



# A Study on World Happiness Data



---

## Team Members:

Ronak Shah  
Hitarth Shah  
Devni Shah  
Vedant Kamat  
Amoghvarsh Kulkarni  
Sai Thanmayi Karpurapu



# Table of contents

01

---

Introduction

02

---

Data Cleaning

03

---

Exploratory  
Data Analysis

04

---

Machine  
Learning

05

---

Conclusion &  
Recommendation





01

---

# Introduction



# Introduction

---

In an ever-evolving world, the pursuit of happiness is a fundamental aspiration for individuals and nations alike.

The World Happiness Index, a comprehensive measure of well-being and life satisfaction, offers a unique lens through which we can analyze and understand the factors that contribute to happiness on a global scale.

Leveraging the power of Python, this project delves into the exploration and analysis of the World Happiness Index dataset, aiming to uncover patterns, trends, and insights that shape the happiness landscape across countries and over time.





# Introduction

---


## Data Source:

<https://worldhappiness.report/data/>

## Mission Objective:

- To identify what is the average happiness throughout all years
- To understand which are the top 10 happiest & least happiest countries
- To analyse the correlation between the variables that define the happiness of a country
- To predict the happiness scores of countries for future years with the given features by using Machine Learning

|                                  |
|----------------------------------|
| Country name                     |
| Year                             |
| Life Ladder                      |
| Log GDP per capita               |
| Social support                   |
| Healthy life expectancy at birth |
| Freedom to make life choices     |
| Generosity                       |
| Perceptions of corruption        |
| Positive affect                  |
| Negative affect                  |



# Data Cleaning



## Handling NaN Values:

### Step 1: Identifying Missing Values

To address missing values in dataset, we first used the `isna()` method for each column. This allowed us to pinpoint which cells contained NaN values.

### Step 2: Filling Missing Values with Mean values

To handle missing values, we opted to fill them with the mean values specific to each country. This approach helped maintain the integrity of the data.

## Removing Rows:

### Identifying and Dropping Rows with NaN Columns:

- To maintain data quality, we identified and removed entire rows where column values particular to the countries had NaN values.
- Removed countries from the dataset which had less than 10 years of data.
- Removed year 2005 and 2006 from the dataset as they had less than 100 rows of data.
- This ensures that we have complete and consistent data for analysis.

```
In [39]: data.isna().sum()
```

```
Out[39]: Country name          0
         year                  0
         Life Ladder           0
         Log GDP per capita     0
         Social support        0
         Healthy life expectancy at birth  0
         Freedom to make life choices      0
         Generosity            0
         Perceptions of corruption         0
         Positive affect          0
         Negative affect          0
         dtype: int64
```



# Descriptive Statistics



|       | year        | Life Ladder | 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 | 1896.000000 | 1896.000000 | 1896.000000        | 1896.000000    | 1896.000000                      | 1896.000000                  | 1896.000000 | 1896.000000               | 1896.000000     | 1896.000000     |
| mean  | 2014.663502 | 5.525224    | 9.442846           | 0.812467       | 63.836927                        | 0.749804                     | -0.002600   | 0.739999                  | 0.652260        | 0.273770        |
| std   | 4.414495    | 1.132399    | 1.131834           | 0.119389       | 6.504014                         | 0.138745                     | 0.160735    | 0.188120                  | 0.107706        | 0.085259        |
| min   | 2007.000000 | 1.281000    | 5.527000           | 0.228000       | 17.360000                        | 0.258000                     | -0.338000   | 0.035000                  | 0.179000        | 0.083000        |
| 25%   | 2011.000000 | 4.669000    | 8.526000           | 0.747750       | 60.130000                        | 0.659000                     | -0.117000   | 0.683000                  | 0.568750        | 0.210000        |
| 50%   | 2015.000000 | 5.488500    | 9.548000           | 0.839000       | 65.325000                        | 0.770000                     | -0.026000   | 0.797500                  | 0.663000        | 0.265000        |
| 75%   | 2018.000000 | 6.375000    | 10.424250          | 0.906000       | 69.006250                        | 0.861000                     | 0.091000    | 0.867250                  | 0.739250        | 0.326000        |
| max   | 2022.000000 | 7.971000    | 11.664000          | 0.987000       | 74.475000                        | 0.985000                     | 0.703000    | 0.983000                  | 0.884000        | 0.607000        |

