

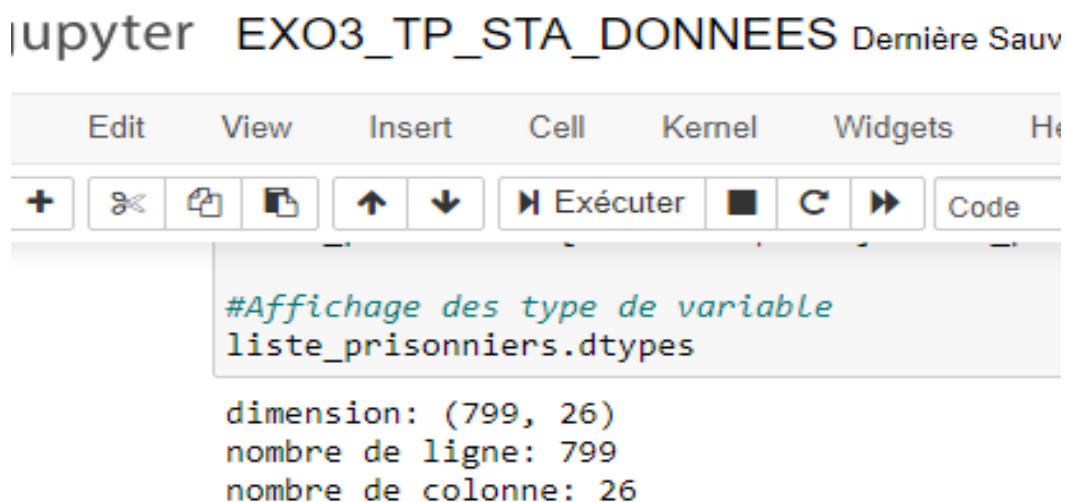
TEAM6 :

MEMBRES DU GROUPE :

CHEIKKH	MBACKE	DIOUF
ROKHAYA		GUEYE
PAPE	MOUSSA	GUEYE

Question 3 : Réponses aux questions

- 1- Enregistrer les données dans un format adapté pour une lecture par la suite avec Python sachant que la première ligne du fichier smp.csv correspond au nom des variables. Vérifier si vous avez une structure de 799 observations et 26 variables.



The screenshot shows a Jupyter Notebook window titled "EX03\_TP\_STA\_DONNEES". The interface includes a menu bar with "Edit", "View", "Insert", "Cell", "Kernel", "Widgets", and "Help". Below the menu is a toolbar with icons for adding, deleting, copying, pasting, and running code. The code cell contains the following text:

```
#Affichage des type de variable  
liste_prisonniers.dtypes
```

The output of the code is displayed below the cell:

```
dimension: (799, 26)  
nombre de ligne: 799  
nombre de colonne: 26
```

2. Changer les types des variables. Vous devez obtenir le résultat suivant :

```
+  🔍  📄  📄  ⬆️  ⬆️  ▶️ Exécuter  ■  ↺  ▶️  Code

nombre de ligne: 799
nombre de colonne: 26

Out[19]: age          float64
prof          category
duree         category
discip        category
n.enfant      float64
n.fratie      int64
ecole         category
separation    category
juge.enfant   category
place         category
abus          category
grav.cons     category
dep.cons      category
ago.cons      category
ptsd.cons     category
alc.cons      category
subst.cons    category
scz.cons      category
char          category
rs            category
ed            category
dr            category
suicide.s     float64
suicide.hr    category
suicide.past  category
dur.interv    float64
dtype: object
```

3.

Calculer la moyenne, la variance, et l'écart type pour chacune des variables suivantes: age, n.enfant, n.fratie, dur.interv. Donner les 3 premiers quantiles pour la variable age.

```
dimension: (799, 26)
nombre de ligne: 799
nombre de colonne: 26
```

	age	n.enfant	n.fratricie	rs	ed	dr \
count	797.000000	773.000000	799.000000	696.000000	692.000000	688.000000
mean	38.899624	1.755498	4.286608	2.057471	1.865607	2.152616
std	13.280978	1.834044	3.441485	0.877948	0.871233	0.829738
min	19.000000	0.000000	0.000000	1.000000	1.000000	1.000000
25%	28.000000	0.000000	2.000000	1.000000	1.000000	1.000000
50%	37.000000	1.000000	3.000000	2.000000	2.000000	2.000000
75%	48.000000	3.000000	6.000000	3.000000	3.000000	3.000000
max	83.000000	13.000000	21.000000	3.000000	3.000000	3.000000

	suicide.s	suicide.hr	suicide.past	dur.interv
count	758.000000	760.000000	785.000000	749.000000
mean	0.794195	0.201316	0.284076	61.891856
std	1.435488	0.401248	0.451261	19.669605
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	48.000000

