stats 1.3

Aim:

To perform exploratory data analysis (EDA) on a dataset related to mental health metrics, including anxiety, stress, and depression, and visualize the relationships and distributions of these metrics.

Objective:

- To analyze the distribution of mental health metrics across different universities, genders, and academic years.
- To visualize relationships between anxiety, stress, and depression.
- To provide summary statistics and generate visualizations, including bar charts, histograms, scatter plots, line graphs, heatmaps, and box plots.

Resources:

- **Software:** Python 3.x, Pandas, Matplotlib, Seaborn.
- **Hardware:** Standard personal computer or laptop.
- **Dataset:** A CSV file containing data on mental health metrics (Anxiety, Stress, Depression) along with gender, university, and academic year information.

Procedure:

1. Load the dataset using Pandas.
2. Preprocess the data by shortening university names and organizing the academic year in the correct order.
3. Group data by gender, university, and academic year to calculate the mean values of anxiety, stress, and depression.
4. Generate and save various plots (bar charts, histograms, scatter plots, line graphs, heatmaps, box plots) to analyze the data.
5. Display summary statistics for the mental health metrics.

Code:

```
import os

import pandas as pd
```

```python
import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

df = pd.read_csv(r'C:\Users\harsh\Downloads\Stats second year 3rd sem\Stress
monitoring model\Raw Data.csv')

# Create output folder

output_folder = 'C:\Users\harsh\Downloads\Learn Statatics From
Scratch\Stress monitoring model\New folder'

os.makedirs(output_folder, exist_ok=True)

# Group data by Gender, University, and Academic Year, calculating mean
values for mental health metrics

grouped_data = df.groupby(['2. Gender', '3. University', '5. Academic
Year'])[['Anxiety Value', 'Stress Value', 'Depression
Value']].mean().reset_index()

# Bar Charts for each grouping column

for col in ['2. Gender', '3. University', '5. Academic Year']:

    fig, axs = plt.subplots(1, 3, figsize=(18, 5))

    for i, metric in enumerate(['Anxiety Value', 'Stress Value', 'Depression
Value']):

        axs[i].bar(grouped_data[col], grouped_data[metric], color=
['skyblue', 'lightgreen', 'lightcoral'][i], edgecolor='black')

        axs[i].set_title(f'Average {metric} by {col}')

        axs[i].set_xlabel(col)

        axs[i].set_ylabel(f'Average {metric}')

        axs[i].tick_params(axis='x', rotation=45)
```

```python
    plt.tight_layout()

    plt.savefig(os.path.join(output_folder, f'bar_chart_{col}.png'))

    plt.close()  # Close the plot to save memory

# Histograms for mental health metrics

df[['Anxiety Value', 'Stress Value', 'Depression Value']].hist(bins=20,
edgecolor='black', figsize=(18, 5))

plt.tight_layout()

plt.savefig(os.path.join(output_folder, 'histograms.png'))

plt.close()

# Scatter Plots to show relationships between metrics

sns.pairplot(df, vars=['Anxiety Value', 'Stress Value', 'Depression Value'],
kind='scatter', plot_kws={'alpha': 0.5})

plt.savefig(os.path.join(output_folder, 'scatter_plots.png'))

plt.close()

# Line Graphs over time (Academic Year)

df['5. Academic Year'] = pd.Categorical(df['5. Academic Year'], categories=
['First Year or Equivalent', 'Second Year or Equivalent', 'Third Year or
Equivalent', 'Fourth Year or Equivalent'], ordered=True)

grouped_time = df.groupby('5. Academic Year')[['Anxiety Value', 'Stress
Value', 'Depression Value']].mean()

grouped_time.plot(marker='o', linestyle='-', figsize=(10, 6))

plt.title('Mental Health Metrics Over Academic Years')

plt.ylabel('Average Value')

plt.tight_layout()
```

```python
plt.savefig(os.path.join(output_folder, 'line_graphs.png'))

plt.close()

# Heatmaps for mental health metrics by Gender and Academic Year

fig, axs = plt.subplots(1, 3, figsize=(18, 5))

for i, metric in enumerate(['Anxiety Value', 'Stress Value', 'Depression
Value']):

    sns.heatmap(df.pivot_table(index='5. Academic Year', columns='2.
Gender', values=metric, aggfunc='mean'), annot=True, cmap="YlGnBu",
ax=axs[i])

    axs[i].set_title(f'{metric} Heatmap')

plt.tight_layout()

plt.savefig(os.path.join(output_folder, 'heatmaps.png'))

plt.close()

# Box Plots for mental health metrics by Gender, University, and Academic
Year

for col in ['2. Gender', '3. University', '5. Academic Year']:

    fig, axs = plt.subplots(1, 3, figsize=(18, 5))

    for i, metric in enumerate(['Anxiety Value', 'Stress Value', 'Depression
Value']):

        sns.boxplot(x=col, y=metric, data=df, ax=axs[i])

        axs[i].set_title(f'{metric} by {col}')

        axs[i].tick_params(axis='x', rotation=45)

    plt.tight_layout()

    plt.savefig(os.path.join(output_folder, f'box_plot_{col}.png'))
```

```
    plt.close()


# Summary Statistics for the mental health metrics


summary_stats = df[['Anxiety Value', 'Stress Value', 'Depression
Value']].describe()


print("\nSummary Statistics:\n", summary_stats)
```

Dataset Information:

- **Columns:**
    - 'Anxiety Value': Numeric, representing anxiety levels.
    - 'Stress Value': Numeric, representing stress levels.
    - 'Depression Value': Numeric, representing depression levels.
    - '2. Gender': Categorical, indicating gender.
    - '3. University': Categorical, indicating the university attended.
    - '5. Academic Year': Categorical, indicating the academic year.
- **Format:** CSV file.