- The highest life ladder value throughout all years : 7.97
- Average happiness throughout all years amongst all countries : 5.52



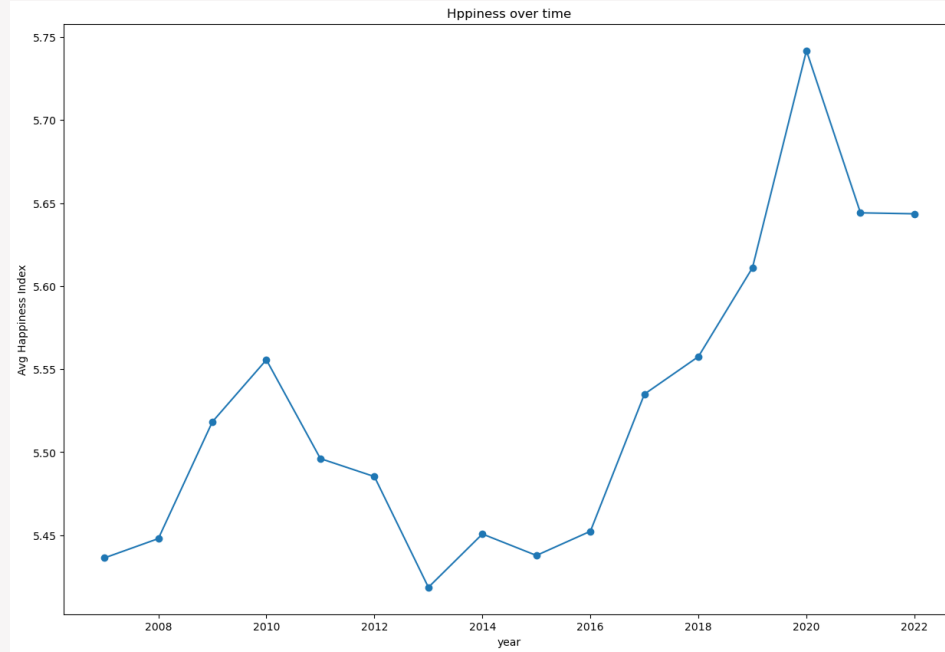


03

## Exploratory Data Analysis



# Happiness Index over the years

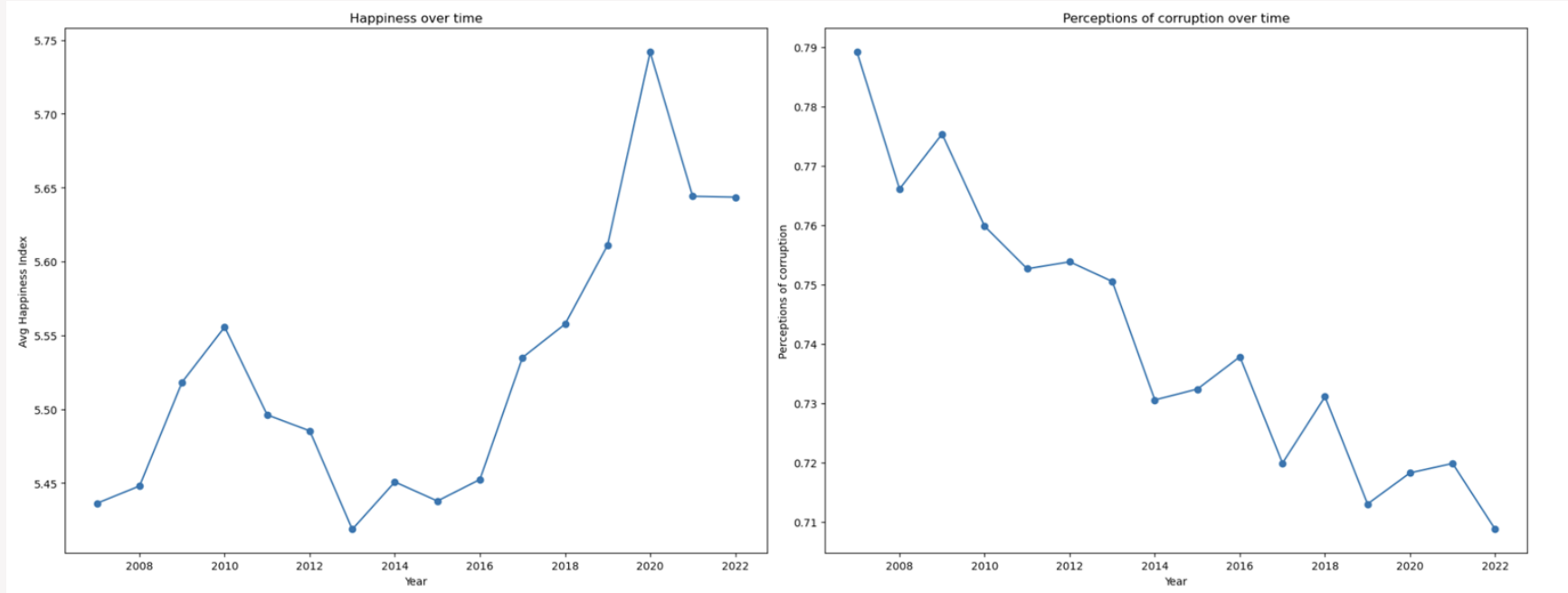


