# Exploring Factors Influencing Global Happiness: A Data Analysis of the World Happiness Report

## I. Introduction

This project delves into the exploration of global happiness by specifically analyzing the "World Happiness Report" dataset for the year 2019. Validated with 4000 upvotes on Kaggle.com, this dataset serves as a comprehensive snapshot of well-being during that particular timeframe. The primary goal is to identify key factors influencing happiness, discern regional variations, and construct a predictive model tailored to the nuances of 2019. By focusing on this specific dataset, the analysis aims to provide targeted insights that contribute to a deeper understanding of the determinants of happiness during that pivotal year.

## Analytical questions

My analytical questions revolve around exploring the determinants of happiness. We aim to investigate

What are the key factors that contribute most to happiness scores worldwide?

Are there geographical or regional variations in the significance of different factors in determining happiness?

Can we build a predictive model to estimate happiness scores based on these determinants?

## Data

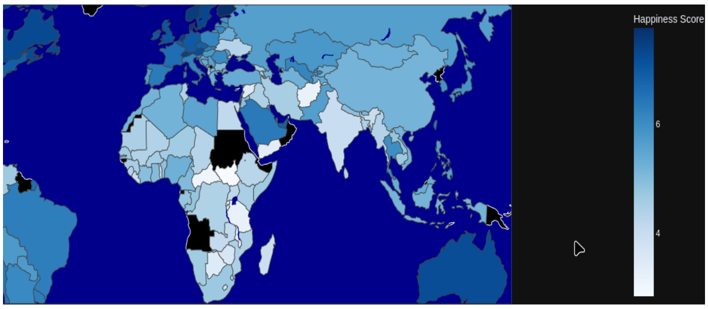
The project utilizes the "World Happiness Report" dataset, consisting of five CSV file (2019), validated with 4000 upvotes on Kaggle.com. This dataset allows a longitudinal analysis of global happiness trends and determinant changes, enhancing its reliability and relevance.

The dataset summarizes the World Happiness Report for 156 countries, presenting overall rank, country names, happiness scores, and key contributing factors such as GDP per capita, social support, health, freedom, generosity, and corruption levels. This data provides a concise yet comprehensive overview of the factors influencing happiness across nations.

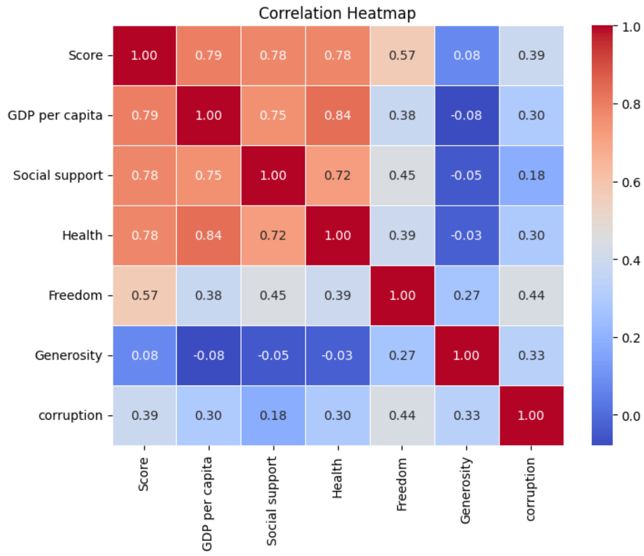
| | Overall rank | Country | Score | GDP per capita | Social support | Health | Freedom | Generosity | corruption |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | Finland | 7.769 | 1.340 | 1.587 | 0.986 | 0.596 | 0.153 | 0.393 |
| 1 | 2 | Denmark | 7.600 | 1.383 | 1.573 | 0.996 | 0.592 | 0.252 | 0.410 |
| 2 | 3 | Norway | 7.554 | 1.488 | 1.582 | 1.028 | 0.603 | 0.271 | 0.341 |
| 3 | 4 | Iceland | 7.494 | 1.380 | 1.624 | 1.026 | 0.591 | 0.354 | 0.118 |
| 4 | 5 | Netherlands | 7.488 | 1.396 | 1.522 | 0.999 | 0.557 | 0.322 | 0.298 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 151 | 152 | Rwanda | 3.334 | 0.359 | 0.711 | 0.614 | 0.555 | 0.217 | 0.411 |
| 152 | 153 | Tanzania | 3.231 | 0.476 | 0.885 | 0.499 | 0.417 | 0.276 | 0.147 |
| 153 | 154 | Afghanistan | 3.203 | 0.350 | 0.517 | 0.361 | 0.000 | 0.158 | 0.025 |
| 154 | 155 | Central African Republic | 3.083 | 0.026 | 0.000 | 0.105 | 0.225 | 0.235 | 0.035 |
| 155 | 156 | South Sudan | 2.853 | 0.306 | 0.575 | 0.295 | 0.010 | 0.202 | 0.091 |

