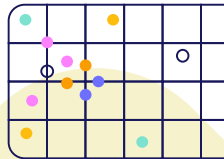
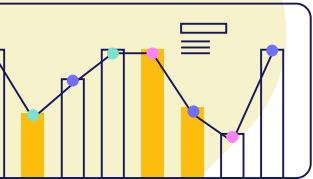
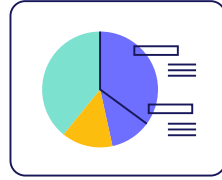


Data Analysis

Student Mental health survey



Team members and role

01.

Youssef kamel

Visualization on python.

02.

Lorna Magdy

Presentation.

03.

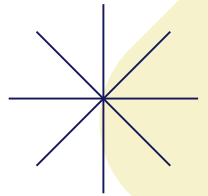
Marina Maged

Objectives and dashboard.

04.

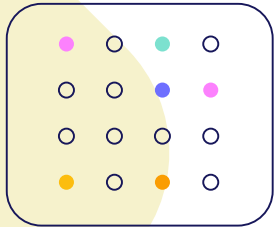
Ebrahiem saied

Objectives and dashboard.



01.

Student Mental Health Survey

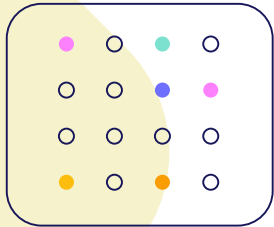


Key aspects covered in the dataset include:

- Demographic details such as gender, age, and university.
- Academic details like degree level, major, academic year, and current CGPA.
- Student's residential status and experiences with discrimination, harassment, or bullying on campus.
- Student's lifestyle factors include frequency of sports engagement and average sleep hours per night.
- Student's satisfaction with their field of study and their perception of academic workload.
- Addressing the academic pressure, financial concerns, and the quality of social relationships on campus.
- Frequency of experiencing depression, anxiety, feelings of isolation, and insecurity about the future.
- Activities that students engage in to relieve stress.

02.

Objectives





Show demographic Charts:

- Gender—are males more than females in IT ?
- Age – younger or older are in this field ?
- University—which university has more students?

Campus-related things:

Are on-campus students getting more grades?

Are on-campus students feeling more isolation?

University-related things:

Is there specific university with most depressed students?



Mental Health Factors and Academic Pressure:

Is there a Correlation between academic pressure and levels of depression, anxiety, and isolation.

Impact of Discrimination:

Analyze how campus discrimination affects mental health outcomes.

Sleep:

Is there a significant relationship between sleep time and academic pressure?



Analysis Approach:

Group the respondents based on their social relationships ratings. Examine how the levels of depression, anxiety, and isolation vary with these groups.

Stress Relief Methods:

What is the most common stress relief activities and their effectiveness on reducing mental health issues.

Hypothesis is the upcoming statement true:

" Stronger social relationships might be associated with lower levels of depression, anxiety, and isolation"?

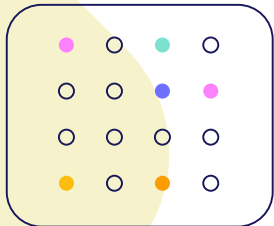


project 3.ipynb



03.

Python code



import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

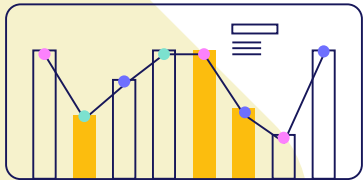
df = pd.read_csv("MentalHealthSurvey.csv")

1. Count duplicates

```
duplicate_count = df.duplicated().sum()  
print(f"Number of duplicate rows: {duplicate_count}")
```

2. Count null values

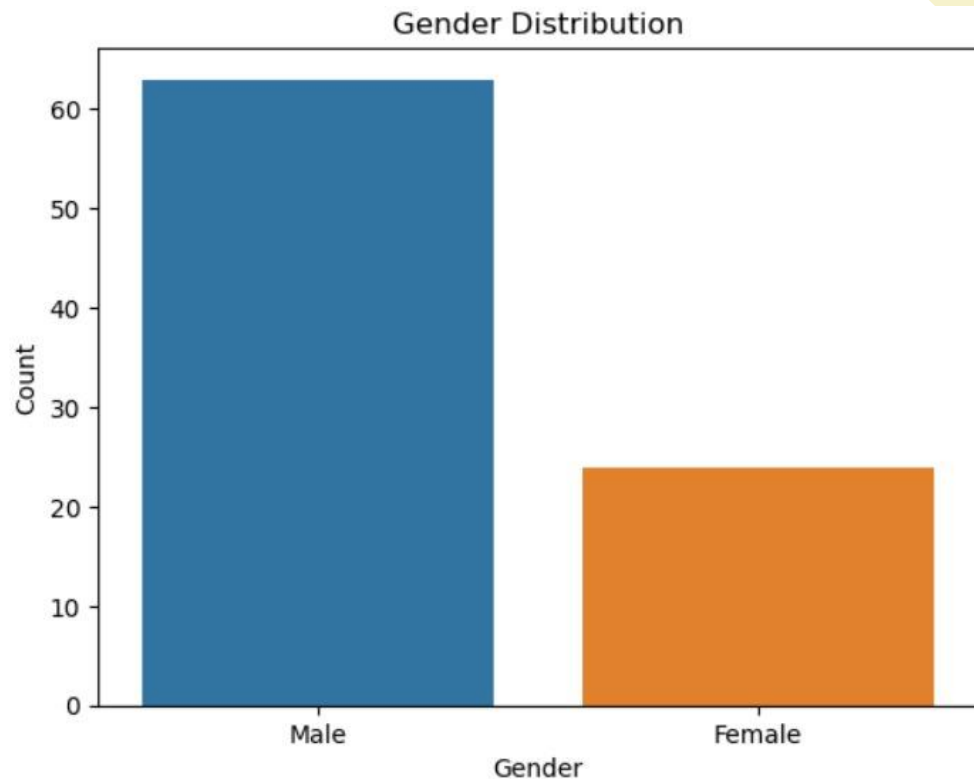
```
null_count = df.isnull().sum()  
print("\nNumber of null values in each column:")  
print(null_count)
```



import seaborn as sns
import matplotlib.pyplot as plt

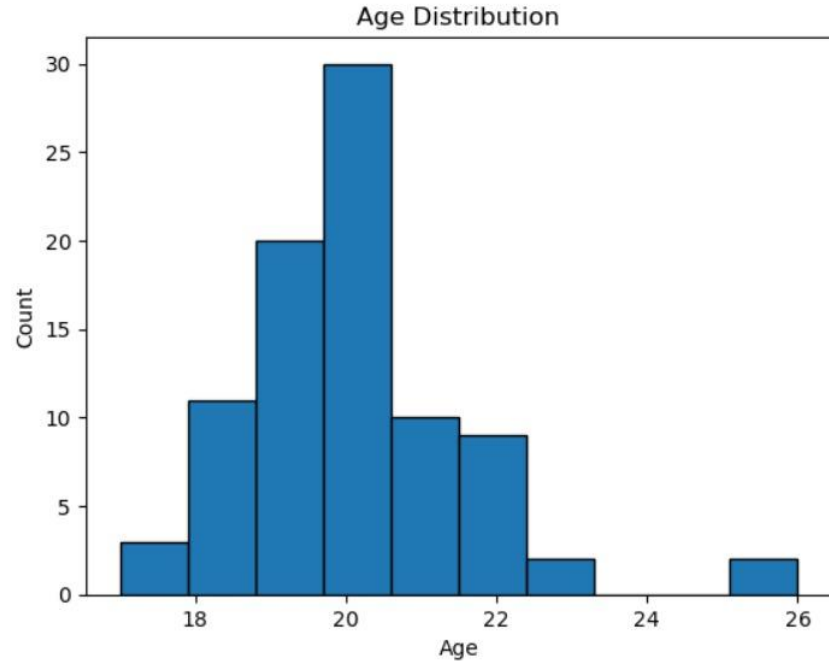
Gender distribution bar chart

```
sns.countplot(x='gender', data=  
plt.title('Gender Distribution')  
plt.xlabel('Gender')  
plt.ylabel('Count')  
plt.show()
```



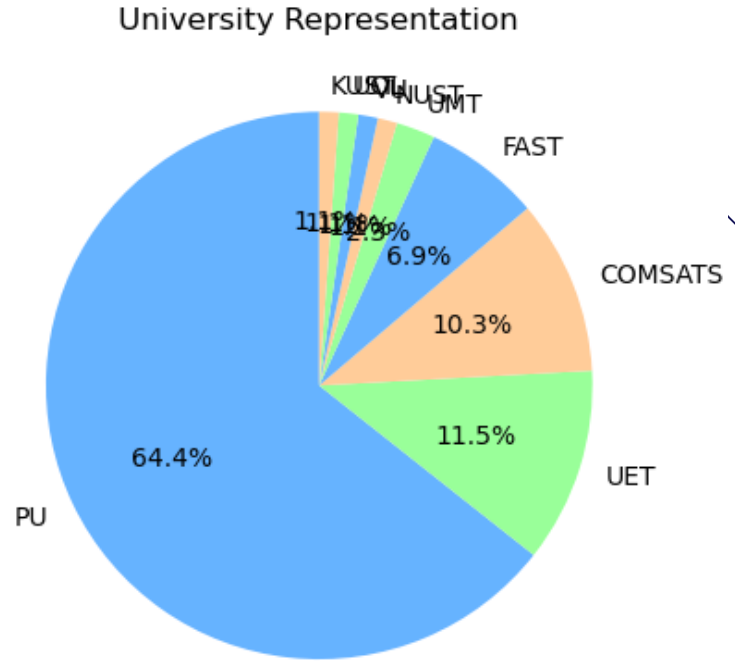
Age distribution histogram

```
plt.hist(df['age'], bins=10,  
edgecolor='black')  
plt.title('Age Distribution')  
plt.xlabel('Age')  
plt.ylabel('Count')  
plt.show()
```