- Overall, It's following an upward trend
- High increase after the recovery from 2008 crisis.
- Falls again during the Covid era

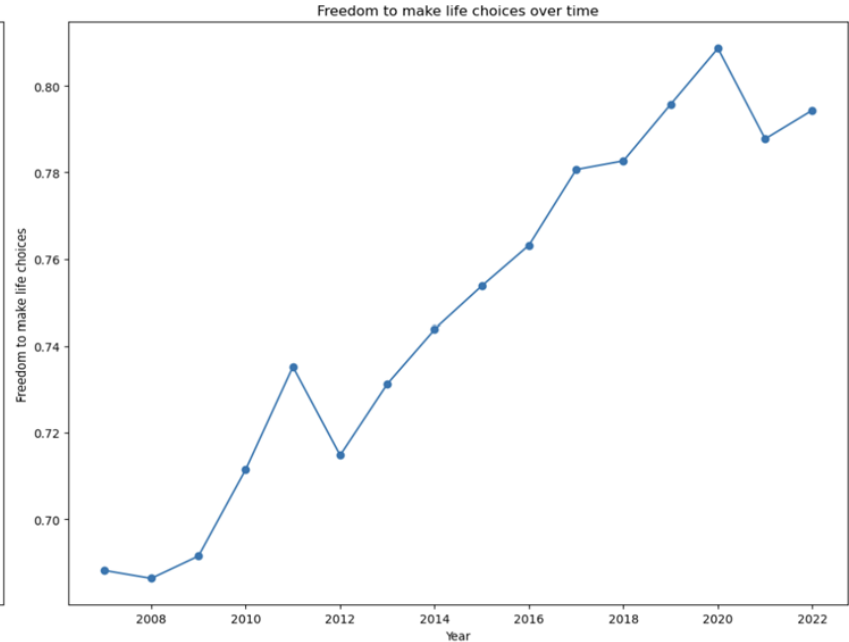
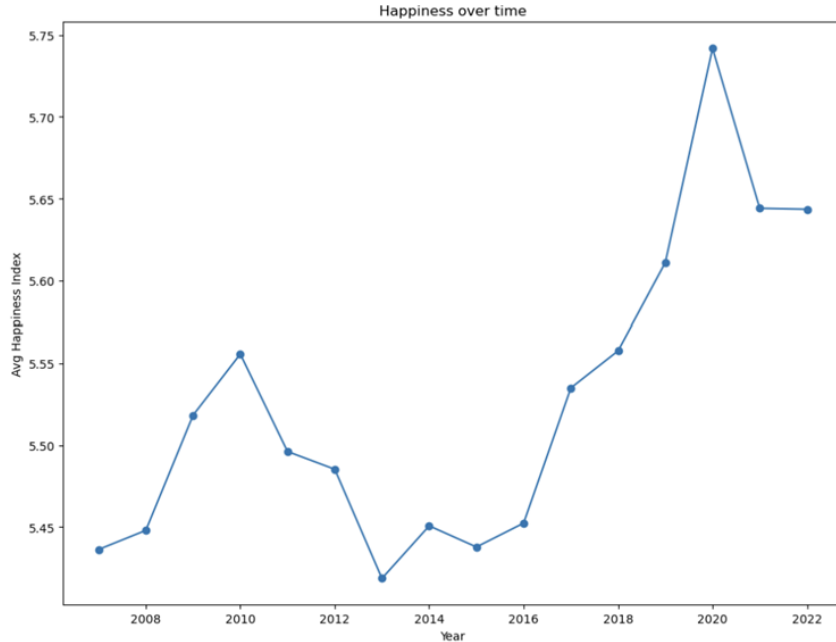
• • •

