

example_logs

April 22, 2022

1 Exploration des logs

Traitements de logs générés par un QCM.

```
[1]: from jupyterhelper import add_notebook_menu
      add_notebook_menu()
```

```
[1]: <IPython.core.display.HTML object>
```

```
[2]: %matplotlib inline
```

1.1 Observations brutes

```
[3]: import os
      names = [os.path.join("logs", _) for _ in os.listdir("logs") if '.log' in _]
      names = names[:1]
      names
```

```
[3]: ['logs\\QCMApp.log']
```

```
[4]: with open(names[0], 'r', encoding="utf-8") as f:
      lines = f.readlines()
      lines[5:10]
```

```
[4]: ['2018-12-12 17:56:29,989,INFO,[DATA],{"msg":"qcm","session":{"alias":"xavierd"},
      ,"client":["167.220.197.38",6274],"game":"simple_french_qcm","qn":"0"}\n',
      '2018-12-12 17:56:33,130,INFO,[DATA],{"msg":"event","session":{"alias":"xavierd"},
      ,"client":["167.220.197.38",6274],"events":["focus:true,game:simple_french_qcm",
      ,qn:0]}\n',
      '2018-12-12 17:56:34,145,INFO,[DATA],{"msg":"event","session":{"alias":"xavierd"},
      ,"client":["167.220.197.38",6274],"events":["focus:true,game:simple_french_qcm",
      ,qn:0]}\n',
      '2018-12-12 17:56:34,224,INFO,[DATA],{"msg":"event","session":{"alias":"xavierd"},
      ,"client":["167.220.196.38",52686],"events":["focus:true,game:simple_french_qc
      m,qn:0]}\n',
      '2018-12-12 17:56:34,255,INFO,[DATA],{"msg":"answer","session":{"alias":"xavier
      d"},"client":["167.220.197.38",6274],"data":{"a0":"on","b":"ok","game":"simple_f
      rench_qcm","qn":"0","next":"1"},"events":["-a0,on"]}\n']
```

```
[5]: from mathenjeu.datalog import enumerate_qcmlog
      obs = list(enumerate_qcmlog(names))
```

```
obs[:5]
```

```
[5]: [{'person_id': 'c241c15008614ea67480',
      'alias': 'xavierd',
      'time': datetime.datetime(2018, 12, 12, 17, 56, 29, 989000),
      'qtime': 'begin'},
      {'person_id': 'c241c15008614ea67480',
      'alias': 'xavierd',
      'time': datetime.datetime(2018, 12, 12, 17, 56, 34, 255000),
      'qtime': 'end',
      'simple_french_qcm-0-a0': 'on',
      'simple_french_qcm-0-b': 'ok',
      'game': 'simple_french_qcm',
      'qn': '0',
      'next': '1',
      'events': '-a0,on',
      'simple_french_qcm-0-nbvisit': 1.0,
      'simple_french_qcm-0-duration': datetime.timedelta(seconds=4,
microseconds=266000)},
      {'person_id': '32606f02fa0df6aac111',
      'alias': 'xavierd',
      'time': datetime.datetime(2018, 12, 12, 17, 56, 34, 302000),
      'qtime': 'begin'},
      {'person_id': '32606f02fa0df6aac111',
      'alias': 'xavierd',
      'time': datetime.datetime(2018, 12, 12, 17, 56, 37, 645000),
      'qtime': 'end',
      'simple_french_qcm-1-a2': 'on',
      'simple_french_qcm-1-b': 'ok',
      'game': 'simple_french_qcm',
      'qn': '1',
      'next': '2',
      'events': '-a2,on',
      'simple_french_qcm-1-nbvisit': 1.0,
      'simple_french_qcm-1-duration': datetime.timedelta(seconds=3,
microseconds=343000)},
      {'person_id': '32606f02fa0df6aac111',
      'alias': 'xavierd',
      'time': datetime.datetime(2018, 12, 12, 17, 56, 37, 677000),
      'qtime': 'begin'}]
```

```
[6]: import pandas
df = pandas.DataFrame(obs)
df.shape
```

```
[6]: (81, 58)
```

```
[7]: df.head().T
```

```
[7]:
```