156 rows × 9 columns

## Happiness Score Map




## Heat map




Certainly! Here's a concise summary of the relationships:

**1. GDP per capita and Score**: Strong positive correlation (0.79), indicating higher GDP per capita is associated with higher happiness scores.

**2. Social support and Score**: Strong positive correlation (0.78), suggesting higher social support correlates with higher happiness scores.

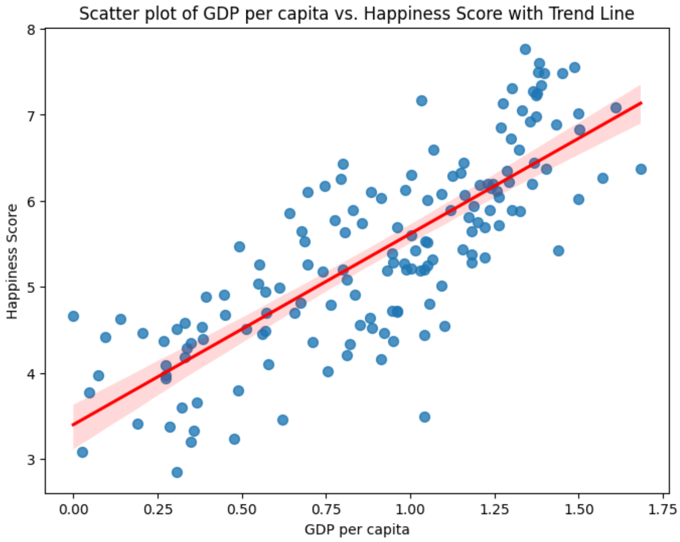
**3. Healthy life expectancy and Score**: Positive correlation (exact coefficient not provided), implying that higher healthy life expectancy tends to be associated with higher happiness scores.

**4. Freedom to make life choices and Score**: Moderate positive correlation (0.58), indicating more freedom is linked to higher happiness scores.

**5. Generosity and Score**: Weak positive correlation (0.08), suggesting a slight connection between generosity and happiness.

**6. Perceptions of corruption and Score**: Moderate positive correlation (0.39), suggesting that higher perceptions of corruption are associated with lower happiness scores.

