



PROJECT REPORT ON

“Social Progress Index Analysis using Python”

SUBMITTED BY:

Ramjan Ali

ET22BTHCS113

SECTION-B

Table Of Contents

Problem Statement.

Introduction.

Dataset-Overview.

Data analysis and Visualization.

Conclusion.

Reference.

Problem Statement:

The objective of this analysis is to explore the relationships between socio-economic indicators and the global happiness score, as provided by the World Happiness Report dataset. The key focus areas include:

1. Identifying correlations between SPI components and happiness scores to determine significant predictors of well-being.
2. Visualizing the influence of social support, GDP, life expectancy, and other factors on happiness through bubble charts.
3. Analyzing regional disparities in SPI component scores to highlight areas for potential policy intervention.
4. Comparing the top 5 happiest countries based on key SPI components using a radar chart.
5. Mapping global patterns of health, social support, and SPI scores using choropleth maps to provide a geographic perspective.

Introduction:

The World Happiness Report provides data on global happiness levels based on various socio-economic factors. This analysis explores the dataset to understand correlations, regional differences, and key determinants of happiness, using visualizations such as bubble charts, heatmaps, radar charts, and choropleth maps.

Dataset Overview:

The dataset includes the following key components:

- Social Support
- Healthy Life Expectancy
- Freedom to Make Life Choices
- Generosity
- Perceptions of Corruption
- Dystopia + Residual
- Logged GDP per Capita
- Ladder Score (Happiness Score)
- Regional Indicator
- Country Name

Data Analysis and Visualizations:

1. Correlation Analysis

A heatmap is generated to analyze the correlation between key components of the Social Progress Index (SPI) and the Happiness Score. The heatmap helps identify strong or weak relationships among these variables.

Code:

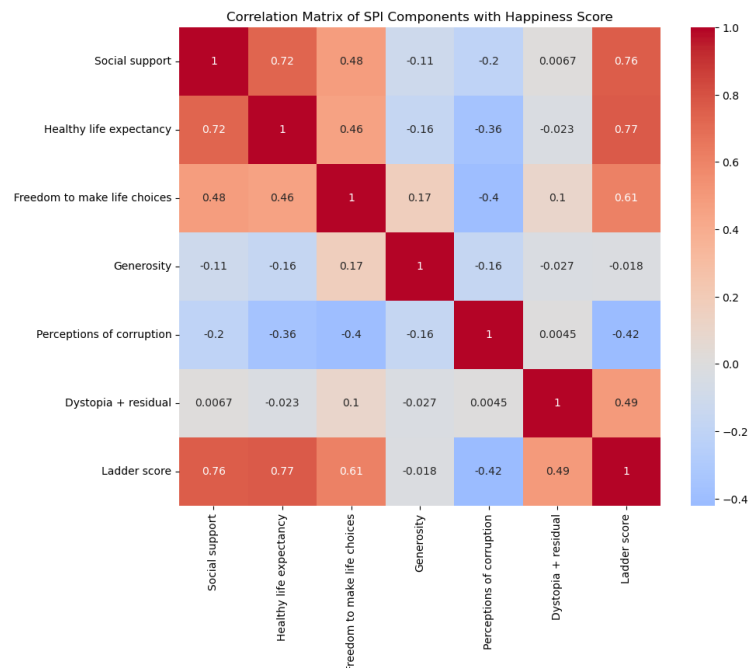
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the data
df = pd.read_csv('C:/Users/Acer/OneDrive/Desktop/Week 1/world-happiness-report 1.csv')

# Define SPI components
spi_components = ['Social support', 'Healthy life expectancy',
                  'Freedom to make life choices', 'Generosity',
                  'Perceptions of corruption', 'Dystopia + residual']