• • •

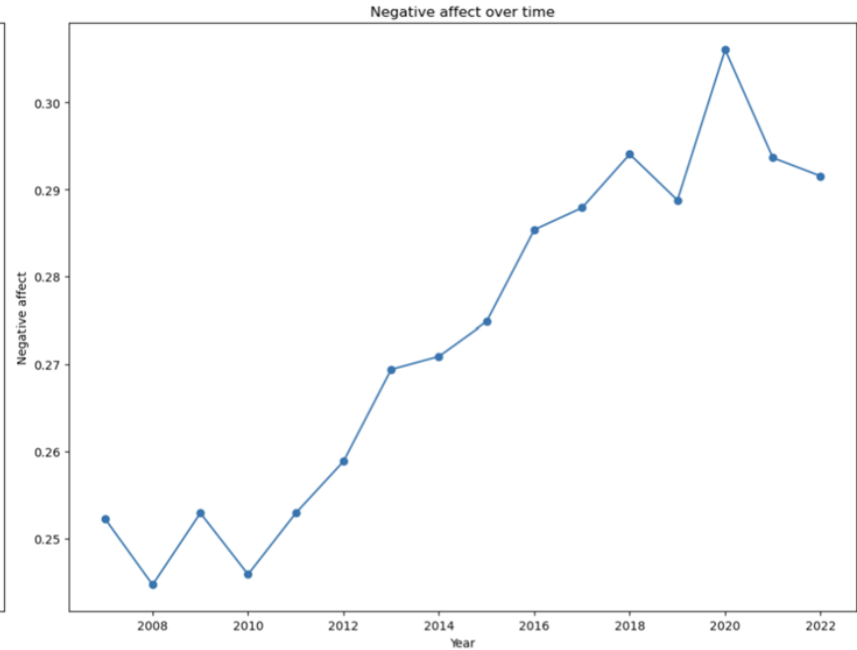
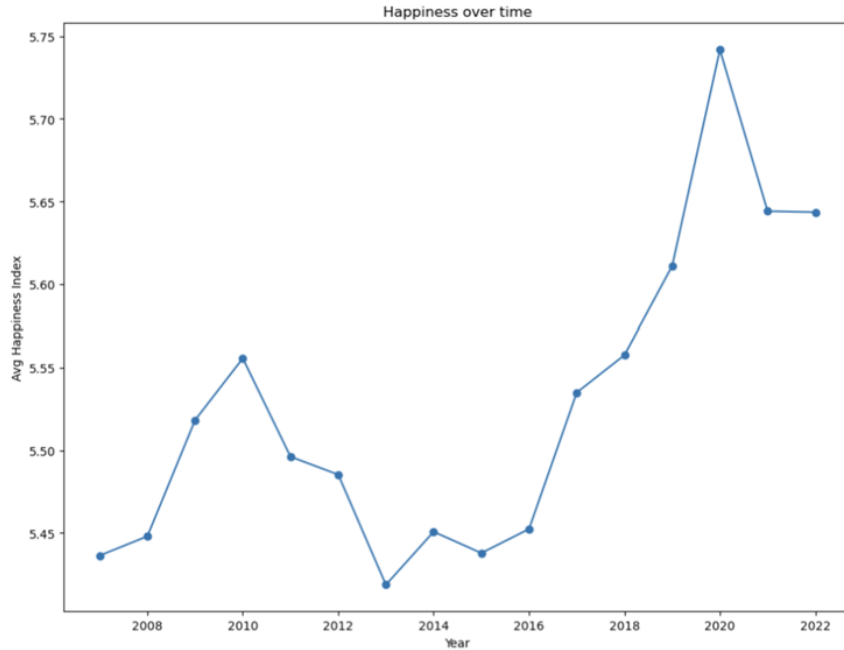
# Happiness V/S Perception of Corruption Over the years



