Mini-projet : Analyse de la dépression chez les étudiants à l'aide des bibliothèques Pandas, NumPy et Matplotlib

Réalisé par : MOHAMED AIT MOUALI
Licence d'Excelence : IOTR

1-📝 Introduction:

Objectif du projet

Ce projet a pour objectif de réaliser une analyse exploratoire du jeu de données intitulé **"Student Depression Dataset"**, qui regroupe des informations sur les étudiants et leur état de santé mentale.
L'analyse porte principalement sur la recherche de tendances ou de facteurs pouvant être associés à la dépression chez les étudiants. Source data : https://www.kaggle.com/datasets/hopesb/student-depression-dataset/code

Bibliothèques utilisées

Pour mener à bien cette étude, nous avons utilisé des bibliothèques Python puissantes et couramment employées en science des données :

- **Pandas** : pour la manipulation et l'analyse des données tabulaires
- **NumPy** : pour les opérations numériques et les calculs statistiques
- **Matplotlib** : pour la visualisation des données à travers des graphiques et des diagrammes

Étapes de l'étude

L'étude s'est articulée autour de plusieurs étapes :

1. L'importation et l'exploration du jeu de données
2. Le nettoyage et la manipulation des données
3. L'analyse statistique descriptive
4. La représentation graphique des résultats

Résultats

Cette approche nous a permis de dégager des observations utiles sur la répartition de la dépression chez les étudiants selon divers critères tels que l'âge, le genre, ou encore les habitudes de vie.

2- 🧩Démarche du mini-projet:

Dans ce mini-projet, nous avons suivi plusieurs étapes clés pour analyser les données du **Student Depression Dataset** :

1. Importation des données

Nous avons chargé le fichier CSV dans un DataFrame à l'aide de la bibliothèque **Pandas** afin de pouvoir manipuler facilement les données.

2. Exploration des données

Nous avons examiné les premières lignes du jeu de données, vérifié les types de variables, identifié les valeurs manquantes et consulté les statistiques descriptives de base.

3. Manipulation des données

Nous avons sélectionné des colonnes spécifiques, filtré les données selon certains critères (par exemple, les étudiants souffrant de dépression), créé de nouvelles colonnes (comme les tranches d'âge), et trié les données pour mieux les analyser.

4. Analyse statistique

Nous avons calculé des mesures telles que la moyenne, la médiane et l'écart-type, et nous avons regroupé les données par genre ou tranche d'âge afin de mieux comprendre la distribution des cas de dépression.

5. Visualisation des données

Enfin, nous avons utilisé la bibliothèque **Matplotlib** pour représenter graphiquement les résultats à l'aide d'histogrammes, de courbes et de boxplots, facilitant ainsi l'interprétation des tendances observées.

3- Code python:

✅ 1. Importation et chargement:

In [5]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Charger le fichier CSV
df = pd.read_csv('Student Depression Dataset.csv')

# Afficher les 5 premières lignes
print(df.head())
```

```
   id  Gender   Age           City Profession  Academic Pressure  \
0   2    Male  33.0  Visakhapatnam    Student                5.0
1   8  Female  24.0      Bangalore    Student                2.0
2  26    Male  31.0       Srinagar    Student                3.0
3  30  Female  28.0       Varanasi    Student                3.0
4  32  Female  25.0         Jaipur    Student                4.0

   Work Pressure  CGPA  Study Satisfaction  Job Satisfaction  \
0            0.0  8.97                 2.0               0.0
1            0.0  5.90                 5.0               0.0
2            0.0  7.03                 5.0               0.0
3            0.0  5.59                 2.0               0.0
4            0.0  8.13                 3.0               0.0

      Sleep Duration Dietary Habits   Degree  \
0          5-6 hours        Healthy  B.Pharm
1          5-6 hours       Moderate      BSc
2  Less than 5 hours        Healthy       BA
3          7-8 hours       Moderate      BCA
4          5-6 hours       Moderate   M.Tech

  Have you ever had suicidal thoughts ?  Work/Study Hours  Financial Stress  \
0                                   Yes               3.0               1.0
1                                    No               3.0               2.0
2                                    No               9.0               1.0
3                                   Yes               4.0               5.0
4                                   Yes               1.0               1.0

  Family History of Mental Illness  Depression
0                               No           1
1                              Yes           0
2                              Yes           0
3                              Yes           1
4                               No           0
```

🔍 2. Exploration des données:

In [6]:
```
# Infos générales sur le DataFrame
print(df.info())

