EM 623 Final Project:

World Happiness Index

Samradnyi Deshmukh, Vineet Kunder, Siddharth Gajare, Ruben Moses Massey

Guided by: Dr. Feng Liu

School of Systems and Enterprises

Stevens Institute of Technology

TABLE OF CONTENT

1) INTRODUCTION

The World Happiness Report 2021 is an annual report that measures happiness levels in countries around the world based on a variety of factors. The report provides valuable insights into the complex nature of well-being and the factors that contribute to it, including economic, social, and environmental factors.

The report includes data on happiness levels in 149 countries, based on surveys conducted by the Gallup World Poll between 2018 and 2020. The dataset contains a range of variables that contribute to happiness and well-being, including GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption.

The World Happiness Report 2021 aims to provide policymakers, researchers, and individuals with the information they need to promote happiness and well-being around the world. The report recognizes that well-being is a complex and multidimensional concept, influenced by a range of factors, including economic, social, and environmental conditions. The report therefore emphasizes the importance of taking a holistic and integrated approach to well-being, and of understanding the interrelationships between different factors that contribute to it.

In this report, we aim to use machine learning algorithms to analyze the World Happiness Report 2021 dataset and gain a deeper understanding of the factors that contribute to happiness and well-being in different countries. We will explore the relationships between these variables and ladder score, which is a measure of happiness, and identify the most important predictors of happiness. This analysis will help us to develop a better understanding of the complex and multidimensional nature of well-being and happiness and develop policies and interventions that promote well-being and happiness on a global scale.

By using machine learning algorithms to analyze the World Happiness Report 2021 dataset, we hope to contribute to the growing body of research on happiness and well-being, and provide insights that can help policymakers, researchers, and individuals to promote happiness and well-being around the world. We hope that our analysis will shed light on the factors that contribute to happiness and well-being in different countries and provide evidence-based guidance for policymakers and practitioners working to improve well-being and happiness in their respective contexts.

1.1) DETAILS OF DATA-SET

The World Happiness Report 2021 dataset contains data on happiness levels in 149 countries around the world, based on surveys conducted by the Gallup World Poll between 2018 and 2020. The dataset includes a range of variables that contribute to happiness and well-being, including:

Country name: The name of the country

Regional indicator: The region to which the country belongs.

Ladder score: The happiness level of the country, based on responses to the Cantril ladder question (0-10 scale)

Standard error of ladder score: The standard error of the ladder score

Upper whisker: The upper end of the confidence interval around the ladder score

Lower whisker: The lower end of the confidence interval around the ladder score

Logged GDP per capita: The natural logarithm of the country's GDP per capita.

Social support: The perceived availability of social support in the country

Healthy life expectancy: The average number of years a person can expect to live in good health.

Freedom to make life choices: The degree to which people feel they have freedom to make life choices.

Generosity: The perceived generosity of the population in the country.

Perceptions of corruption: The perceived level of corruption in the country.

Each of these variables provides valuable information about the factors that contribute to happiness and well-being in different countries. By analyzing these variables using machine learning algorithms, we can gain a deeper understanding of the complex and multidimensional nature of happiness and well-being and develop policies and interventions that promote well-being and happiness on a global scale.

1.2) DATA CLEANING

Reading the dataset: We read the csv file “World Happiness Report 2021” and save it in a variable dataset which is in data frame format.

```
#reading and loading the dataset in the jupyter notebook
data_set= pd.read_csv("C:/Users/vinee/Downloads/world-happiness-report-2021 final excel.csv")
data_set.head()
```

	Country name	Regional indicator	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Explained by: Log GDP per capita	Explained by: Social support
0	Finland	Western Europe	7.842	0.032	7.904	7.780	10.775	0.954	72.0	0.949	-0.098	0.186	1.446	1.106
1	Denmark	Western Europe	7.620	0.035	7.687	7.552	10.933	0.954	72.7	0.946	0.030	0.179	1.502	1.108
2	Switzerland	Western Europe	7.571	0.036	7.643	7.500	11.117	0.942	74.4	0.919	0.025	0.292	1.566	1.079
3	Iceland	Western Europe	7.554	0.059	7.670	7.438	10.878	0.983	73.0	0.955	0.160	0.673	1.482	1.172
4	Netherlands	Western Europe	7.464	0.027	7.518	7.410	10.932	0.942	72.4	0.913	0.175	0.338	1.501	1.079

Refining the Columns: Since the data set is raw and a big one there are certain columns which we won’t require for further analysis, hence we have selected columns in variable “data_set_columns” and copied that in our previously defined variable “dataset”.

Renaming the Columns: For easy use of columns, we have renamed it accordingly and saved in new variable “df_happy1”:

Country name: country_name

Regional indicator: regional_indicator

Ladder score: happiness_score

Logged GDP per capita: logged_percapita

Social support: social_support

Healthy life expectancy: health_life

Freedom to make life choices: freedom_lifechoices

Generosity: generosity

Perceptions of corruption: perception_corruption

```
data_set_columns = ['Country name','Regional indicator','Ladder score','Logged GDP per capita','Social support','Healthy life exp
data_set = data_set[data_set_columns].copy()
df_happy1 = data_set.rename({'Country name':'country_name','Regional indicator':'regional_indicator','Ladder score':'happiness_sc
df_happy1.head()
```

	country_name	regional_indicator	happiness_score	logged_percapita	social_support	health_life	freedom_lifechoices	generosity	perception_corruption
0	Finland	Western Europe	7.842	10.775	0.954	72.0	0.949	-0.098	0.186
1	Denmark	Western Europe	7.620	10.933	0.954	72.7	0.946	0.030	0.179
2	Switzerland	Western Europe	7.571	11.117	0.942	74.4	0.919	0.025	0.292
3	Iceland	Western Europe	7.554	10.878	0.983	73.0	0.955	0.160	0.673
4	Netherlands	Western Europe	7.464	10.932	0.942	72.4	0.913	0.175	0.338

1.3) EXPLORATORY DATA ANALYSIS.