# Happiness V/S Freedom to make Life choices Over the years

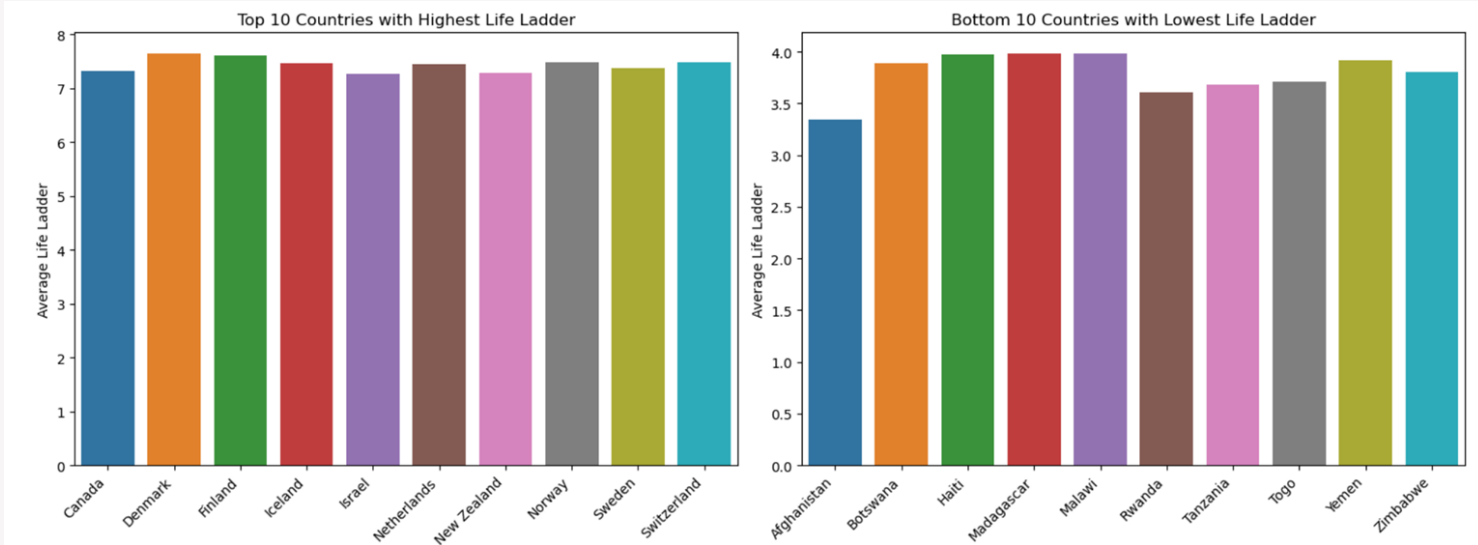


# Happiness V/S Negative Affect Over the years



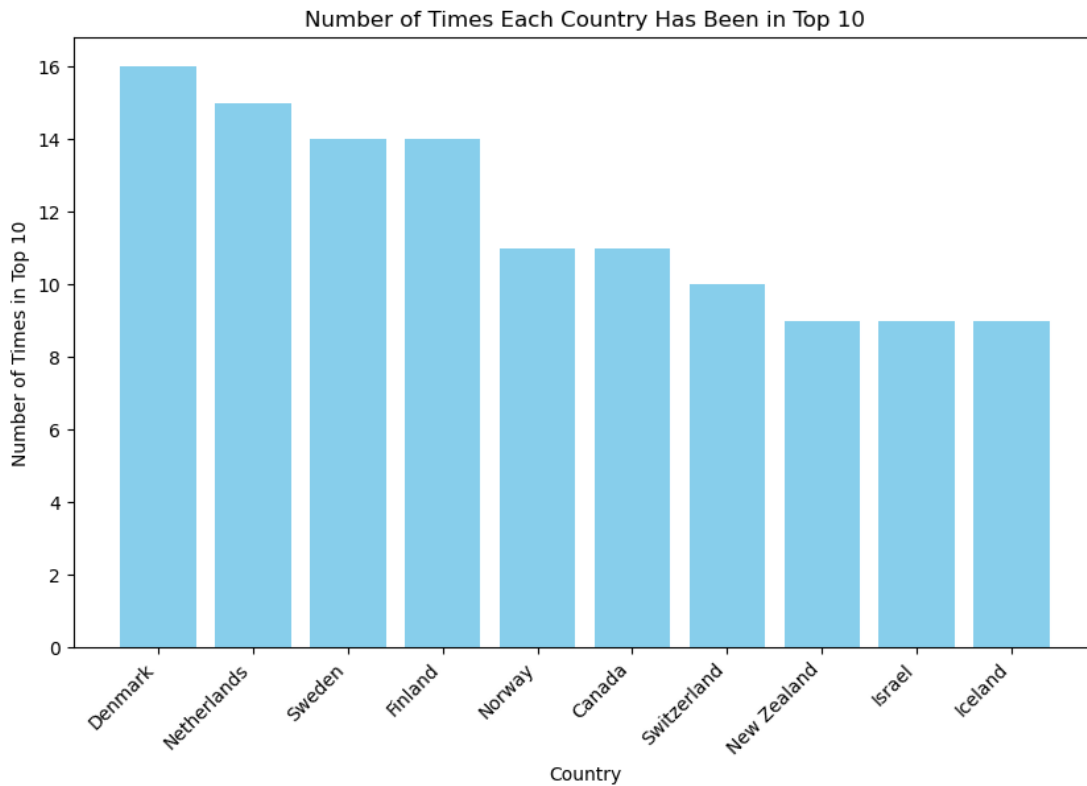