# Statistiques de base
print(df.describe())

# Valeurs manquantes
print(df.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2318 entries, 0 to 2317
Data columns (total 18 columns):
 #   Column                                 Non-Null Count  Dtype
---  ------                                 --------------  -----
 0   id                                     2318 non-null   int64
 1   Gender                                 2318 non-null   object
 2   Age                                    2318 non-null   float64
 3   City                                   2318 non-null   object
 4   Profession                             2318 non-null   object
 5   Academic Pressure                      2318 non-null   float64
 6   Work Pressure                          2318 non-null   float64
 7   CGPA                                   2318 non-null   float64
 8   Study Satisfaction                     2318 non-null   float64
 9   Job Satisfaction                       2318 non-null   float64
 10  Sleep Duration                         2318 non-null   object
 11  Dietary Habits                         2318 non-null   object
 12  Degree                                 2318 non-null   object
 13  Have you ever had suicidal thoughts ?  2318 non-null   object
 14  Work/Study Hours                       2318 non-null   float64
 15  Financial Stress                       2318 non-null   float64
 16  Family History of Mental Illness       2318 non-null   object
 17  Depression                             2318 non-null   int64
dtypes: float64(8), int64(2), object(8)
memory usage: 326.1+ KB
None
                 id          Age  Academic Pressure  Work Pressure  \
count   2318.000000  2318.000000        2318.000000         2318.0
mean    5955.878775    25.754530           3.159620            0.0
std     3431.914932     4.884392           1.387244            0.0
min        2.000000    18.000000           1.000000            0.0
25%     3018.000000    21.000000           2.000000            0.0
50%     6008.500000    25.000000           3.000000            0.0
75%     8954.750000    30.000000           4.000000            0.0
max    11822.000000    42.000000           5.000000            0.0

              CGPA  Study Satisfaction  Job Satisfaction  Work/Study Hours  \
count  2318.000000         2318.000000       2318.000000       2318.000000
mean      7.686311            2.931838          0.001294          7.195427
std       1.461859            1.361222          0.062311          3.712736
min       5.060000            1.000000          0.000000          0.000000
25%       6.370000            2.000000          0.000000          4.000000
50%       7.800000            3.000000          0.000000          8.000000
75%       8.950000            4.000000          0.000000         10.000000
max      10.000000            5.000000          3.000000         12.000000

       Financial Stress   Depression
count       2318.000000  2318.000000
mean           3.133305     0.590595
std            1.454208     0.491830
min            1.000000     0.000000
25%            2.000000     0.000000
50%            3.000000     1.000000
75%            4.000000     1.000000
max            5.000000     1.000000
id                                       0
Gender                                   0
Age                                      0
City                                     0
Profession                               0
Academic Pressure                        0
Work Pressure                            0
CGPA                                     0
Study Satisfaction                       0
Job Satisfaction                         0
Sleep Duration                           0
Dietary Habits                           0
Degree                                   0
Have you ever had suicidal thoughts ?    0
Work/Study Hours                         0
Financial Stress                         0
Family History of Mental Illness         0
Depression                               0
dtype: int64
```

🛠️ 3. Manipulation des données:

In [9]:
```
# Sélection de colonnes spécifiques
print(df[['Age', 'Gender', 'Depression']])

# Filtrer : les étudiants dépressifs
depressed_students = df[df['Depression'] == 'Yes']
print(depressed_students)

# Créer une nouvelle colonne : "Age_Group"
df['Age_Group'] = pd.cut(df['Age'], bins=[15, 18, 22, 30], labels=['Teen', 'Young Adult', 'Adult'])
print(df[['Age', 'Age_Group']])

# Trier les données par Age décroissant
print(df.sort_values(by='Age', ascending=False))
```

```
       Age  Gender  Depression
0     33.0    Male           1
1     24.0  Female           0
2     31.0    Male           0
3     28.0  Female           1
4     25.0  Female           0
...    ...     ...         ...
2313  20.0    Male           0
2314  29.0    Male           0
2315  33.0  Female           0
2316  29.0  Female           0
2317  21.0    Male           1

[2318 rows x 3 columns]
Empty DataFrame
Columns: [id, Gender, Age, City, Profession, Academic Pressure, Work Pressure, CGPA, Study Satisfaction, Job Satisfaction, Sleep Duration, Dietary Habits, Degree, Have you ever had suicidal thoughts ?, Work/Study Hours, Financial Stres
s, Family History of Mental Illness, Depression]
Index: []
       Age    Age_Group
0     33.0          NaN
1     24.0        Adult
2     31.0          NaN
3     28.0        Adult
4     25.0        Adult
...    ...          ...
2313  20.0  Young Adult
2314  29.0        Adult
2315  33.0          NaN
2316  29.0        Adult
2317  21.0  Young Adult