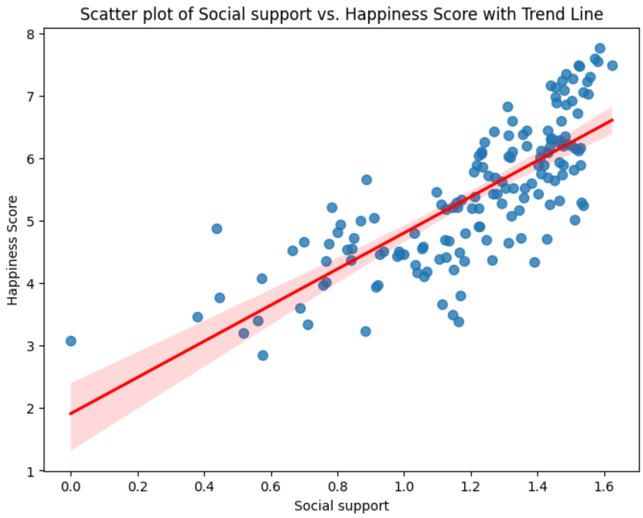
## Scatter plots




**GDP vs Happiness Score**

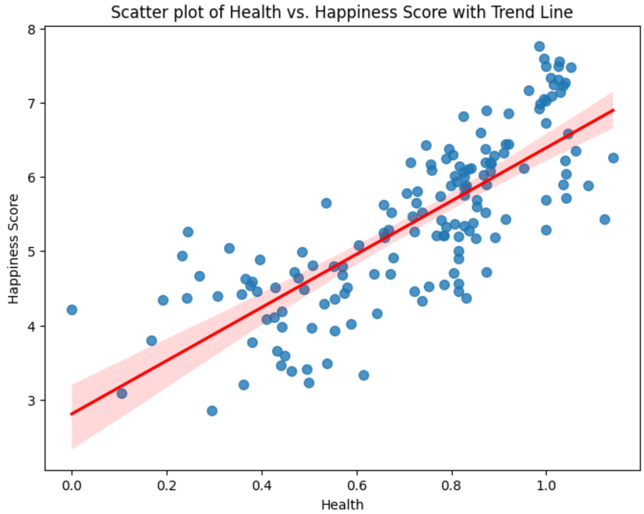
The trend line indicates a positive correlation between GDP per capita and happiness score, implying that countries with higher GDP per capita generally have higher happiness scores. However, the scattered data points around the trend line suggest that there are additional factors influencing happiness beyond just economic prosperity.

**Social Support vs Happiness Score**




*The trend line illustrates an overall positive correlation between social support and happiness scores, indicating that individuals with greater social support tend to have higher happiness scores. The majority of data points are concentrated in the upper right quadrant, signifying that most people experience both high social support and high happiness scores.*

**Health vs Happiness Score**

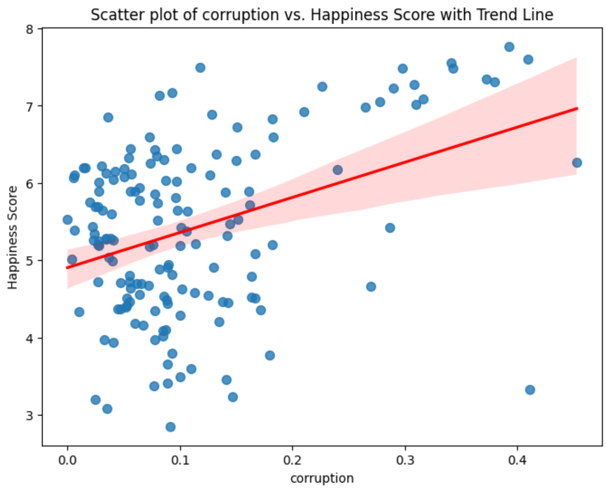



The trend line indicates a positive correlation between health and happiness scores, implying that nations with better health tend to exhibit higher happiness scores.

**Freedom vs Happiness Score**


The scatter plot implies that freedom is a significant determinant of happiness. Nations with greater levels of freedom generally exhibit higher happiness scores.

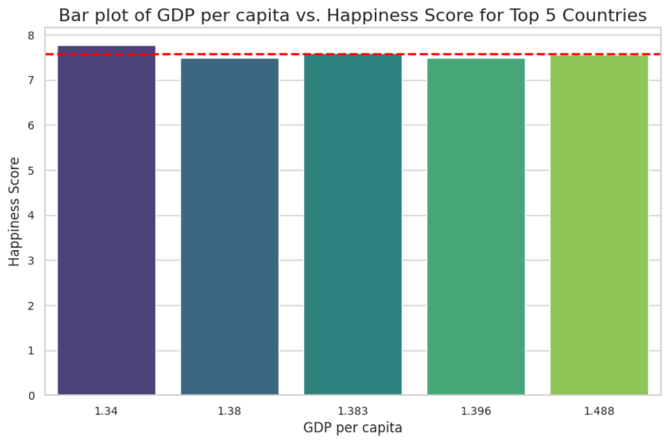
### Curruption vs  Happiness Score




The trend line illustrates a negative correlation between corruption and happiness scores, indicating that countries with elevated levels of corruption typically experience lower happiness scores. Despite the trend, there is some variability in the data points around the trend line.

