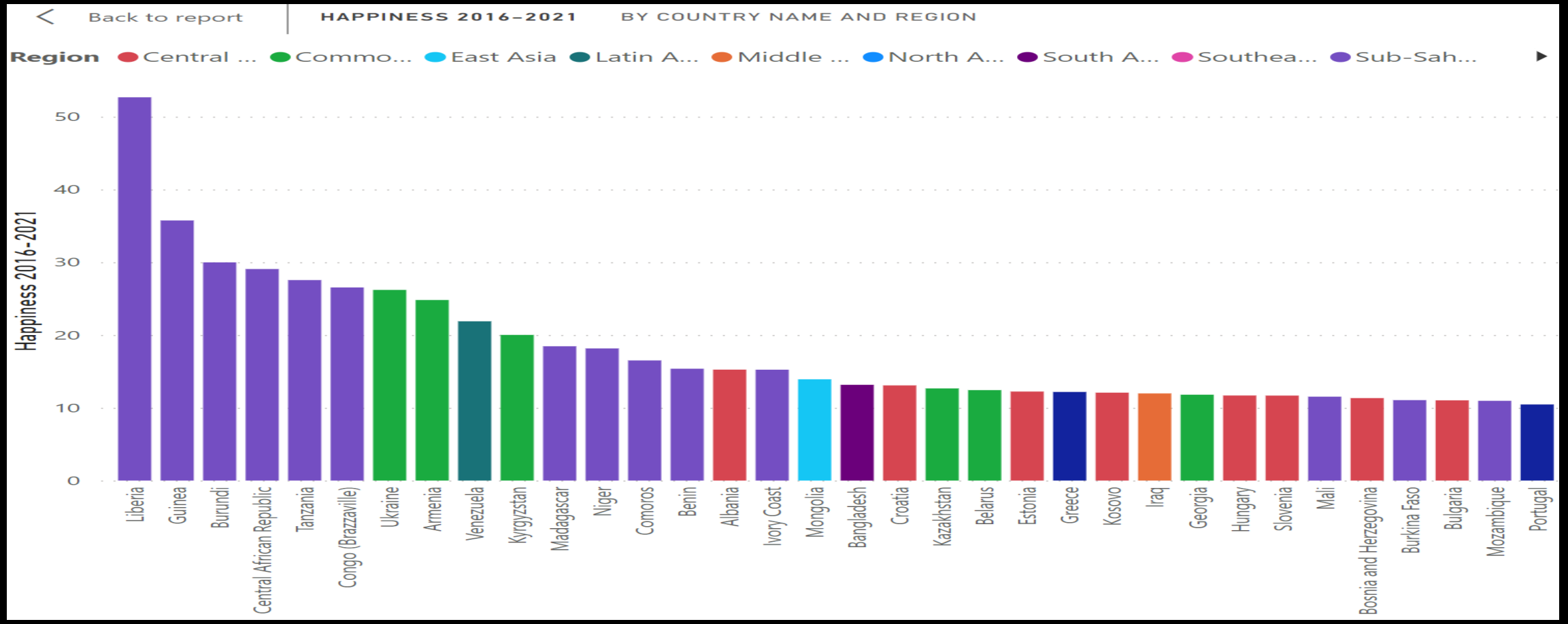
HAPPIEST AND UNHAPPIEST COUNTRIES IN THE WORLD





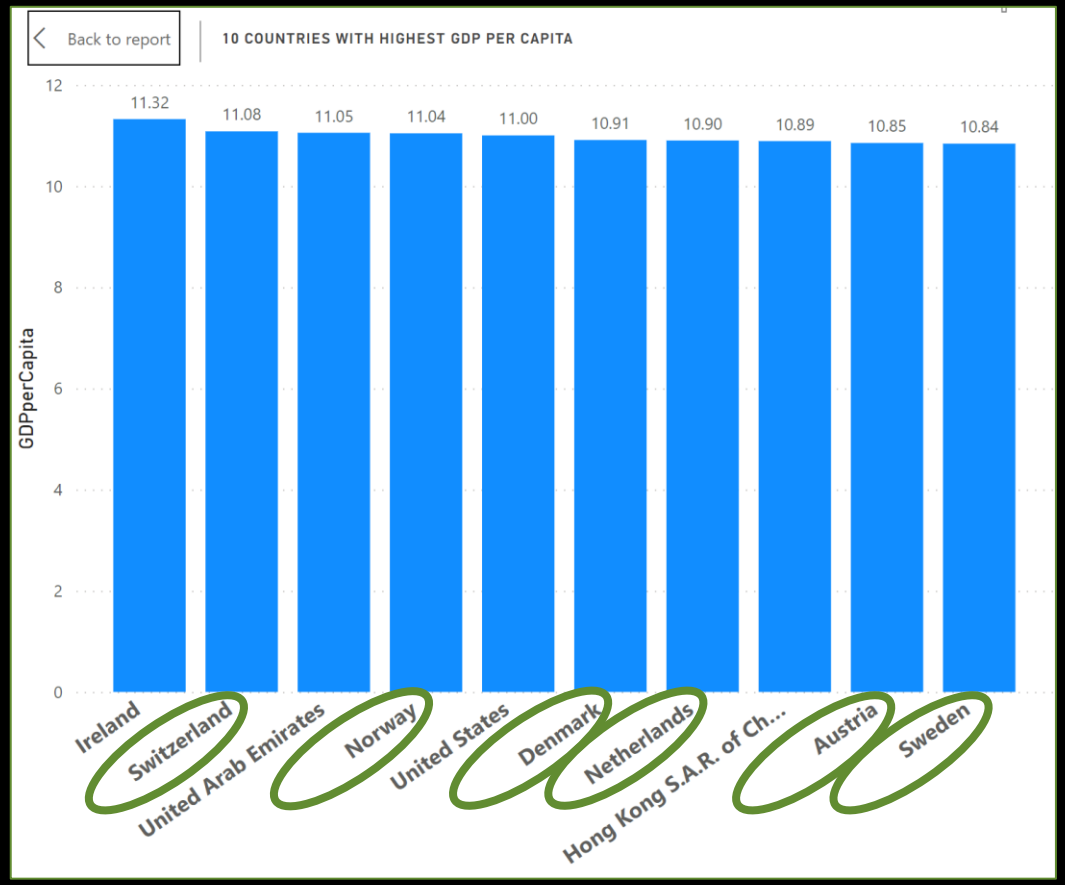
- The Happiest countries are mostly from Western Europe
- The Unhappiest countries are mostly from Sub-Saharan Africa