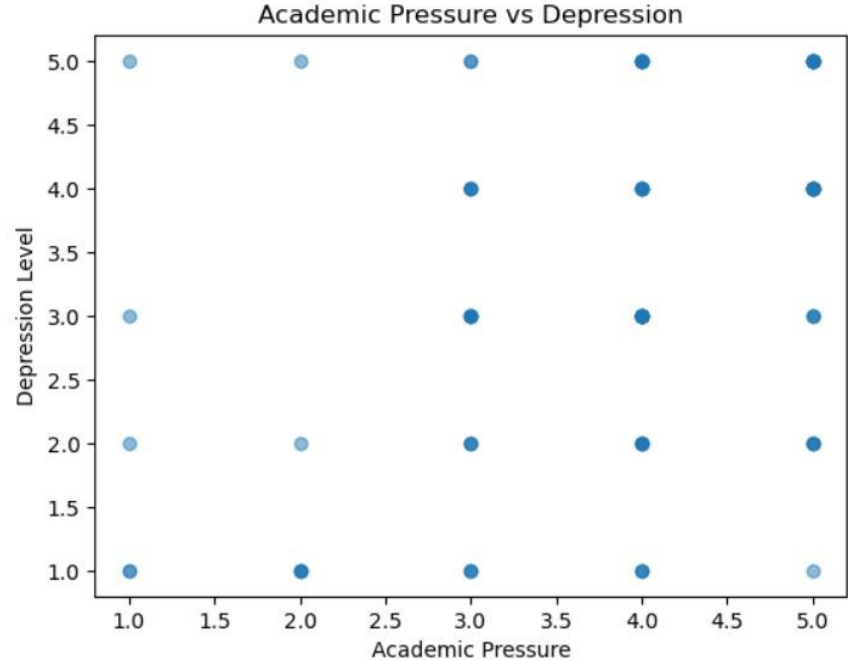
University distribution pie chart

```
df['university'].value_counts().plot.pie(autopct='%1.1f%%',
startangle=90,
colors=['#66b3ff','#99ff99','#ffcc99'])
plt.title('University Representation')
plt.ylabel('')
plt.show()
```



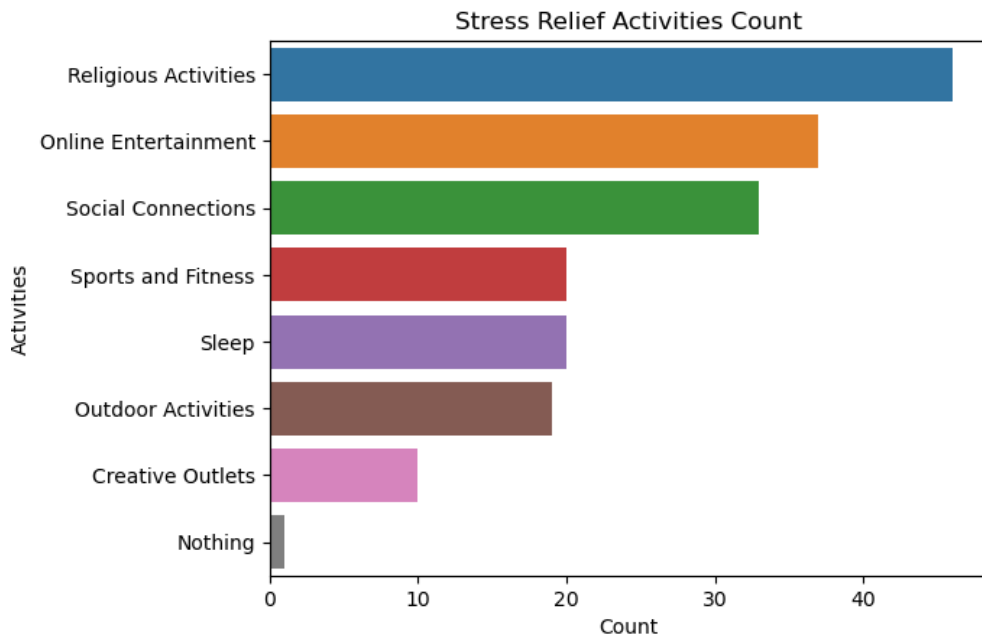
Scatter plot: Academic pressure vs Depression

```
plt.scatter(df['academic_pressure'],  
            df['depression'], alpha=0.5)  
plt.title('Academic Pressure vs  
Depression')  
plt.xlabel('Academic Pressure')  
plt.ylabel('Depression Level')  
plt.show()
```



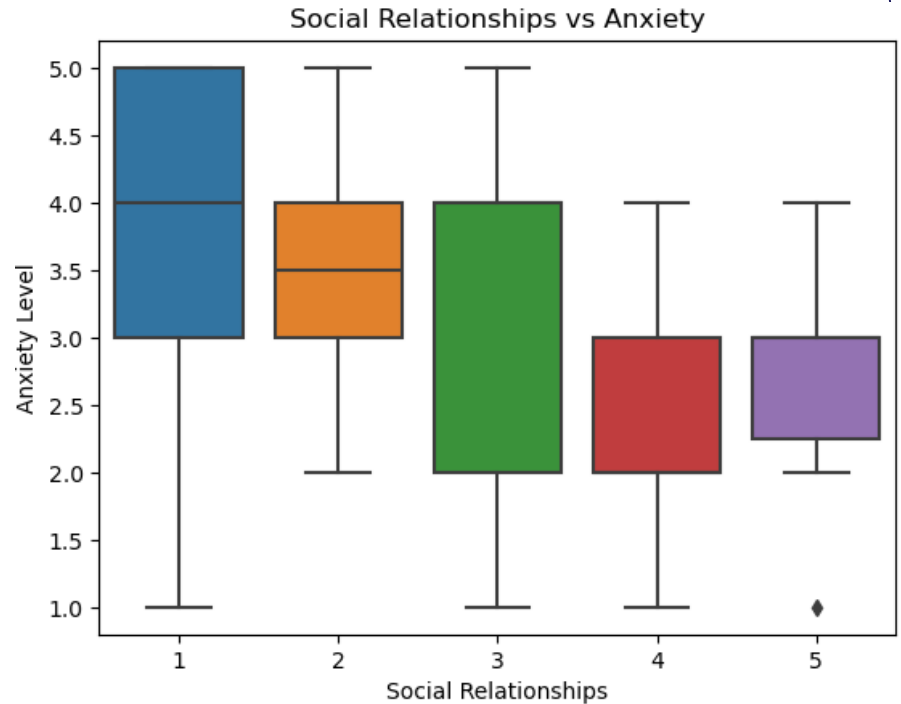
Stress relief activities countplot

```
df['stress_relief_activities'] =  
df['stress_relief_activities'].str.split(', ')  
activities = df['stress_relief_activities'].explode()  
sns.countplot(y=activities, order=activities.value_counts().index)  
plt.title('Stress Relief Activities Count')  
plt.xlabel('Count')  
plt.ylabel('Activities')  
plt.show()
```



Social relationships vs Anxiety

```
sns.boxplot(x='social_relationships', y='anxiety', data=df)
plt.title('Social Relationships vs Anxiety')
plt.xlabel('Social Relationships')
plt.ylabel('Anxiety Level')
plt.show()
```



Violin plot: Future insecurity vs Stress relief activities

```
sns.violinplot(x='stress_relief_activities', y='future_insecurity',  
data=df.explode('stress_relief_activities'))
```

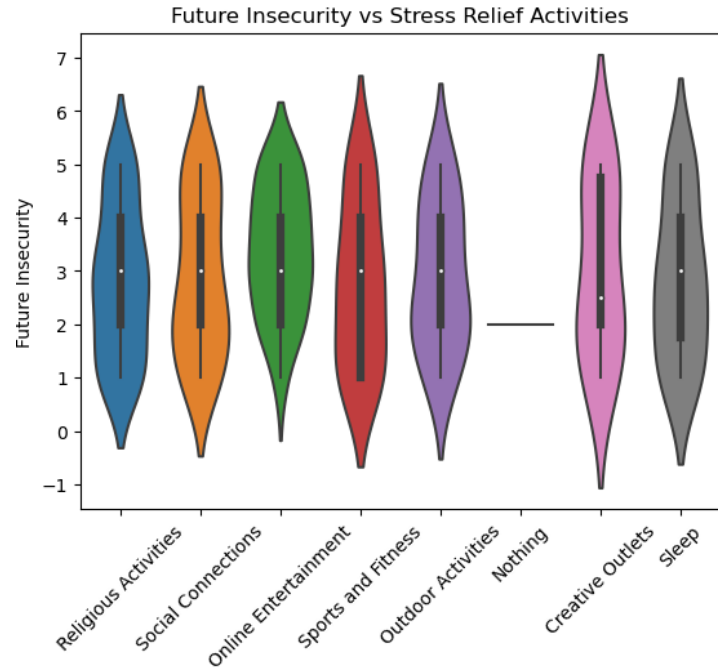
```
plt.title('Future Insecurity vs Stress Relief Activities')
```

```
plt.xlabel('Stress Relief Activities')
```

```
plt.ylabel('Future Insecurity')
```