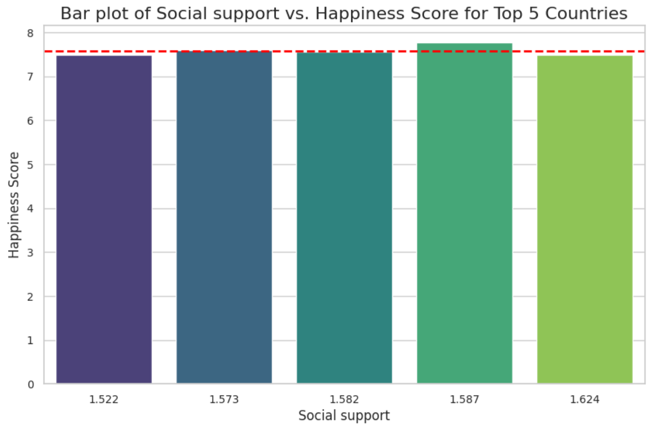
## Comparison of variables among the top five countries based on their Happiness Scores.

### GDP and Happiness Score of Top 5 Countries



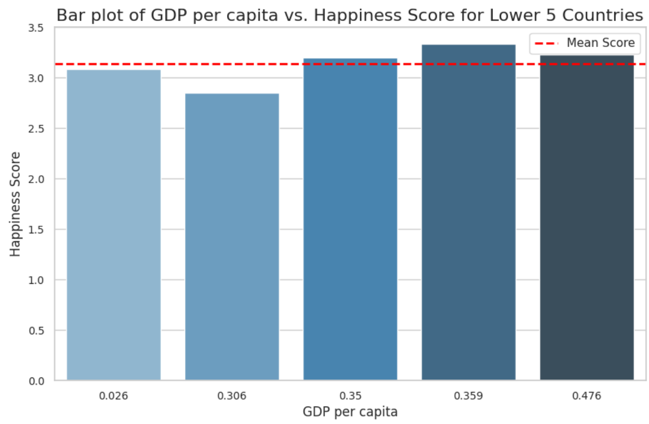

Happier countries often have more money. A quick look shows that places with a higher amount of money (GDP) tend to have happier people. This suggests that having more money contributes a lot to the well-being of citizens.

### Social Status and Happiness Score Top 5 Countries



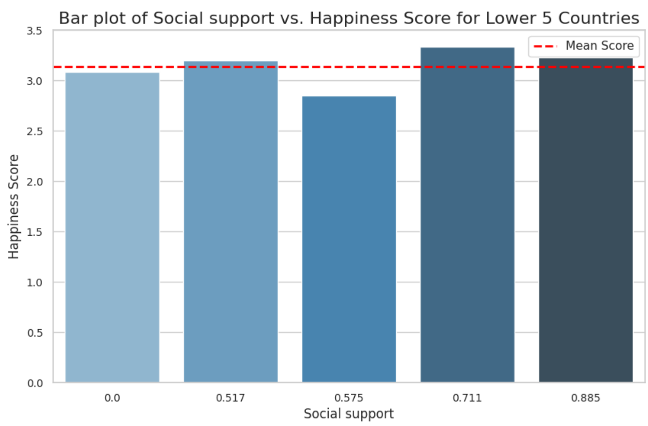

Looking at how social status affects happiness, we see that countries with good social conditions usually have happier people. This includes things like having access to education, healthcare, and fair sharing of income. It seems that when countries have policies that support social progress, it makes people happier.

### GDP and Happiness Score Of lower 5 Countries




Nations grappling with lower GDP often experience diminished happiness scores. Economic challenges and constrained financial resources contribute to a sense of dissatisfaction among the population.

### Social Statues and Happiness Score Of lower 5 Countries



Countries with lower social status indicators also tend to exhibit lower happiness scores. Inadequate access to education, healthcare, and disparities in income distribution can collectively impact the overall well-being of citizens, contributing to a less content populace within these nations.

# Regression Analysis

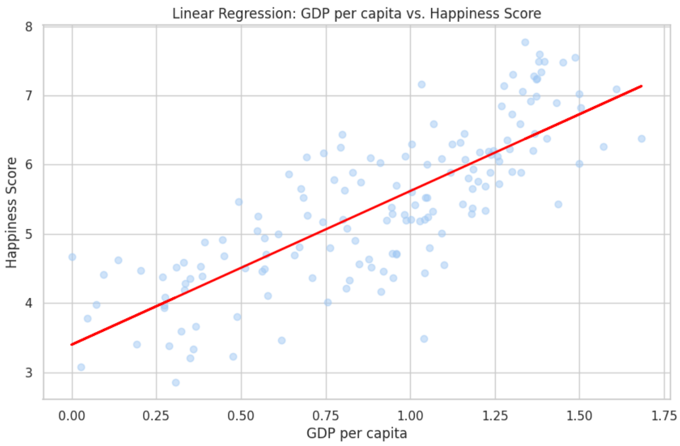
### Regression Analysis of Score and Health

