```
import numpy as np
import pandas as pd



# I dataframe, nella realtà, sono molto grandi. Questo significa che
non si filtra in base
# alla posizione, ma si filtra in base ad una condizione. I filtri ci
permettono di selezionare
# le righe in base a delle condizioni applicate alle colonne

# Vediamo questi filtri nel dettaglio, considerando il nostro solito
Df tips



tips = pd.read_csv("tips.csv")
tips.head()
```

```
   total_bill   tip     sex smoker  day    time  size
price_per_person  \
0       16.99  1.01  Female     No  Sun  Dinner     2
8.49
1       10.34  1.66    Male     No  Sun  Dinner     3
3.45
2       21.01  3.50    Male     No  Sun  Dinner     3
7.00
3       23.68  3.31    Male     No  Sun  Dinner     2
11.84
4       24.59  3.61  Female     No  Sun  Dinner     4
6.15

           Payer Name         CC Number Payment ID
0  Christy Cunningham  3560325168603410    Sun2959
1      Douglas Tucker  4478071379779230    Sun4608
2      Travis Walters  6011812112971322    Sun4458
3    Nathaniel Harris  4676137647685994    Sun5260
4        Tonya Carter  4832732618637221    Sun2251
```

```
# VOGLIAMO ESTRARRE TUTTE LE RIGHE I CUI TOTAL_BILL SONO > 40



mask = tips["total_bill"] > 40

tips[mask]
```

```
     total_bill    tip     sex smoker   day    time  size
price_per_person  \
59        48.27   6.73    Male     No   Sat  Dinner     4
```

```
12.07
95        40.17   4.73    Male    Yes   Fri  Dinner     4
10.04
102       44.30   2.50  Female    Yes   Sat  Dinner     3
14.77
142       41.19   5.00    Male     No  Thur   Lunch     5
8.24
156       48.17   5.00    Male     No   Sun  Dinner     6
8.03
170       50.81  10.00    Male    Yes   Sat  Dinner     3
16.94
182       45.35   3.50    Male    Yes   Sun  Dinner     3
15.12
184       40.55   3.00    Male    Yes   Sun  Dinner     2
20.27
197       43.11   5.00  Female    Yes  Thur   Lunch     4
10.78
212       48.33   9.00    Male     No   Sat  Dinner     4
12.08

          Payer Name          CC Number Payment ID
59       Brian Ortiz  6596453823950595    Sat8139
95     Aaron Bentley   180026611638690    Fri9628
102    Heather Cohen   379771118886604    Sat6240
142     Eric Andrews  4356531761046453   Thur3621
156    Ryan Gonzales  3523151482063321    Sun7518
170    Gregory Clark  5473850968388236    Sat1954
182     Jose Parsons  4112207559459910    Sun2337
184      Stephen Cox  3547798222044029    Sun5140
197      Brooke Soto  5544902205760175   Thur9313
212  Alex Williamson      676218815212    Sat4590


# VOGLIAMO ESTRARRE TUTTE LE RIGHE IL CUI SESSO SIA PARI A M


mask = tips["sex"] == "Male"
tips[mask]

     total_bill   tip   sex smoker  day    time  size
price_per_person  \
1         10.34  1.66  Male     No  Sun  Dinner     3
3.45
2         21.01  3.50  Male     No  Sun  Dinner     3
7.00
3         23.68  3.31  Male     No  Sun  Dinner     2
11.84
5         25.29  4.71  Male     No  Sun  Dinner     4
6.32
```

```
6          8.77  2.00  Male     No  Sun  Dinner     2
4.38
..          ...   ...   ...    ...  ...     ...   ...                ..
.
236       12.60  1.00  Male    Yes  Sat  Dinner     2
6.30
237       32.83  1.17  Male    Yes  Sat  Dinner     2
16.42
239       29.03  5.92  Male     No  Sat  Dinner     3
9.68
241       22.67  2.00  Male    Yes  Sat  Dinner     2
11.34
242       17.82  1.75  Male     No  Sat  Dinner     2
8.91

             Payer Name         CC Number Payment ID
1        Douglas Tucker  4478071379779230    Sun4608
2        Travis Walters  6011812112971322    Sun4458
3      Nathaniel Harris  4676137647685994    Sun5260
5            Erik Smith   213140353657882    Sun9679
6    Kristopher Johnson  2223727524230344    Sun5985
..                  ...               ...        ...