# Top 10 and Lowest 10 Countries

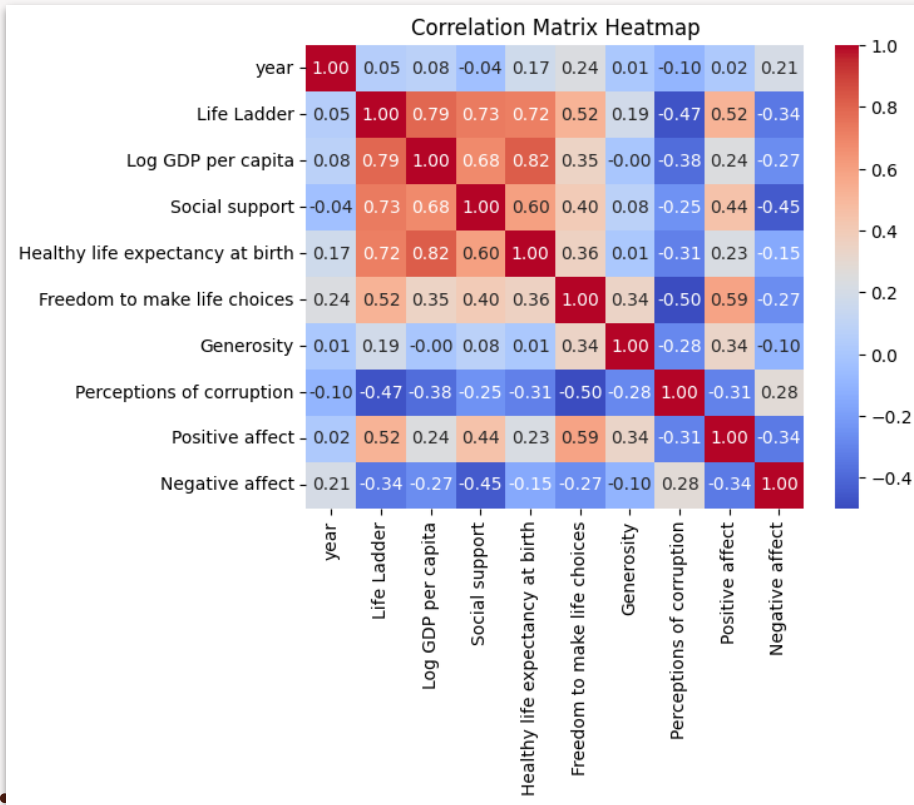


- Average happiness of the top 10 countries is around 7.5
  - Average happiness of the lowest 10 countries ranges from 3.4 to 5
- 