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Score   R-squared:                       0.608
Model:                            OLS   Adj. R-squared:                  0.606
Method:                 Least Squares   F-statistic:                     239.1
Date:                Sun, 03 Dec 2023   Prob (F-statistic):           3.79e-33
Time:                        17:29:14   Log-Likelihood:                -164.48
No. Observations:                 156   AIC:                             333.0
Df Residuals:                     154   BIC:                             339.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.8068      0.177     15.837      0.000       2.457       3.157
Health         3.5854      0.232     15.462      0.000       3.127       4.043
==============================================================================
Omnibus:                        6.324   Durbin-Watson:                   1.140
Prob(Omnibus):                  0.042   Jarque-Bera (JB):                3.543
Skew:                          -0.148   Prob(JB):                        0.170
Kurtosis:                       2.324   Cond. No.                         6.41
==============================================================================
```

According to the findings, the model indicates that 'Health' is an important and reliable predictor of the 'Score.' This means that the health factor has a meaningful impact on determining the overall score. Additionally, the model as a whole fits well, which is evident from the R-squared and F-statistic values. These statistics suggest that the model effectively captures and explains the variation in the data, making it a reliable tool for understanding the relationship between health and the overall score.

### Regression Analysis of Score and GDP

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Score   R-squared:                       0.630
Model:                            OLS   Adj. R-squared:                  0.628
Method:                 Least Squares   F-statistic:                     262.5
Date:                Sun, 03 Dec 2023   Prob (F-statistic):           4.32e-35
Time:                        17:29:14   Log-Likelihood:                -159.97
No. Observations:                 156   AIC:                             323.9
Df Residuals:                     154   BIC:                             330.0
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               3.3993      0.135     25.120      0.000       3.132       3.667
GDP per capita      2.2181      0.137     16.202      0.000       1.948       2.489
==============================================================================
Omnibus:                        1.139   Durbin-Watson:                   1.378
Prob(Omnibus):                  0.566   Jarque-Bera (JB):                1.244
Skew:                          -0.177   Prob(JB):                        0.537
Kurtosis:                       2.742   Cond. No.                         4.77
==============================================================================
```



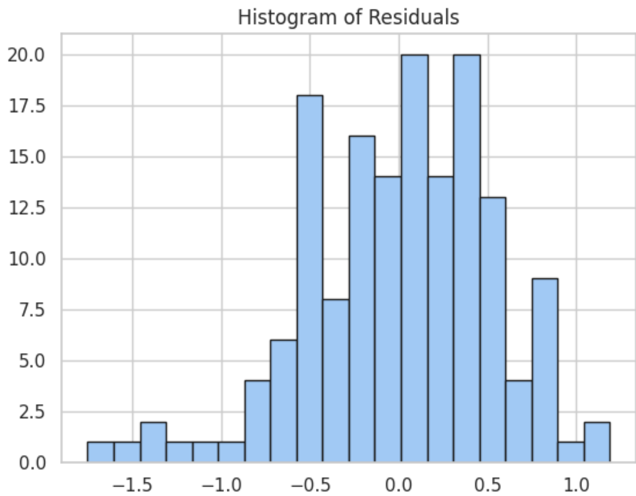

The Analysis indicates a noteworthy positive connection between GDP per capita and happiness scores. In simpler terms, as the GDP per capita of a country goes up, the happiness scores of its citizens also tend to go up.