	0	\
person_id	c241c15008614ea67480	
alias	xavierd	
time	2018-12-12 17:56:29.989000	
qtime	begin	

simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	NaN
qn	NaN
next	NaN
events	NaN
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaN
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaN
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaN
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaN
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaN
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaN
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaN
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN

simple_french_qcm-4-a3	NaN
	1 \
person_id	c241c15008614ea67480
alias	xavierd
time	2018-12-12 17:56:34.255000
qtime	end
simple_french_qcm-0-a0	on
simple_french_qcm-0-b	ok
game	simple_french_qcm
qn	0
next	1
events	-a0,on
simple_french_qcm-0-nbvisit	1
simple_french_qcm-0-duration	0 days 00:00:04.266000
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaN
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaN
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaN
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaN
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaN
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaN
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN

simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

	2 \
person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:56:34.302000
qtime	begin
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	NaN
qn	NaN
next	NaN
events	NaN
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaN
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaN
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaN
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaN
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaN
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaN

simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaT
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

	3 \
person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:56:37.645000
qtime	end
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	simple_french_qcm
qn	1
next	2
events	-a2,on
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaT
simple_french_qcm-1-a2	on
simple_french_qcm-1-b	ok
simple_french_qcm-1-nbvisit	1
simple_french_qcm-1-duration	0 days 00:00:03.343000
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaT
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaT
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaT
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaT
simple_french_qcm-6-a3	NaN

simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaT
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaT
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaT
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

4

person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:56:37.677000
qtime	begin
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	NaN
qn	NaN
next	NaN
events	NaN
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaT
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaT
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaT
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaT
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaT

simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaT
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaT
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaT
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaT
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

```
[8]: set(df.alias)
```

```
[8]: {'Clemence', 'thierry-d', 'xavierd', 'xavierg'}
```

1.2 Préparation des données

```
[9]: df2 = df[df.qtime == 'end'].copy()
df2.head().T
```

```
[9]:
```

person_id	c241c15008614ea67480	1 \
alias	xavierd	
time	2018-12-12 17:56:34.255000	
qtime	end	
simple_french_qcm-0-a0	on	
simple_french_qcm-0-b	ok	
game	simple_french_qcm	
qn	0	
next	1	
events	-a0,on	
simple_french_qcm-0-nbvisit	1	
simple_french_qcm-0-duration	0 days 00:00:04.266000	
simple_french_qcm-1-a2	NaN	

simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaN
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaN
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaN
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaN
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaN
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaN
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

3 \

person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:56:37.645000
qtime	end
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN

game	simple_french_qcm
qn	1
next	2
events	-a2,on
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaT
simple_french_qcm-1-a2	on
simple_french_qcm-1-b	ok
simple_french_qcm-1-nbvisit	1
simple_french_qcm-1-duration	0 days 00:00:03.343000
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaT
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaT
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaT
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaT
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaT
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaT
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaT
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

	5 \
person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:56:44.427000
qtime	end
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	simple_french_qcm
qn	2
next	3
events	-a2,on
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaN
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	on
simple_french_qcm-2-b	ok
simple_french_qcm-2-nbvisit	1
simple_french_qcm-2-duration	0 days 00:00:06.750000
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaN
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaN
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaN
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaN
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaN
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN

simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

	7 \
person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:56:54.317000
qtime	end
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	simple_french_qcm
qn	3
next	4
events	-a2,on-a2,on-a3,on-a2,on
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaN
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaN
simple_french_qcm-3-a2	on
simple_french_qcm-3-a3	on
simple_french_qcm-3-b	ok
simple_french_qcm-3-nbvisit	0.5
simple_french_qcm-3-duration	1 days 00:00:00
simple_french_qcm-4-a2	NaN
simple_french_qcm-4-b	NaN
simple_french_qcm-4-nbvisit	NaN
simple_french_qcm-4-duration	NaN
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN
simple_french_qcm-6-duration	NaN
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaN
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN

simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaN
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