236       Matthew Myers  3543676378973965    Sat5032
237        Thomas Brown  4284722681265508    Sat2929
239       Michael Avila  5296068606052842    Sat2657
241          Keith Wong  6011891618747196    Sat3880
242        Dennis Dixon     4375220550950      Sat17

[157 rows x 11 columns]


# VOGLIAMO ESTRARRE LE RIGHE IN CORRISPONDENZA DELLE QUALI total_bill
> 30 E sex=="Male"


mask = (tips["total_bill"]>30) & (tips["sex"]=="Male")
tips[mask]

     total_bill    tip   sex smoker   day    time  size
price_per_person  \
23        39.42   7.58  Male     No   Sat  Dinner     4
9.86
39        31.27   5.00  Male     No   Sat  Dinner     3
10.42
44        30.40   5.60  Male     No   Sun  Dinner     4
7.60
47        32.40   6.00  Male     No   Sun  Dinner     4
8.10
56        38.01   3.00  Male    Yes   Sat  Dinner     4
```

```
9.50
59        48.27   6.73  Male     No   Sat  Dinner     4
12.07
83        32.68   5.00  Male    Yes  Thur   Lunch     2
16.34
95        40.17   4.73  Male    Yes   Fri  Dinner     4
10.04
112       38.07   4.00  Male     No   Sun  Dinner     3
12.69
141       34.30   6.70  Male     No  Thur   Lunch     6
5.72
142       41.19   5.00  Male     No  Thur   Lunch     5
8.24
156       48.17   5.00  Male     No   Sun  Dinner     6
8.03
167       31.71   4.50  Male     No   Sun  Dinner     4
7.93
170       50.81  10.00  Male    Yes   Sat  Dinner     3
16.94
173       31.85   3.18  Male    Yes   Sun  Dinner     2
15.92
175       32.90   3.11  Male    Yes   Sun  Dinner     2
16.45
179       34.63   3.55  Male    Yes   Sun  Dinner     2
17.32
180       34.65   3.68  Male    Yes   Sun  Dinner     4
8.66
182       45.35   3.50  Male    Yes   Sun  Dinner     3
15.12
184       40.55   3.00  Male    Yes   Sun  Dinner     2
20.27
187       30.46   2.00  Male    Yes   Sun  Dinner     5
6.09
207       38.73   3.00  Male    Yes   Sat  Dinner     4
9.68
210       30.06   2.00  Male    Yes   Sat  Dinner     3
10.02
212       48.33   9.00  Male     No   Sat  Dinner     4
12.08
237       32.83   1.17  Male    Yes   Sat  Dinner     2
16.42

                Payer Name         CC Number Payment ID
23          Lance Peterson  3542584061609808     Sat239
39       Mr. Brandon Berry  6011525851069856    Sat6373
44             Todd Cooper      503846761263    Sun2274
47           James Barnes  3552002592874186    Sun9677
56   James Christensen DDS   349793629453226    Sat8903
59            Brian Ortiz  6596453823950595    Sat8139
```

```
83            Daniel Murphy  5356177501009133   Thur8801
95            Aaron Bentley   180026611638690    Fri9628
112              Jeff Lopez  3572865915176463     Sun591
141          Steven Carlson  3526515703718508   Thur1025
142            Eric Andrews  4356531761046453   Thur3621
156           Ryan Gonzales  3523151482063321    Sun7518
167          Michael Lawson  3566285921227119    Sun3719
170           Gregory Clark  5473850968388236    Sat1954
173             Scott Perez  3577115550328507    Sun9335
175         Nathan Reynolds   370307040837149    Sun5109
179            Brian Bailey   346656312114848    Sun9851
180        James Hebert DDS      676168737648    Sun7544
182            Jose Parsons  4112207559459910    Sun2337
184             Stephen Cox  3547798222044029    Sun5140
187           David Barrett  4792882899700988    Sun9987
207           Ricky Ramirez   347817964484033    Sat4505
210           Shawn Mendoza    30184049218122    Sat8361
212         Alex Williamson      676218815212    Sat4590
237            Thomas Brown  4284722681265508    Sat2929


# Possiamo anche vedere un esempio in cui si utilizza l'operatore or
|. Questo significa che,
# date due condizioni, otterremo true se è vera solo la prima, solo la
seconda, o entrambe.