### Regression Analysis of All columns

```
...
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Score   R-squared:                       0.779
Model:                            OLS   Adj. R-squared:                  0.770
Method:                 Least Squares   F-statistic:                     87.62
Date:                Sun, 03 Dec 2023   Prob (F-statistic):           2.40e-46
Time:                        17:29:15   Log-Likelihood:                -119.76
No. Observations:                 156   AIC:                             253.5
Df Residuals:                     149   BIC:                             274.9
Df Model:                           6
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               1.7952      0.211      8.505      0.000       1.378       2.212
GDP per capita      0.7754      0.218      3.553      0.001       0.344       1.207
Social support      1.1242      0.237      4.745      0.000       0.656       1.592
Health              1.0781      0.335      3.223      0.002       0.417       1.739
Freedom             1.4548      0.375      3.876      0.000       0.713       2.197
Generosity          0.4898      0.498      0.984      0.327      -0.494       1.473
corruption          0.9723      0.542      1.793      0.075      -0.099       2.044
==============================================================================
Omnibus:                        8.188   Durbin-Watson:                   1.648
Prob(Omnibus):                  0.017   Jarque-Bera (JB):                7.971
Skew:                          -0.498   Prob(JB):                       0.0186
...
===================================================================================
```

This model shows that when we consider things like GDP per capita, Social support, Health, Freedom, Generosity, and Corruption together, they help explain a big part of why happiness scores differ ("Score"). This means that these factors combined affect how happy a country is.But, when we look at just one thing at a time, like "Generosity" and "Corruption," it's not as clear how important they are. The uncertainty comes from the p-values linked to these factors, suggesting that individually, their impact on happiness scores might not be very certain.

### Histogram of Resduals




The histogram of residuals in your graph indicates that these residuals are roughly distributed in a normal fashion, centered around a mean of zero. This implies that the model aligns well with the data, suggesting a good fit.

## Null Hypothesis (H0): There is no statistically significant influence of the other columns on the score

```
                             Test for Constraints
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
c0             0.7754      0.218      3.553      0.001       0.344       1.207
c1             1.1242      0.237      4.745      0.000       0.656       1.592
```

Given that the p-values for c0 and c1 are both less than the typical significance level of 0.05, you would reject the null hypotheses. The rejection of these null hypotheses suggests that the coefficients for c0 and c1 are statistically significantly different from zero, and these predictors are likely to have a significant effect on the dependent variable in your regression model.

## Durbin-Watson Test Statistic

Durbin-Watson Test Statistic

```python
dw_test_stat = sm.stats.stattools.durbin_watson(model.resid)
print(f'Durbin-Watson Test Statistic: {dw_test_stat}')
```

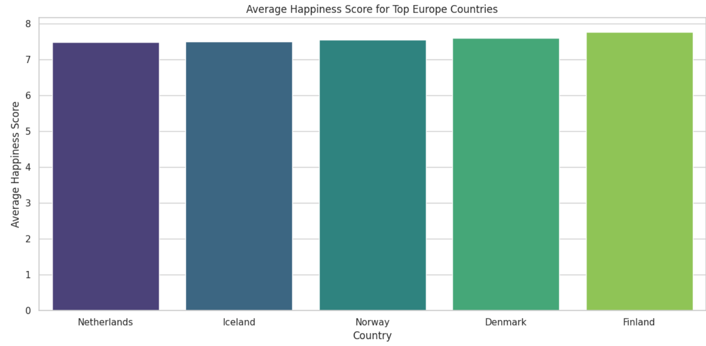
```
Durbin-Watson Test Statistic: 1.6483585453130485
```

The Durbin-Watson test statistic, around 1.65, indicates a tendency toward positive autocorrelation in the residuals of the regression model as it is less than 2.

## Are there geographical or regional variations in the significance of different factors in determining happiness?

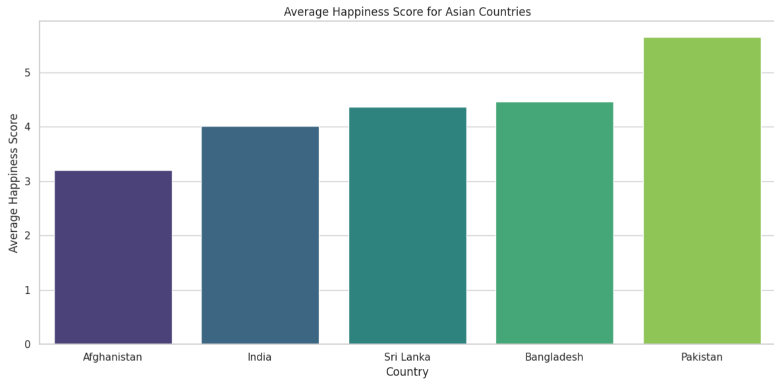
### Happiness Score of 5 Europe Countries

```
Country
Netherlands    7.488
Iceland        7.494
Norway         7.554
Denmark        7.600
Finland        7.769
Name: Score, dtype: float64
```