	9
person_id	32606f02fa0df6aac111
alias	xavierd
time	2018-12-12 17:57:04.052000
qtime	end
simple_french_qcm-0-a0	NaN
simple_french_qcm-0-b	NaN
game	simple_french_qcm
qn	4
next	5
events	-a2,on
simple_french_qcm-0-nbvisit	NaN
simple_french_qcm-0-duration	NaN
simple_french_qcm-1-a2	NaN
simple_french_qcm-1-b	NaN
simple_french_qcm-1-nbvisit	NaN
simple_french_qcm-1-duration	NaN
simple_french_qcm-2-a2	NaN
simple_french_qcm-2-b	NaN
simple_french_qcm-2-nbvisit	NaN
simple_french_qcm-2-duration	NaN
simple_french_qcm-3-a2	NaN
simple_french_qcm-3-a3	NaN
simple_french_qcm-3-b	NaN
simple_french_qcm-3-nbvisit	NaN
simple_french_qcm-3-duration	NaN
simple_french_qcm-4-a2	on
simple_french_qcm-4-b	ok
simple_french_qcm-4-nbvisit	1
simple_french_qcm-4-duration	0 days 00:00:09.688000
simple_french_qcm-5-a0	NaN
simple_french_qcm-5-a1	NaN
simple_french_qcm-5-a2	NaN
simple_french_qcm-5-b	NaN
simple_french_qcm-5-nbvisit	NaN
simple_french_qcm-5-duration	NaN
simple_french_qcm-6-a3	NaN
simple_french_qcm-6-b	NaN
simple_french_qcm-6-nbvisit	NaN

simple_french_qcm-6-duration	NaT
simple_french_qcm-7-a2	NaN
simple_french_qcm-7-b	NaN
simple_french_qcm-7-nbvisit	NaN
simple_french_qcm-7-duration	NaT
simple_french_qcm-8-ANS	NaN
simple_french_qcm-8-b	NaN
simple_french_qcm-8-nbvisit	NaN
simple_french_qcm-8-duration	NaT
simple_french_qcm-3-a0	NaN
simple_french_qcm-6-a2	NaN
simple_french_qcm-1-a1	NaN
simple_french_qcm-4-a0	NaN
simple_french_qcm-6-a5	NaN
simple_french_qcm-7-a0	NaN
simple_french_qcm-0-a1	NaN
-a1	NaN
on-a2	NaN
on	NaN
simple_french_qcm-4-a3	NaN

```
[10]: cols = ['alias'] + [c for c in df2.columns if "simple_french_qcm" in c and '-a' in c]
df_question = df2[cols]
cols = ['alias'] + [c for c in df2.columns if "simple_french_qcm" in c and '-b' in c]
df_bouton = df2[cols]
cols = ['alias'] + [c for c in df2.columns if "simple_french_qcm" in c and '-nb' in c]
df_visit = df2[cols]
cols = ['alias'] + [c for c in df2.columns if "simple_french_qcm" in c and '-ANS' in c]
df_ans = df2[cols]
cols = ['alias'] + [c for c in df2.columns if "simple_french_qcm" in c and '-dur' in c]
df_dur = df2[cols]
```

```
[11]: df_dur.head().T
```

```
[11]:
```

	1	3 \
alias	xavierd	xavierd
simple_french_qcm-0-duration	0 days 00:00:04.266000	NaT
simple_french_qcm-1-duration	NaT	0 days 00:00:03.343000
simple_french_qcm-2-duration	NaT	NaT
simple_french_qcm-3-duration	NaT	NaT
simple_french_qcm-4-duration	NaT	NaT
simple_french_qcm-5-duration	NaT	NaT
simple_french_qcm-6-duration	NaT	NaT
simple_french_qcm-7-duration	NaT	NaT
simple_french_qcm-8-duration	NaT	NaT

	5	7 \
alias	xavierd	xavierd
simple_french_qcm-0-duration	NaT	NaT
simple_french_qcm-1-duration	NaT	NaT
simple_french_qcm-2-duration	0 days 00:00:06.750000	NaT
simple_french_qcm-3-duration	NaT	1 days 00:00:00
simple_french_qcm-4-duration	NaT	NaT

simple_french_qcm-5-duration	NaT	NaT
simple_french_qcm-6-duration	NaT	NaT
simple_french_qcm-7-duration	NaT	NaT
simple_french_qcm-8-duration	NaT	NaT