Exploratory Data Analysis (EDA) is the process of examining and analyzing data sets to summarize their main characteristics, often with visual methods. EDA helps to identify patterns, relationships, and anomalies in data and to gain a deeper understanding of the data.

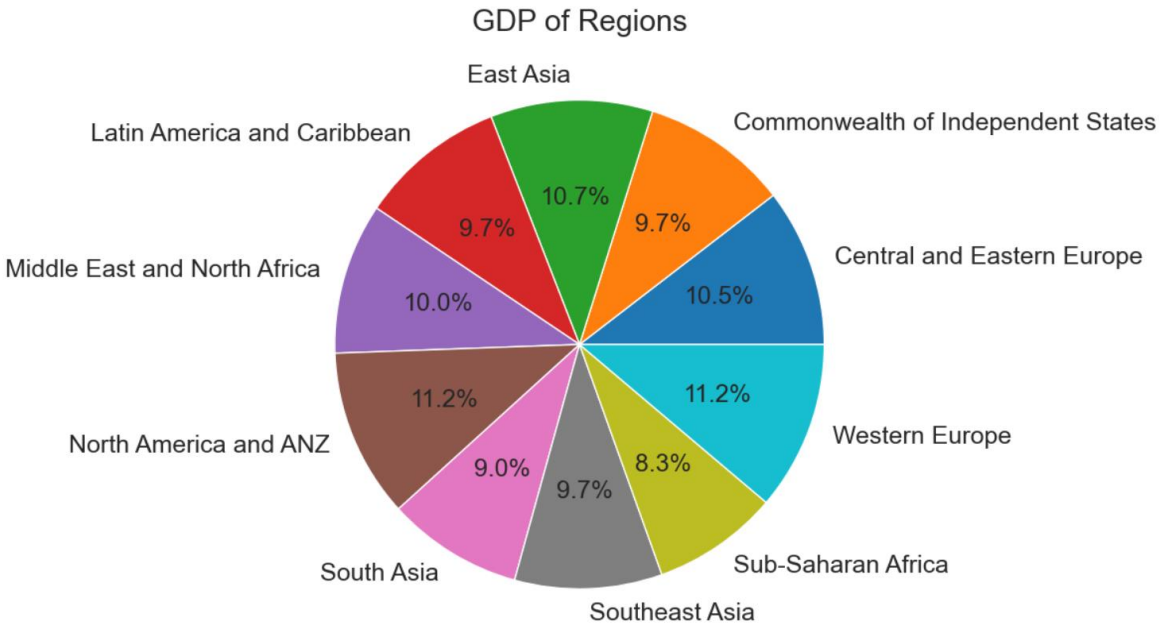
In [10]:
```
#Viewing the changes to the dataset made
df_happy1.head()
#type(data_set)
df_happy1.shape
df_happy1.info()
df_happy1.describe()
#df_happy1.isnull().sum() #checking whether any column has null or blank; if Boolean False appears in table which infers no null
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149 entries, 0 to 148
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   country_name          149 non-null    object 
 1   regional_indicator    149 non-null    object 
 2   happiness_score       149 non-null    float64
 3   logged_percapita      149 non-null    float64
 4   social_support        149 non-null    float64
 5   health_life           149 non-null    float64
 6   freedom_lifechoices   149 non-null    float64
 7   generosity            149 non-null    float64
 8   perception_corruption  149 non-null    float64
dtypes: float64(7), object(2)
memory usage: 10.6+ KB
```

In this we understand the type of each column and if there are any null values in any of the columns.

2) DATA ANALYSIS

2.1) GDP of regions - Pie Chart



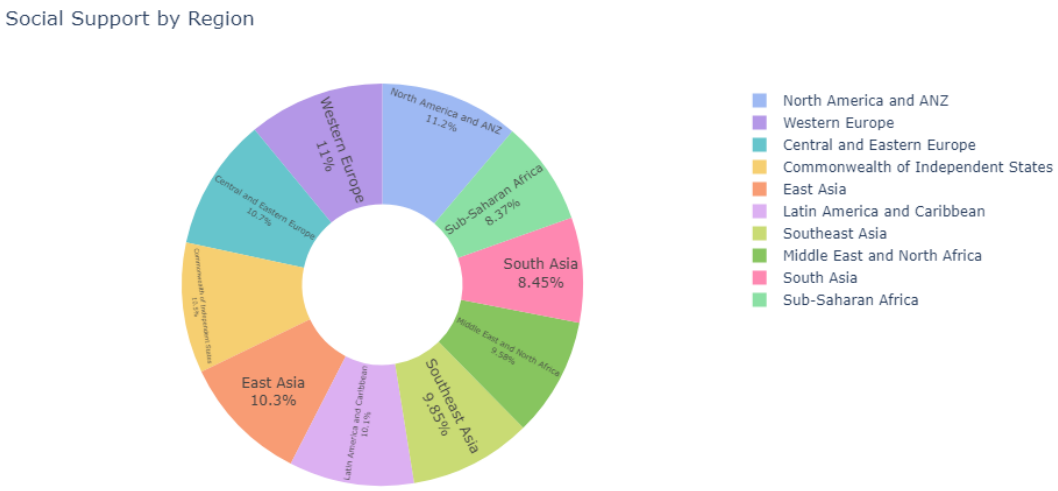
The pie chart shows the percentage breakdown of GDP across different regions using data from the World Happiness Report 2021. The regions included in the chart are Western Europe, North America and ANZ, Latin America and Caribbean, Central and Eastern Europe, East Asia, Southeast Asia, South Asia, Middle East and North Africa, and Sub-Saharan Africa

The region with the highest percentage of GDP is North America and ANZ along with Western Europe at 11.2%. The region with the lowest percentage of GDP is Sub Saharan Africa, which accounts for only 8.3% of the total GDP.