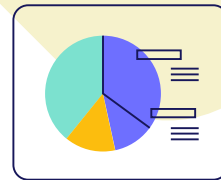
```
plt.xticks(rotation=45)
```

```
plt.show()
```



Create a function to convert CGPA ranges to their averages

```
def convert_cgpa_range(cgpa_range):  
    lower, upper = map(float, cgpa_range.split('-'))  
    return (lower + upper) / 2
```



Apply the function to the 'cgpa' column

```
df['cgpa'] = df['cgpa'].apply(convert_cgpa_range)
```

Create a function to convert sleep hours ranges to their averages

```
def convert_sleep_range(sleep_range):  
    lower, upper = map(float, sleep_range.replace(' hrs', '').split('-'))  
    return (lower + upper) / 2
```

Apply the function to the 'average_sleep' column

```
df['average_sleep'] = df['average_sleep'].apply(convert_sleep_range)
```

CGPA based on residential status (On-Campus vs Off-Campus)

```
sns.boxplot(x='residential_status', y='cgpa', data=df)
```

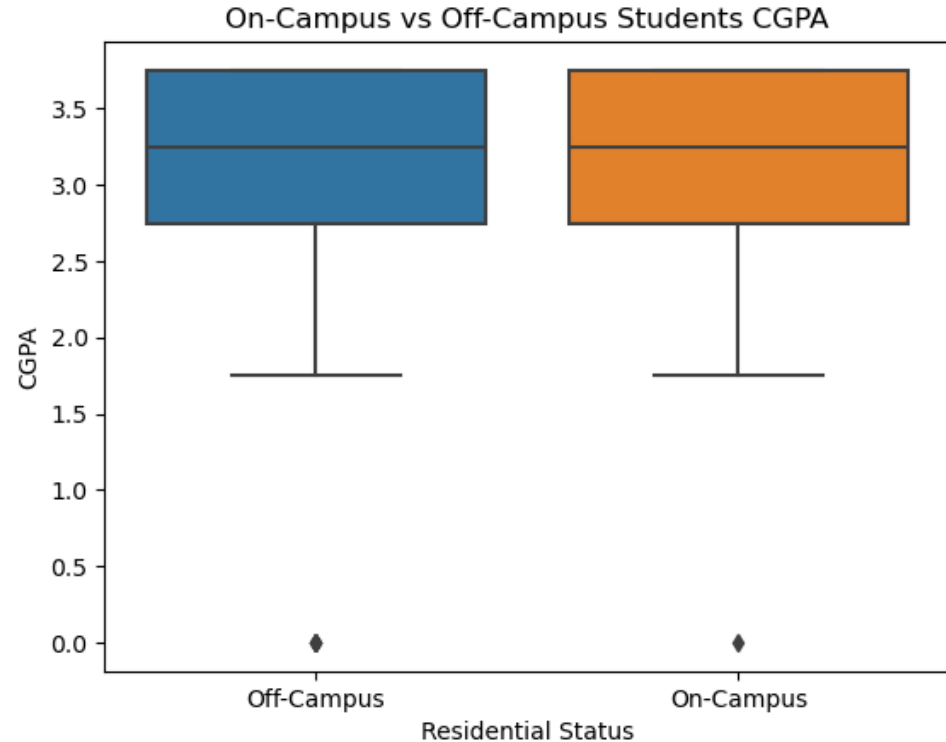
```
plt.title('On-Campus vs Off-
```

○ Campus Students CGPA')

```
plt.xlabel('Residential Status')
```

```
plt.ylabel('CGPA')
```

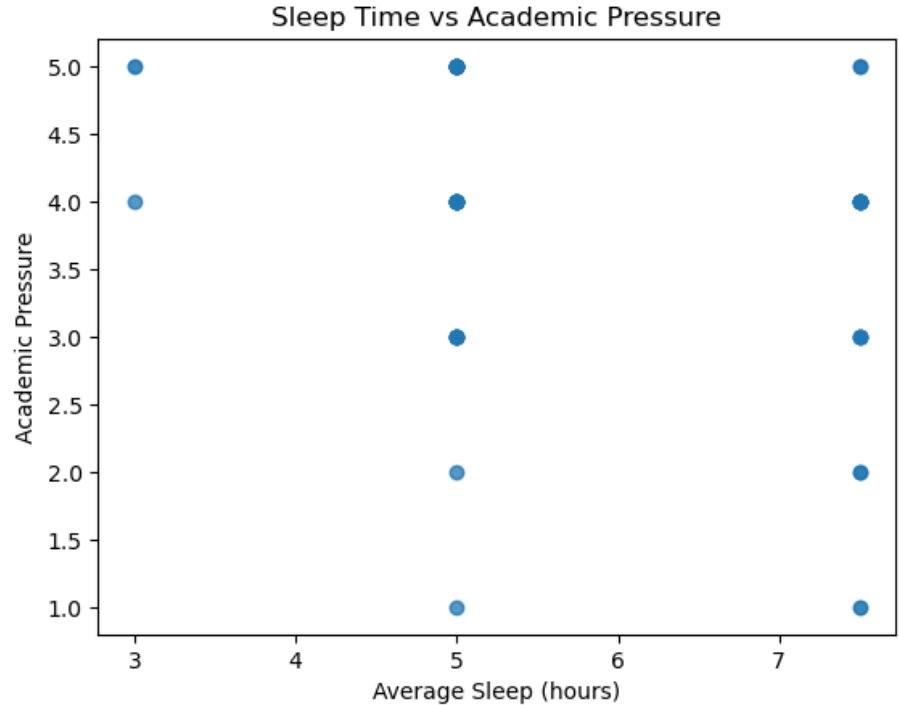
```
plt.show()
```



Sleep time vs Academic Pressure scatter plot

```
plt.scatter(df['average_sleep'],  
df['academic_pressure'],  
alpha=0.5)
```

- plt.title('Sleep Time vs Academic Pressure')
- plt.xlabel('Average Sleep (hours)')
- plt.ylabel('Academic Pressure')
- plt.show()



Social relationships vs Depression, Anxiety, Isolation

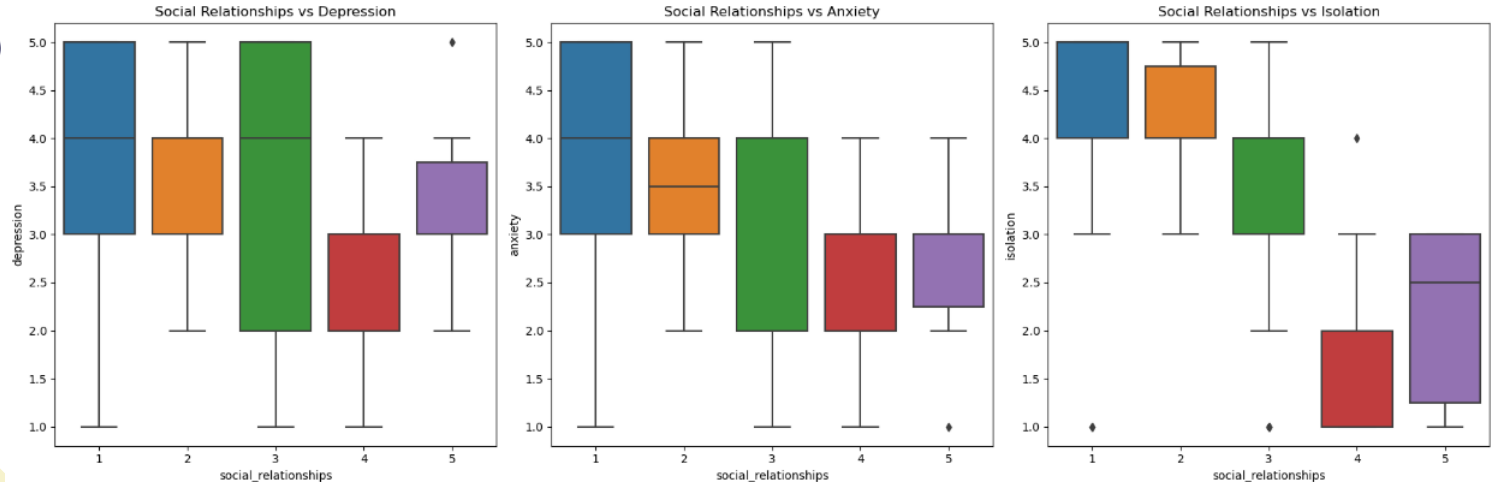
```
fig, axes = plt.subplots(1, 3, figsize=(18, 6))
```

```
sns.boxplot(x='social_relationships', y='depression', data=df, ax=axes[0])  
axes[0].set_title('Social Relationships vs Depression')
```

```
sns.boxplot(x='social_relationships', y='anxiety', data=df, ax=axes[1])  
axes[1].set_title('Social Relationships vs Anxiety')
```

```
sns.boxplot(x='social_relationships', y='isolation', data=df, ax=axes[2])  
axes[2].set_title('Social Relationships vs Isolation')
```

```
plt.tight_layout()  
plt.show()
```



