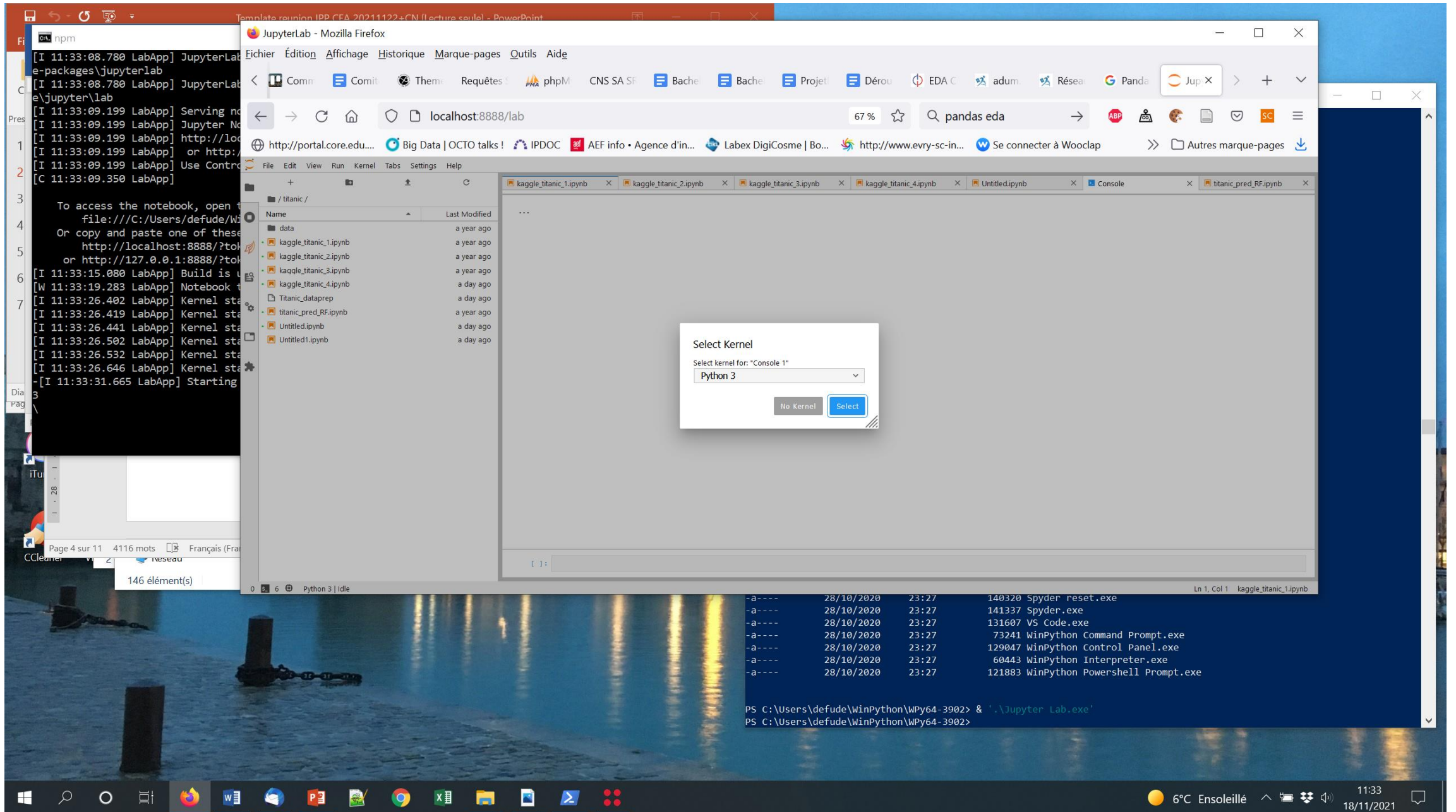


Jupyter lab, Pandas



Répertoire : C:\Users\defude\WinPython\WPy64-3902

Mode	LastWriteTime	Length	Name
d-----	26/10/2020 20:26		n
d-----	20/11/2020 14:56		notebooks
d-----	05/11/2020 15:07		python-3.9.0.amd64
d-----	28/10/2020 23:27		scripts
d-----	17/11/2021 14:00		settings
d-----	28/10/2020 23:29		t
-a----	05/11/2020 15:04	1886796	get-pip.py
-a----	28/10/2020 23:27	60440	IDLE (Python GUI).exe
-a----	28/10/2020 23:27	60439	IDLEX.exe
-a----	28/10/2020 23:27	142368	IPython Qt Console.exe
-a----	28/10/2020 23:27	75292	Jupyter Lab.exe
-a----	28/10/2020 23:27	75297	Jupyter Notebook.exe
-a----	16/03/2019 19:55	1543	license.txt
-a----	28/10/2020 23:27	145433	Pyzo.exe
-a----	28/10/2020 23:27	152094	Qt Assistant.exe
-a----	28/10/2020 23:27	144412	Qt Designer.exe
-a----	28/10/2020 23:27	150044	Qt Linguist.exe
-a----	28/10/2020 23:27	140320	Spyder reset.exe
-a----	28/10/2020 23:27	141337	Spyder.exe
-a----	28/10/2020 23:27	131607	VS Code.exe
-a----	28/10/2020 23:27	73241	WinPython Command Prompt.exe
-a----	28/10/2020 23:27	129047	WinPython Control Panel.exe
-a----	28/10/2020 23:27	60443	WinPython Interpreter.exe
-a----	28/10/2020 23:27	121883	WinPython Powershell Prompt.exe

PS C:\Users\defude\WinPython\WPy64-3902> & '.\Jupyter Lab.exe'

PS C:\Users\defude\WinPython\WPy64-3902> ls .\notebooks\Data\

Répertoire : C:\Users\defude\WinPython\WPy64-3902\notebooks\Data

Mode	LastWriteTime	Length	Name
d-----	05/11/2020 15:29		titanic
-a----	25/10/2021 11:28	9033957	olist_customers_dataset.csv

JupyterLab - Mozilla Firefox

FichierÉditionAffichageHistoriqueMarque-pagesOutilsAide

