

On importe la librairie `pandas` qui est une boîte à outils

```
In [ ]: import pandas as pd
```

On importe le dataset `jd_ventes_idf.csv`

```
In [ ]: df = pd.read_csv("/content/jd_ventes_idf.csv" , sep =";")
```

On affiche les 10 premières lignes du dataset

```
In [ ]: df.head(10)
```

```
Out[ ]:
```

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
0	1	2025-01-02	JD Les Halles	Paris	Nike Air Max	Blanc	4	129.99
1	2	2025-01-02	JD Italie 2	Paris	Adidas Campus	Noir	2	99.99
2	3	2025-01-02	JD Rosny 2	Rosny	Puma RS-X	Bleu	3	119.99
3	4	2025-01-03	JD Les Halles	Paris	Nike Dunk	Blanc	5	149.99
4	5	2025-01-03	JD Vélizy	Velizy	New Balance 530	Gris	2	139.99
5	6	2025-01-03	JD Créteil Soleil	Creteil	Adidas Campus	Bleu	3	99.99
6	7	2025-01-03	JD Boulogne	Boulogne	Nike Air Max	Rouge	1	129.99
7	8	2025-01-04	JD Les Halles	Paris	Puma RS-X	Bleu	4	119.99
8	9	2025-01-04	JD Rosny 2	Rosny	Nike Dunk	Noir	2	149.99
9	10	2025-01-04	JD Italie 2	Paris	Nike Air Max	Blanc	3	129.99

Pour afficher la fin de la liste

```
In [ ]: df.tail(15)
```

Out []:

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
15	16	2025-01-06	JD Créteil Soleil	Creteil	New Balance 530	Gris	3	139.99
16	17	2025-01-07	JD Boulogne	Boulogne	Nike Dunk	Blanc	2	149.99
17	18	2025-01-07	JD Rosny 2	Rosny	Adidas Campus	Bleu	2	99.99
18	19	2025-01-07	JD Vélizy	Velizy	Nike Air Max	Blanc	5	129.99
19	20	2025-01-08	JD Les Halles	Paris	Nike Dunk	Noir	3	149.99
20	21	2025-01-08	JD Italie 2	Paris	New Balance 530	Gris	2	139.99
21	22	2025-01-08	JD Créteil Soleil	Creteil	Puma RS-X	Bleu	4	119.99
22	23	2025-01-08	JD Boulogne	Boulogne	Adidas Campus	Noir	1	99.99
23	24	2025-01-08	JD Rosny 2	Rosny	Nike Air Max	Rouge	3	129.99
24	25	2025-01-09	JD Vélizy	Velizy	Nike Dunk	Blanc	4	149.99
25	26	2025-01-09	JD Les Halles	Paris	New Balance 530	Gris	3	139.99
26	27	2025-01-09	JD Italie 2	Paris	Nike Air Max	Blanc	2	129.99
27	28	2025-01-09	JD Rosny 2	Rosny	Puma RS-X	Bleu	2	119.99
28	29	2025-01-09	JD Boulogne	Boulogne	Nike Dunk	Noir	3	149.99
29	30	2025-01-09	JD Créteil Soleil	Creteil	Nike Air Max	Rouge	4	129.99

On va afficher les infos generales du dataset

```
In [ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id_vente         30 non-null     int64
1   date             30 non-null     object
2   magasin          30 non-null     object
3   ville            30 non-null     object
4   modele           30 non-null     object
5   couleur          30 non-null     object
6   quantite         30 non-null     int64
7   prix_unitaire    30 non-null     float64
dtypes: float64(1), int64(2), object(5)
memory usage: 2.0+ KB
```

float= nb a virgule,int= nb entier object= objet du texte

```
In [ ]: df.shape
```

```
Out[ ]: (30, 8)
```

```
In [ ]: df.describe()
```

```
Out[ ]:
```

	id_vente	quantite	prix_unitaire
count	30.000000	30.000000	30.000000
mean	15.500000	2.800000	129.656667
std	8.803408	1.095445	16.914253
min	1.000000	1.000000	99.990000
25%	8.250000	2.000000	119.990000
50%	15.500000	3.000000	129.990000
75%	22.750000	3.750000	139.990000
max	30.000000	5.000000	149.990000

describe = y aura que les donnee quantitative statistique descriptive mean = moyenne std= standard de deviation (ecart type) min= valeur minimum 25%= 1er cartile 50%= medianne 75%=3eme cart tile max

```
In [ ]: df['modele'].unique()
```

```
Out[ ]: array(['Nike Air Max', 'Adidas Campus', 'Puma RS-X', 'Nike Dunk',
          'New Balance 530'], dtype=object)
```

liste valeur unique de la colonne choisi

unique affiche toute les valeurs que prend une colonne

```
In [ ]: df['ville'].value_counts()
```