Mental health outcomes by discrimination

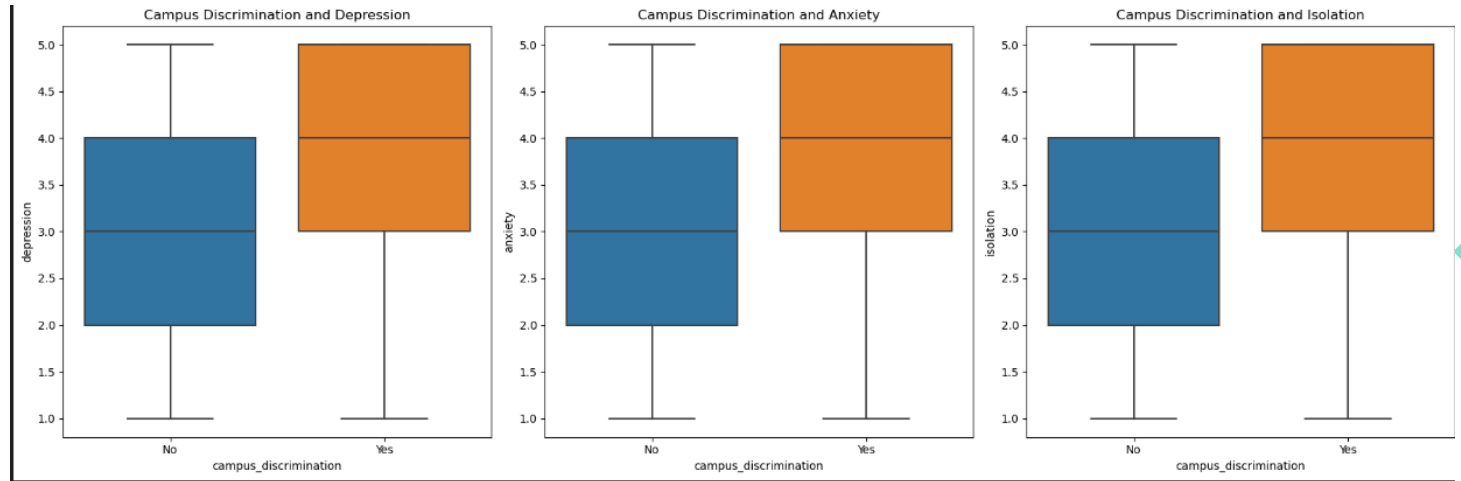
```
fig, axes = plt.subplots(1, 3, figsize=(18, 6))
```

```
sns.boxplot(x='campus_discrimination', y='depression', data=df, ax=axes[0])  
axes[0].set_title('Campus Discrimination and Depression')
```

```
sns.boxplot(x='campus_discrimination', y='anxiety', data=df, ax=axes[1])  
axes[1].set_title('Campus Discrimination and Anxiety')
```

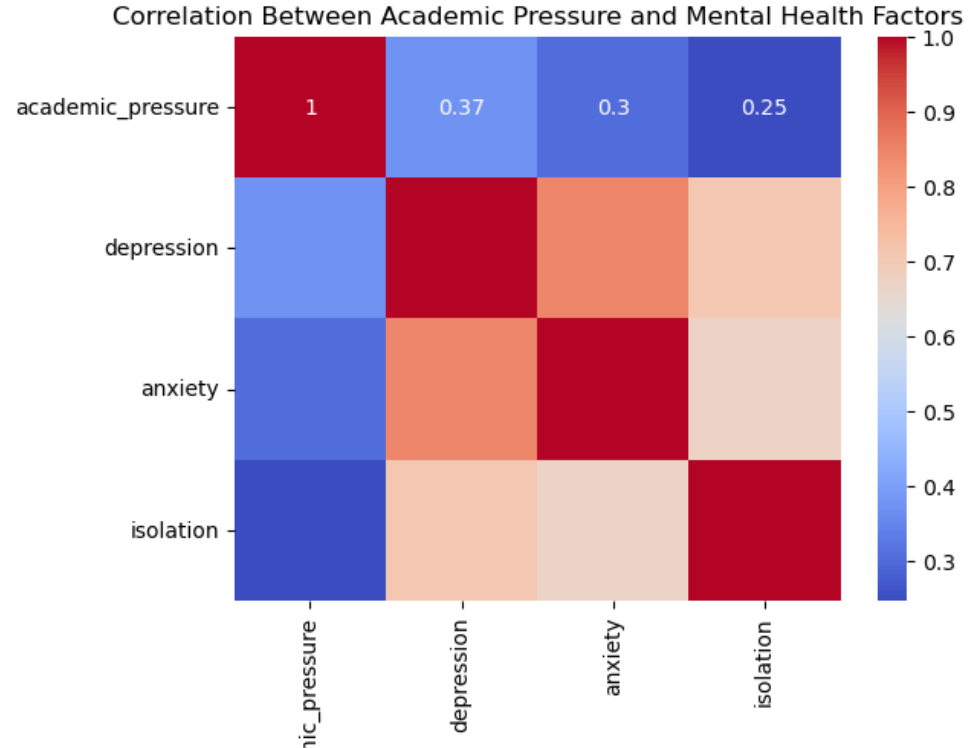
```
sns.boxplot(x='campus_discrimination', y='isolation', data=df, ax=axes[2])  
axes[2].set_title('Campus Discrimination and Isolation')
```

```
plt.tight_layout()  
plt.show()
```



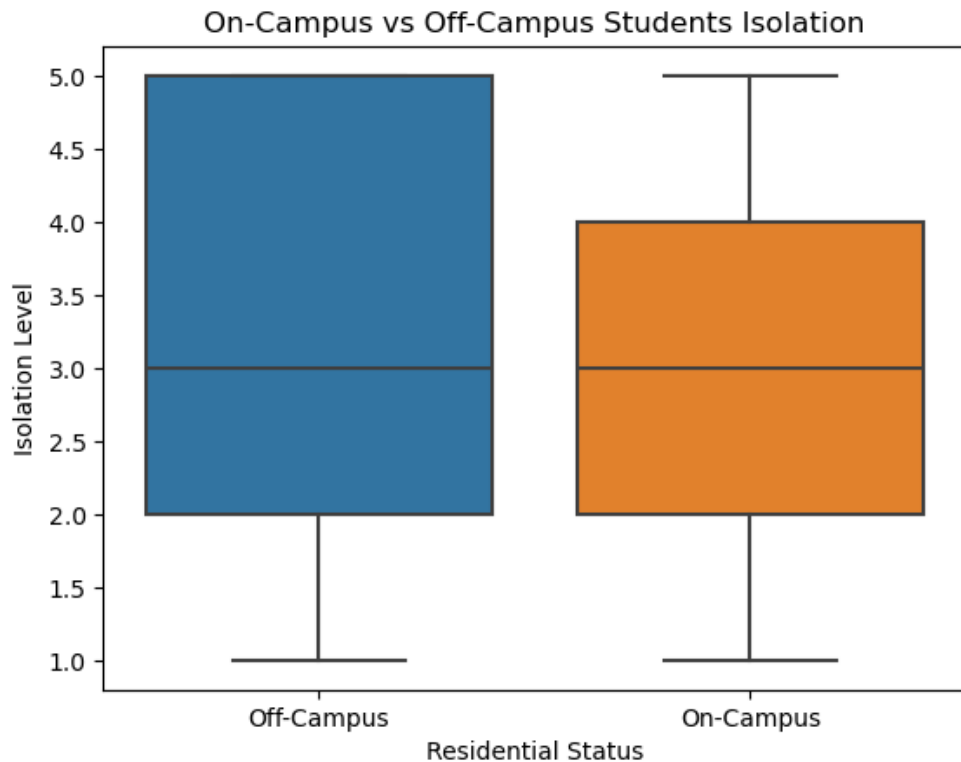
Correlation heatmap for mental health factors and academic pressure

```
corr_matrix =  
df[['academic_pressure',  
    'depression', 'anxiety',  
    'isolation']].corr()  
sns.heatmap(corr_matrix,  
    annot=True, cmap='coolwarm')  
plt.title('Correlation Between  
Academic Pressure and Mental  
Health Factors')  
plt.show()
```



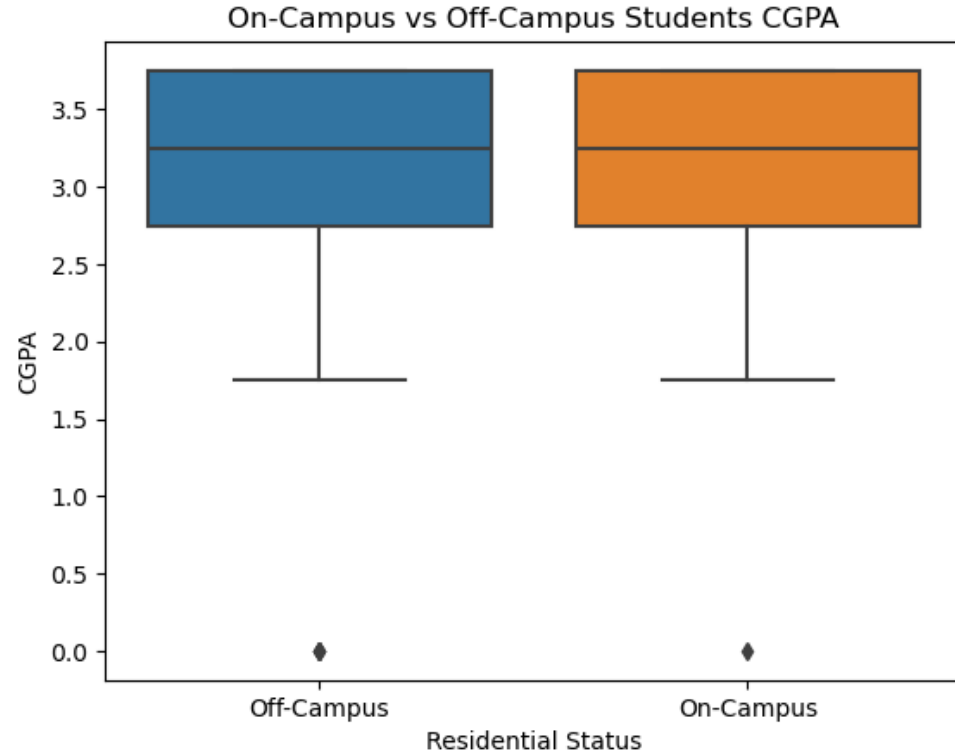
Isolation based on residential status

```
sns.boxplot(x='residential_status', y='isolation', data=df)  
plt.title('On-Campus vs Off-Campus Students Isolation')  
plt.xlabel('Residential Status')  
plt.ylabel('Isolation Level')  
plt.show()
```



