# DATA ANALYST TP1

RAGUIG Mohammed Amine

```python
import numpy as n
#1
week1_sleep=[6,7,8,5,9]
week2_sleep=[7,6,8,7,6]
#2
days_vector=['Monday','Tuesday','Wednesday','Thursday','Friday']
```

```python
#3
daily_difference=[]
for i, j in zip(week1_sleep, week2_sleep):
    daily_difference.append(i - j)
print(daily_difference)
```

```
C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\Pycha
[-1, 1, 0, -2, 3]

Process finished with exit code 0
```

```python
#4

total_week1=sum(week1_sleep)

total_week2=sum(week2_sleep)

average_week1=n.average(week1_sleep)
average_week2=n.average(week2_sleep)
print(average_week1)
print(average_week2)
```

```
C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\
7.0
6.8

Process finished with exit code 0
```

```python
#5
slept_more_first_week= True if (average_week1 > average_week2) else False

print(slept_more_first_week)

```

```
C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\PycharmProjects\DataAn
True

Process finished with exit code 0
```

```python
#6

for i,j in zip(week1_sleep,days_vector):
    if j=='Wednesday':
        sleep_wednesday_week1=i
print(sleep_wednesday_week1)
sleep_wednesday_week2=[]

for i,j in zip(week2_sleep,days_vector):
    if j=='Tuesday' or j=='Wednesday' or j=='Thursday':
        sleep_wednesday_week2.append(i)
print(sleep_wednesday_week2)

```

```
C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\PycharmP
8
[6, 8, 7]

Process finished with exit code 0
```

```python
#7
sleep_enough_week1=[]
for i,j in zip(week1_sleep,days_vector):
    if i>=8:
        sleep_enough_week1.append(j)
print(sleep_enough_week1)

```

```
C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\PycharmProjects\DataAnalyste\Ex1.py
['Wednesday', 'Friday']

Process finished with exit code 0
```

```python
#8
sleep_successful_days_week1=[]
for i,j in zip(week1_sleep,days_vector):
    if i>=8:
        sleep_successful_days_week1.append(i)
print(sleep_successful_days_week1)

```

```
C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\PycharmProjects\DataAnalyste\Ex1.py
[8, 9]

Process finished with exit code 0
```

```python
#1
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

```

```python
#2
df=pd.read_csv('StudentsPerformance.csv')
print(df.head())

```

```
Ex2

C:\Users\pc\PycharmProjects\DataAnalyste\.venv\Scripts\python.exe C:\Users\pc\PycharmProjects\DataAnalyste\Ex2.py
   gender race/ethnicity  ... reading score writing score
0  female        group B  ...            72            74
1  female        group C  ...            90            88
2  female        group B  ...            95            93
3    male        group A  ...            57            44
4    male        group C  ...            78            75

[5 rows x 8 columns]

Process finished with exit code 0
```

- Categorical: gender and race
- Numerical: reading score and writing score

```python
#3
print("--------------------3----------------------")
print(df.info())
print("++++++++++Describe++++++++++++++++++")
print(df.describe())

```

```
Ex2

--------------------3----------------------
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   gender                       1000 non-null   object
 1   race/ethnicity               1000 non-null   object
 2   parental level of education  1000 non-null   object
 3   lunch                        1000 non-null   object
 4   test preparation course      1000 non-null   object
 5   math score                   1000 non-null   int64
```

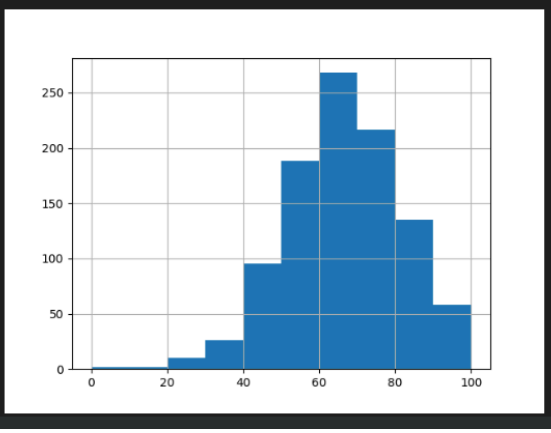
```
print("\n-----------------4-------------------------------\n")
print(df.sample(10))
```

Ex2

```
-----------------4-------------------------------

     gender race/ethnicity  ... reading score writing score
125  female        group B  ...            95            86
137    male        group E  ...            55            56
635    male        group A  ...            74            64
57     male        group D  ...            55            49
610    male        group D  ...            60            63
701  female        group B  ...            67            72
853    male        group E  ...            67            61
325  female        group C  ...            90            94
915  female        group E  ...            70            66
210    male        group D  ...            79            79

[10 rows x 8 columns]
```

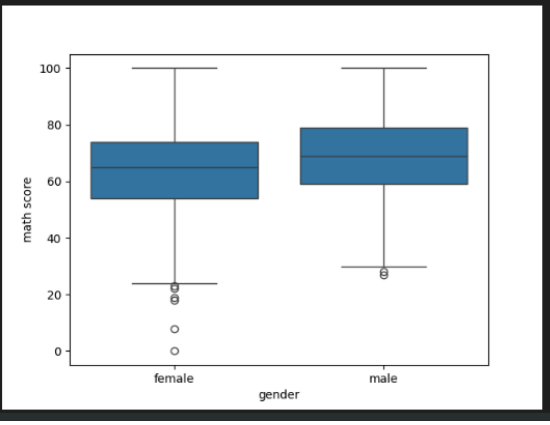
```
#1
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#2
df=pd.read_csv('StudentsPerformance.csv')



print("\n----------------5-------------------------")
df['math score'].hist(bins=10)
plt.show()
```

640x480 PNG (32-bit color) 15,8 kB

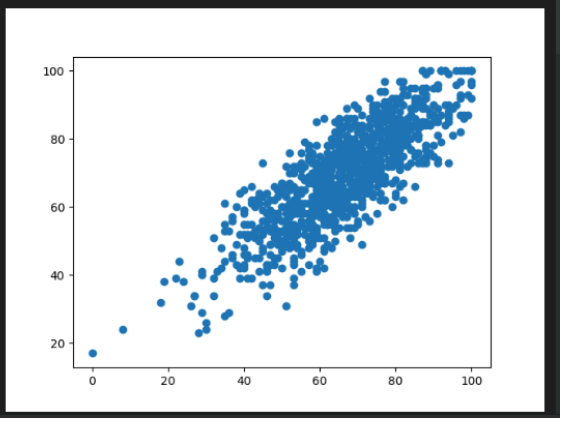



```
#1
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#2
df=pd.read_csv('StudentsPerformance.csv')



print("\n----------------6-------------------------")
sns.boxplot(x='gender', y='math score', data=df)
plt.show()
```

640x480 PNG (32-bit color) 13,3 kB



```
import matplotlib.pyplot as plt
import seaborn as sns
#2
df=pd.read_csv('StudentsPerformance.csv')



print("\n----------------7------------------------------\n")
plt.scatter(df['math score'], df['reading score'])
plt.show()
```




```
import matplotlib.pyplot as plt
import seaborn as sns
#2
df=pd.read_csv('StudentsPerformance.csv')



print("\n----------------8------------------------------\n")
correlation_mx = df[['math score', 'reading score', 'writing score']].corr()
sns.heatmap(correlation_mx)
plt.show()
```