# Correlation Matrix
plt.figure(figsize=(10, 8))
correlation_matrix = df[spi_components + ['Ladder score']].corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0)
plt.title('Correlation Matrix of SPI Components with Happiness Score')
plt.show()
```

Output:



2. Bubble Charts Analysis:

Bubble charts visually depict the relationships between variables, with bubble sizes representing an additional dimension of data. Three bubble charts are created to analyze:

- Happiness vs Social Support
- Freedom vs Corruption
- GDP vs Happiness with Life Expectancy as bubble size

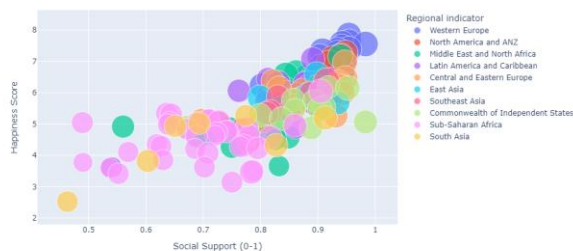
Code:

```
import plotly.express as px

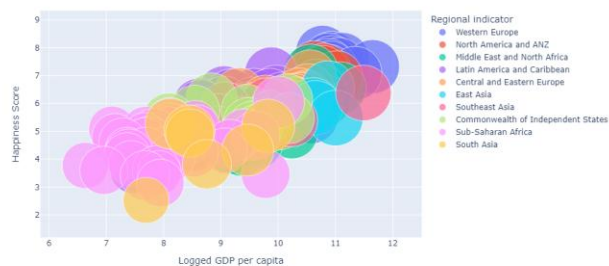
# Bubble Chart 1: Happiness vs Social Support
fig1 = px.scatter(
    df,
    x='Social support',
    y='Ladder score',
    size='Logged GDP per capita',
    color='Regional indicator',
    hover_name='Country name',
    title='Happiness vs. Social Support (Bubble Size = GDP per capita)',
    size_max=30
)
fig1.show()
```

Output:

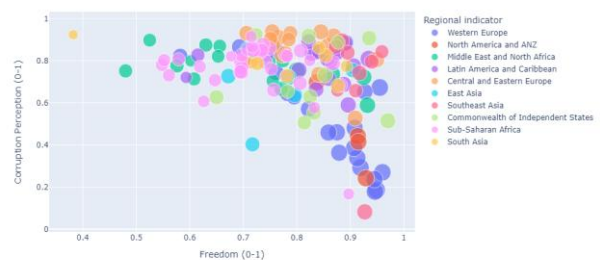
Happiness vs. Social Support (Bubble Size = GDP per capita)



Happiness vs GDP with Life Expectancy Bubble Size



Freedom vs. Corruption (Bubble Size = Happiness Score)



3. Global Maps Analysis:

Choropleth maps provide a geographical representation of data distribution across countries. Three maps are generated to visualize:

- Healthy Life Expectancy
- Social Support Index
- Composite SPI Score

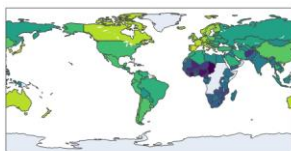
Code:

```
import plotly.express as px

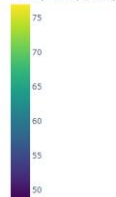
# Map 1: Healthy Life Expectancy
map1 = px.choropleth(
    df,
    locations='Country name',
    locationmode='country names',
    color='Healthy life expectancy',
    hover_name='Country name',
    color_continuous_scale='Viridis',
    title='Global Healthy Life Expectancy (Years)'
)
map1.show()
```

Output:

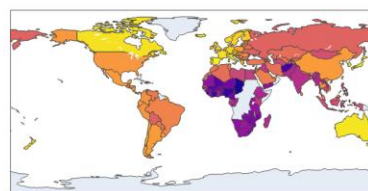
Global Healthy Life Expectancy (Years)



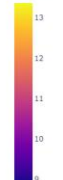
Healthy life expectancy



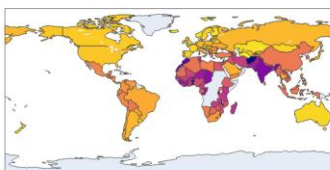
Composite Social Progress Index (SPI) Score



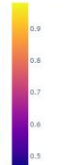
SPI Score



Global Social Support Index



Social support



4. Radar Chart Analysis:

The radar chart provides a visual comparison of key SPI components for the top 5 happiest countries. Each axis represents a component, allowing for multi-dimensional analysis.

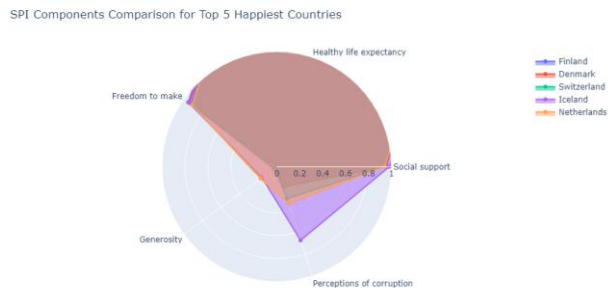
Code:

```
import plotly.graph_objects as go

# Top 5 Countries
top_countries = df.nlargest(5, 'Ladder score')

radar_fig = go.Figure()
for i, row in top_countries.iterrows():
    radar_fig.add_trace(go.Scatterpolar(
        r=[
            row['Social support'],
            row['Healthy life expectancy'],
            row['Freedom to make life choices'],
            row['Generosity'],
            row['Perceptions of corruption'],
            row['Social support']
        ],
        theta=[
            'Social support', 'Healthy life expectancy',
            'Freedom to make', 'Generosity',
            'Perceptions of corruption', 'Social support'
        ],
        fill='toself',
        name=row['Country name']
    ))
radar_fig.update_layout(
    polar=dict(radialaxis=dict(visible=True, range=[0, 1])),
    showlegend=True,
    title='SPI Components Comparison for Top 5 Happiest Countries'
)
radar_fig.show()
```

Output:



5. Regional Analysis:

The final visualization provides a comparative bar chart of average SPI component scores across regions. This analysis highlights regional disparities and strengths in various SPI dimensions.

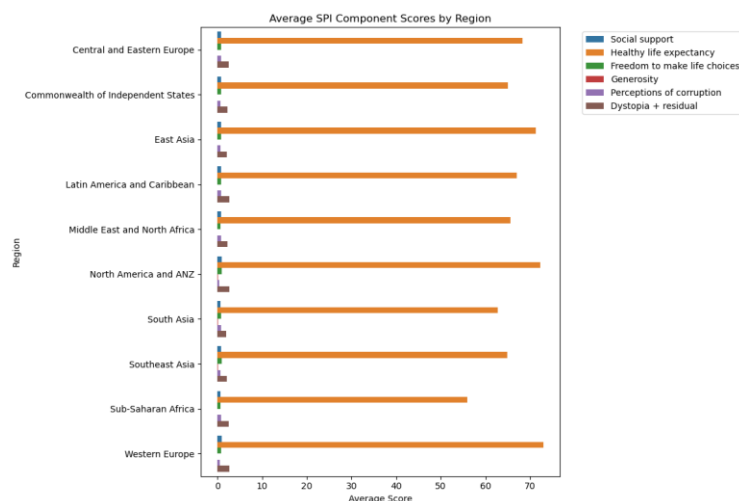
Code:

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

# Regional Analysis
regional_means = df.groupby('Regional indicator')[spi_components].mean().reset_index()
melted_regional = pd.melt(
    regional_means,
    id_vars=['Regional indicator'],
    value_vars=spi_components,
    var_name='SPI Component',
    value_name='Average Score'
)

plt.figure(figsize=(12, 8))
sns.barplot(
    data=melted_regional,
    x='Average Score',
    y='Regional indicator',
    hue='SPI Component'
)
plt.title('Average SPI Component Scores by Region')
plt.xlabel('Average Score')
plt.ylabel('Region')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

Output:



Conclusion:

The World Happiness Report data provides a comprehensive perspective on global happiness, emphasizing the impact of socio-economic factors such as social support, life expectancy, and freedom of choice. The visualizations reveal significant regional disparities and highlight areas for potential policy intervention to enhance overall happiness and well-being.

References:

1. World Happiness Report Dataset: The dataset utilized in this analysis was obtained from the World Happiness Report, which provides global data on socio-economic indicators influencing happiness scores. [Source: World Happiness Report](#)
2. Plotly Documentation: The visualizations, including bubble charts, choropleth maps, and radar charts, were created using Plotly. Source: Plotly
3. Seaborn and Matplotlib: The correlation matrix and bar charts were developed using Seaborn and Matplotlib for enhanced data visualization. Source: Seaborn, [Source: Matplotlib](#)