⌄ **Exercise 1**

```
import os
from zipfile import ZipFile
import pandas as pd
```

```
# Set the Kaggle API key environment variables
os.environ['KAGGLE_USERNAME'] = 'kaorisuzuki'
os.environ['KAGGLE_KEY'] = 'bab3558185555761accf622c855aa4be'
```

```
# Download the World Happiness Report dataset
!kaggle datasets download -d unsdsn/world-happiness
```

⇥ Dataset URL: https://www.kaggle.com/datasets/unsdsn/world-happiness
   License(s): CC0-1.0
   Downloading world-happiness.zip to /content
     0% 0.00/36.8k [00:00<?, ?B/s]
   100% 36.8k/36.8k [00:00<00:00, 52.8MB/s]

```
# Unzip the dataset
with ZipFile('world-happiness.zip', 'r') as zip_ref:
    zip_ref.extractall('world_happiness')
```

```
# Load and display the dataset (adjust the file name based on the year you want)
data = pd.read_csv('WHR2024.csv')
print(data.head())
```

⇥
```
   Country name  Ladder score  upperwhisker  lowerwhisker  ¥
0     Finland          7.741        7.815         7.667
1     Denmark          7.583        7.665         7.500
2     Iceland          7.525        7.618         7.433
3      Sweden          7.344        7.422         7.267
4      Israel          7.341        7.405         7.277

   Explained by: Log GDP per capita  Explained by: Social support  ¥
0                            1.844                         1.572
1                            1.908                         1.520
2                            1.881                         1.617
3                            1.878                         1.501
4                            1.803                         1.513

   Explained by: Healthy life expectancy  ¥
0                                  0.695
1                                  0.699
2                                  0.718
3                                  0.724
4                                  0.740

   Explained by: Freedom to make life choices  Explained by: Generosity  ¥
0                                        0.859                     0.142
1                                        0.823                     0.204
2                                        0.819                     0.258
3                                        0.838                     0.221
4                                        0.641                     0.153

   Explained by: Perceptions of corruption  Dystopia + residual
0                                    0.546                 2.082
1                                    0.548                 1.881
2                                    0.182                 2.050
3                                    0.524                 1.658
4                                    0.193                 2.298
```

```
# Create a histogram for the "Ladder score" (happiness score)
import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Create a histogram for the "Ladder score" (happiness score)
plt.figure(figsize=(8, 6))
sns.histplot(data['Ladder score'], bins=20, kde=True)
plt.title('Distribution of Happiness Scores')
plt.xlabel('Happiness Score')
```

```
plt.ylabel('Frequency')
plt.show()
```


Distribution of Happiness Scores
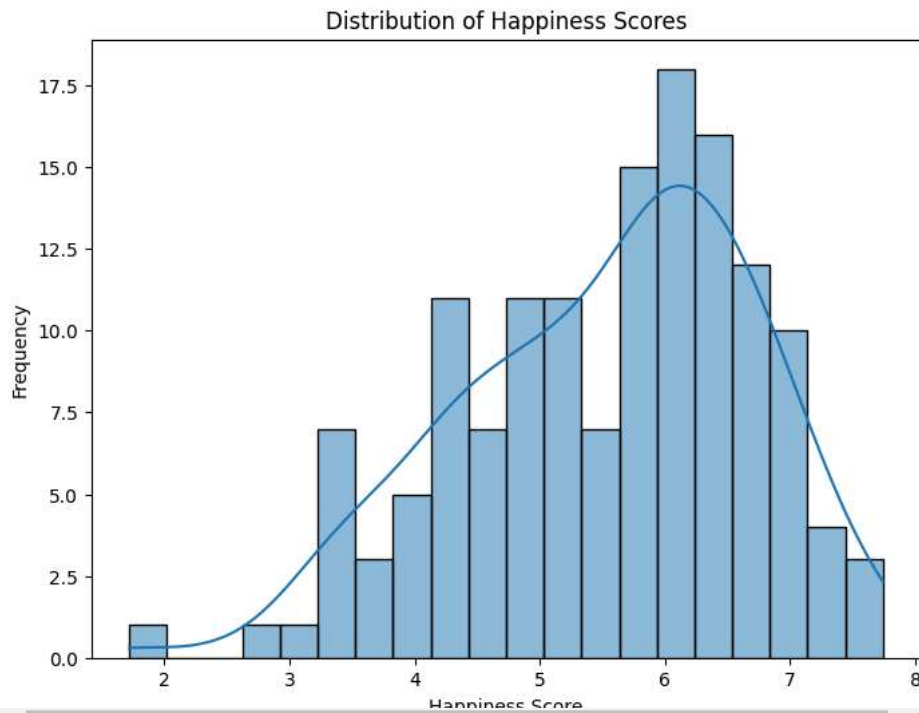
The distribution of happiness scores is roughly normal, with a peak around 6. There are some countries with very high and very low happiness scores, but most countries cluster around the middle.