# Quindi, supponiamo di voler selezionare tutte le righe in
corrispondenza delle quali il sesso sia
# F OPPURE smoker sia "Yes"


mask = (tips["sex"] == "Female") | (tips["smoker"] == "Yes")

tips[mask]

     total_bill   tip     sex smoker   day    time  size
price_per_person  \
0         16.99  1.01  Female     No   Sun  Dinner     2
8.49
4         24.59  3.61  Female     No   Sun  Dinner     4
6.15
11        35.26  5.00  Female     No   Sun  Dinner     4
8.82
14        14.83  3.02  Female     No   Sun  Dinner     2
7.42
16        10.33  1.67  Female     No   Sun  Dinner     3
3.44
..          ...   ...     ...    ...   ...     ...   ...
...
237       32.83  1.17    Male    Yes   Sat  Dinner     2
```

```
16.42
238       35.83  4.67  Female     No   Sat  Dinner     3
11.94
240       27.18  2.00  Female    Yes   Sat  Dinner     2
13.59
241       22.67  2.00    Male    Yes   Sat  Dinner     2
11.34
243       18.78  3.00  Female     No  Thur  Dinner     2
9.39

              Payer Name         CC Number Payment ID
0    Christy Cunningham  3560325168603410    Sun2959
4          Tonya Carter  4832732618637221    Sun2251
11         Diane Macias  4577817359320969    Sun6686
14        Vanessa Jones    30016702287574    Sun3848
16     Elizabeth Foster  4240025044626033    Sun9715
..                  ...               ...        ...
237        Thomas Brown  4284722681265508    Sat2929
238      Kimberly Crane      676184013727    Sat9777
240      Monica Sanders  3506806155565404    Sat1766
241          Keith Wong  6011891618747196    Sat3880
243     Michelle Hardin  3511451626698139    Thur672

[147 rows x 11 columns]
```

```
# POSSIAMO ANCHE LAVORARE SU UNA SOLA COLONNA: VEDIAMO QUI UN ESEMPIO.

# CONSIDERIAMO LA VARIABILE time. PER PRIMA COSA VEDIAMO QUALI SONO LE
OCCORRENZE DI QUESTA VARIABILE
# CATEGORICA
```

```
tips['time'].value_counts()
```

```
time
Dinner    176
Lunch      68
Name: count, dtype: int64
```

```
# SONO SOLO DINNER O LOUNCH. L'ESEMPIO NON VERREBBE BENE
```

```
# PROVIAMO CON LA VARIABILE day:
```

```
tips['day'].value_counts()
```

```
day
Sat     87
Sun     76
Thur    62
Fri     19
Name: count, dtype: int64
```

```
# VOGLIAMO ESTRARRE LE RIHJE PER LE QUALI LA COLONNA day ASSUME I VALORI Sun oppure Fri
```

```
mask = (tips["day"]=="Sun") | (tips["day"]=="Fri")
tips[mask]
```

```
     total_bill   tip     sex smoker  day    time  size
price_per_person  \
0         16.99  1.01  Female     No  Sun  Dinner     2
8.49
1         10.34  1.66    Male     No  Sun  Dinner     3
3.45
2         21.01  3.50    Male     No  Sun  Dinner     3
7.00
3         23.68  3.31    Male     No  Sun  Dinner     2
11.84
4         24.59  3.61  Female     No  Sun  Dinner     4
6.15
..          ...   ...     ...    ...  ...     ...   ...
...
222        8.58  1.92    Male    Yes  Fri   Lunch     1
8.58
223       15.98  3.00  Female     No  Fri   Lunch     3
5.33
224       13.42  1.58    Male    Yes  Fri   Lunch     2
6.71
225       16.27  2.50  Female    Yes  Fri   Lunch     2
8.14
226       10.09  2.00  Female    Yes  Fri   Lunch     2
5.04

              Payer Name         CC Number Payment ID
0    Christy Cunningham  3560325168603410    Sun2959
1        Douglas Tucker  4478071379779230    Sun4608
2        Travis Walters  6011812112971322    Sun4458
3      Nathaniel Harris  4676137647685994    Sun5260
4          Tonya Carter  4832732618637221    Sun2251
..                  ...               ...        ...