# Number of times each Country has been in top 10



# Correlation Matrix

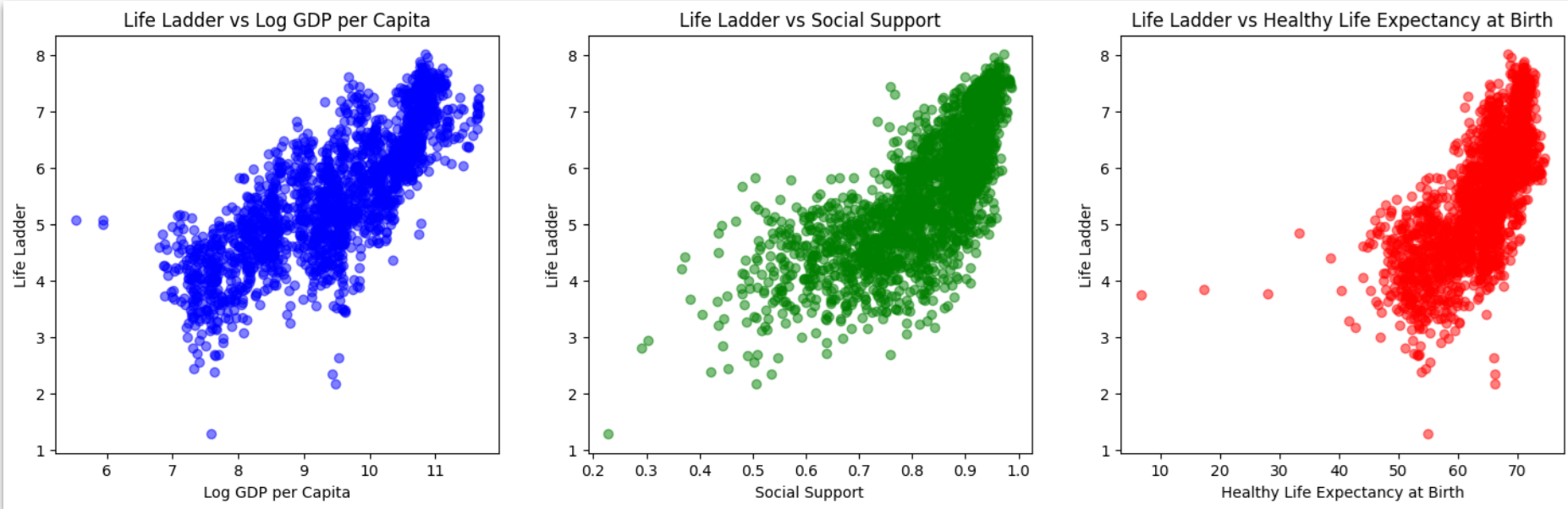


All the independent variables are not highly correlated.

High correlation between Happiness Index and :

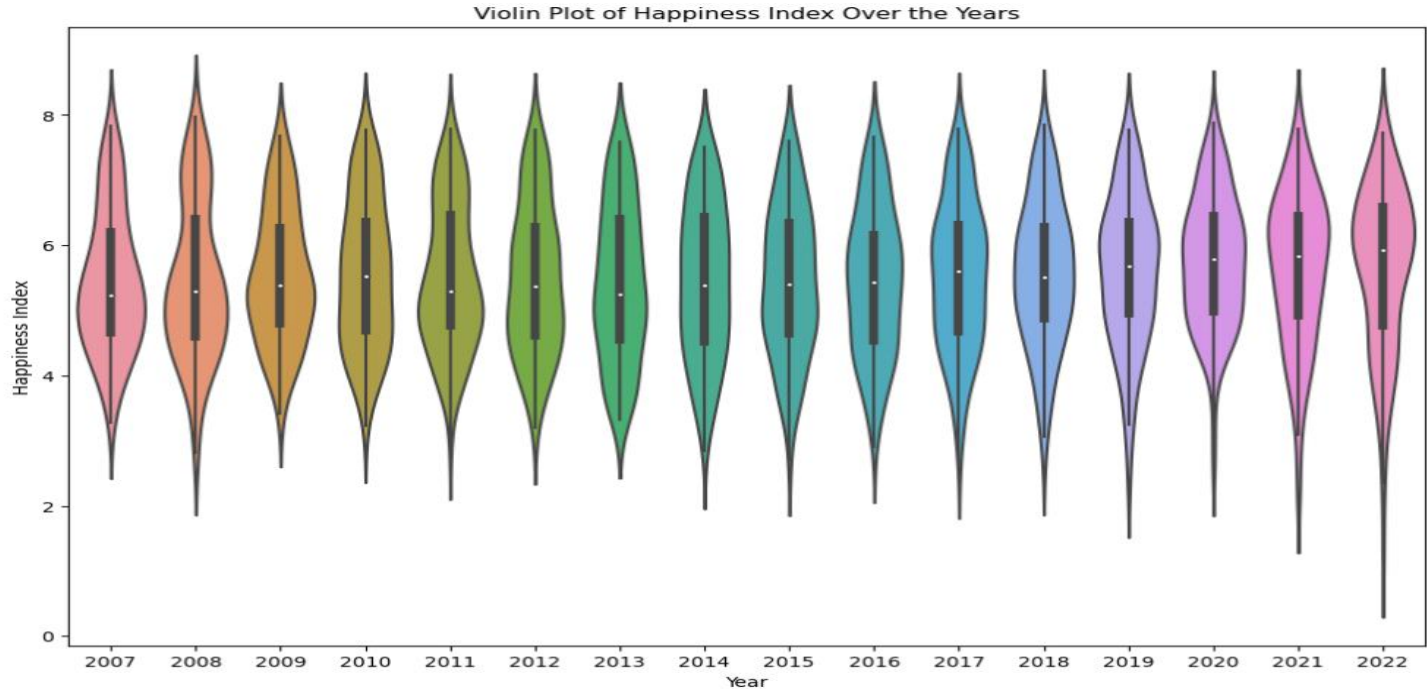
1. Log GDP per capita
2. Social Support
3. Healthy life expectancy