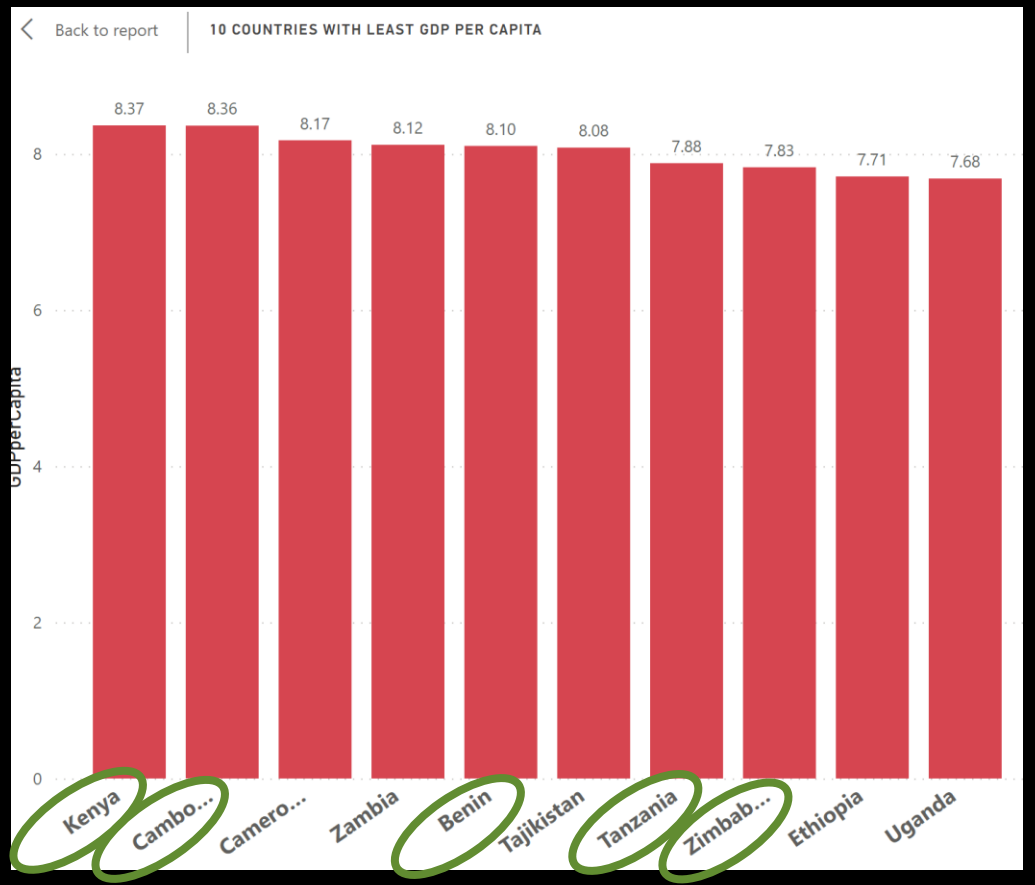
COUNTRIES THAT BECAME HAPPIER IN LAST FIVE YEARS



- There are 104 countries that are happier than they were five years back
- Most of these countries with highest percentage increase are from Sub Saharan Africa

