University of Kasdi Merbah Ouargla
Faculty of New Information and Communication Technologies
Department of Computer Science and Information Technology
Master 1 Artificial intelligence and data science

Programming for Data Scientists

Solution of `TP Exam`

Student ……..……

Exercise 1

Suppose that we have the following randomly generated array A.

```
import numpy as np
A = np.random.randint(1,23, size=(5,5))
A
```

```
A = [[ 2 17 21 22  6]
     [ 3  9 19  1 22]
     [15  8 21 19 20]
     [22  2 15 12  5]
     [17  2 16  5 11]]
```

Fill the next table:

Editor	Console
`B = A[2:4, :]` `print(B)`	`[[15  8 21 19 20]` ` [22  2 15 12  5]]`
`C = A[1, np.newaxis]` `print(C)`	`C is:` `[[ 3  9 19  1 22]]`

`print(A[::-1,[0,3]])`	`[[17  5]`
	` [22 12]`
	` [15 19]`
	` [ 3  1]`
	` [ 2 22]]`

`D = np.concatenate((B,C), axis=0)`
`print(D)`
	`D is:`
	`[[15  8 21 19 20]`
	` [22  2 15 12  5]`
	` [ 3  9 19  1 22]]`

`E = np.concatenate([A, D])[::-2]`
`print(E)`
	`[[ 3  9 19  1 22]`
	` [15  8 21 19 20]`
	` [22  2 15 12  5]`
	` [ 3  9 19  1 22]]`

Exercise 2

We have the following Pandas' DataFrame **df_1**:

index		Name	WBC
Type	**Rank**		
A+	**1st**	Billie	9500
	3rd	Jasper	6400
	2nd	Lawson	5200
A-	**3rd**	Declan	7850
	1st	Maison	10055
	2nd	Davis	8050
B+	**2nd**	Saoirse	8110
	3rd	Maddison	6200
	1st	Kylie	9550
B-	**1st**	Rosalie	11000
	2nd	Hamish	11550
	3rd	Ellis	4800

WBC: White Blood Count .

1. Extract the people of the 3rd rank in A+ and B- types. (using loc, 1 single ins)

```
players_df.loc[["A+", "B-"], '3rd', :]
```

Using **df_1**, create a new DataFrame **df_2** composed of only A+ and B+. (using loc, 1 single ins)

```
df_2 = players_df.loc[["A+", "A-"]]
```

2. Split **df_1** into two families A and B, A must contain A+ and A-, B must contain B+ and B-, the new index will be (Family, Type, Rank).

```
players_df.loc[["A+", "A-"], 'Family']= 'A'
players_df.loc[["B+", "B-"], 'Family']= 'B'
players_df = players_df.reset_index()
players_df = players_df.set_index(['Family','Type','Rank'])
```

3. Display the WBC mean of the A and B. (1 single ins)

```
players_df.groupby("Family")['WBC'].mean()
```

4. We want to add a new column to the DataFrame which called **Compact**, this column will contain all the data of each row according to the following formula:
   {Family}_{Type}_{Rank}_{WBC}_{Name}

   Using **groupby("Type").apply(yourFuncName),** add the new column to **df_1.**

```
def add_compact(x):
  ranks = np.array(["_".join(r) for r in x.index.values])
  compact = np.array([v1+"_"+v2+"_"+v3 for v1,v2,v3 in zip(ranks,
x.WBC.values.astype(str), x.Name.values)])
  x['Compact'] = compact
  return x

players_df.groupby("Type").apply(add_compact)
```