

NOVA

IMS

Information
Management
School

Programming for Data Science

5th Session: Pandas

Flávio Pinheiro

fpinheiro@novaims.unl.pt

Maria Almeida

malmeida@novaims.unl.pt

Nuno Alpalhão

nalpalhao@novaims.unl.pt

What's next?

Class topics

 **Quick refresh of last week**

 **More pandas**

Pandas DataFrames



Pandas Series:

```
grades = [20,15,17,18,20,16]
```

```
students = ["Han Solo", "Ron Weasley", "Leonard Hofstadter", "Jerry Smith", "Mildred Ratched", "Piper Chapman"]
```

```
pd.Series(grades, index = students)
```

```
Han Solo      20
Ron Weasley   15
Leonard Hofstadter 17
Jerry Smith   18
Mildred Ratched 20
Piper Chapman 16
dtype: int64
```

Pandas DataFrames:

```
df = pd.read_csv("/Users/rizzoli/Desktop/Nova Ims/STATS/winequality-red.csv")
```

```
df.head()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978



Pandas Series



Pandas Series:

```
grades = pd.Series(grades, index = students)
```

```
grades['Jerry Smith']
```

18

```
grades[grades == 17]
```

```
Leonard Hofstader    17
dtype: int64
```

```
grades[grades > 16]
```

```
Han Solo            20
Leonard Hofstader  17
Jerry Smith         18
Mildred Ratched     20
dtype: int64
```

Class topics

 Quick refresh of last week

 **More pandas**

PD Final Notes

applymap() method only works on a pandas dataframe where function is applied on every element individually.

apply() method can be applied both to series and dataframes where function can be applied both series and individual elements based on the type of function provided.

map() method only works on a pandas series where type of operation to be applied depends on argument passed as a function, dictionary or a list.

- [apply\(\)](#) is used to apply a function **along an axis of the DataFrame** or **on values of Series**.
- [applymap\(\)](#) is used to apply a function to a DataFrame elementwise.
- [map\(\)](#) is used to substitute each value in a Series with another value.

Apply

```
def custom_sum(row):
    return row.sum()

df['D'] = df.apply(custom_sum, axis=1)
```

	A	B	C	D
Row 1	1	10	20	31
Row 2	2	20	40	62
Row 3	3	30	60	93
Row 4	4	40	80	124

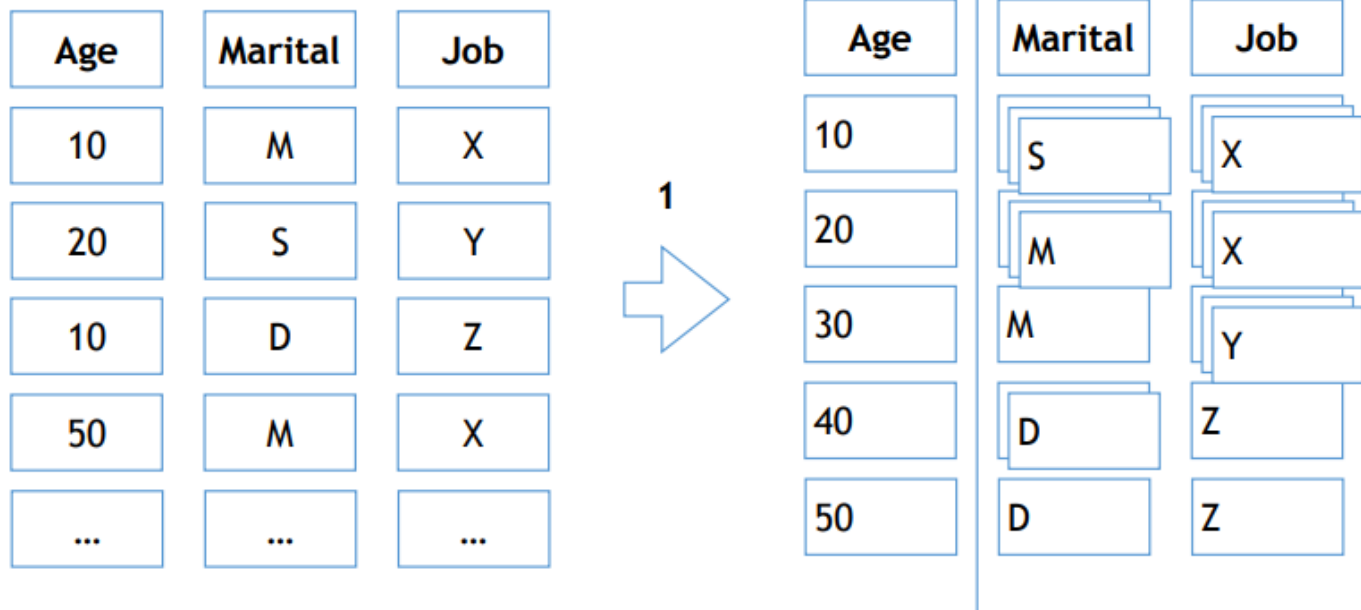
```
df.loc['Row 5'] = df.apply(custom_sum, axis=0)
```

	A	B	C
Row 1	1	10	20
Row 2	2	20	40
Row 3	3	30	60
Row 4	4	40	80
Row 5	10	100	200

.groupby()