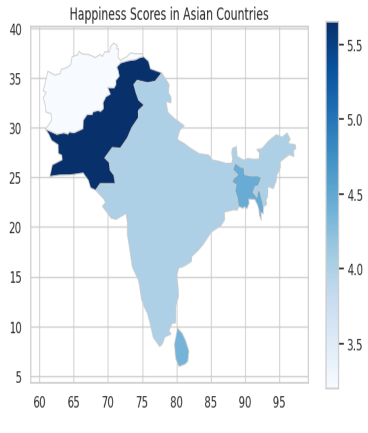


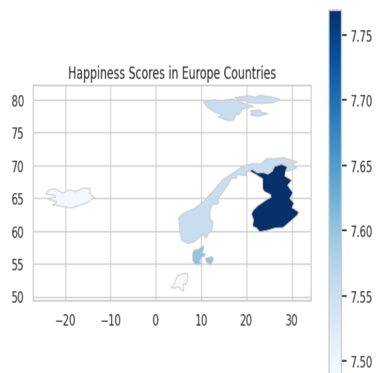

### Happiness Score of 5 Asian Countries

```
Country
Afghanistan    3.203
India          4.015
Sri Lanka      4.366
Bangladesh     4.456
Pakistan       5.653
Name: Score, dtype: float64
```




### Happiness Score map



Geographical or regional variations play a significant role in determining happiness, with distinct priorities observed between Asians and Europeans. Notably, Asians tend to prioritize financial security and freedom, whereas Europeans place a higher emphasis on community

## Predictive models

Machine Learning (ML) regression model is employed to predict the 'Score' based on features such as 'GDP per capita', 'Social support', 'Health', and 'Freedom'. The hyperparameters of the ML model are tuned using a grid search, and the best model is evaluated using Mean Squared Error (MSE) and R-squared ($R^2$) metrics.

## Linear Regression

```
Mean Squared Error: 0.42024794847379593
R-squared: 0.5961963009874782
```

## GradientBoostingRegressor

```
Best Hyperparameters: {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50}
Mean Squared Error: 0.3620533913701818
R-squared: 0.6521137123780065
```

## SVM Model

```
Best Hyperparameters: {'C': 10, 'epsilon': 0.01, 'kernel': 'rbf'}
Mean Squared Error: 0.2824629317157399
R-squared: 0.7285898073388235
```

Hyperparameters ('C', 'epsilon', 'kernel') represent the best configuration found during a hyperparameter tuning process. The model's performance is assessed using the Mean Squared Error (lower is better) and R-squared (closer to 1 is better). The results suggest a reasonably good fit of the SVM regression model to the data.

## Predictions

```
best_model.fit(X_train, y_train)
new_sample_data = np.array([1.092, 1.513, 0.815, 0.311])
new_sample = new_sample_data.reshape(1, -1)
new_prediction = best_model.predict(new_sample)
print("Prediction:", new_prediction[0])
```

```
Prediction: 5.673166692727373
```

A new set of sample data (`new_sample_data`) representing features 'GDP per capita', 'Social support', 'Health', and 'Freedom' is created.

The new sample data is reshaped to fit the model's input requirements.

The trained model is used to predict the 'Score' for the new sample.

The predicted 'Score' for the new sample is printed.


## Abstract

This report explores the factors influencing happiness scores across countries, utilizing statistical analyses, scatter plots, and regression models. Key findings reveal strong positive correlations between GDP per capita, social support, healthy life expectancy, and freedom to make life choices with happiness scores. Meanwhile, generosity and perceptions of corruption exhibit weaker connections. Scatter plots illustrate the relationships visually, highlighting nuances beyond the identified correlations. The top and bottom five countries' comparisons emphasize the impact of GDP and social status on happiness. Regression analyses confirm health and GDP as significant predictors, while combining all factors explains variations in happiness scores. Geographical variations in priorities are evident between Asian and European countries. Lastly, machine learning models, including Linear Regression, GradientBoostingRegressor, and SVM, provide predictive insights. The report concludes with the application of these models to new sample data, showcasing their ability to estimate happiness scores based on key features.