Conclusions

Conclusions par rapport à la problématique initiale

Comparaison par rapport les données en ligne que disséminent Tunisie.

1. chargement et traitement des données

```
# Charger le dataset
df = pd.read_csv('Student_Lifestyle.csv')
print("Dimensions :", df.shape)
df.head()
```

```
Dimensions : (2000, 7)
```

	Gender	Age	Study Hours	Sleep Hours	Activity Level	Diet Quality	Exam Score
0	Male	20	6.9	7.4	Medium	Poor	54.1
1	Female	23	5.3	7.3	High	Middle Class	49.2
2	Male	21	8.0	6.8	Medium	Middle Class	71.8
3	Male	19	6.1	7.5	Medium	Middle Class	62.6
4	Female	22	7.4	7.1	Medium	Middle Class	75.2

```
# Vérifier les types de données
print("Types de données :")
df.dtypes
```

```
Types de données :
Gender            object
Age                int64
Study Hours      float64
Sleep Hours      float64
Activity Level    object
Diet Quality      object
Exam Score       float64
dtype: object
```

```
# Vérifier les valeurs manquantes
print("Valeurs manquantes :")
df.isna().sum()
```

```
Valeurs manquantes :
Gender            0
Age               0
Study Hours       0
Sleep Hours       0
Activity Level    0
Diet Quality      0
Exam Score        0
dtype: int64
```

```
# Statistiques descriptives
print("Statistiques descriptives :")
df.describe()
```

```
Statistiques descriptives :
       Age  Study Hours  Sleep Hours  Exam Score
count  ...         ...          ...         ...
mean   ...         ...          ...         ...
std    ...         ...          ...         ...
min    ...         ...          ...         ...
25%    ...         ...          ...         ...
50%    ...         ...          ...         ...
75%    ...         ...          ...         ...
max    ...         ...          ...         ...
```

2. Exploration des données

Maintenant que les données sont nettoyées et numériques, nous pouvons commencer l'exploration. Nous allons d'abord examiner la distribution de la variable cible (Exam Score), puis explorer les relations entre les différentes variables et la réussite aux examens.

```
# Distribution de la variable cible
plt.figure(figure_size=(10,6))
for col in df.columns:
    if df[col].dtype == 'object':
        print(f"{col} :")
        print(df[col].value_counts())
        print(f"Moyenne : {df[col].mean()}")
        print(f"Médiane : {df[col].median()}")
        print(f"Ecart-type : {df[col].std()}")
        print(f"Min : {df[col].min()}")
        print(f"Max : {df[col].max()}")
        print(f"Valeurs manquantes : {df[col].isna().sum()}")
        print(f"Nombre d'individus : {df[col].count()}")
        print(f"Nombre d'individus différents : {df[col].nunique()}")
        print(f"Mode : {df[col].mode()[0]}")
        print(f"Fréquence du mode : {df[col].value_counts()[df[col].mode()[0]]}")
        print(f"Pourcentage du mode : {df[col].value_counts()[df[col].mode()[0]]/df[col].count()*100}%")
        print(f"Valeurs uniques : {df[col].unique()}")
        print(f"Nombre de valeurs uniques : {df[col].nunique
```