<ComitThémRequêtesPMAphpMCNS SA SFBachelBachelProjetDérouEDA CAdumRésealJupXmatplotlib>+v

localhost:8888/lab67%★🔍 idas créer frame vide →ABP👤🍪📄🔒SC☰

http://portal.core.edu...Big Data | OCTO talks !IPDOCABF AEF info • Agence d'in...hiku Labex DigiCosme | Bo...http://www.evry-sc-in...>>📁 Autres marque-pages📄

FileEditViewRunKernelTabsSettingsHelp

📁 / titanic /

Name	Last Modified
data	a year ago
• kaggle_titanic_1.ipynb	a year ago
• kaggle_titanic_2.ipynb	a year ago
• kaggle_titanic_3.ipynb	a year ago
• kaggle_titanic_4.ipynb	a day ago
📄 Titanic_dataprep	a day ago
• titanic_pred_RF.ipynb	a year ago
• Untitled.ipynb	a day ago
• Untitled1.ipynb	an hour ago
• Untitled2.ipynb	seconds ago

kaggle_titanic_1.ipynbkaggle_titanic_2.ipynbkaggle_titanic_3.ipynkaggle_titanic_4.ipynbUntitled.ipynbUntitled1.ipynbUntitled2.ipynbConsole 1titanic_pred_RF.ip

Python 3

```
[4]: import pandas as pd
data = pd.read_csv("../Data/titanic/train.csv")
data.columns

[4]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
         'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
         dtype='object')

[5]: data['Age']

[5]: 0      22.0
     1      38.0
     2      26.0
     3      35.0
     4      35.0
     ...
    886     27.0
    887     19.0
    888      NaN
    889     26.0
    890     32.0
     Name: Age, Length: 891, dtype: float64

[6]: data['Sex'].value_counts()

[6]: male      577
     female   314
     Name: Sex, dtype: int64

[7]: data_by_sex = data.groupby('Sex')
data_by_sex.size()

[7]: Sex
     female    314
     male      577
     dtype: int64

[ ]: data_by_sex = data.groupby('Sex')
data_by_sex.groupby(['Pclass', 'Sex']).size()
```