[2318 rows x 2 columns]
         id  Gender   Age         City Profession  Academic Pressure  \
1076   5569  Female  42.0       Rajkot    Student                2.0
201     978  Female  39.0       Kalyan    Student                5.0
1330   6852    Male  36.0       Mumbai    Student                5.0
307    1493  Female  35.0       Meerut    Student                5.0
374    1864  Female  34.0     Varanasi    Student                5.0
...     ...     ...   ...          ...        ...                ...
860    4417    Male  18.0      Kolkata    Student                5.0
832    4284  Female  18.0       Bhopal    Student                3.0
835    4309    Male  18.0         Agra    Student                3.0
1394   7157  Female  18.0     Ludhiana    Student                2.0
2257  11519    Male  18.0  Vasai-Virar    Student                1.0

      Work Pressure  CGPA  Study Satisfaction  Job Satisfaction  \
1076            0.0  9.03                 5.0               0.0
201             0.0  6.51                 4.0               0.0
1330            0.0  7.10                 5.0               0.0
307             0.0  5.32                 2.0               0.0
374             0.0  5.72                 2.0               0.0
...             ...   ...                 ...               ...
860             0.0  6.37                 3.0               0.0
832             0.0  9.94                 4.0               0.0
835             0.0  6.83                 4.0               0.0
1394            0.0  5.56                 4.0               0.0
2257            0.0  7.25                 5.0               0.0

         Sleep Duration Dietary Habits    Degree  \
1076  More than 8 hours       Moderate  Class 12
201           7-8 hours       Moderate    M.Tech
1330  Less than 5 hours       Moderate       MSc
307           7-8 hours       Moderate       PhD
374           7-8 hours        Healthy     M.Com
...                 ...            ...       ...
860   Less than 5 hours       Moderate  Class 12
832           5-6 hours        Healthy  Class 12
835           7-8 hours      Unhealthy  Class 12
1394          7-8 hours       Moderate  Class 12
2257  Less than 5 hours       Moderate  Class 12

     Have you ever had suicidal thoughts ?  Work/Study Hours  \
1076                                   Yes               3.0
201                                    Yes               7.0
1330                                   Yes               0.0
307                                     No               3.0
374                                    Yes               9.0
...                                    ...               ...
860                                    Yes               6.0
832                                    Yes               9.0
835                                    Yes              10.0
1394                                    No              12.0
2257                                    No              12.0

      Financial Stress Family History of Mental Illness  Depression Age_Group
1076               1.0                              Yes           0       NaN
201                2.0                               No           1       NaN
1330               2.0                              Yes           0       NaN
307                4.0                              Yes           0       NaN
374                1.0                               No           0       NaN
...                ...                              ...         ...       ...
860                5.0                              Yes           1      Teen
832                4.0                               No           1      Teen
835                4.0                               No           1      Teen
1394               5.0                               No           0      Teen
2257               2.0                               No           0      Teen

[2318 rows x 19 columns]
```

📊 4. Analyse statistique:

In [14]:
```
# Moyenne, médiane, écart-type
print("Moyenne d'âge :", df['Age'].mean())
print("Médiane d'âge :", df['Age'].median())
print("Écart-type d'âge :", df['Age'].std())

# Grouper par Genre et calculer la moyenne d'âge
print(df.groupby('Gender')['Age'].mean())

# Grouper par Age_Group et compter les cas de dépression
print(df.groupby('Age_Group')['Depression'].value_counts())
```

```
Moyenne d'âge : 25.75452976704055
Médiane d'âge : 25.0
Écart-type d'âge : 4.884391894260112
Gender
Female    25.524085
Male      25.941406
Name: Age, dtype: float64
Age_Group    Depression
Teen         1              92
             0              34
Young Adult  1             389
             0             189
Adult        1             685
             0             427
Name: count, dtype: int64
C:\Users\DELL\AppData\Local\Temp\ipykernel_8416\856318151.py:10: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or ob
served=True to adopt the future default and silence this warning.
  print(df.groupby('Age_Group')['Depression'].value_counts())
```

📈 5. Visualisation des données:

In [15]:
```
# Histogramme de l'âge
df['Age'].hist(bins=10)
plt.title("Répartition de l'âge")
plt.xlabel("Âge")
plt.ylabel("Nombre d'étudiants")
plt.show()

# Courbe : nombre de cas de dépression par âge
df.groupby('Age')['Depression'].apply(lambda x: (x == 'Yes').sum()).plot()
plt.title("Cas de dépression par âge")
plt.xlabel("Âge")
plt.ylabel("Nombre de cas")
plt.show()

# Boxplot : âge selon la dépression
df.boxplot(column='Age', by='Depression')
plt.title("Âge par statut de dépression")
plt.suptitle("")
plt.xlabel("Dépression")
plt.ylabel("Âge")
plt.show()
```