La variance de toutes les variables :

```
print(liste_prisonniers.age_classe.value_counts('NaN'))
```

50%	0.000000	0.000000	0.000000	60.000000
75%	1.000000	0.000000	1.000000	75.000000
max	5.000000	1.000000	1.000000	120.000000

```
variance: age 176.384384
```

```
duree 0.753804
```

```
discip 0.178417
```

```
n.enfant 3.363716
```

```
n.fratricie 11.843818
```

```
ecole 0.955673
```

```
separation 0.244318
```

```
juge.enfant 0.200558
```

```
place 0.176530
```

```
abus 0.200871
```

```
grav.cons 2.726128
```

```
dep.cons 0.239639
```

```
ago.cons 0.138924
```

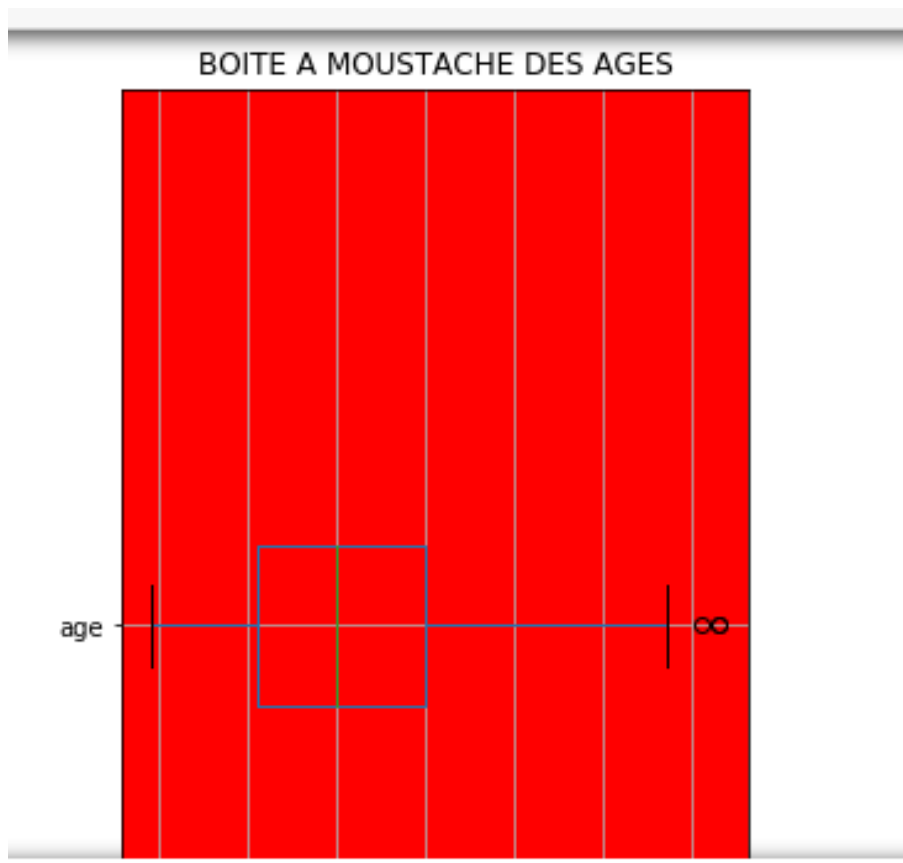
```
ptsd.cons 0.169852
```

```
alc.cons 0.151897
```

```
subst.cons 0.195175
```

```
scz.cons 0.075875
```

4. Tracer le boxplot pour la variable age. Quelles conclusions en tirez-vous?



5. Afficher les données pour les agriculteurs qui ont plus de 2 enfants.

```
dtype: float64
variance: 176.38438428024685
```

	age	prof	duree	discip	n.enfant	n.fratrerie	ecole	separation	\
14	64.0	agriculteur	NaN	0.0	3.0	2	1.0	0.0	
311	42.0	agriculteur	4.0	0.0	3.0	6	1.0	0.0	
390	36.0	agriculteur	4.0	1.0	3.0	4	3.0	1.0	
441	79.0	agriculteur	5.0	0.0	5.0	6	2.0	0.0	

	juge.enfant	place	...	subst.cons	scz.cons	char	rs	ed	dr	suicide.s
14	0.0	0.0	...	0	0	1.0	1.0	1.0	3.0	0.0
311	0.0	0.0	...	0	0	2.0	1.0	3.0	2.0	3.0
390	1.0	1.0	...	1	0	1.0	NaN	3.0	1.0	0.0
441	0.0	0.0	...	0	0	1.0	2.0	1.0	1.0	0.0