Exercise 2

```
import pandas as pd
# Load and display the dataset (adjust the file name based on the year you want)
data = pd.read_csv('WHR2024.csv')
print(data.head())
```

```
   Country name  Ladder score  upperwhisker  lowerwhisker  ¥
0       Finland         7.741         7.815         7.667
1       Denmark         7.583         7.665         7.500
2       Iceland         7.525         7.618         7.433
3        Sweden         7.344         7.422         7.267
4        Israel         7.341         7.405         7.277

   Explained by: Log GDP per capita  Explained by: Social support  ¥
0                             1.844                         1.572
1                             1.908                         1.520
2                             1.881                         1.617
3                             1.878                         1.501
4                             1.803                         1.513

   Explained by: Healthy life expectancy  ¥
0                                  0.695
1                                  0.699
2                                  0.718
3                                  0.724
4                                  0.740

   Explained by: Freedom to make life choices  Explained by: Generosity  ¥
0                                       0.859                     0.142
1                                       0.823                     0.204
2                                       0.819                     0.258
3                                       0.838                     0.221
4                                       0.641                     0.153

   Explained by: Perceptions of corruption  Dystopia + residual
0                                    0.546                2.082
1                                    0.548                1.881
2                                    0.182                2.050
3                                    0.524                1.658
4                                    0.193                2.298
```

```
import seaborn as sns
import matplotlib.pyplot as plt


# Select relevant numeric columns for correlation: Life Ladder (Ladder score), Log GDP per capita, Social support, Healthy life expectancy
selected_columns = ['Ladder score', 'Explained by: Log GDP per capita', 'Explained by: Social support', 'Explained by: Healthy life expectancy']
data_selected = data[selected_columns]

# Compute the correlation matrix
correlation_matrix = data_selected.corr()

# Create a heatmap to visualize the correlations with annotations
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", cbar=True, square=True)
plt.title("Correlation Heatmap: Happiness Score and Related Factors", fontsize=14)
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.tight_layout()
plt.show()
```


Correlation Heatmap: Happiness Score and Related Factors
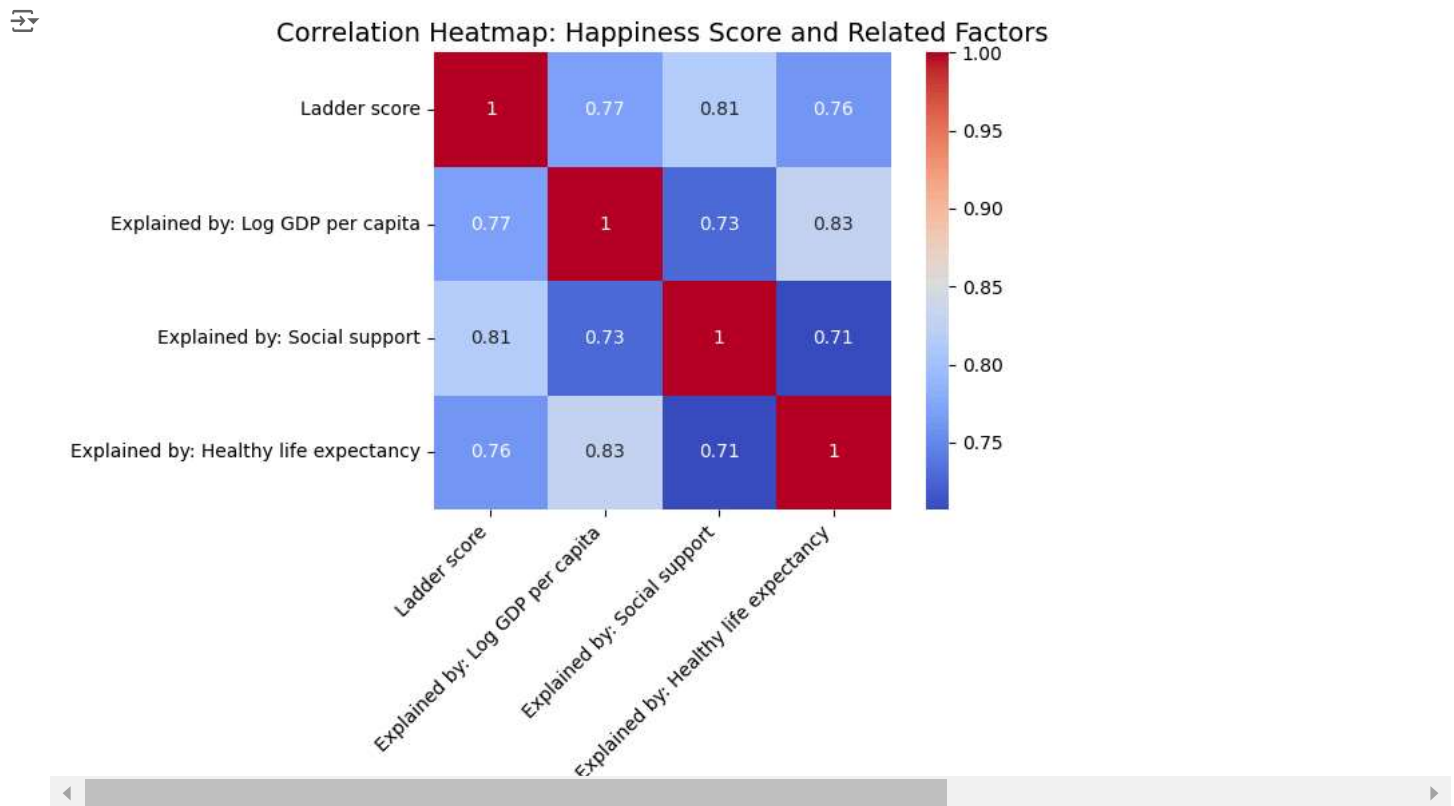
Exercise 3

```
import pandas as pd
import matplotlib.pyplot as plt


# Load the datasets
data_2015 = pd.read_csv('world_happiness/2015.csv')
data_2016 = pd.read_csv('world_happiness/2016.csv')
data_2017 = pd.read_csv('world_happiness/2017.csv')
data_2018 = pd.read_csv('world_happiness/2018.csv')
data_2019 = pd.read_csv('world_happiness/2019.csv')


# Add a 'Year' column to each dataframe
data_2015['Year'] = 2015
data_2016['Year'] = 2016
data_2017['Year'] = 2017
data_2018['Year'] = 2018
data_2019['Year'] = 2019
```

```
# Concatenate all datasets
data = pd.concat([data_2015, data_2016, data_2017, data_2018, data_2019], ignore_index=True)