222      Jason Lawrence  3505302934650403    Fri6624
223         Mary Rivera  5343428579353069    Fri6014
```

```
224   Ronald Vaughn DVM   341503466406403    Fri5959
225      Whitney Arnold  3579111947217428    Fri6665
226          Ruth Weiss  5268689490381635    Fri6359

[95 rows x 11 columns]
```

```
# VEDIAMO SE è VERO CHE RIMANGONO SOLO LE OCCORRENZE Sun E Fri:

solo_sun_fri = tips[mask]
solo_sun_fri["day"].value_counts()
```

```
day
Sun    76
Fri    19
Name: count, dtype: int64
```

```
# CVD


# UN ALTRO MODO PER UTILIZZARE OR | QUANDO SI LAVORA SU CONDIZIONI
APPARTENENTI ALLA STESSA VARIABILE
# E' L'OPERATORE isin. RIVEDIAMO L'ULTIMO ESEMPIO


tips
```

```
     total_bill   tip     sex smoker   day    time  size
price_per_person  \
0         16.99  1.01  Female     No   Sun  Dinner     2
8.49
1         10.34  1.66    Male     No   Sun  Dinner     3
3.45
2         21.01  3.50    Male     No   Sun  Dinner     3
7.00
3         23.68  3.31    Male     No   Sun  Dinner     2
11.84
4         24.59  3.61  Female     No   Sun  Dinner     4
6.15
..          ...   ...     ...    ...   ...     ...   ...
...
239       29.03  5.92    Male     No   Sat  Dinner     3
9.68
240       27.18  2.00  Female    Yes   Sat  Dinner     2
13.59
241       22.67  2.00    Male    Yes   Sat  Dinner     2
11.34
242       17.82  1.75    Male     No   Sat  Dinner     2
8.91
```

```
243       18.78  3.00  Female     No  Thur  Dinner     2
9.39

             Payer Name         CC Number Payment ID
0    Christy Cunningham  3560325168603410    Sun2959
1        Douglas Tucker  4478071379779230    Sun4608
2        Travis Walters  6011812112971322    Sun4458
3      Nathaniel Harris  4676137647685994    Sun5260
4          Tonya Carter  4832732618637221    Sun2251
..                  ...                ...        ...
239       Michael Avila  5296068606052842    Sat2657
240      Monica Sanders  3506806155565404    Sat1766
241          Keith Wong  6011891618747196    Sat3880
242        Dennis Dixon     4375220550950      Sat17
243     Michelle Hardin  3511451626698139    Thur672

[244 rows x 11 columns]



mask = tips["day"].isin(["Sun","Fri"])
tips[mask]

     total_bill   tip     sex smoker  day    time  size
price_per_person  \
0         16.99  1.01  Female     No  Sun  Dinner     2
8.49
1         10.34  1.66    Male     No  Sun  Dinner     3
3.45
2         21.01  3.50    Male     No  Sun  Dinner     3
7.00
3         23.68  3.31    Male     No  Sun  Dinner     2
11.84
4         24.59  3.61  Female     No  Sun  Dinner     4
6.15
..          ...   ...     ...    ...  ...     ...   ...
...
222        8.58  1.92    Male    Yes  Fri   Lunch     1
8.58
223       15.98  3.00  Female     No  Fri   Lunch     3
5.33
224       13.42  1.58    Male    Yes  Fri   Lunch     2
6.71
225       16.27  2.50  Female    Yes  Fri   Lunch     2
8.14
226       10.09  2.00  Female    Yes  Fri   Lunch     2
5.04

             Payer Name         CC Number Payment ID
0    Christy Cunningham  3560325168603410    Sun2959
```

```
1        Douglas Tucker  4478071379779230    Sun4608
2        Travis Walters  6011812112971322    Sun4458
3      Nathaniel Harris  4676137647685994    Sun5260
4          Tonya Carter  4832732618637221    Sun2251
..                  ...                ...       ...
222      Jason Lawrence  3505302934650403    Fri6624
223         Mary Rivera  5343428579353069    Fri6014
224   Ronald Vaughn DVM   341503466406403    Fri5959
225      Whitney Arnold  3579111947217428    Fri6665
226          Ruth Weiss  5268689490381635    Fri6359

[95 rows x 11 columns]
```

```
solo_sun_fri = tips[mask]
solo_sun_fri["day"].value_counts()
```

```
day
Sun    76
Fri    19
Name: count, dtype: int64
```

```
# Il risultato è lo stesso
```