0 9 Python 3 | IdleSaving completedMode: CommandLn 2, Col 39Untitled2.ipynb

📁 / titanic /

Name	Last Modified
📁 data	a year ago
📄 kaggle_titanic_1.ipynb	a year ago
📄 kaggle_titanic_2.ipynb	a year ago
📄 kaggle_titanic_3.ipynb	a year ago
📄 kaggle_titanic_4.ipynb	a day ago
📄 Titanic_dataprep	a day ago
📄 titanic_pred_RF.ipynb	a year ago
📄 Untitled.ipynb	a day ago
📄 Untitled1.ipynb	an hour ago
📄 Untitled2.ipynb	a minute ago

```
male      577
dtype: int64
```

```
[8]: data_by_sex = data.groupby('Sex')
data.groupby(['Pclass', 'Sex']).size()
```

```
[8]: Pclass  Sex      count
1    female    94
     male     122
2    female    76
     male     108
3    female   144
     male     347
dtype: int64
```

```
[9]: data['name-age-periode'] = pd.cut(data['Age'], bins=[0, 12, 19, 55, 100],
labels=['children', 'young', 'adult', 'old'])
data[data.columns[-3:]].head(6)
```

```
[9]:
```

	Cabin	Embarked	name-age-periode
0	NaN	S	adult
1	C85	C	adult
2	NaN	S	adult
3	C123	S	adult
4	NaN	S	adult
5	NaN	Q	NaN

```
[10]: xx=('children', 'young', 'adult', 'old')
yy=data.groupby(['name-age-periode']).size()
```

```
[ ]:
```

S'entraîner sur jupyter et pandas

- refaire <https://github.com/ue12/python-numerique/tree/master/notebooks> (partie pandas)

Exploration de données
pandas, dateprep.eda

JupyterLab - Mozilla Firefox

FichierÉditionAffichageHistoriqueMarque-pagesOutilsAide

ThemeRequêtesPMAphpMCNS SA SFBachelBachelProjetDéroutEDACadumRésealTélémaJupAmazon

localhost:8888/lab67%illeur telemetre laser

http://portal.core.edu...Big Data | OCTO talks !IPDOCABFAEF info • Agence d'in...Labex DigiCosme | Bo...http://www.evry-sc-in...

Autres marque-pages

FileEditViewRunKernelTabsSettingsHelp

/ titanic /

Name	Last Modified
data	a year ago
kaggle_titanic_1.ipynb	a year ago
kaggle_titanic_2.ipynb	a year ago
kaggle_titanic_3.ipynb	a year ago
kaggle_titanic_4.ipynb	a day ago
Titanic_dataprep	a day ago
titanic_pred_RF.ipynb	a year ago
Untitled.ipynb	a day ago
Untitled1.ipynb	3 hours ago

kaggle_titanic_1.ipynbkaggle_titanic_2.ipynbkaggle_titanic_3.ipynbkaggle_titanic_4.ipynbUntitled.ipynbUntitled1.ipynbConsole 1titanic_pred_RF.ipynb

Python 3

```
[6]: import pandas as pd
data = pd.read_csv("../Data/olist_customers_dataset.csv")
data.describe()
```

```
[6]:
```

	customer_zip_code_prefix
count	99441.000000
mean	35137.474583
std	29797.938996
min	1003.000000
25%	11347.000000
50%	24416.000000
75%	58900.000000
max	99990.000000

```
[7]: data.isnull().sum()
```

```
[7]: customer_id          0
customer_unique_id    0
customer_zip_code_prefix  0
customer_city          0
customer_state         0
dtype: int64
```

0 Python 3 | IdleMode: CommandLn 1, Col 1Untitled1.ipynb

JupyterLab - Mozilla Firefox

FichierÉditionAffichageHistoriqueMarque-pagesOutilsAide

<📁CommComitThèmeRequêtesPMAphpMCNS SA SFBachelBachelProjetDérouEDACadumRésealGAPandaJupX>+v

←→↺🏠🛡️📄localhost:8888/lab67 %☆🔍pandas eda→ABP👤🍪📄📧SC☰

🌐http://portal.core.edu....🌐Big Data | OCTO talks !🌐IPDOC📄AEF info • Agence d'in...🌐Labex DigiCosme | Bo...🌐http://www.evry-sc-in...🌐Se connecter à Wooclap>>📁Autres marque-pages📄

FileEditViewRunKernelTabsSettingsHelp

📁+📁🔄

📁 / titanic /

Name	Last Modified
📁 data	a year ago
📄 kaggle_titanic_1.ipynb	a year ago
📄 kaggle_titanic_2.ipynb	a year ago
📄 kaggle_titanic_3.ipynb	a year ago
📄 kaggle_titanic_4.ipynb	a day ago
📄 Titanic_dataprep	a day ago
📄 titanic_pred_RF.ipynb	a year ago
📄 Untitled.ipynb	a day ago
📄 Untitled1.ipynb	seconds ago

kaggle_titanic_1.ipynb ×kaggle_titanic_2.ipynb ×kaggle_titanic_3.ipynb ×kaggle_titanic_4.ipynb ×Untitled.ipynb ×Untitled1.ipynb ×Console 1 ×titanic_pred_RF.ipynb ×