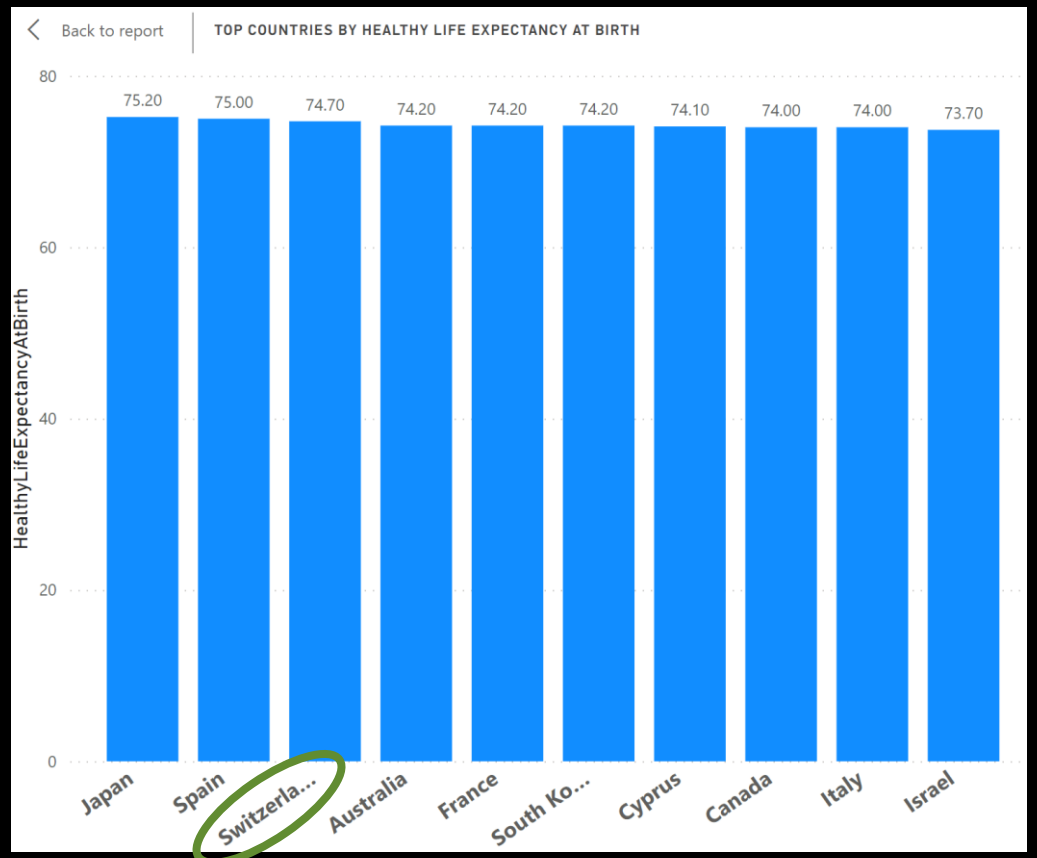
COUNTRIES WITH HIGHEST AND LEAST GDP PER CAPITA

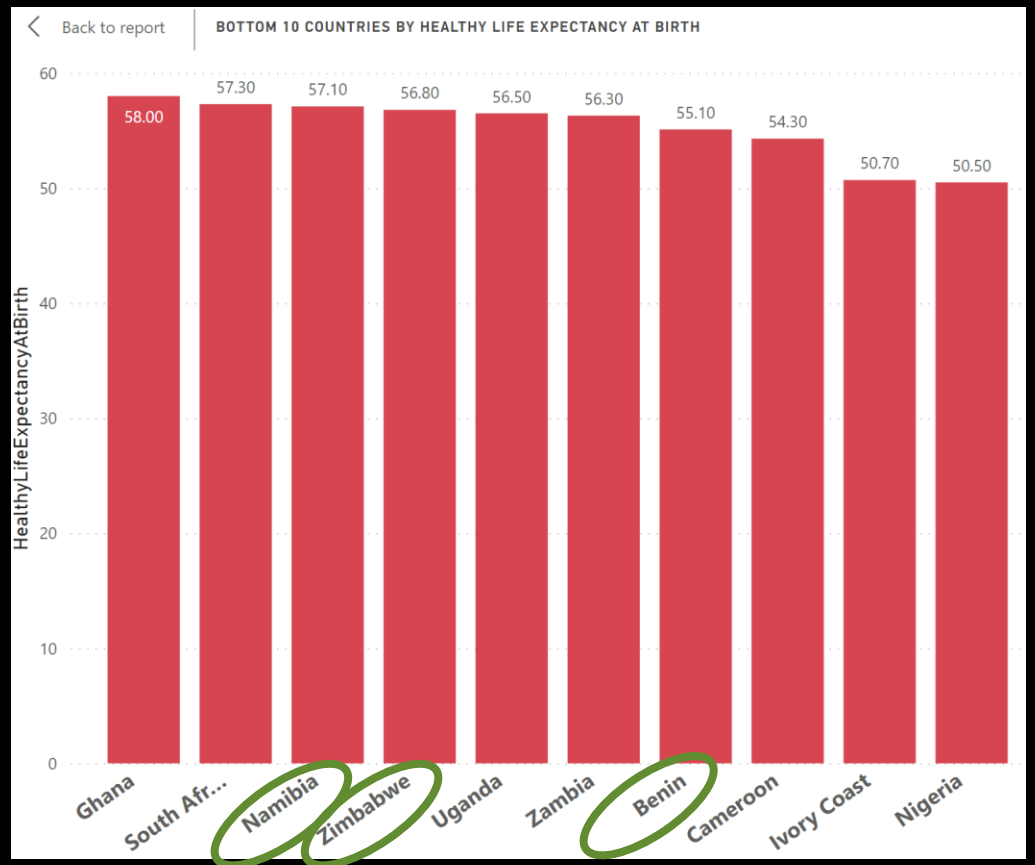




- 6 countries with highest GDP per Capita are also the Happiest countries
- 5 countries with least GDP per Capita are also the unhappiest countries