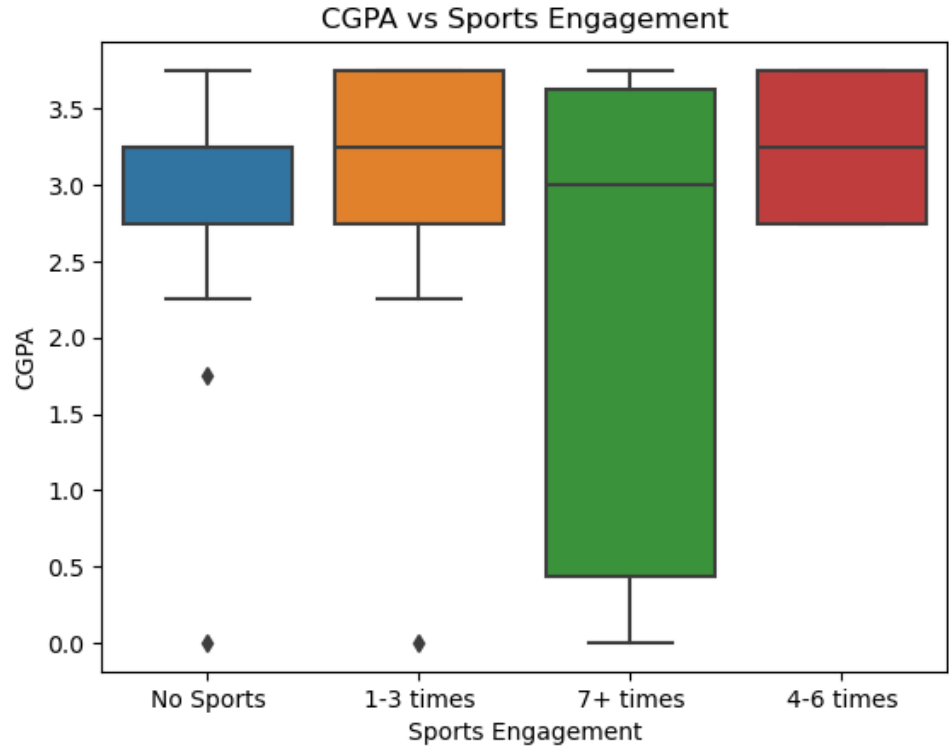
CGPA based on residential status

```
sns.boxplot(x='residential_status', y='cgpa', data=df)  
plt.title('On-Campus vs Off-Campus Students CGPA')  
plt.xlabel('Residential Status')  
plt.ylabel('CGPA')  
plt.show()
```



Sports engagement vs CGPA

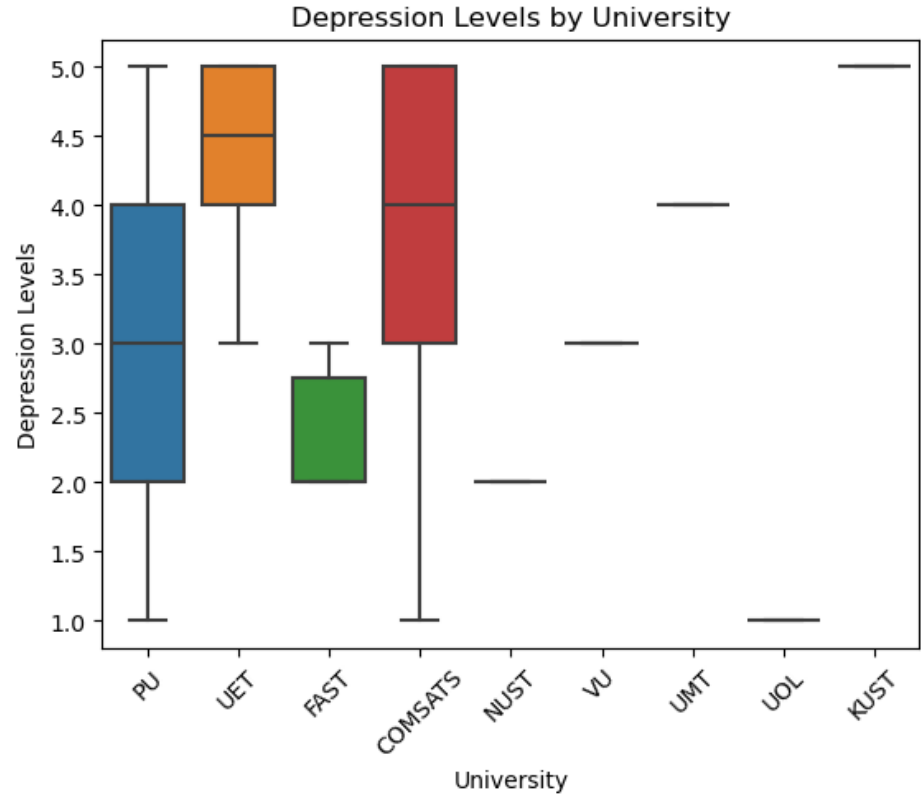
```
sns.boxplot(x='sports_engagem  
ent', y='cgpa', data=df)  
plt.title('CGPA vs Sports  
Engagement')  
○ plt.xlabel('Sports Engagement')  
plt.ylabel('CGPA')  
plt.show()
```



Depression levels by university

```
sns.boxplot(x='university',  
y='depression', data=df)  
plt.title('Depression Levels by  
University')
```

- plt.xlabel('University')
- plt.ylabel('Depression Levels')
- plt.xticks(rotation=45)
- plt.show()



import pandas as pd
import matplotlib.pyplot as plt

Assuming df is your DataFrame

Convert columns to numeric, coercing errors

```
df['depression'] = pd.to_numeric(df['depression'], errors='coerce')  
df['anxiety'] = pd.to_numeric(df['anxiety'], errors='coerce')  
df['isolation'] = pd.to_numeric(df['isolation'], errors='coerce')
```

Drop rows with NaN values in specified columns

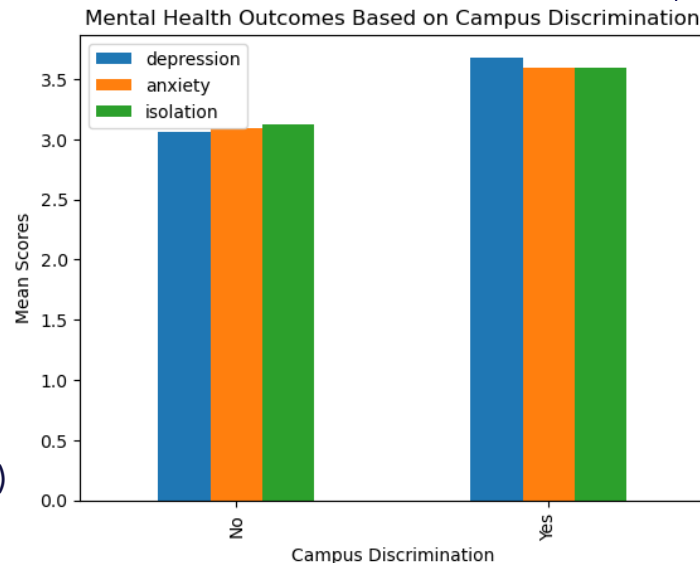
```
df.dropna(subset=['depression', 'anxiety', 'isolation'], inplace=True)
```

Group by 'campus_discrimination' and calculate the mean

```
df_grouped = df.groupby('campus_discrimination')[['depression', 'anxiety',  
'isolation']].mean().reset_index()
```