📄+✂️📄📄▶️⏏️🔄⏩CodePython 3

```
[5]: from dataprep.eda import create_report
report = create_report(data)
report
```

[5]:

DataPrep Report

OverviewVariablesInteractionsCorrelationsMissing Values

Overview

Dataset Statistics

Number of Variables	5
Number of Rows	99441
Missing Cells	0
Missing Cells (%)	0.0%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	29.6 MB
Average Row Size in Memory	312.3 B
Variable Types	Categorical: 4 Numerical: 1

Dataset Insights

customer_id	has a high cardinality: 99441 distinct values	High Cardinality
customer_unique_id	has a high cardinality: 96096 distinct values	High Cardinality
customer_city	has a high cardinality: 4119 distinct values	High Cardinality
customer_id	has constant length 32	Constant Length
customer_unique_id	has constant length 32	Constant Length
customer_state	has constant length 2	Constant Length
customer_id	has all distinct values	Unique

08Python 3 | IdleSaving completedMode: EditLn 1, Col 1Untitled1.ipynb

Travail à faire

- S'entraîner avec dataprep avec un des jeux de données fournis dans l'outil
 - `from dataprep.datasets import get_dataset_names`
 - `Get_data_set_names()`
- Choisir un jeu de données de e-commerce (un des fichiers csv)
- En faire l'exploration avec `dataprep.eda`
- Déterminer deux points précis de l'analyse (corrélation entre colonnes, ...) et en faire l'analyse détaillée avec `dataprep.eda`
- Livrer le notebook jupyter (.ipynb)

Visualisation de données

matplotlib, seaborn

JupyterLab - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

localhost:8888/lab

67 %

andas créer frame vide →

Se connecter à Wooclap

Autres marque-pages

File Edit View Run Kernel Tabs Settings Help

/ titanic /

Name	Last Modified
data	a year ago
kaggle_titanic_1.ipynb	a year ago
kaggle_titanic_2.ipynb	a year ago
kaggle_titanic_3.ipynb	a year ago
kaggle_titanic_4.ipynb	a day ago
pandas-prep-visu.ipynb	seconds ago
Titanic_dataprep	a day ago
titanic_pred_RF.ipynb	a year ago
Untitled.ipynb	a day ago
Untitled1.ipynb	an hour ago
Untitled2.ipynb	seconds ago

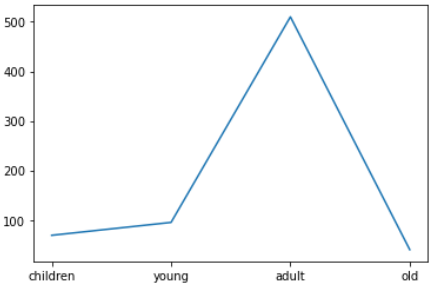
kaggle_titanic_1.ipynb kaggle_titanic_2.ipynb kaggle_titanic_3.ipynb kaggle_titanic_4.ipynb Untitled.ipynb Untitled1.ipynb pandas-prep-visu.x Console 1 titanic_pred_RF.ipynb

```
yy=data.groupby(['name-age-periode']).size()

[11]: import matplotlib.pyplot as plt

plt.plot(xx,yy)

plt.show()
```

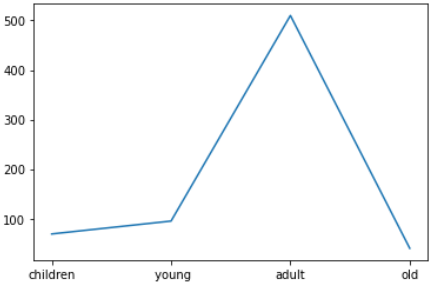


```
[12]: plt.plot(yy)
```

C:\Users\defude\WinPython\WPy64-3902\python-3.9.0.amd64\lib\site-packages\matplotlib\axes_base.py:405: FutureWarning: Support for multi-dimensional indexing (e.g. `obj[:, None]`) is deprecated and will be removed in a future version. Convert to a numpy array before indexing instead.

```
x = x[:, np.newaxis]

[12]: [<matplotlib.lines.Line2D at 0x159bdfd8850>]
```



```
[ ]: data.hist()
```