COUNTRIES WITH HIGHEST AND LEAST HEALTHY LIFE EXPECTANCY AT BIRTH

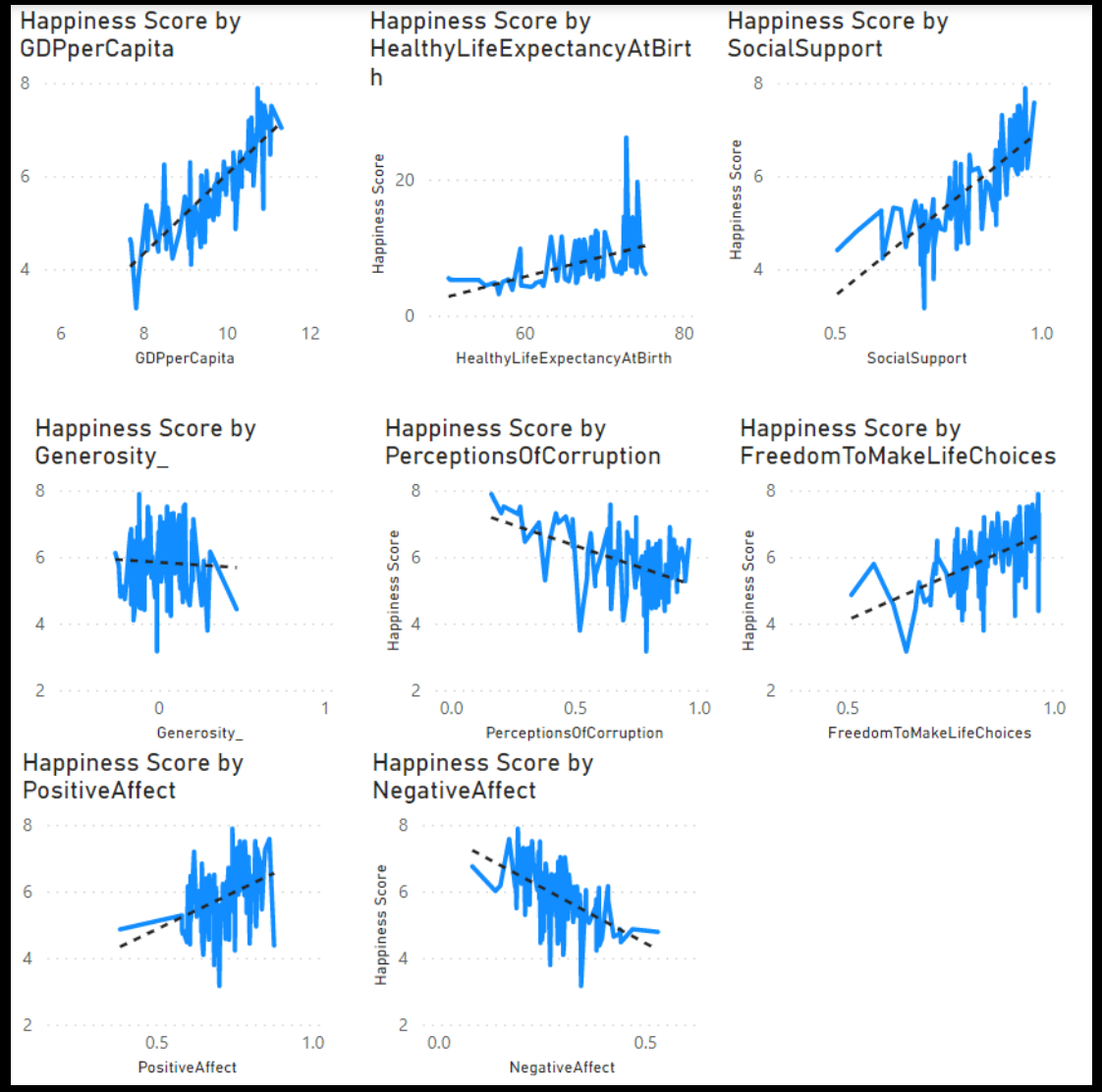




- Only 1 country with highest Healthy Life Expectancy is also the Happiest country
- 3 countries with least Heathy Life Expectancy are the unhappiest countries