# Scatter Plots





# Violin Plots



Major world events like the economic crisis of 2007-08, COVID-19, Ukraine war after 2019 has resulted in happiness ranging very widely due to economic, political instabilities in the world.



04

---

# Machine Learning



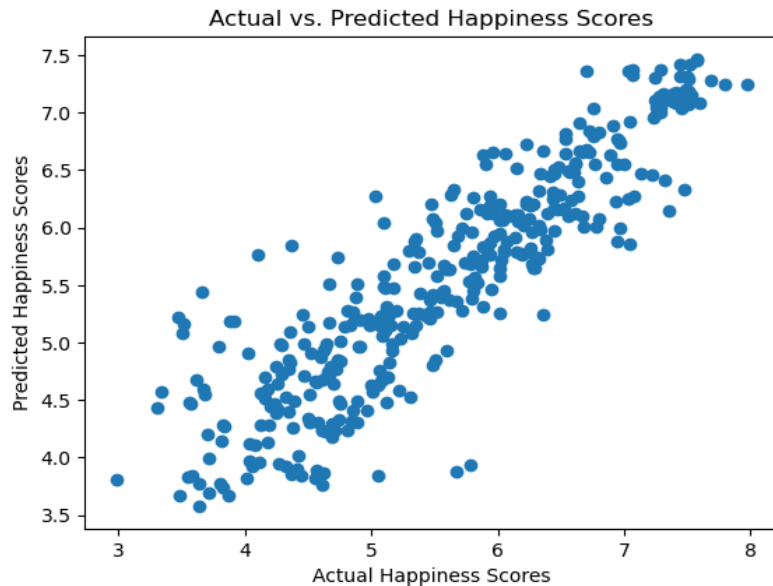
# Steps for Regression Modeling

- Defined features (X) and target variable (Y) i.e Happiness score
- Split the dataset into training and testing sets
- Train the model using the training data
- Made predictions on the testing set
- Displayed model outcomes using scatter plot
- Assessed model performance using metrics like squared mean error and R squared value



# Linear Regression

Mean Squared Error: 0.25344450551980696  
R-squared: 0.7959102303568403



# Random Forest Regression

Mean Squared Error: 0.15149844272973684  
R-squared: 0.8780037380783027





05

## Conclusion



# Conclusion

- Average happiness of the top 10 countries is around **7.5**
- Overall, Average Happiness of years is following an **upward trend**
- **Economic, social, and health factors** are key influencers of national happiness.
- Happiness **trends vary by region**, reflecting socio-economic and political dynamics.
- **Regression** models can predict happiness with **significant accuracy**.
- The findings encourage further research in happiness and well-being metrics.

# Suggestions

- **Implement** universal **health care** to ensure access to quality medical services for all citizens.
- Governments should focus more on **increasing their GDP per capita** by increasing Investment and Infrastructure.
- Promote policies that encourage **work-life balance**, including **flexible work hours** and parental leave.





# Thank You