The data used in the pie chart may have some limitations, such as missing data or outliers. Moreover, the pie chart only shows the percentage breakdown of GDP across different regions, and does not provide information on other important factors that may influence economic development, such as education, healthcare, or political stability.

The findings in the pie chart have several implications for economic development and social welfare. For example, regions with higher GDP may have better access to resources and infrastructure, which can contribute to better living standards and higher levels of well-being. On the other hand, regions with lower GDP may face greater challenges in addressing poverty, inequality, and other social issues.

2.2) Donut Graph



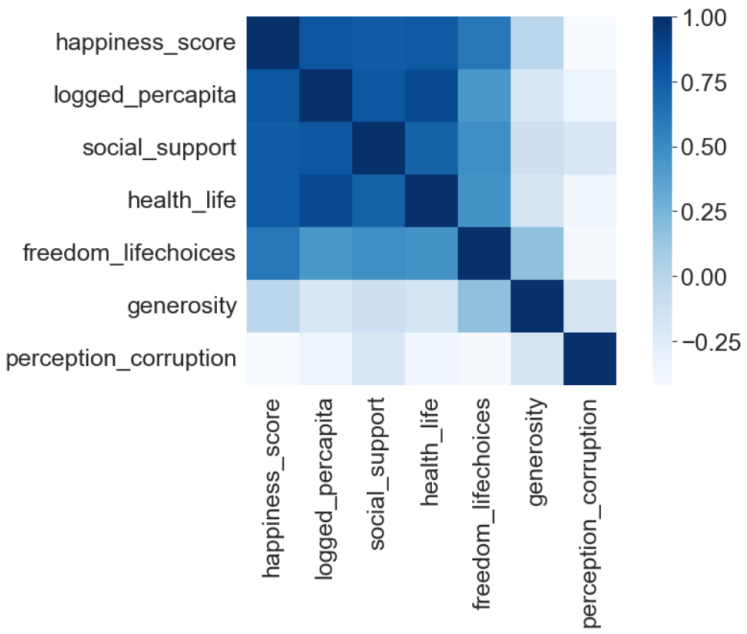
The donut graph visually presents the distribution of social support across different regions.. It offers insights into the varying levels of social support experienced by individuals in each region, highlighting the importance of social connections and networks in overall well-being.

High Social Support Regions: North America and ANZ region has the highest social support score at 11.2%. This region demonstrates a strong sense of community, close-knit relationships, and robust support systems.

Low Social Support Regions: Sub Saharan Africa region has the lowest social support score at 8.37%. In this region individuals may face challenges in accessing social networks, experiencing a sense of isolation, or lacking strong support systems.

By analyzing the donut graph representing social support by region, we can gain insights into the importance of social connections and networks in overall well-being. These insights can guide policymakers and stakeholders in designing strategies and interventions that strengthen social support systems and promote a sense of belonging and community for individuals within their respective regions.

2.3) Correlation Matrix.



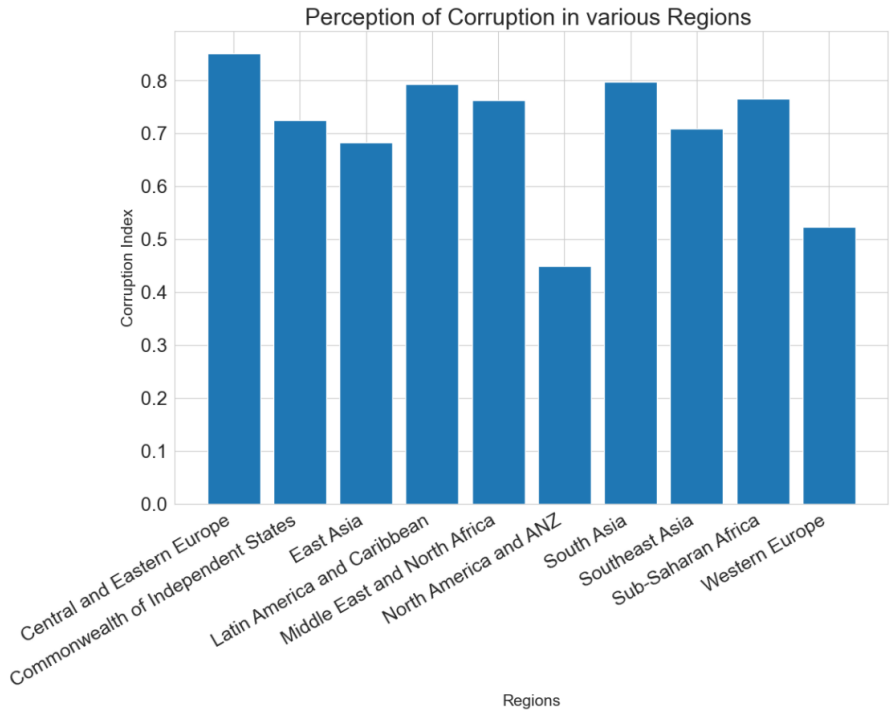
Correlation between variables: The correlation matrix shows the correlation coefficients between different variables in the dataset. Correlation coefficients range from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. Positive correlations indicate that as one variable increases, the other variable tends to increase as well, while negative correlations indicate that as one variable increases, the other variable tends to decrease.

Strong positive correlations: From the correlation matrix, you can observe that there are several variables that have strong positive correlations with the happiness score. For example, GDP per capita, social support, and life expectancy all have correlation coefficients greater than 0.7, indicating a strong positive relationship with happiness.