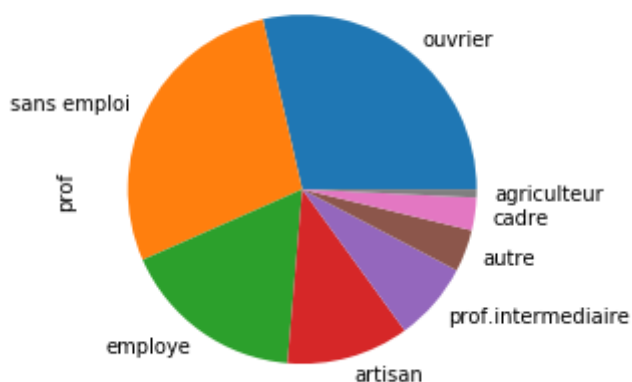
	suicide.hr	suicide.past	dur.interv
14	0.0	0.0	80.0
311	1.0	0.0	NaN
390	0.0	0.0	NaN
441	0.0	0.0	85.0

6. Calculer les fréquences des modalités de la variable prof. Quelle est la catégorie modale?

```
[4 rows x 26 columns]
ouvrier          0.286255
sans emploi      0.279950
employe          0.170240
artisan          0.113493
prof.intermediaire 0.073140
autre            0.039092
cadre            0.030265
agriculteur      0.007566
Name: prof, dtype: float64
count           8.000000
mean            99.125000
std             87.398002
min              6.000000
25%            29.250000
50%            74.000000
75%           156.750000
max           227.000000
Name: prof, dtype: float64
ouvrier          227
```

7. Tracer le diagramme circulaire de la variable profession

```
(40.0, 50.0]      0.238215
Name: age_classe, dtype: float64
```



8. Donner les moyennes des âges par profession :

```

cadre  agriculteur  total
0  3.026482      0.75662  100.0
moyenne des ages des ouvriers: 37.39647577092511
moyenne des ages des sans emploi: 35.8963963963964
moyenne des ages des employes: 38.71111111111111
moyenne des ages des artisans: 45.11111111111114
moyenne des ages des professions intermediaires: 43.258620689655174
moyenne des ages pour autres: 34.935483870967744
moyenne des ages des cadres: 50.083333333333336
moyenne des ages des agriculteurs: 48.833333333333336
ouvrier          227
sans emploi      222
employe          135
artisan          90

```

9. Donner la table des effectifs pour les variables prof incluant les "NaN".

```

ouvrier          227
sans emploi      222
employe          135
artisan          90
prof.intermediaire  58
autre            31
cadre            24
agriculteur       6
Name: prof, dtype: int64
prof 6
age 2
prof 6
duree 223
discip 6
n.enfant 26
ecole 5
separation 11
juge.enfant 5
place 7
abus 7