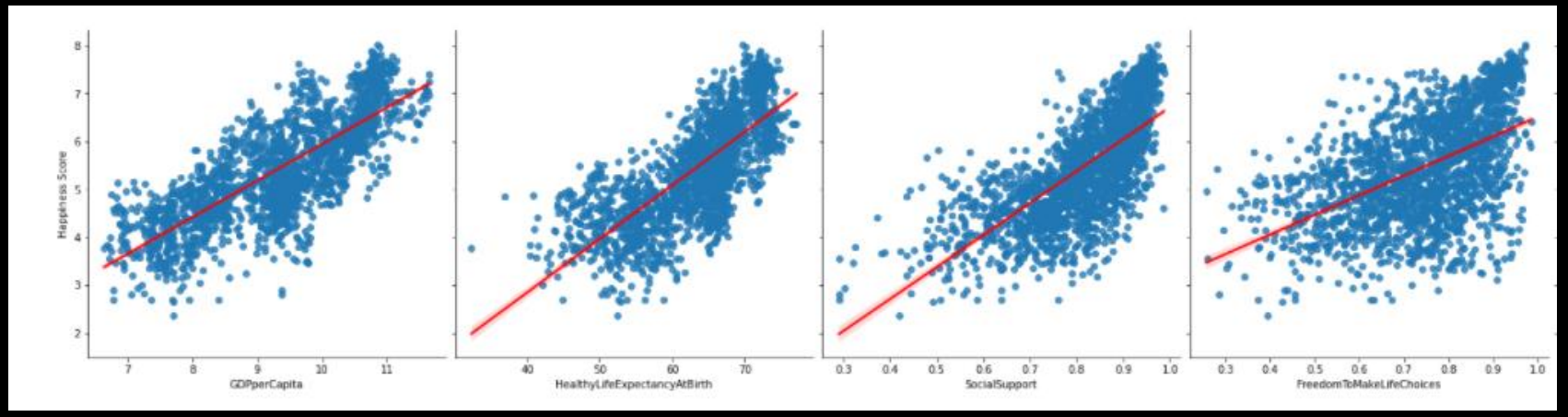
DATA ANALYSIS

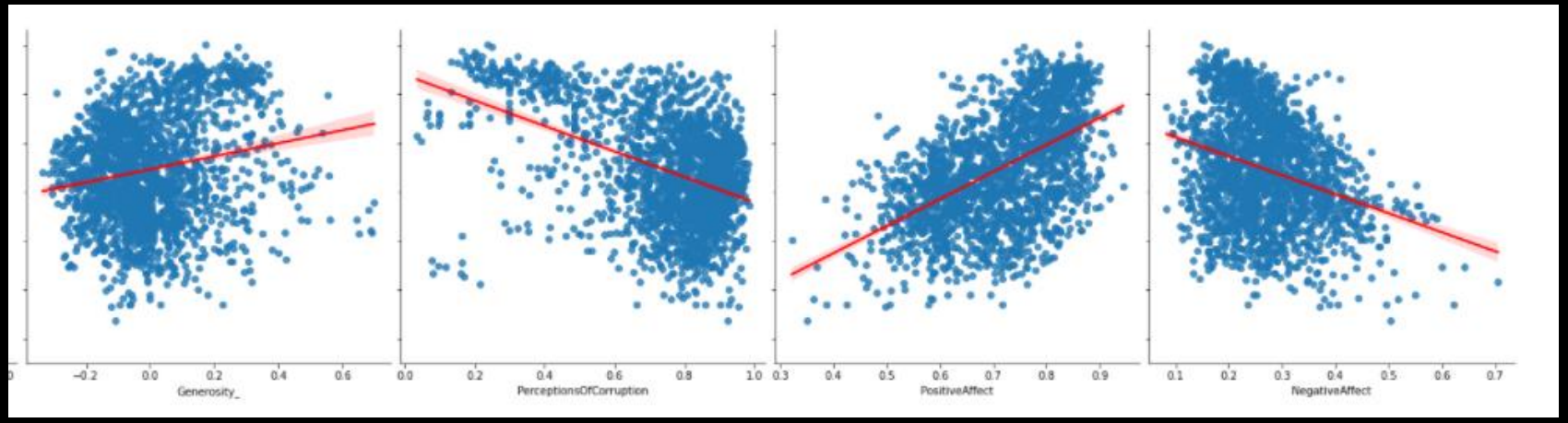
CORRELATION OF ALL FEATURES WITH HAPPINESS SCORE - 2020