Weak or negative correlations: There are also some variables that have weak or negative correlations with happiness. For example, the generosity and Freedom to make life choices variables have weak correlations with happiness, while the Perception of corruption variable has a negative correlation with happiness.

It's worth noting that some variables in the dataset are highly correlated with each other, indicating multicollinearity. For example, GDP per capita and healthy life expectancy have a correlation coefficient of 0.8, suggesting that they might be measuring similar aspects of well-being.

2.4) Perception of corruption among regions - Bar Chart.



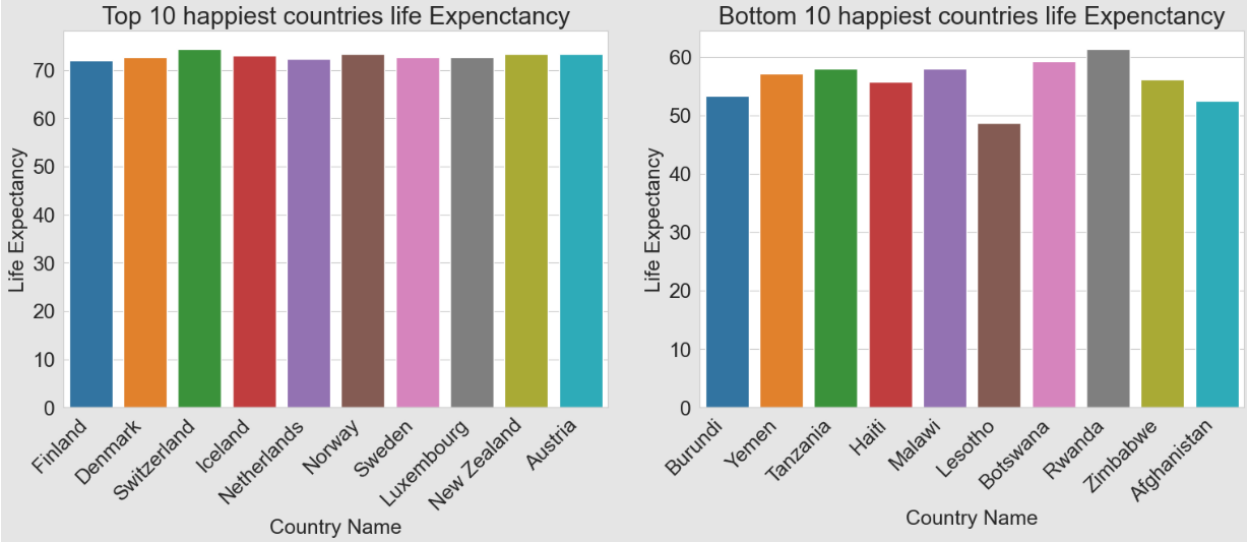
The bar graph provides a visual representation of the perception of corruption in different regions. It serves as a valuable tool for comparing and understanding the varying levels of corruption perception across regions.

The graph highlights notable regional differences in the perception of corruption, shedding light on areas where corruption is believed to be more prevalent or less prevalent. By examining the graph, we can identify regions with the highest perception of corruption, indicating a pressing need for targeted efforts to address corruption-related issues in those areas. Similarly, regions with the lowest perception of corruption can serve as positive examples or models for implementing effective anti-corruption measures.

For instance, the Central and Eastern Europe Region exhibits the highest perception of corruption among all regions, indicating a significant challenge that needs to be addressed urgently. On the other hand, the North America and ANZ Region demonstrates the lowest perception of corruption, suggesting the presence of robust anti-corruption measures and a high level of public trust in institutions.

It is crucial to note that the bar graph reflects the perception of corruption rather than the actual occurrence. It represents the opinions and beliefs of individuals within each region and may not always align with the objective reality of corruption levels. Factors such as media coverage, public awareness, and cultural context can influence the perception of corruption.

2.5) Top and Bottom happiest countries and their Life expectancy.



The bar graph visually presents the life expectancy of the top 10 happiest and unhappiest countries. It allows us to compare and understand the differences in life expectancy among these countries, providing insights into the relationship between happiness and life expectancy.

Top 10 Happiest Countries:

The graph showcases the life expectancy of the top 10 happiest countries. These countries have consistently ranked high in happiness scores, reflecting positive well-being and satisfaction among their populations. From the graph, we can observe the variations in life expectancy among these countries.

Highest Life Expectancy: Among the top 10 happiest countries, Switzerland has the highest life expectancy, indicating the presence of favorable healthcare systems, quality of life, and other contributing factors that promote longevity.

Bottom 10 happiest Countries:

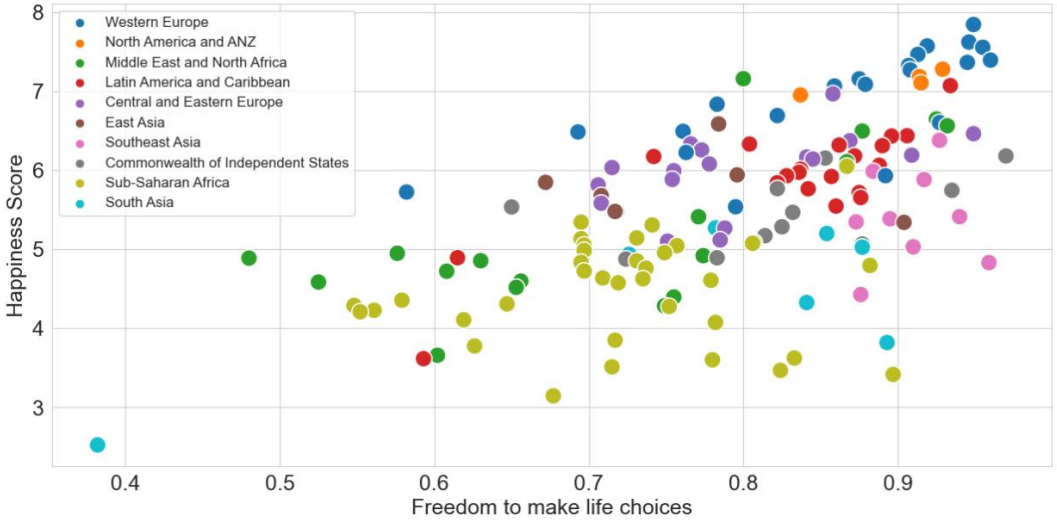
The graph also presents the life expectancy of the top 10 unhappiest countries. These countries have consistently ranked lower in happiness scores, indicating challenges in well-being and life satisfaction. Analyzing their life expectancy provides insights into potential factors contributing to lower happiness levels.