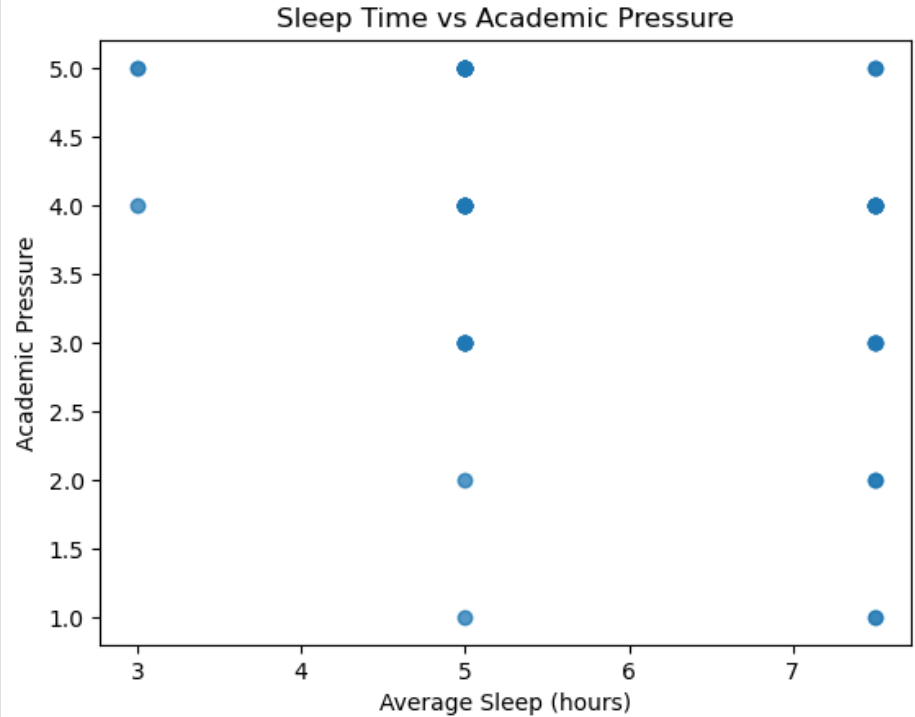
Plot the results

```
df_grouped.plot(x='campus_discrimination', kind='bar')  
plt.title('Mental Health Outcomes Based on Campus Discrimination')  
plt.xlabel('Campus Discrimination')  
plt.ylabel('Mean Scores')  
plt.show()
```



Sleep time vs Academic Pressure

```
plt.scatter(df['average_sleep'],  
df['academic_pressure'],  
alpha=0.5)  
plt.title('Sleep Time vs Academic  
Pressure')  
plt.xlabel('Average Sleep  
(hours)')  
plt.ylabel('Academic Pressure')  
plt.show()
```





Assuming df is your DataFrame

Check if the column 'stress_relief_activities' exists

```
if 'stress_relief_activities' in df.columns:
```

```
# Get value counts and check if it's not empty
```

```
activity_counts = df['stress_relief_activities'].value_counts()
```

```
if not activity_counts.empty:
```

Most common stress relief activities

```
activity_counts.plot(kind='bar')
```

```
plt.title('Most Common Stress Relief Activities')
```

```
plt.xlabel('Activity')
```

```
plt.ylabel('Count')
```

```
plt.show()
```

```
else:
```

```
print("The 'stress_relief_activities' column is empty.")
```

else:

```
print("The 'stress_relief_activities' column is missing.")
```

Check if the required columns exist for boxplot

```
if 'stress_relief_activities' in df.columns and 'depression' in df.columns:
```

```
# Create the boxplot if data is available
```

```
if not df['stress_relief_activities'].dropna().empty and not df['depression'].dropna().empty:
```

```
sns.boxplot(x='stress_relief_activities', y='depression', data=df)
```

```
plt.title('Effectiveness of Stress Relief Activities on Depression')
```

```
plt.xlabel('Stress Relief Activity')
```

```
plt.ylabel('Depression Levels')
```

```
plt.xticks(rotation=45)
```

```
plt.show()
```

```
else:
```

```
print("The required columns have no valid data for the boxplot.")
```

else:

```
print("The required columns for the boxplot are missing.")
```



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

Ensure the relevant columns are numeric, coercing errors to NaN

```
df['depression'] = pd.to_numeric(df['depression'], errors='coerce')
df['anxiety'] = pd.to_numeric(df['anxiety'], errors='coerce')
df['isolation'] = pd.to_numeric(df['isolation'], errors='coerce')
```

Drop rows with NaN values in the relevant columns

```
df_cleaned = df.dropna(subset=['depression', 'anxiety', 'isolation',
                                'social_relationships'])
```

Group by social relationships and check mental health outcomes

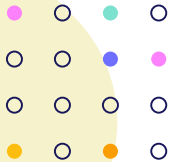
```
df_grouped =
df_cleaned.groupby('social_relationships').mean()[['depression',
                                                    'anxiety', 'isolation']].reset_index()
```

Check if the grouped DataFrame is not empty before plotting

```
if not df_grouped.empty:
    df_grouped.plot(x='social_relationships', kind='bar')
    plt.title('Social Relationships vs Mental Health Factors')
    plt.xlabel('Social Relationships Rating')
    plt.ylabel('Average Levels of Depression, Anxiety, and Isolation')
    plt.show()
else:
    print("No valid data to plot.")
```

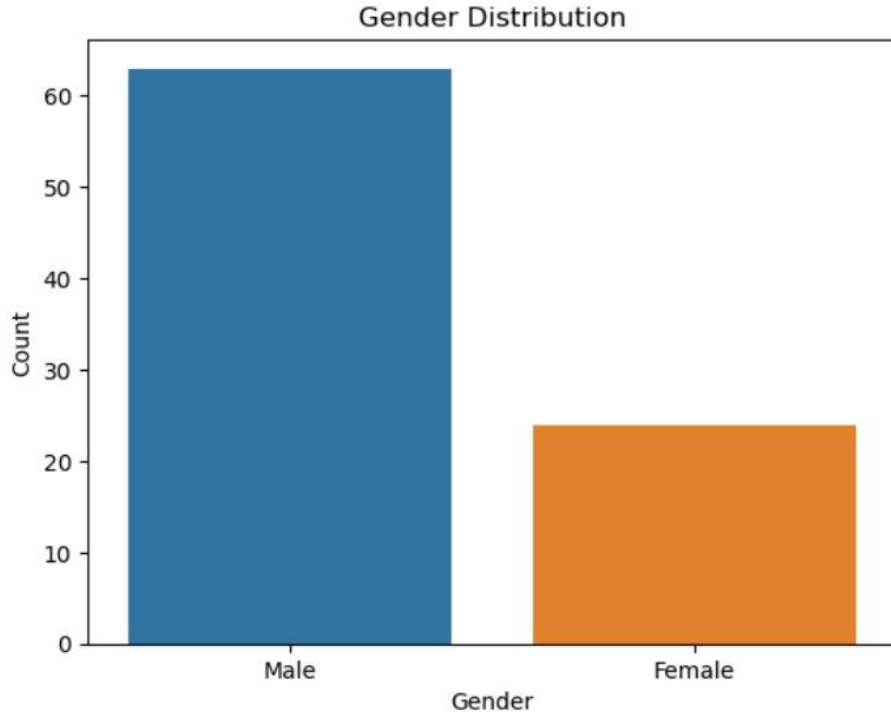
04.

OBJECTIVES ANSWERS

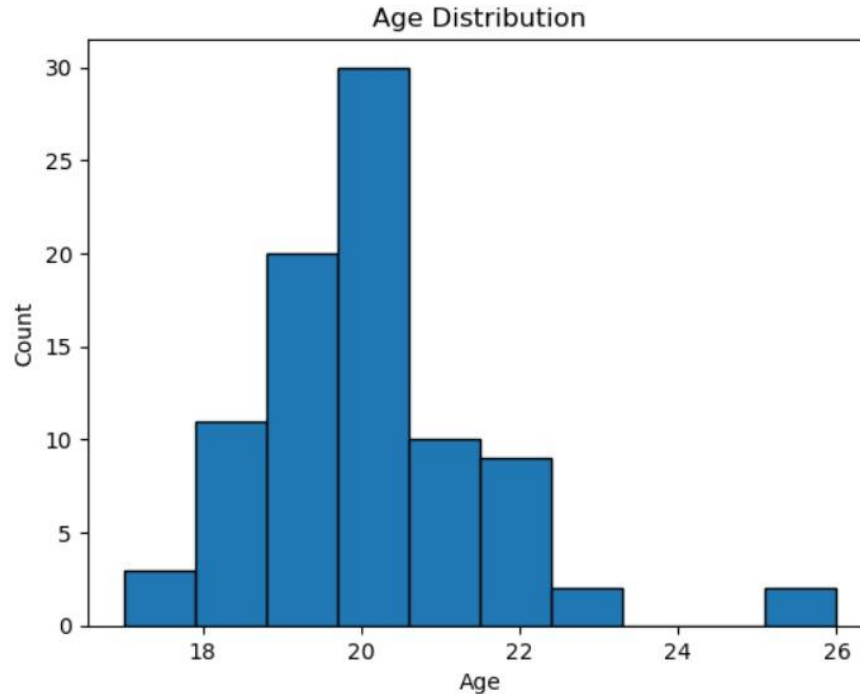
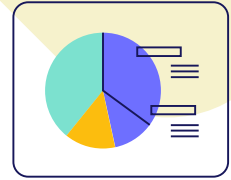


Gender—are males more than females in IT ?