All features except Generosity seem to be correlated to happiness score

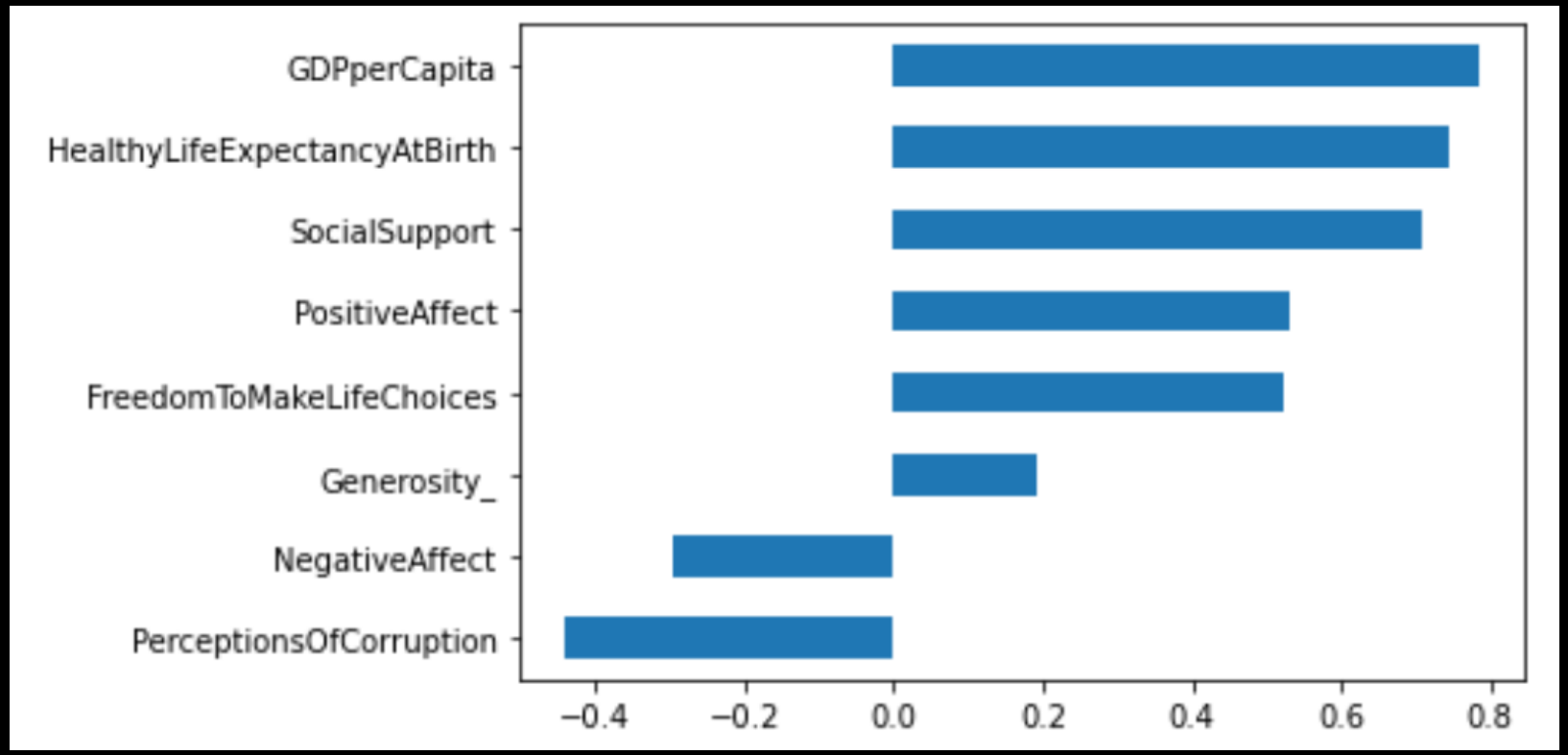
CORRELATION OF ALL FEATURES WITH HAPPINESS SCORE





- All features have correlation with happiness score
- Positive – GDP per Capita, Healthy Life Expectancy, Social Support, Freedom to make life choices, Generosity, Positive Affect
- Negative – Perceptions of corruption, Negative affect

CORRELATION COEFFICIENT OF ALL FEATURES WITH HAPPINESS SCORE



- GDP per Capita, Healthy life expectancy at birth and Social support are highly, positively correlated to happiness score

REGRESSION MODEL BUILDING

1. OLS REGRESSION METHOD

```
X = df[['GDPperCapita', 'HealthyLifeExpectancyAtBirth', 'SocialSupport',
        'FreedomToMakeLifeChoices','Generosity_', 'PerceptionsOfCorruption', 'PositiveAffect', 'NegativeAffect']]
y = df['Happiness Score']
```

In [5]:

```
import statsmodels.api as sm
X = sm.add_constant(X)
ols_model = sm.OLS(y, X)
fitted_model = ols_model.fit()
fitted_model.params
fitted_model.summary()
```

Out[5]:

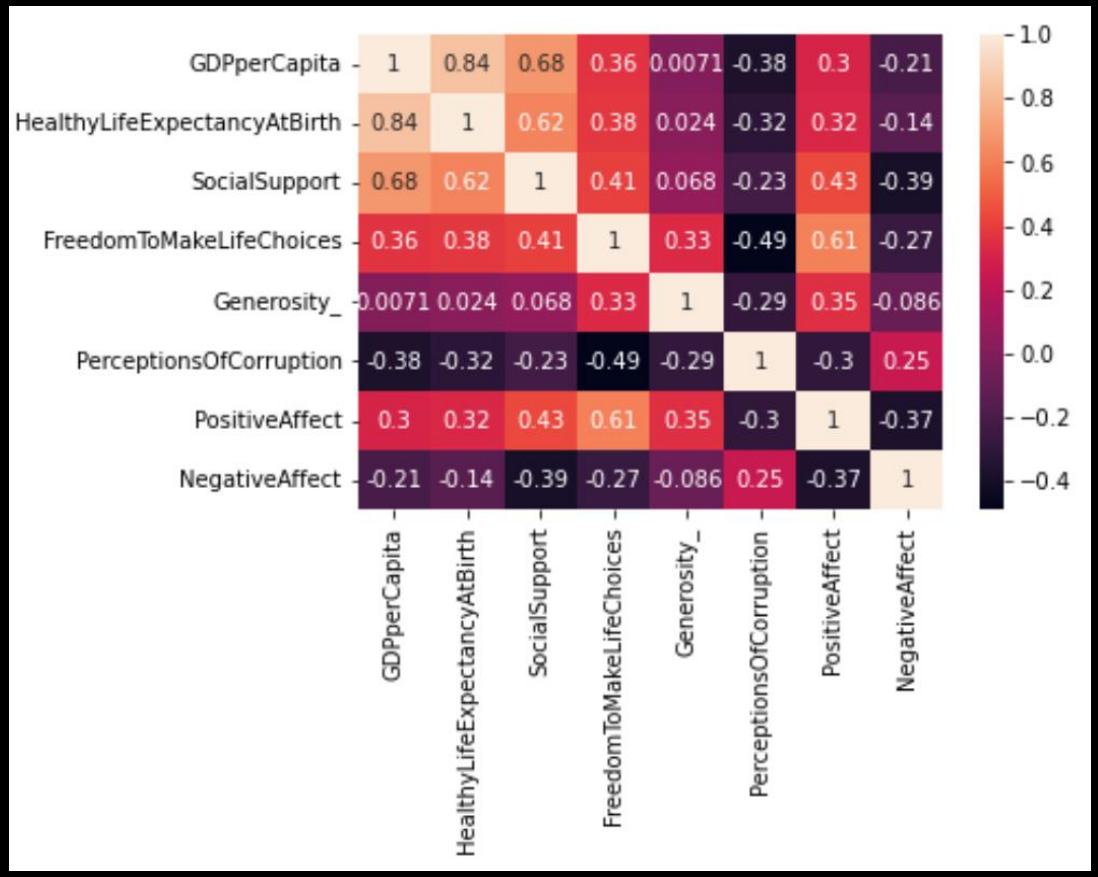
OLS Regression Results

Dep. Variable:	Happiness Score	**R-squared:**	0.762
Model:	OLS	**Adj. R-squared:**	0.762
Method:	Least Squares	**F-statistic:**	778.5
Date:	Fri, 25 Mar 2022	**Prob (F-statistic):**	0.00
Time:	20:52:39	**Log-Likelihood:**	-1577.5
No. Observations:	1949	**AIC:**	3173.
Df Residuals:	1940	**BIC:**	3223.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.6174	0.178	-14.704	0.000	-2.967	-2.268
GDPperCapita	0.3579	0.022	16.073	0.000	0.314	0.402
HealthyLifeExpectancyAtBirth	0.0295	0.003	9.420	0.000	0.023	0.036
SocialSupport	1.9648	0.162	12.134	0.000	1.647	2.282
FreedomToMakeLifeChoices	0.3705	0.123	3.010	0.003	0.129	0.612
Generosity_	0.4204	0.086	4.880	0.000	0.251	0.589
PerceptionsOfCorruption	-0.5755	0.081	-7.070	0.000	-0.735	-0.416
PositiveAffect	1.9826	0.159	12.488	0.000	1.671	2.294
NegativeAffect	0.0253	0.168	0.150	0.880	-0.305	0.356

Omnibus:	43.827	**Durbin-Watson:**	1.910
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	52.830
Skew:	-0.298	**Prob(JB):**	3.37e-12
Kurtosis:	3.544	**Cond. No.**	1.14e+03

CHECKING FOR MULTICOLLINEARITY



GDP per Capita and Healthy life expectancy at birth are highly correlated to each other

2. LINEAR REGRESSION METHOD

In [11]:
```
#Split the Dataset into training and testing set
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.3,random_state=3)

#Instanciate the model
model = LinearRegression()

#Fit the model
model.fit(x_train, y_train)

#read the model score
model.score(x_test, y_test)
```

Out[11]:
```
0.7513193701606652
```

FEATURE IMPORTANCE

```
perm_score = permutation_importance(model, x_train, y_train, n_repeats=40)
perm_score
```

Out[15]:

	features	importance
0	GDPperCapita	0.27711
1	SocialSupport	0.105154
2	PositiveAffect	0.074875
3	HealthyLifeExpectancyAtBirth	0.05196
4	PerceptionsOfCorruption	0.028234
5	Generosity_	0.00477
6	FreedomToMakeLifeChoices	0.00304
7	NegativeAffect	0.000301

GDP per Capita is most important feature and Negative affect is the least important feature

3. DIFFERENT MODELS USING DIFFERENT NUMBER OF FEATURES:

```
#Model 1 with Top 1 Feature
```

```
Out[16]: 0.6086722003586218
```

```
#Model 2 with Top 2 Feature
```

```
Out[17]: 0.6444258270041014
```

```
#Model 3 with Top 3 Feature
```

```
Out[18]: 0.6832358705791305
```

```
#Model 4 with Top 4 Feature
```

```
Out[19]: 0.7181010274173195
```

```
#Model 5 with Top 5 Feature
```

```
Out[20]: 0.7311216305329913
```

```
#Model 6 with Top 6 Feature
```

```
Out[21]: 0.7297239084879932
```

```
#Model 7 with Top 7 Feature
```

```
Out[22]: 0.7517063329486765
```