Lowest Life Expectancy: Within the top 10 unhappiest countries, Lesotho has the lowest life expectancy. This highlights a critical area of concern that may require targeted interventions to improve healthcare access and overall well-being.

It's important to note that life expectancy is influenced by various factors beyond happiness scores. While happiness and life expectancy may have some correlation, additional variables like healthcare systems, socio-economic conditions, and public health policies play significant roles.

By analyzing the life expectancy of the top 10 happiest and unhappiest countries, we can gain insights into the well-being and potential areas for improvement within these nations. This information can inform policymakers and stakeholders in implementing targeted strategies to enhance healthcare, improve quality of life, and address the underlying factors that influence happiness and life expectancy.

2.6) Happiness score based on freedom to make life choices.



The scatter plot visualizes the relationship between happiness scores and freedom to make life choices across different countries. It provides a visual representation of how these two variables are associated and offers insights into the connection between individual freedom and well-being.

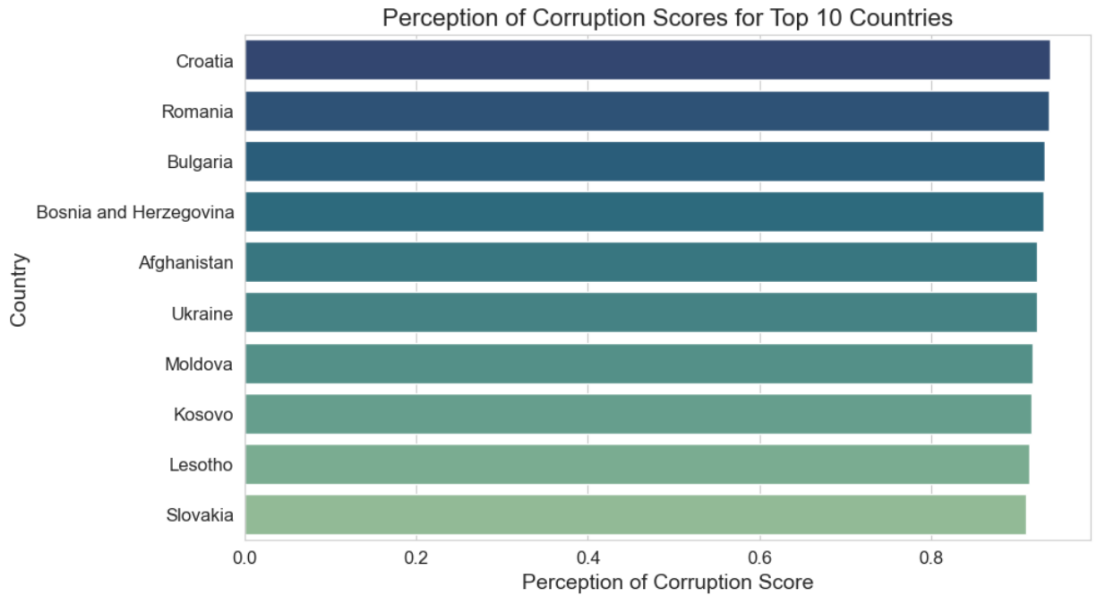
Positive Correlation: The scatter plot demonstrates a positive correlation between happiness scores and freedom to make life choices. Countries where individuals have greater freedom to make life choices tend to have higher happiness scores. This suggests that personal autonomy and the ability to make decisions aligned with one's values and preferences positively impact overall well-being.

Linear Trend: The scatter plot may show a linear trend, indicating a direct relationship between happiness scores and freedom to make life choices. As the level of freedom increases, there is a tendency for happiness scores to increase as well, and vice versa. This suggests that the more freedom individuals have in making life choices, the higher their reported happiness tends to be.

Variations within the Relationship: While a positive correlation exists, the scatter plot also reveals variations within the relationship between happiness scores and freedom to make life choices. Not all countries with high levels of freedom exhibit correspondingly high happiness scores, and vice versa. This indicates that other factors, such as social support, economic conditions, and cultural aspects, also contribute to overall happiness levels.

Outliers: The scatter plot may display outliers—countries that deviate from the general pattern. These outliers indicate countries that exhibit unexpectedly high or low happiness scores relative to their level of freedom to make life choices. Examining these outliers can provide valuable insights into the unique factors influencing happiness in those specific countries.

2.7) Perception of Corruption in Countries.



The horizontal bar graph provides a visual representation of the perception of corruption in different countries, offering insights into the levels of perceived corruption and transparency across nations.