	9
alias	xavierd
simple_french_qcm-0-duration	NaT
simple_french_qcm-1-duration	NaT
simple_french_qcm-2-duration	NaT
simple_french_qcm-3-duration	NaT
simple_french_qcm-4-duration	0 days 00:00:09.688000
simple_french_qcm-5-duration	NaT
simple_french_qcm-6-duration	NaT
simple_french_qcm-7-duration	NaT
simple_french_qcm-8-duration	NaT

```
[12]: import numpy

def aggotnan_serie(values):
    res = []
    for v in values:
        if isinstance(v, float) and numpy.isnan(v):
            continue
        if pandas.isnull(v):
            continue
        if v in ('ok', 'on'):
            v = 1
        elif v == 'skip':
            v = 1000
        res.append(v)
    if len(res) > 0:
        if isinstance(res[0], str):
            r = ",".join(str(_) for _ in res)
        else:
            if len(res) == 1:
                r = res[0]
            else:
                try:
                    r = sum(res)
                except:
                    r = 0
    else:
        r = numpy.nan
    return r

def aggotnan(values):
    if isinstance(values, pandas.core.series.Series):
        r = aggotnan_serie(values)
        return r
    else:
        res = []
```

```

        for col in values.columns:
            val = list(values[col])
            res.append(aggnotnan_serie(val))
        df = pandas.DataFrame(res, columns)
        return df

gr_ans = df_ans.groupby("alias").agg(aggnotnan)
gr_ans

```

```

[12]:          simple_french_qcm-8-ANS
alias
Clemence          Prout
thierry-d    Astérix et Cléopâtre
xavierd          thalès
xavierg

```

```

[13]: gr_dur = df_dur.groupby("alias").agg(aggnotnan)
gr_dur

```

```

[13]:          simple_french_qcm-0-duration simple_french_qcm-1-duration \
alias
Clemence          0 days 00:00:16.530000      0 days 00:00:14.010000
thierry-d          0 days 00:00:06.904000              0
xavierd            0 days 00:00:04.266000      0 days 00:00:03.343000
xavierg            0 days 00:00:02.920000              NaN

          simple_french_qcm-2-duration simple_french_qcm-3-duration \
alias
Clemence          0 days 00:00:28.765000      0 days 00:00:19.492000
thierry-d          0 days 00:00:31.978000      0 days 00:00:19.246000
xavierd            0 days 00:00:06.750000      1 days 00:00:00
xavierg            NaT          0 days 00:00:09.323000

          simple_french_qcm-4-duration simple_french_qcm-5-duration \
alias
Clemence          0 days 00:03:19.593000      0 days 00:00:11.740000
thierry-d          0 days 00:00:21.230000      0 days 00:00:10.153000
xavierd            0 days 00:00:09.688000      1 days 00:00:00
xavierg            0          0 days 00:00:02.247000

          simple_french_qcm-6-duration simple_french_qcm-7-duration \
alias
Clemence          0 days 00:00:21.868000      0 days 00:00:20.923000
thierry-d          0 days 00:00:20.314000      0 days 00:00:17.141000
xavierd            1 days 00:00:00      0 days 00:00:06.390000
xavierg            0 days 00:00:00.750000      0 days 00:00:01.188000

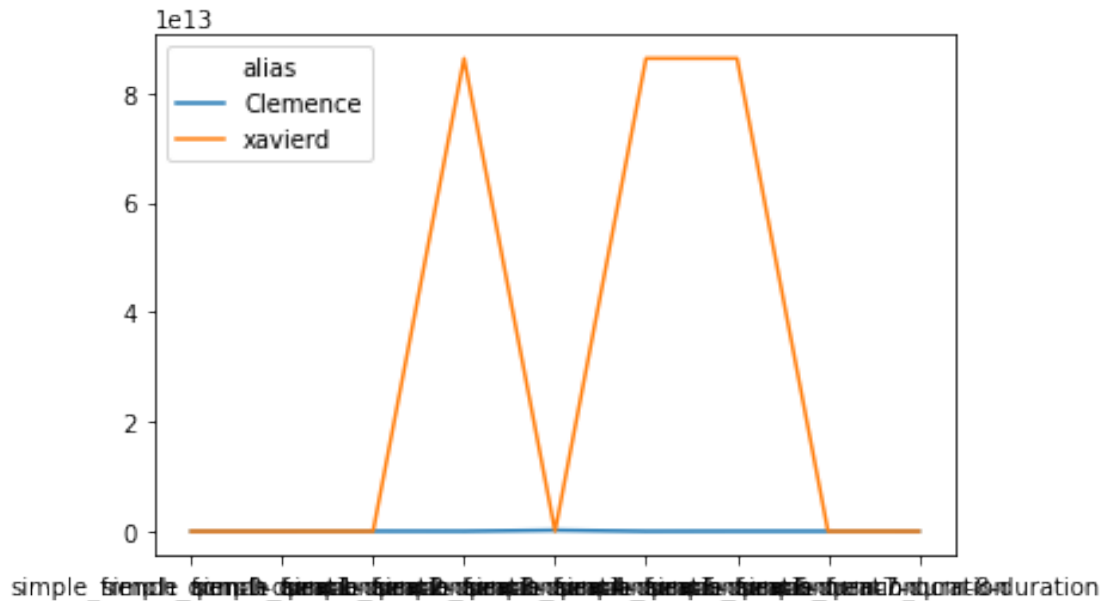
          simple_french_qcm-8-duration
alias
Clemence          0 days 00:00:14.483000
thierry-d          0 days 00:03:02.506000
xavierd            0 days 00:00:04.390000
xavierg            0 days 00:00:01.422000