```
df['age'] = df['age'].map(lambda x: int(x/10)*10)
df.groupby('age')['marital'].value_counts(normalize=True).unstack().fillna(0)
```

1



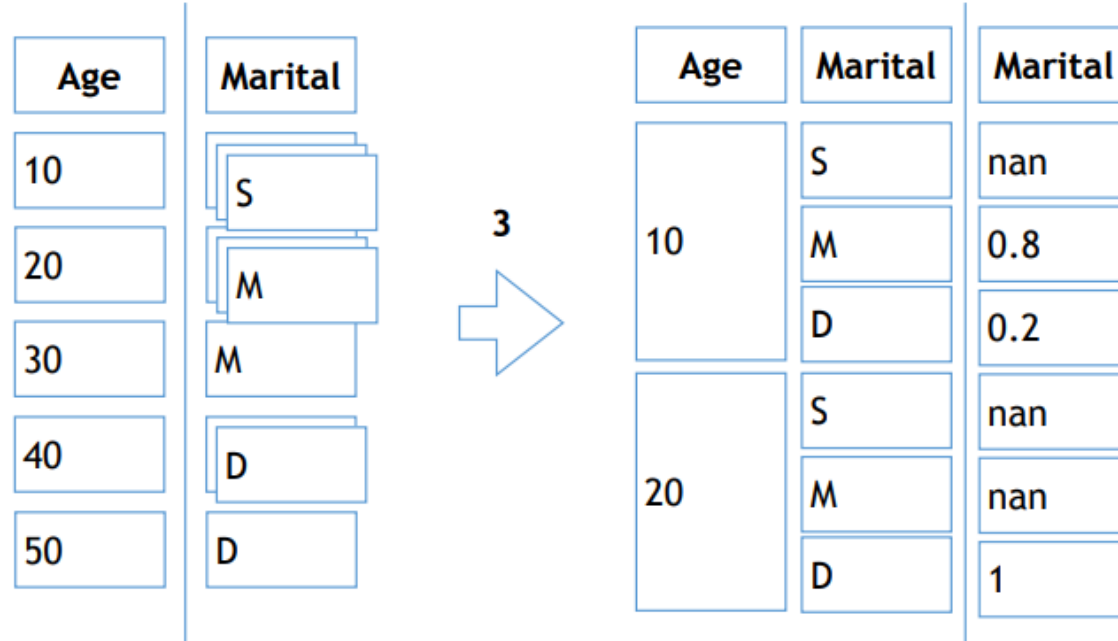
.groupby()

```
df['age'] = df['age'].map(lambda x: int(x/10)*10)
df.groupby('age')['marital'].value_counts(normalize=True).unstack().fillna(0)
```

1

2

3



End