```

10. Donner le nombre de "Nan" pour chaque variable.

localhost:8890/notebooks/Desktop/Dossier%20CheikhMbacke/TP%20STA%20AVEC%20PY

jupyter EXO3\_TP\_STA\_DONNEES Dernière Sauvegarde : il y a une heure

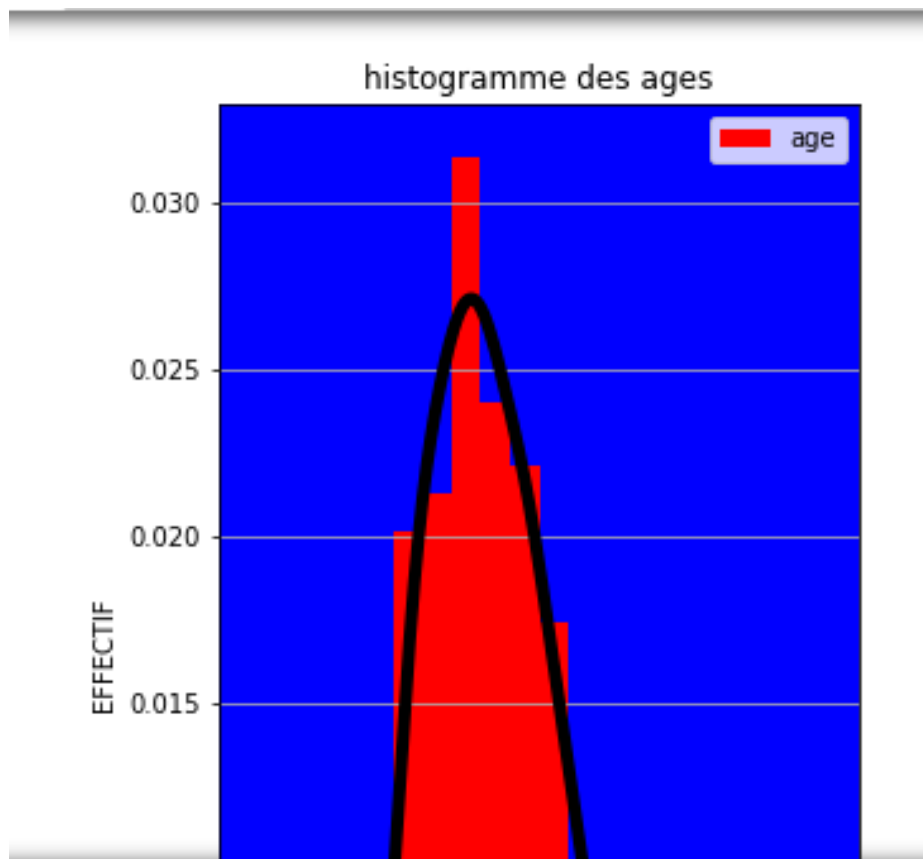
File Edit View Insert Cell Kernel Widgets Help

+ ↩ ↲ ↳ ↵ ⬆ ⬇ ▶ Exécuter ■ ↺ ▶ Code ▼

```
# fréquences des modalités de la nouvelle variable age_classe
print(liste_prisonniers.age_classe.value_counts('NaN'))
```

```
prof 6
age 2
prof 6
duree 223
discip 6
n.enfant 26
ecole 5
separation 11
juge.enfant 5
place 7
abus 7
grav.cons 4
char 96
rs 103
ed 107
dr 111
suicide.s 41
suicide.hr 39
suicide.past 14
dur.interv 50
```

11. Supprimer toutes les lignes contenant des "Nan". Voir script
12. Tracer l'histogramme et la densité de la variable age sur la même figure.



13.

Discretisez la variable age. Pour ce faire, on ajoutera une variable dans le Data Frame des données une nouvelle variable nommée age\_classe. Cette variable aura 4 classes :  $[\min(\text{age}), Q1]$ ,  $]Q1, Q2]$ ,  $]Q2, Q3]$ ,  $]Q3, \max(\text{age})]$ . ou  $Q1$ ,  $Q2$ ,  $Q3$  sont respectivement les 3 premiers quantiles de la variable age,  $\min(\text{age})$  et  $\max(\text{age})$  respectivement la plus petite et la plus grande valeur de la variable age.

	age	prof	duree	discip	n.enfant	n.fratrerie	ecole	separation	\
14	64.0	agriculteur	NaN	0.0	3.0	2	1.0	0.0	
311	42.0	agriculteur	4.0	0.0	3.0	6	1.0	0.0	
390	36.0	agriculteur	4.0	1.0	3.0	4	3.0	1.0	
441	79.0	agriculteur	5.0	0.0	5.0	6	2.0	0.0	

	juge.enfant	place	...	subst.cons	scz.cons	char	rs	ed	dr	suicide.s	\
14	0.0	0.0	...	0	0	1.0	1.0	1.0	3.0	0.0	
311	0.0	0.0	...	0	0	2.0	1.0	3.0	2.0	3.0	
390	1.0	1.0	...	1	0	1.0	NaN	3.0	1.0	0.0	
441	0.0	0.0	...	0	0	1.0	2.0	1.0	1.0	0.0	

	suicide.hr	suicide.past	dur.interv
14	0.0	0.0	80.0

14. Donner les fréquences des modalités de la nouvelle variable age\_classe.



```

...
793  (18.999, 31.0]
795  (40.0, 50.0]
796  (18.999, 31.0]
797  (31.0, 40.0]
798  (50.0, 83.0]
Name: age_classe, Length: 403, dtype: category
Categories (4, interval[float64]): [(18.999, 31.0] < (31.0, 40.0] < (40.0, 50.0] < (50.0, 83.0]]
(18.999, 31.0]    0.265509
(31.0, 40.0]     0.258065
(50.0, 83.0]     0.238213
(40.0, 50.0]     0.238213
Name: age_classe, dtype: float64

```

Travaillons vite et bien car le temps c'est la vie