```



```
[14]: gr_dur.T.plot();
```

c:\python372_x64\lib\site-packages\pandas\plotting_matplotlib\core.py:1235:
 UserWarning: FixedFormatter should only be used together with FixedLocator
 ax.set_xticklabels(xticklabels)



```
[15]: gr_question = df_question.groupby("alias").agg(aggnotnan)
      gr_question.T
```

```
[15]: alias      Clemence  thierry-d  xaviera  xaviere
simple_french_qcm-0-a0      1.0        1.0      1.0      NaN
simple_french_qcm-1-a2      NaN        2.0      1.0      NaN
simple_french_qcm-2-a2      1.0        1.0      1.0      NaN
simple_french_qcm-3-a2      NaN        NaN      1.0      1.0
simple_french_qcm-3-a3      NaN        NaN      1.0      1.0
simple_french_qcm-4-a2      NaN        1.0      1.0      1.0
simple_french_qcm-5-a0      1.0        NaN      1.0      NaN
simple_french_qcm-5-a1      NaN        1.0      1.0      NaN
simple_french_qcm-5-a2      NaN        NaN      1.0      NaN
simple_french_qcm-6-a3      NaN        NaN      1.0      NaN
simple_french_qcm-7-a2      NaN        1.0      1.0      NaN
simple_french_qcm-3-a0      1.0        1.0      NaN      NaN
simple_french_qcm-6-a2      NaN        1.0      NaN      NaN
simple_french_qcm-1-a1      1.0        NaN      NaN      NaN
simple_french_qcm-4-a0      1.0        NaN      NaN      NaN
simple_french_qcm-6-a5      1.0        NaN      NaN      NaN
simple_french_qcm-7-a0      1.0        NaN      NaN      NaN
simple_french_qcm-0-a1      NaN        NaN      NaN      1.0
simple_french_qcm-4-a3      NaN        NaN      NaN      1.0
```

```
[16]: gr_bouton = df_bouton.groupby("alias").agg(aggnotnan)
      gr_bouton.T
```

```
[16]: alias          Clemence  thierry-d  xavierd  xavierg
simple_french_qcm-0-b      1.0         1.0      1.0      1.0
simple_french_qcm-1-b      1.0        1001.0      1.0      NaN
simple_french_qcm-2-b      1.0         1.0      1.0      NaN
simple_french_qcm-3-b      1.0         1.0      1.0    1000.0
simple_french_qcm-4-b      1.0         1.0      1.0       2.0
simple_french_qcm-5-b      1.0         1.0      1.0    1000.0
simple_french_qcm-6-b      1.0         1.0      1.0       1.0
simple_french_qcm-7-b      1.0         1.0      1.0       1.0
simple_french_qcm-8-b      1.0         1.0      1.0       1.0
```

```
[17]: gr_visit = df_visit.groupby("alias").agg(aggnotnan)
      gr_visit.T
```