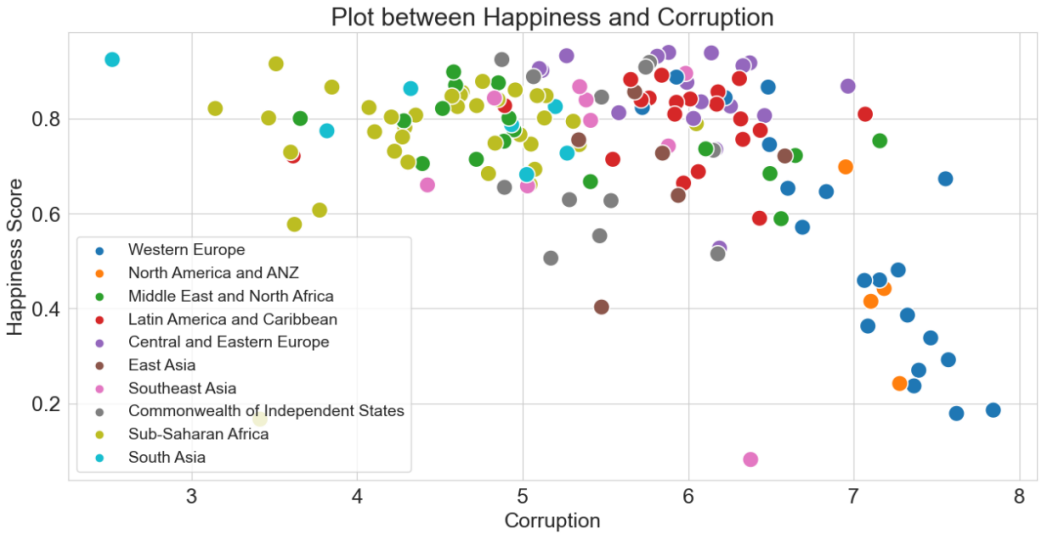
Country Comparison: The horizontal bar graph allows for easy comparison of the perception of corruption across different countries. By examining the lengths of the bars, one can identify countries with higher and lower perceived levels of corruption.

High Corruption Perception: Croatia has the highest Perception of Corruption.This country faces significant challenges in terms of transparency, accountability, and public trust in their institutions. Identifying countries with higher corruption perception can serve as a starting point for targeted efforts to combat corruption and promote integrity.

Low Corruption Perception: Slovakia has the least Perception of Corruption among the top 10. This country is perceived to have higher levels of transparency, lower corruption instances, and greater public trust in their institutions. Examining countries with lower corruption perception can provide insights into successful anti-corruption measures and serve as positive examples for others.

By analyzing the horizontal bar graph representing the perception of corruption in countries, we can gain insights into the perceived transparency and accountability within nations. These insights can guide policymakers, governments, and stakeholders in formulating anti-corruption strategies, promoting good governance, and fostering public trust in institutions to combat corruption effectively.

2.8) Happiness and Corruption Plot.



The scatter plot visualizes the relationship between happiness and corruption perceptions across different countries. It provides insights into the potential association between these two variables and sheds light on how corruption perception relates to overall happiness levels.

Inverse Relationship: The scatter plot may reveal an inverse relationship between happiness and corruption perceptions. As corruption perceptions increase, happiness scores tend to decrease, indicating that higher levels of perceived corruption are associated with lower levels of happiness. This suggests that corruption can have detrimental effects on overall well-being and societal satisfaction.

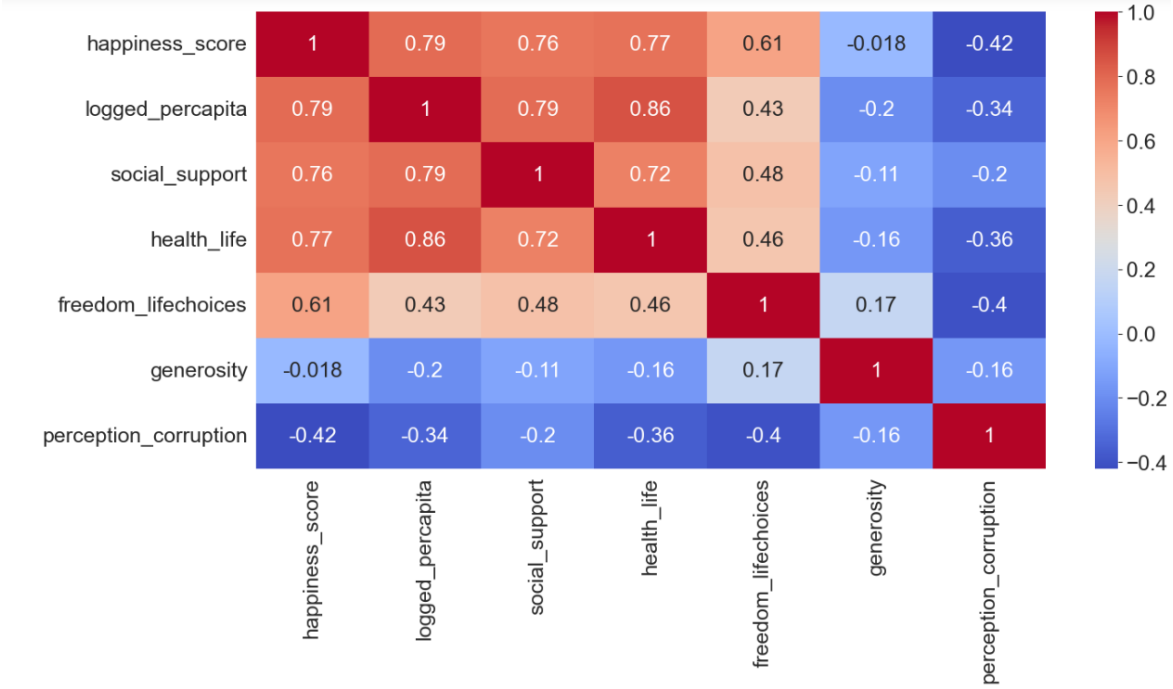
Scatter Distribution: The scatter plot displays the distribution of countries across the happiness and corruption perception spectrum. It allows for the identification of countries that deviate from the general pattern and provides insights into unique cases and potential outliers.