0 9 Python 3 | Idle

Saving completed

Mode: Command Ln 5, Col 11 pandas-prep-visu.ipynb

JupyterLab - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

localhost:8888/lab

67 %

andas créer frame vide →

Se connecter à Woodlap

Autres marque-pages

File Edit View Run Kernel Tabs Settings Help

/ titanic /

Name	Last Modified
data	a year ago
kaggle_titanic_1.ipynb	a year ago
kaggle_titanic_2.ipynb	a year ago
kaggle_titanic_3.ipynb	a year ago
kaggle_titanic_4.ipynb	a day ago
pandas-prep-visu.ipynb	2 minutes ago
Titanic_dataprep	a day ago
titanic_pred_RF.ipynb	a year ago
Untitled.ipynb	a day ago
Untitled1.ipynb	an hour ago
Untitled2.ipynb	2 minutes ago

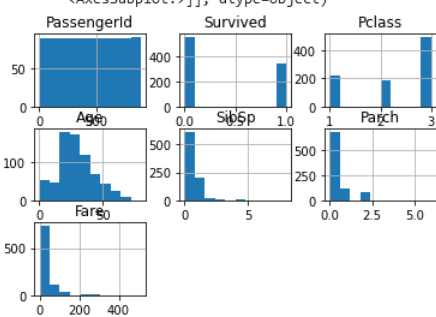
kaggle_titanic_1.ipynb X kaggle_titanic_2.ipynb X kaggle_titanic_3.ipynb X kaggle_titanic_4.ipynb X Untitled.ipynb X Untitled1.ipynb X pandas-prep-visu. pandas-prep-visu. Console 1 titanic_pred_RF.ipynb X

Code

Python 3

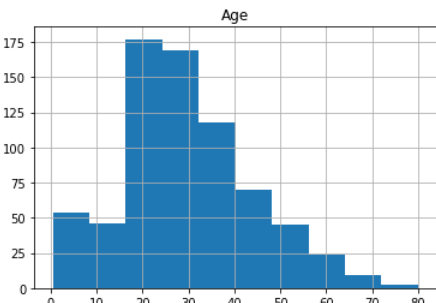
```
[13]: data.hist()
```

```
[13]: array([[<AxesSubplot:title={'center':'PassengerId'}>,<AxesSubplot:title={'center':'Survived'}>,<AxesSubplot:title={'center':'Pclass'}>,<AxesSubplot:title={'center':'Age'}>,<AxesSubplot:title={'center':'SibSp'}>,<AxesSubplot:title={'center':'Parch'}>,<AxesSubplot:title={'center':'Fare'}>,<AxesSubplot:title={'center':'Fare'}>], dtype=object)]
```



```
[14]: data.hist('Age')
```

```
[14]: array([[<AxesSubplot:title={'center':'Age'}>]], dtype=object)
```



0 Python 3 | Idle

Saving completed

Mode: Command Ln 1, Col 26 pandas-prep-visu.ipynb

JupyterLab - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

localhost:8888/lab

67 %

andas créer frame vide →

http://portal.core.edu... Big Data | OCTO talks ! IPDOC AEF info • Agence d'in... Labex DigiCosme | Bo... http://www.evry-sc-in... Se connecter à Wooclap

Autres marque-pages

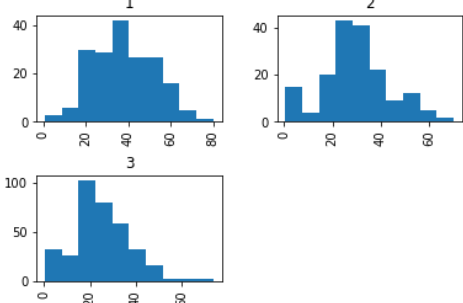
File Edit View Run Kernel Tabs Settings Help

/ titanic /

Name	Last Modified
data	a year ago
kaggle_titanic_1.ipynb	a year ago
kaggle_titanic_2.ipynb	a year ago
kaggle_titanic_3.ipynb	a year ago
kaggle_titanic_4.ipynb	a day ago
pandas-prep-visu.ipynb	a minute ago
Titanic_dataprep	a day ago
titanic_pred_RF.ipynb	a year ago
Untitled.ipynb	a day ago
Untitled1.ipynb	an hour ago
Untitled2.ipynb	3 minutes ago