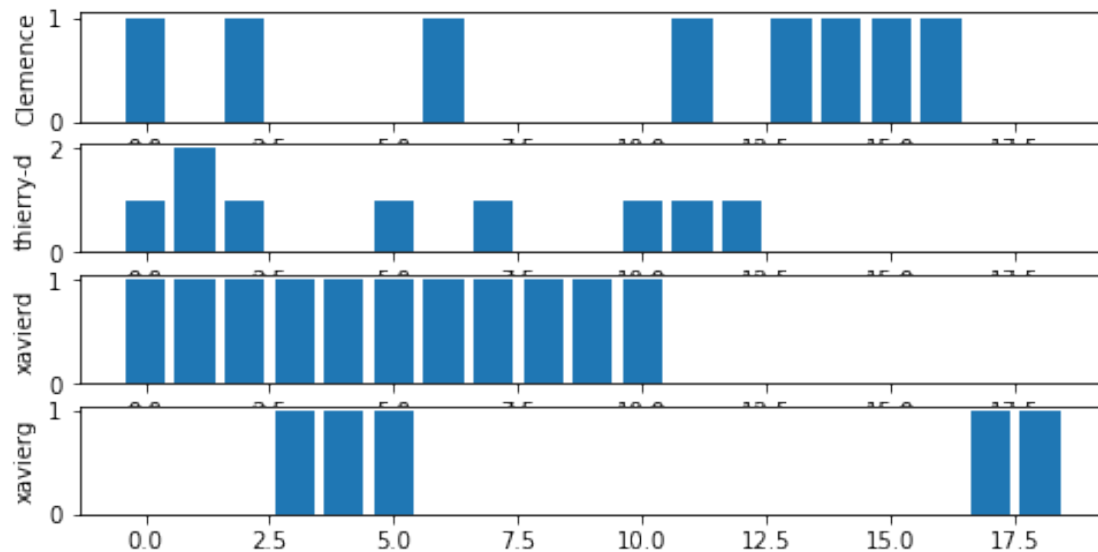
```
[17]: alias          Clemence  thierry-d  xavierd  xavierg
simple_french_qcm-0-nbvisit      1.0         1.0      1.0      1.0
simple_french_qcm-1-nbvisit      1.0         3.0      1.0      NaN
simple_french_qcm-2-nbvisit      1.0         2.0      1.0      NaN
simple_french_qcm-3-nbvisit      1.0         1.0      0.5      3.0
simple_french_qcm-4-nbvisit      1.0         1.0      1.0      5.0
simple_french_qcm-5-nbvisit      1.0         1.0      0.5      2.0
simple_french_qcm-6-nbvisit      1.0         1.0      0.5      1.0
simple_french_qcm-7-nbvisit      1.0         1.0      1.0      1.0
simple_french_qcm-8-nbvisit      1.0         1.0      1.0      1.0
```

1.3 Histogrammes

```
[18]: nonan_question = gr_question.fillna(0)
```

```
[19]: import matplotlib.pyplot as plt

fig, ax = plt.subplots(nonan_question.shape[0], 1,
                        figsize=(8, nonan_question.shape[0]))
for i in range(0, nonan_question.shape[0]):
    ax[i].set_ylabel(nonan_question.index[i])
    ax[i].bar(list(range(nonan_question.shape[1])),
              nonan_question.iloc[i,:])
```



1.4 Clustering

```
[20]: nonan_question = gr_question.fillna(0)
```

```
[21]: from sklearn.cluster import KMeans
      km = KMeans(n_clusters=2)
```

```
[22]: km.fit(nonan_question)
```

```
[22]: KMeans(n_clusters=2)
```

```
[23]: pred = km.predict(nonan_question)
      pred
```

```
[23]: array([1, 0, 0, 1])
```

```
[24]: solution = pandas.DataFrame(data=pred, columns=["cluster"], index=nonan_question.index)
```

```
[25]: solution
```

```
[25]:      cluster
alias
Clemence      1
thierry-d     0
xavierd       0
xaviern       1
```

1.5 ACP

```
[26]: nonan_question.shape
```

```
[26]: (4, 19)
```

```
[27]: from sklearn.decomposition import PCA
acp = PCA(n_components=2, svd_solver='arpack')
acp.fit(nonan_question)
```