Outliers: The scatter plot may include outliers—countries that exhibit unexpected combinations of happiness and corruption perceptions. These outliers represent countries that defy the general trend, suggesting that factors beyond corruption perception alone contribute to happiness levels. Analyzing these outliers can offer valuable insights into the complex interplay between corruption, happiness, and other societal factors.

Correlation Strength: By examining the dispersion of data points on the scatter plot, one can assess the strength of the correlation between happiness and corruption perceptions. A tighter cluster of points indicates a stronger correlation, while a more scattered distribution suggests a weaker or more nuanced relationship.

3) Machine learning models and Performance.

3.1) Random Forest Regressor



The Random Forest Regressor model was effective in predicting the happiness score of countries. The model achieved an R-squared value of 0.78, indicating a good fit. The predicted happiness scores for 2022 were added to the dataset, and the correlation between the features and the target variable was visualized using a heatmap.

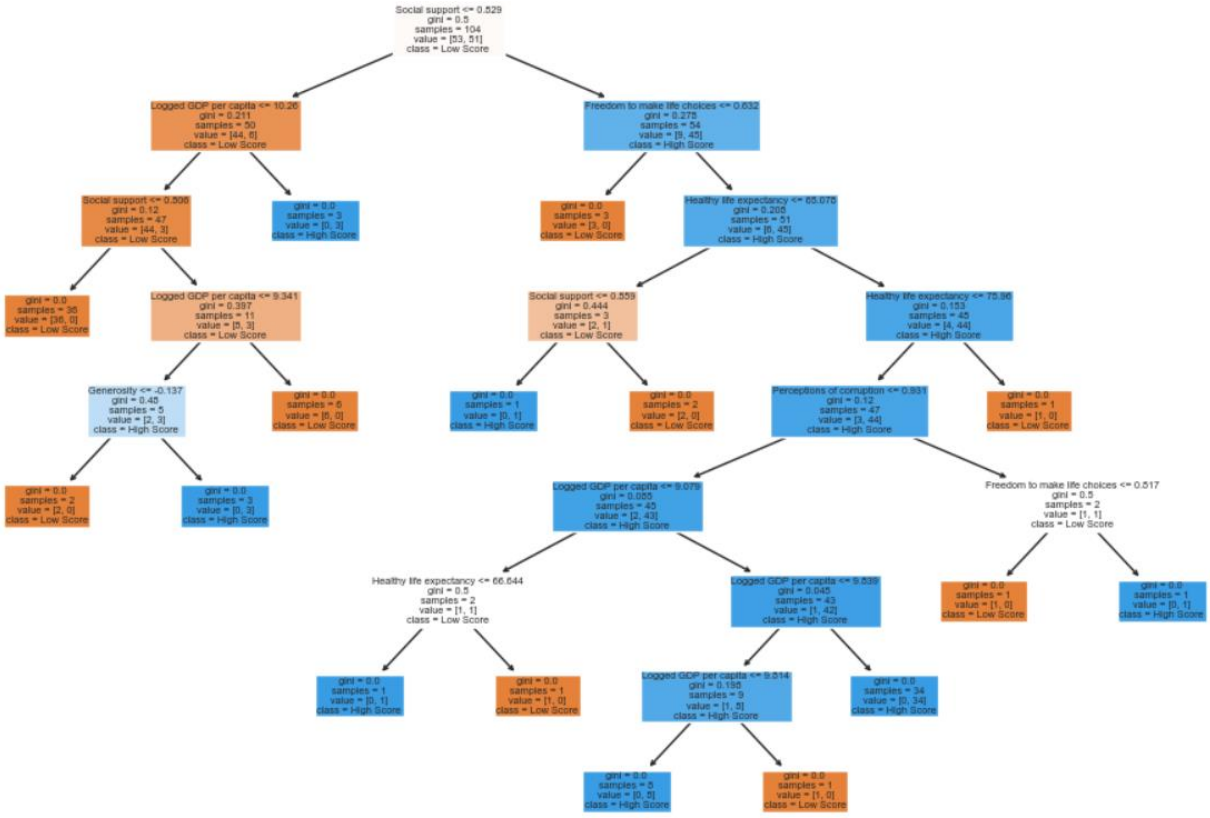
The heatmap showed that all features have a positive correlation with the happiness score except for perceptions of corruption, which has a negative correlation. Among the features, logged GDP per capita has the highest correlation with the happiness score, followed by social support and healthy life expectancy. These findings suggest that economic and social factors have a significant impact on the happiness of individuals.

In conclusion, we successfully predicted the happiness score of countries for the year 2022 using a Random Forest Regressor model. The model achieved a good fit and provided insights into the factors that influence the happiness of individuals. Our findings suggest that economic and social factors play a significant role in determining the happiness of individuals. Our study can help policymakers and organizations understand the factors that contribute to the happiness of individuals and take appropriate measures to promote happiness.

3.2) Decision Tree Model

The World Happiness Report dataset's decision tree was created to evaluate and comprehend the elements that affect happiness levels. It strives to identify the most crucial elements and the connections between them in order to forecast happiness levels or classify nations according to their level of happiness. The World Happiness Report dataset's numerous properties or attributes are taken into account by the decision tree model. These characteristics may include GDP per capita, social support, the ability to choose one's course in life, and life expectancy.Based on the splits in the tree, the decision tree model creates decision rules. These guidelines shed light on the circumstances or arrangements of characteristics that affect people's levels of happiness. For instance, the model might demonstrate that nations with high levels of social support, low perceptions of corruption, and high GDP per capita tend to. The accuracy given by Decision Tree Model 88.88%

```
Accuracy: 0.8888888888888888
```



3.3) Random Forest Model.