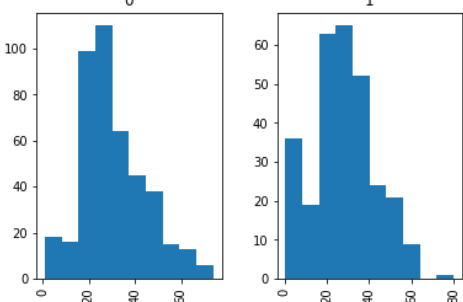
[16]: data.hist('Age', 'Pclass')

[16]: array([[<AxesSubplot:title={'center':'1'}>,<AxesSubplot:title={'center':'2'}>],<AxesSubplot:title={'center':'3'}>], dtype=object)



[17]: data.hist('Age', 'Survived')

[17]: array([[<AxesSubplot:title={'center':'0'}>,<AxesSubplot:title={'center':'1'}>], dtype=object)



[]:

Python 3 | Idle

Saving completed

Mode: Command Ln 1, Col 17 pandas-prep-visu.ipynb

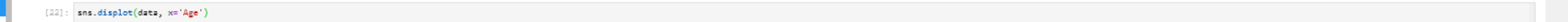
Name	Last Modified
data	a year ago
kaggle_titanic_1.ipynb	a year ago
kaggle_titanic_2.ipynb	a year ago
kaggle_titanic_3.ipynb	a year ago
kaggle_titanic_4.ipynb	a day ago
pandas-prep-visu.ipynb	a minute ago
Titanic_dataprep	a day ago
titanic_pred_RF.ipynb	a year ago
Untitled.ipynb	a day ago
Untitled1.ipynb	2 hours ago
Untitled2.ipynb	36 minutes ago

kaggle_titanic_1.ipynb x kaggle_titanic_2.ipynb x kaggle_titanic_3.ipynb x kaggle_titanic_4.ipynb x Untitled.ipynb x Untitled1.ipynb x pandas-prep-visu.ipynb Console 1 x titanic_pred_RF.ipynb x Python 3

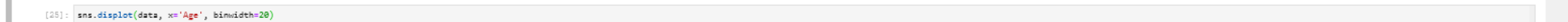
```
[27]: import seaborn as sns
sns.pairplot(data, hue='Survived')
```

```
[27]: <seaborn.axisgrid.PairGrid at 0x159c269b678>
```

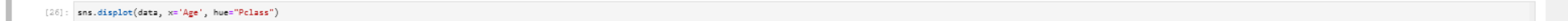




```
[22]: <seaborn.axisgrid.FacetGrid at 0x159c2290f70>
```



```
[25]: <sesborn.xmisgrid.FacetGrid at 0x159c229ed90>
```



```
[26]: <seaborn.axisgrid.FacetGrid at 0x159c21b4550>
```

```
[26]: <seaborn.axisgrid.FacetGrid at 0x159c21b4550>
```


JupyterLab - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

localhost:8888/lab

47 %

seaborn

http://portal.core.edu... Big Data | OCTO talks ! IPDOC AEF info • Agence d'in... Labex DigiCosme | Bo... http://www.evry-sc-in... Se connecter à Wooclap

File Edit View Run Kernel Tabs Settings Help

/ titanic /

Name	Last Modified
data	a year ago
kaggle_titanic_1.ipynb	a year ago
kaggle_titanic_2.ipynb	a year ago
kaggle_titanic_3.ipynb	a year ago
kaggle_titanic_4.ipynb	a day ago
pandas-prep-visu.ipynb	a minute ago
Titanic_dataprep	a day ago
titanic_pred_RF.ipynb	a year ago
Untitled.ipynb	a day ago
Untitled1.ipynb	2 hours ago
Untitled2.ipynb	38 minutes ago

Code

```
[26]: sns.displot(data, x='Age', hue='Pclass')  
[26]: <seaborn.axisgrid.FacetGrid at 0x159c21b4558>
```

Count

Age

Pclass

- 1
- 2
- 3

[1:]

0 9 Python 3 | Idle

Saving completed

Mode: Command Ln 1, Col 27 pandas-prep-visu.ipynb

Travail à faire

- S'entraîner sur matplotlib (<https://github.com/ue12/python-numerique/tree/master/notebooks> partie matplotlib) et sur seaborn (<https://seaborn.pydata.org/tutorial.html>)
- En se mettant du point de vue de l'administrateur du site de ecommerce, choisir 2 ou 3 visualisations intéressantes
- Les mettre en œuvre soit avec matplotlib, soit avec seaborn
- Livrer le notebook