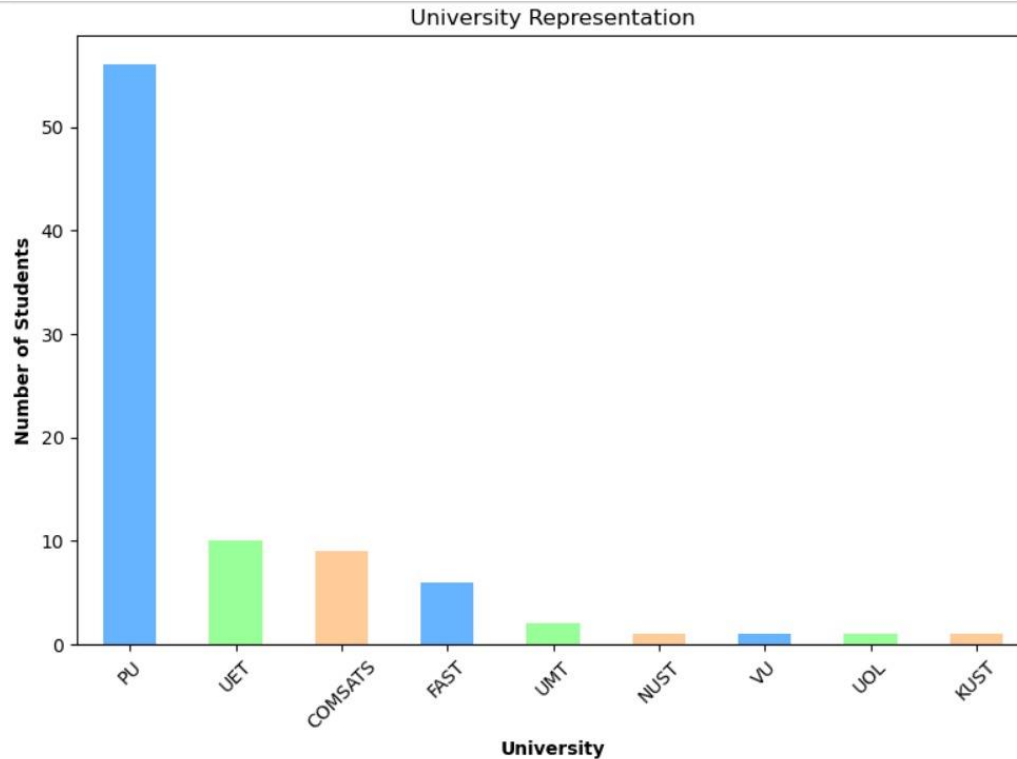
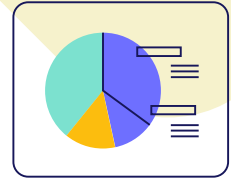
**All students
are IT**



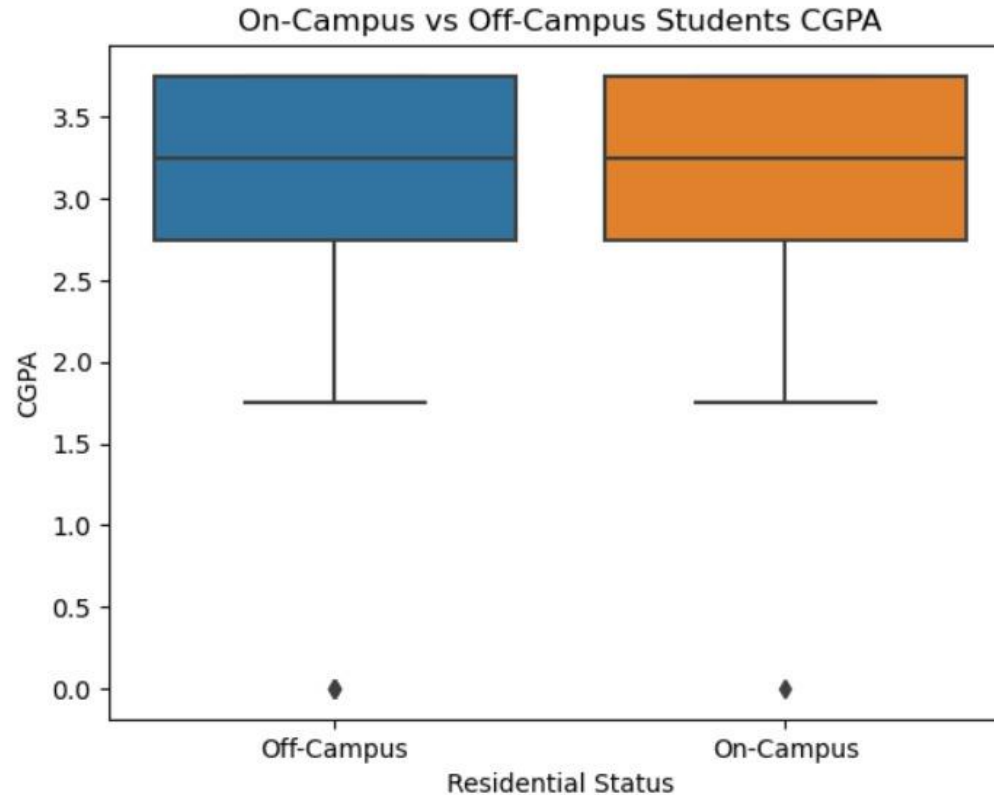
Age – younger or older are in this field ?



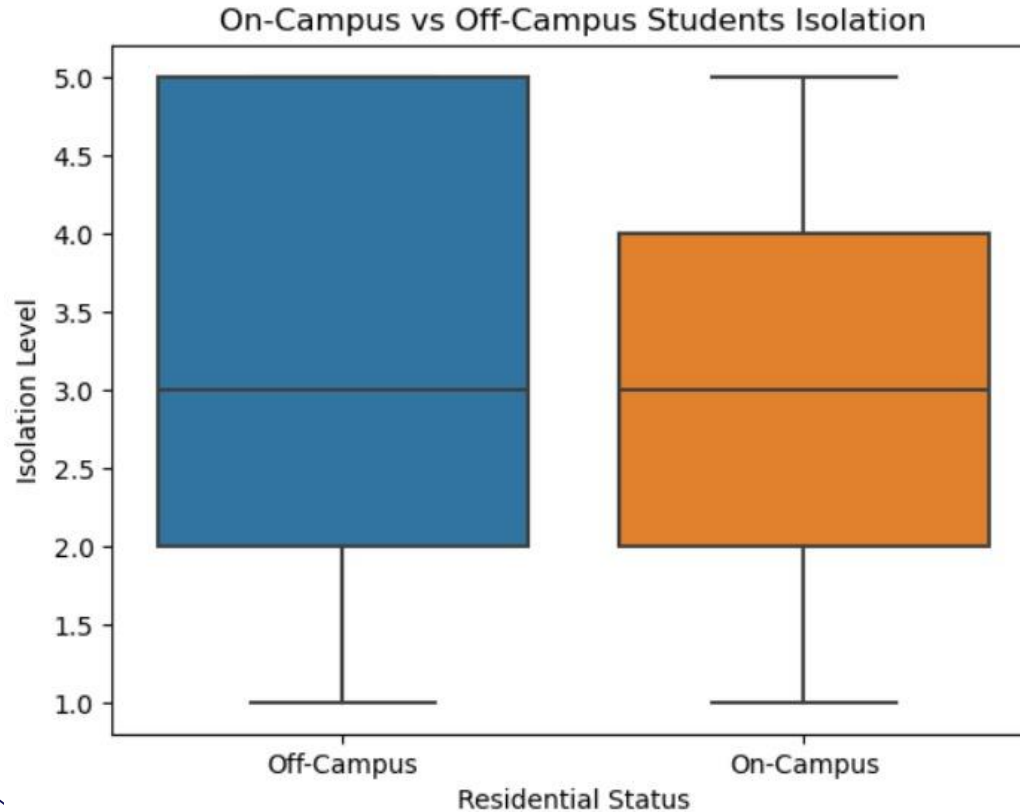
University—which university has more students?



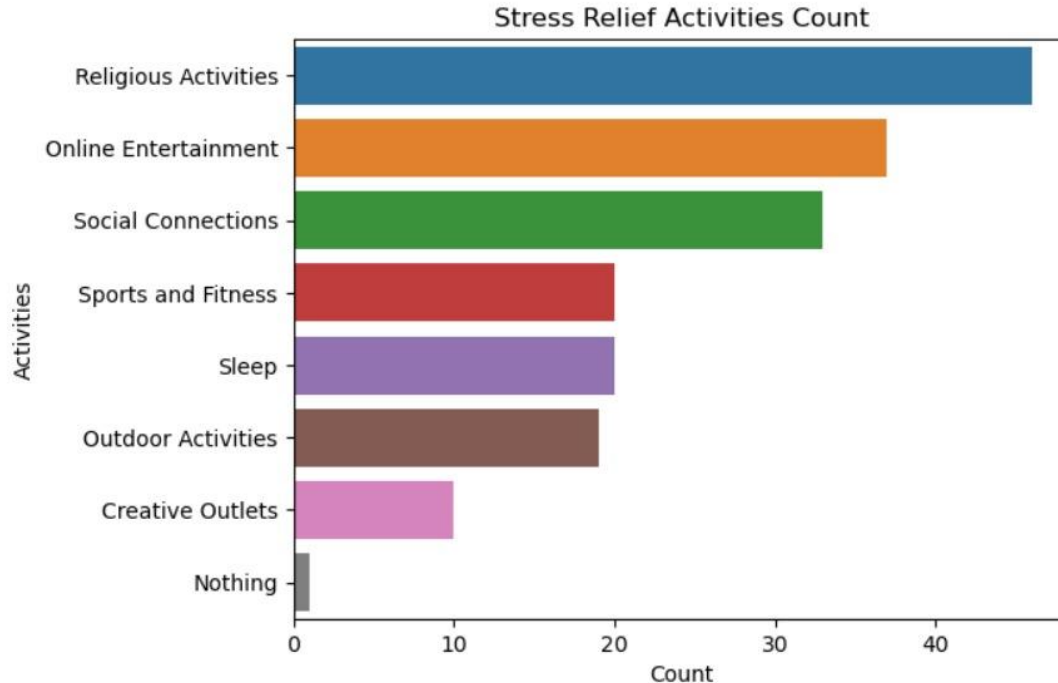
Are on-campus students getting more grades?