# Ensure column name consistency for 'Country' and 'Happiness Score'
# Use 'Country name' or 'Country' depending on the dataset's structure, and adjust the score column name if needed.
if 'Country name' in data.columns:
    country_column = 'Country name'
elif 'Country' in data.columns:
    country_column = 'Country'

if 'Happiness Score' in data.columns:
    score_column = 'Happiness Score'
elif 'Score' in data.columns:
    score_column = 'Score'


# Select the country of interest
country_name = 'United States'


# Filter the data for the selected country
country_data = data[data[country_column] == country_name]


# Plot the trend of happiness score over the years
plt.figure(figsize=(10, 6))
plt.plot(country_data['Year'], country_data[score_column], marker='o', linestyle='-', color='b')

# Add labels, title, and markers
plt.xlabel('Year')
plt.ylabel('Happiness Score')
plt.title(f'Happiness Score Trend for {country_name} (2015-2019)')
plt.grid(True)

# Show the plot
plt.show()
```
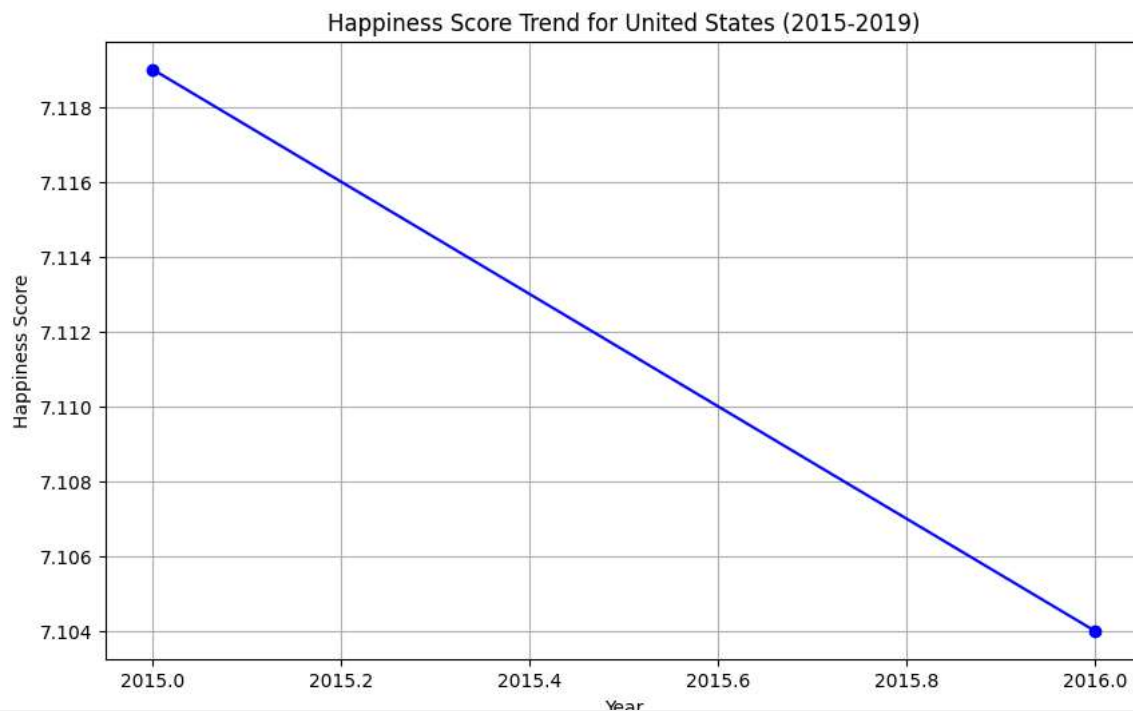

Happiness Score Trend for United States (2015-2019)

```
# 1. Happiness Score Trend Analysis:
# The plot shows the happiness score of the United States over the years from 2015 to 2019.
# The score remains relatively stable, with slight fluctuations, showing no drastic increase or decrease.
# Overall, the trend suggests the happiness score has not changed significantly over this period.


# 2. Happiness Scores Across Countries:
# Some countries, such as Finland and Denmark, consistently rank high in happiness scores across the years.
```

```
# Countries like Afghanistan and Syria have consistently lower scores, indicating significant well-being challenges.
# There is a noticeable gap between high-ranking and low-ranking countries, often influenced by economic and social factors.


# 3. Key Factor Analysis (Log GDP per capita, Social support):
# Log GDP per capita: Countries with higher GDP, such as Switzerland and Norway, tend to have higher happiness scores. Lower-GDP countries, like some in S
# Social Support: Countries with strong social support systems (e.g., Nordic countries) score consistently higher in happiness. Countries with weaker soci
```