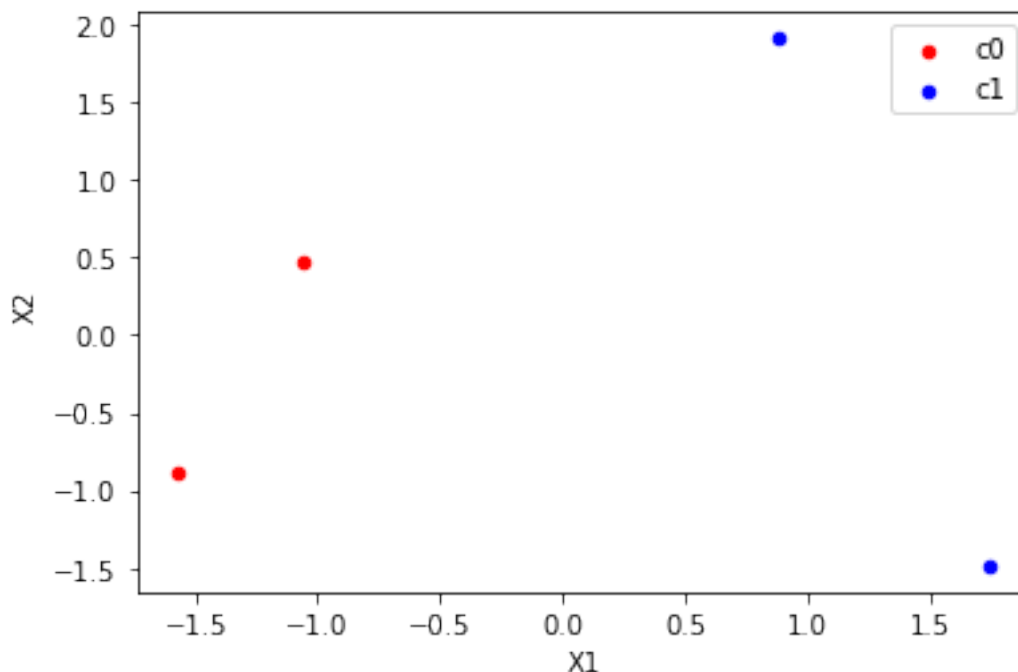
```
[27]: PCA(n_components=2, svd_solver='arpack')
```

```
[28]: coord = acp.transform(nonan_question)
data = pandas.DataFrame(data=coord, columns=['X1', 'X2'], index=nonan_question.index)
data["cluster"] = solution
data
```

```
[28]:
```

	X1	X2	cluster
alias			
Clemence	1.745710	-1.484954	1
thierry-d	-1.574330	-0.883546	0
xavierd	-1.057572	0.461806	0
xavierg	0.886193	1.906695	1

```
[29]: import matplotlib.pyplot as plt
fig, ax = plt.subplots(1, 1)
colors = ['red', 'blue', 'orange', 'green']
for i in range(0, 2):
    data[data.cluster==i].plot(x="X1", y="X2",
                               kind="scatter",
                               ax=ax, label="c%d" % i,
                               color=colors[i])
```