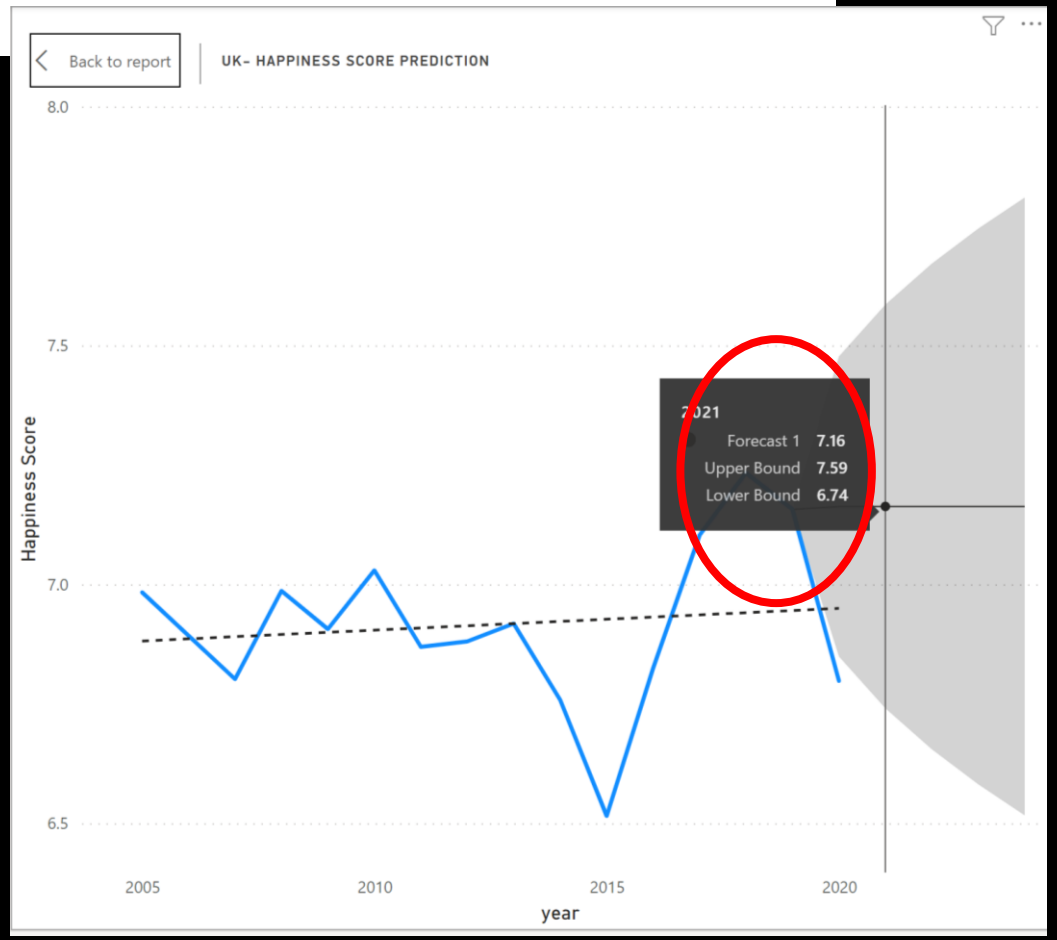
Best model when Negative affect is dropped

PREDICTING HAPPINESS SCORE OF UK FOR 2021

```
In [25]: # Using values for UK for 2021 to predict 2021 UK Happiness and comapre with actual value
         model.predict([[10.715, 75, 1.143, 0.597, 0.289, 0.329, 0.76]])

Out[25]: array([7.35763563])
```

Approx 6% more than the Actual value (6.9425)



HAPPINESS SCORE OF A COUNTRY WITH BEST HISTORICAL VALUES FOR ALL FEATURES

In [23]:
```
a = df['GDPperCapita'].max()
b = df['HealthyLifeExpectancyAtBirth'].max()
c = df['SocialSupport'].max()
d = df['FreedomToMakeLifeChoices'].max()
e = df['Generosity_'].max()
f = df['PerceptionsOfCorruption'].min()
g = df['PositiveAffect'].max()
max_values =[a, b, c, d, e, f, g]
max_values
```

Out[23]:
```
[11.64816856,
 77.0999984741211,
 0.987343490123749,
 0.985177755355835,
 0.69809877872467,
 0.035197988152504,
 0.943620622158051]
```

In [24]:
```
model.predict([[a, b, c, d, e, f, g]])
```

Out[24]:
```
array([8.2501354])
```

CONCLUSIONS

- Happiness score is highest for the region North America and ANZ and least for South Asia in 2021
- Difference in Happiness score of the happiest and unhappiest countries in most in Middle East and North Africa and least in North America and ANZ
- The Happiest countries are mostly from Western Europe and the unhappiest countries are mostly from Sub-Saharan Africa
- Happiness has increased around the world with 102 countries being happier than they were 5 years back and those with highest percentage increase are mostly from Central Europe
- GDP per Capita, Healthy life expectancy at birth and Social support are highly and positively correlated to happiness score
- GDP per Capita and Social Support are most important features for regression modelling
- Best regression model is the one using all features except negative affect
- It is possible to predict a country's Happiness score with 95% confidence if values for all the features are known
- It is also possible to predict a country's Happiness score with 95% confidence using time series analysis in Power BI

THANK YOU!

DO YOU HAVE ANY QUESTIONS?