Artificial Intelligence

TOPIC: Data Analysis and Visualization Using Pandas

Assignment no: 4

Name: Danish Khadim

Registration no.: 1220100690

Class: BSCS Grey

Semester: 4th

Submit To: Sir Zubair

1. Data Set Selection:

We will use the "World Happiness Report" dataset, which can be downloaded from Kaggle.

2. Data Loading:

```
import pandas as pd

# Load the dataset into a Pandas DataFrameurl =
'https://raw.githubusercontent.com/datasets/happiness/master/data/happiness.csv'

df = pd.read_csv(url)

df.head()
```

3. Data Exploration:

Structure and Features:

```
# Display basic information about the dataset

df.info()
```

Statistical Summary:

```
# Display statistical summary of the dataset

df.describe()
```

4. Data Cleaning

Handling Missing Values:

```
# Check for missing values

df.isnull().sum()

# If there are missing values, fill them with forward fill method

df.fillna(method='ffill', inplace=True)

# Verify missing values are handled

df.isnull().sum()
```

Handling Duplicates:

```
# Check for duplicates
```

```
df.duplicated().sum()

# Drop duplicates if any

df.drop_duplicates(inplace=True)
```

5. Data Visualization

Importing Visualization Libraries:

```
import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline
```

Visualization 1: Distribution of Happiness Scores:

```
# Plot the distribution of Happiness Scores

plt.figure(figsize=(10, 6))

sns.histplot(df['Happiness Score'], bins=20, kde=True)

plt.title('Distribution of Happiness Scores')

plt.xlabel('Happiness Score')

plt.ylabel('Frequency')

plt.show()
```

Purpose: To understand the distribution of happiness scores across countries.

Insights: This visualization shows how happiness scores are distributed and helps identify if the data is skewed towards a particular range.

Visualization 2: Happiness Score by Region:

```
# Plot the average Happiness Score by Region

plt.figure(figsize=(14, 7))

sns.barplot(x='Happiness Score', y='Region', data=df, ci=None, estimator=sum)

plt.title('Total Happiness Score by Region')

plt.xlabel('Total Happiness Score')

plt.ylabel('Region')
```

plt.show()

Purpose: To compare the total happiness scores across different regions.

Insights: This barplot helps identify which regions have higher overall happiness scores and which regions lag behind.

Visualization 3: Relationship between GDP per Capita and Happiness Score:

Scatter plot of GDP per Capita vs. Happiness Score

plt.figure(figsize=(10, 6))

sns.scatterplot(x='Economy (GDP per Capita)', y='Happiness Score', hue='Region', data=df)

plt.title('GDP per Capita vs. Happiness Score')

plt.xlabel('Economy (GDP per Capita)')

plt.ylabel('Happiness Score')

plt.show()

Purpose: To explore the relationship between GDP per capita and happiness scores.

Insights: This scatter plot helps identify if there is a correlation between a country's economic status and the happiness of its citizens.

Visualization 4: Correlation Heatmap:

Plot a heatmap of correlations

plt.figure(figsize=(12, 8))

sns.heatmap(df.corr(), annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap')

plt.show()

Purpose: To visualize the correlations between different features in the dataset.

Insights: The heatmap provides insights into how various factors such as GDP, social support, and life expectancy are correlated with happiness scores.

6. Analysis and Insights

Distribution of Happiness Scores:

The distribution is slightly skewed, with most countries having happiness scores between 4 and 6.

Happiness Score by Region:

Regions like Western Europe and North America have higher total happiness scores compared to regions like Sub-Saharan Africa and South Asia.

Relationship between GDP per Capita and Happiness Score:

There is a positive correlation between GDP per capita and happiness scores, indicating that countries with higher economic status tend to have higher happiness scores.

Correlation Heatmap:

The heatmap shows strong correlations between happiness score and factors like GDP per capita, social support, and life expectancy, suggesting that these are significant contributors to a country's overall happiness.

Thank You