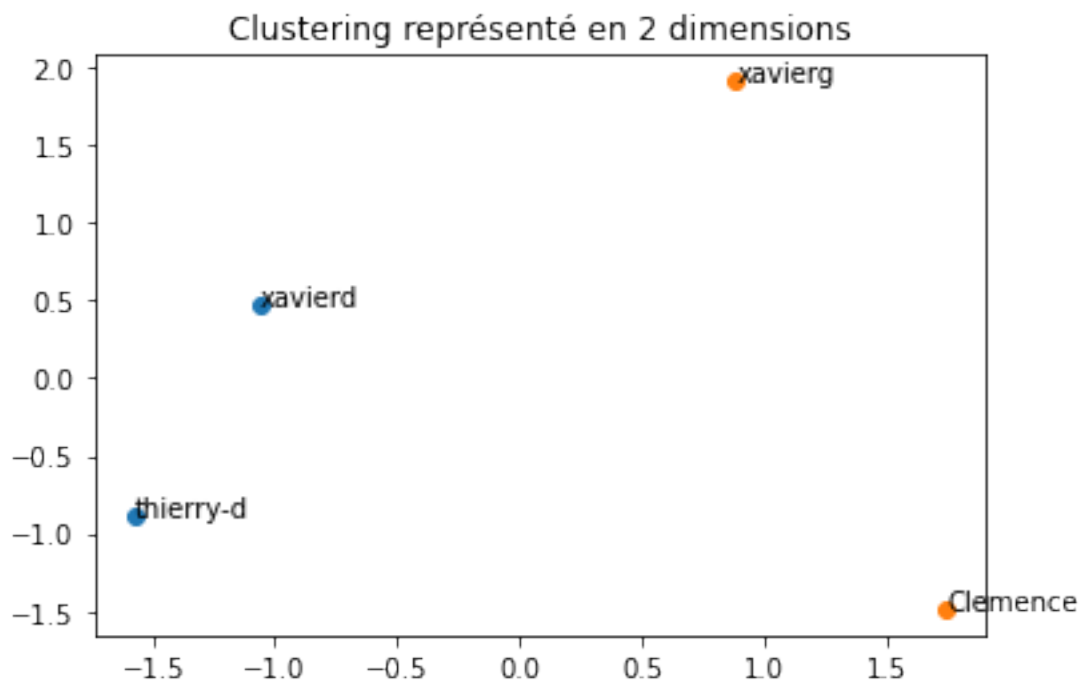
```
[30]: data.columns
```

```
[30]: Index(['X1', 'X2', 'cluster', 'weight'], dtype='object')
```

```
[31]: data.loc[data.cluster == 0, 'X1'].values, data.loc[data.cluster == 0, 'X2'].values
```

```
[31]: (array([-1.57433033, -1.05757229]), array([-0.88354622,  0.46180558]))
```

```
[32]: data["weight"] = 10
# Plot miles per gallon against horsepower with other semantics
fig, ax = plt.subplots(1, 1)
ax.plot(data.loc[data.cluster == 0, 'X1'].values,
        data.loc[data.cluster == 0, 'X2'].values, 'o', label='c0')
ax.plot(data.loc[data.cluster == 1, 'X1'].values,
        data.loc[data.cluster == 1, 'X2'].values, 'o', label='c1')
ind = list(data.index)
for i in range(0, data.shape[0]):
    ax.text(data.iloc[i, 0], data.iloc[i, 1], ind[i])
ax.set_title('Clustering représenté en 2 dimensions');
```



1.6 Prediction

```
[33]: nonan_question.T
```

```
[33]: alias
      Clemence  thierry-d  xavierd  xavierng
simple_french_qcm-0-a0      1.0      1.0      1.0      0.0
simple_french_qcm-1-a2      0.0      2.0      1.0      0.0
simple_french_qcm-2-a2      1.0      1.0      1.0      0.0
simple_french_qcm-3-a2      0.0      0.0      1.0      1.0
simple_french_qcm-3-a3      0.0      0.0      1.0      1.0
simple_french_qcm-4-a2      0.0      1.0      1.0      1.0
simple_french_qcm-5-a0      1.0      0.0      1.0      0.0
simple_french_qcm-5-a1      0.0      1.0      1.0      0.0
simple_french_qcm-5-a2      0.0      0.0      1.0      0.0
simple_french_qcm-6-a3      0.0      0.0      1.0      0.0
simple_french_qcm-7-a2      0.0      1.0      1.0      0.0
simple_french_qcm-3-a0      1.0      1.0      0.0      0.0
simple_french_qcm-6-a2      0.0      1.0      0.0      0.0
simple_french_qcm-1-a1      1.0      0.0      0.0      0.0
simple_french_qcm-4-a0      1.0      0.0      0.0      0.0
simple_french_qcm-6-a5      1.0      0.0      0.0      0.0
simple_french_qcm-7-a0      1.0      0.0      0.0      0.0
simple_french_qcm-0-a1      0.0      0.0      0.0      1.0
simple_french_qcm-4-a3      0.0      0.0      0.0      1.0
```

```
[34]: xcols = nonan_question.columns[:15]
      ycol = nonan_question.columns[16]
      ycol, set(nonan_question[ycol])
```

```
[34]: ('simple_french_qcm-7-a0', {0.0, 1.0})
```

```
[35]: from sklearn.model_selection import train_test_split
      X_train, X_test, y_train, y_test = train_test_split(nonan_question[xcols],
      ↪nonan_question[ycol], test_size=0.5)
```

```
[36]: from sklearn.ensemble import RandomForestClassifier
      clr = RandomForestClassifier()
      clr.fit(X_train, y_train)
```

```
[36]: RandomForestClassifier()
```

```
[37]: from sklearn.metrics import confusion_matrix
      confusion_matrix(y_test, clr.predict(X_test))
```

```
[37]: array([[1, 1],
      [0, 0]], dtype=int64)
```

```
[38]:
```

```
[39]:
```