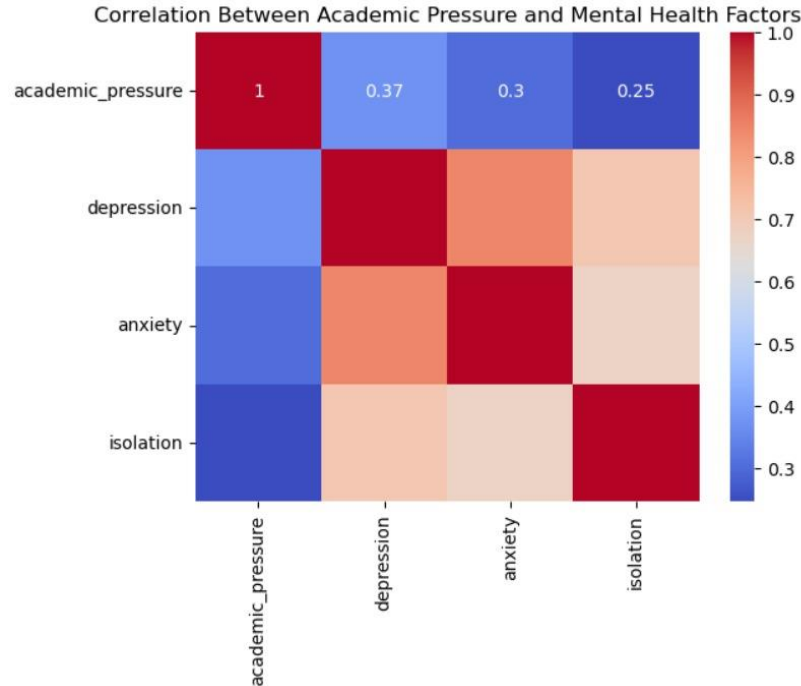
Are on-campus students feeling more isolation?



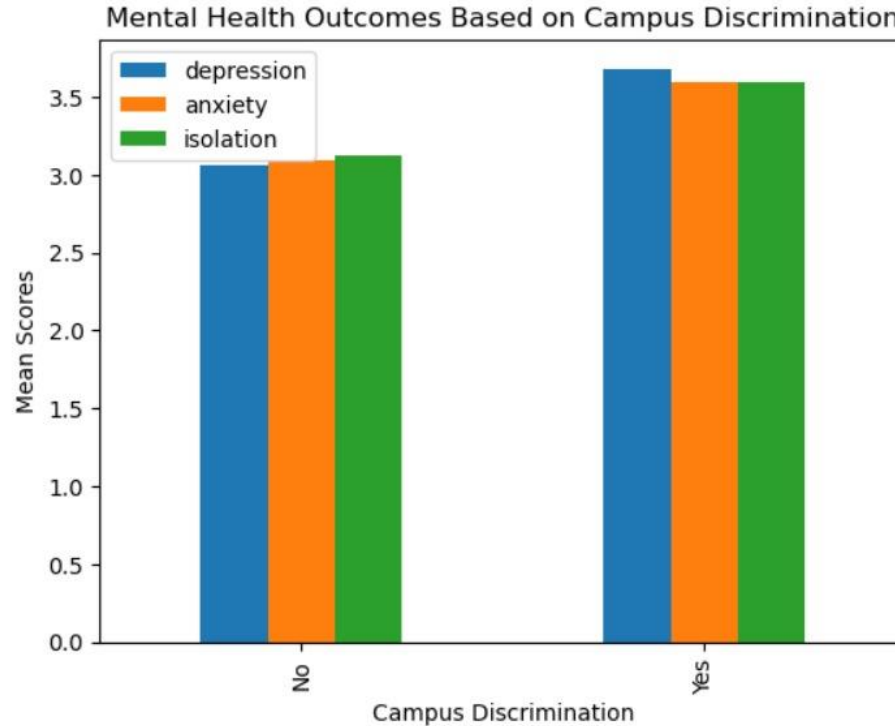
Is there specific university with most depressed students?



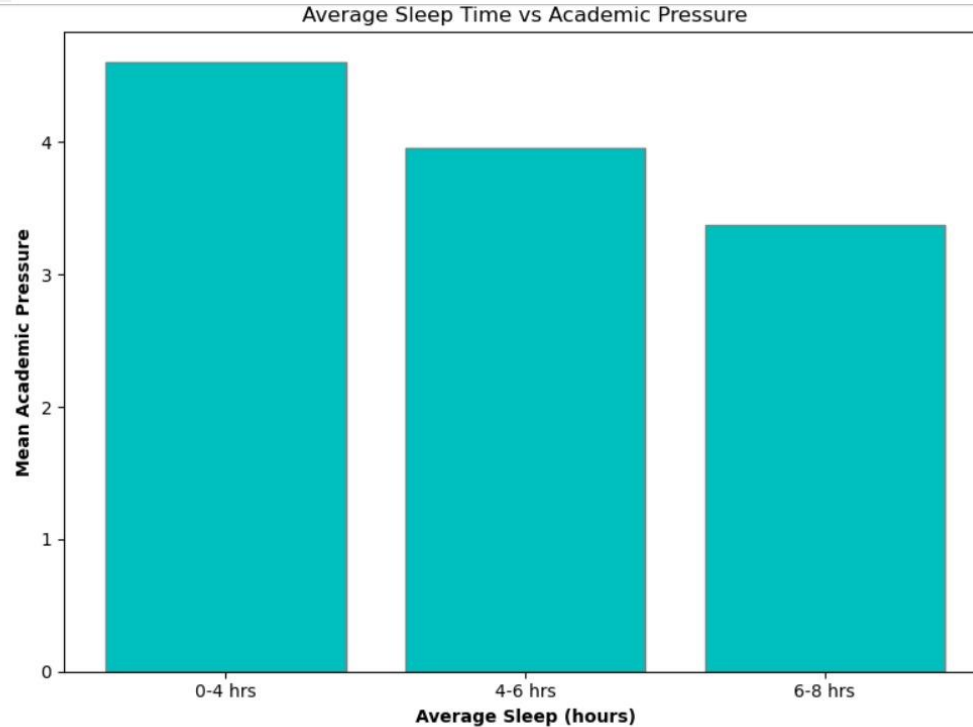
Is there a Correlation between academic pressure and levels of depression, anxiety, and isolation.



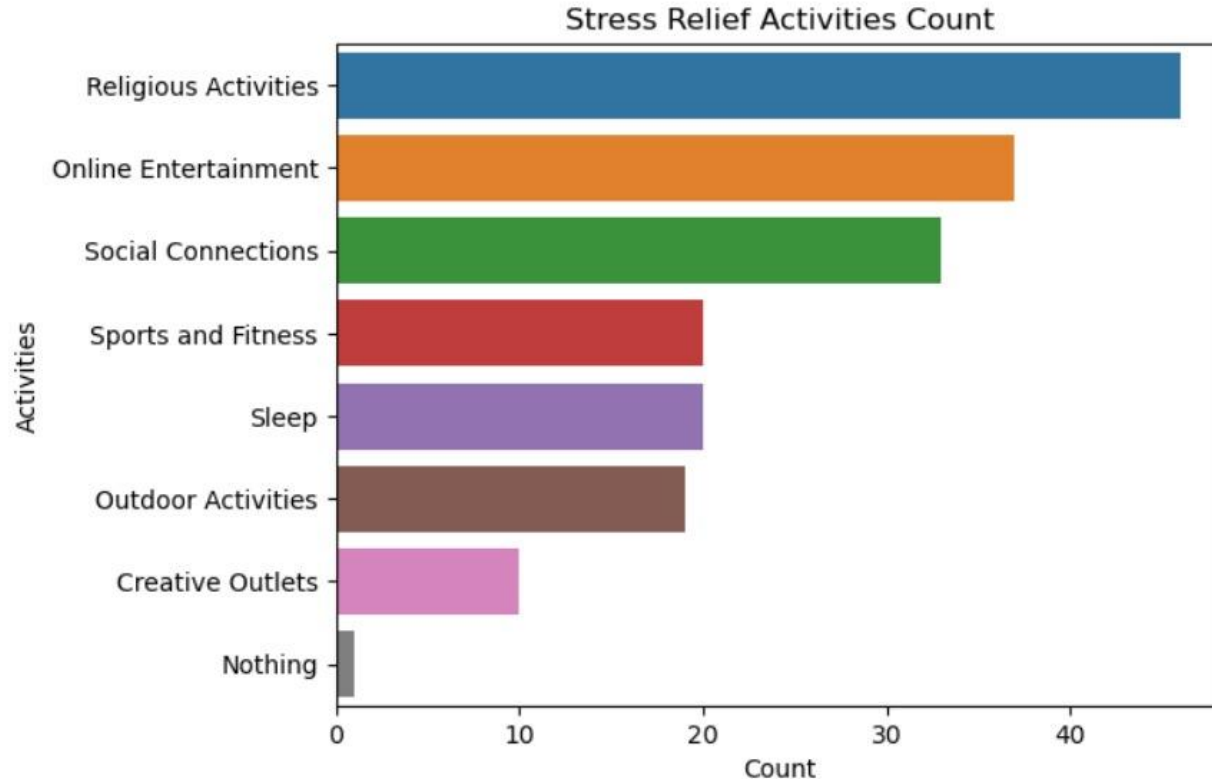
Analyze how campus discrimination affects mental health outcomes.



Is there a significant relationship between sleep time and academic pressure?



What is the most common stress relief activities and their effectiveness on reducing mental health issues.



" Stronger social relationships might be associated with lower levels of depression, anxiety, and isolation"?