The decision tree for the World Happiness Report dataset was developed to assess and comprehend the factors that influence happiness levels. In order to predict happiness levels or categorize countries based on their degree of happiness, it works to determine the most important components and the relationships among them. The decision tree model considers many aspects or attributes of the World Happiness Report dataset. The ability to select one's path in life, social support, GDP per capita, and life expectancy are a few examples of these traits. The decision tree model generates decision rules based on the splits in the tree. These recommendations give light on the situations or combinations of traits that influence people's degrees of happiness. For example, the model could show that countries with high levels of social support, low perceptions of corruption, and high GDP per capita. R2 is 76.20%

```
# Random Forest
# Select features and target variable
X = df_happy1[['logged_percapita', 'social_support', 'health_life', 'freedom_lifechoices', 'generosity', 'perception_corruption']]
y = df_happy1['happiness_score']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

# Create a Random Forest Regressor object
clf = RandomForestRegressor()

# Train the model using the training sets
clf.fit(X_train, y_train)

# Predict the target variable for the test set
y_pred = clf.predict(X_test)

# Model Accuracy, how well does the regressor predict the target variable?
print("R2 score:", metrics.r2_score(y_test, y_pred))
```

```
R2 score: 0.7620990574636601
```

3.4) K-Nearest Neighbors (KNN) Model.

The World Happiness Report dataset's KNN model was developed to study and comprehend the elements that affect happiness levels. Based on how closely a country's feature values resemble those of its neighbors, it seeks to predict happiness ratings or categorize nations into distinct happiness categories.By predicting a country's happiness level based on the feature values of its k closest neighbors, the KNN model employs a nearest neighbor technique. The task at hand and the dataset are used to determine the value of k. The model captures local patterns in the data by taking the commonalities between nations into account..The KNN model relies on a distance metric to determine the similarity between feature vectors of different countries. It's important to evaluate the performance of the KNN model to ensure its reliability and generalizability. Metrics such as accuracy, precision, recall, and F1 score can assess how well the model predicts happiness levels or classifies countries into happiness categories. Model evaluation helps determine the effectiveness of the KNN model in capturing patterns and relationships in the World Happiness Report data.

```
# Importing required libraries

# Select features and target variable
X = df_happy1[['logged_percapita', 'social_support', 'health_life', 'freedom_lifechoices', 'generosity', 'perception_corruption']
y = df_happy1['happiness_score']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

# Create a KNN Regressor object
clf = KNeighborsRegressor(n_neighbors=5)

# Train the model using the training sets
clf.fit(X_train, y_train)

# Predict the target variable for the test set
y_pred = clf.predict(X_test)

# Model Accuracy, how well does the regressor predict the target variable?
print("R2 score:", metrics.r2_score(y_test, y_pred))
```

```
R2 score: 0.6550073969870142
```

4) Discussion

The report found that the most important factors that contribute to happiness are GDP per capita, social support, healthy life expectancy, freedom to make life choices, and generosity. These factors are all important for ensuring that people have the necessities of life, such as food, shelter, and healthcare. They also help to create a sense of community and belonging, which are essential for happiness. The report also found that there is a strong correlation between happiness and GDP per capita. This suggests that economic development can play a role in promoting happiness. However, the report also found that other factors, such as social support and healthy life expectancy, are just as important. The report concludes by calling for policies that promote happiness and well-being. These policies could include investing in education and healthcare, promoting social inclusion, and reducing corruption. These policies could help to create a world where everyone can live a happy and fulfilling life. In addition to the factors mentioned in the report, there are a few other factors that can contribute to happiness. These include having strong relationships with family and friends, feeling a sense of purpose in life, and being able to make a difference in the world. Happiness is a complex concept, and there is no single answer to the question of what makes people happy. However, the World Happiness Report provides valuable insights into the factors that contribute to happiness. These insights can be used to develop policies that promote happiness and well-being for people around the world.

5) Summary

- The top 10 happiest countries in the world in 2021 were Finland, Denmark, Iceland, Switzerland, Netherlands, Norway, Sweden, New Zealand, Austria, and Luxembourg.
- The bottom 10 happiest countries in the world in 2021 were Afghanistan, Lebanon, Yemen, South Sudan, Syria, Central African Republic, Tanzania, Madagascar, Guinea, and Haiti.
- The most important factors that contribute to happiness are GDP per capita, social support, healthy life expectancy, freedom to make life choices, and generosity.
- There is a strong correlation between happiness and GDP per capita. Countries with higher GDP per capita tend to be happier.
- Social support is also an important factor that contributes to happiness. People who feel supported by their friends, family, and community tend to be happier.
- Healthy life expectancy is another important factor that contributes to happiness. People who live longer and healthier lives tend to be happier.
- Freedom to make life choices is also an important factor that contributes to happiness. People who feel like they have control over their own lives tend to be happier.
- Generosity is also an important factor that contributes to happiness. People who are generous with their time, money, and resources tend to be happier.
- The World Happiness Report 2021 found that happiness levels have increased in most countries over the past few decades. However, there is still significant variation in happiness levels between countries. The report attributes this variation to a number of factors, including economic development, social support, health, and governance.
- The World Happiness Report 2021 concludes by calling for policies that promote happiness and well-being. These policies could include investing in education and healthcare, promoting social inclusion, and reducing corruption.

6) References

- **https://worldhappiness.report/ed/2021/**
- **https://worldpopulationreview.com/country-rankings/happiest-countries-in-the-world**
- **https://en.wikipedia.org/wiki/World_Happiness_Report**
- **https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a**
- **https://datascience.stackexchange.com/questions/9275/is-r2-an-appropriate-evaluation-metric-for-k-nearest-neighbors**