Out []:

	count
ville	
Paris	11
Rosny	5
Creteil	5
Boulogne	5
Velizy	4

dtype: int64

nb de fois que viens chaque valeurs

5 premiere valeur commande model

```
In [ ]: df['modele'].value_counts()
```

Out []:

	count
modele	
Nike Air Max	8
Nike Dunk	7
Adidas Campus	5
Puma RS-X	5
New Balance 530	5

dtype: int64

mtn on va voir les modeles qui sont que à paris

le data set s appelle df et le filtre qu on veut voir c dfville vu qu on veut la colonne ville

```
In [ ]: df[df['ville'] == "Paris"]
```

Out []:

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
0	1	2025-01-02	JD Les Halles	Paris	Nike Air Max	Blanc	4	129.99
1	2	2025-01-02	JD Italie 2	Paris	Adidas Campus	Noir	2	99.99
3	4	2025-01-03	JD Les Halles	Paris	Nike Dunk	Blanc	5	149.99
7	8	2025-01-04	JD Les Halles	Paris	Puma RS-X	Bleu	4	119.99
9	10	2025-01-04	JD Italie 2	Paris	Nike Air Max	Blanc	3	129.99
13	14	2025-01-06	JD Les Halles	Paris	Adidas Campus	Noir	3	99.99
14	15	2025-01-06	JD Italie 2	Paris	Puma RS-X	Bleu	1	119.99
19	20	2025-01-08	JD Les Halles	Paris	Nike Dunk	Noir	3	149.99
20	21	2025-01-08	JD Italie 2	Paris	New Balance 530	Gris	2	139.99
25	26	2025-01-09	JD Les Halles	Paris	New Balance 530	Gris	3	139.99
26	27	2025-01-09	JD Italie 2	Paris	Nike Air Max	Blanc	2	129.99

In []: `df[df['magasin']== "JD Les Halles"]`

Out []:

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
0	1	2025-01-02	JD Les Halles	Paris	Nike Air Max	Blanc	4	129.99
3	4	2025-01-03	JD Les Halles	Paris	Nike Dunk	Blanc	5	149.99
7	8	2025-01-04	JD Les Halles	Paris	Puma RS-X	Bleu	4	119.99
13	14	2025-01-06	JD Les Halles	Paris	Adidas Campus	Noir	3	99.99
19	20	2025-01-08	JD Les Halles	Paris	Nike Dunk	Noir	3	149.99
25	26	2025-01-09	JD Les Halles	Paris	New Balance 530	Gris	3	139.99

ca c pour avoir les 5 premiere ligne des halles

```
In [ ]: df[df['magasin']== "JD Les Halles"].head()
```

```
Out [ ]:
```

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
0	1	2025-01-02	JD Les Halles	Paris	Nike Air Max	Blanc	4	129.99
3	4	2025-01-03	JD Les Halles	Paris	Nike Dunk	Blanc	5	149.99
7	8	2025-01-04	JD Les Halles	Paris	Puma RS-X	Bleu	4	119.99
13	14	2025-01-06	JD Les Halles	Paris	Adidas Campus	Noir	3	99.99
19	20	2025-01-08	JD Les Halles	Paris	Nike Dunk	Noir	3	149.99

```
In [ ]: df[df['prix_unitaire']< 120]
```

```
Out [ ]:
```

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
1	2	2025-01-02	JD Italie 2	Paris	Adidas Campus	Noir	2	99.99
2	3	2025-01-02	JD Rosny 2	Rosny	Puma RS-X	Bleu	3	119.99
5	6	2025-01-03	JD Créteil Soleil	Creteil	Adidas Campus	Bleu	3	99.99
7	8	2025-01-04	JD Les Halles	Paris	Puma RS-X	Bleu	4	119.99
13	14	2025-01-06	JD Les Halles	Paris	Adidas Campus	Noir	3	99.99
14	15	2025-01-06	JD Italie 2	Paris	Puma RS-X	Bleu	1	119.99
17	18	2025-01-07	JD Rosny 2	Rosny	Adidas Campus	Bleu	2	99.99
21	22	2025-01-08	JD Créteil Soleil	Creteil	Puma RS-X	Bleu	4	119.99
22	23	2025-01-08	JD Boulogne	Boulogne	Adidas Campus	Noir	1	99.99
27	28	2025-01-09	JD Rosny 2	Rosny	Puma RS-X	Bleu	2	119.99

```
In [ ]: df[df['quantite']<2]
```

```
Out [ ]:
```

	id_vente	date	magasin	ville	modele	couleur	quantite	prix_unitaire
6	7	2025-01-03	JD Boulogne	Boulogne	Nike Air Max	Rouge	1	129.99
14	15	2025-01-06	JD Italie 2	Paris	Puma RS-X	Bleu	1	119.99
22	23	2025-01-08	JD Boulogne	Boulogne	Adidas Campus	Noir	1	99.99

```
In [ ]: df[(df['magasin']=='JD Créteil Soleil') & (df['couleur']=='Blanc')]
```

```
-----
-
NameError                                Traceback (most recent call last)
/tmp/ipython-input-3735562689.py in <cell line: 0>()
----> 1 df[(df['magasin']=='JD Créteil Soleil') & (df['couleur']=='Blanc')]

NameError: name 'df' is not defined
```

```
In [ ]: df[(df['magasin'] == "JD Créteil Soleil") & (df['couleur'] == "Blanc")]
```

```
-----
-
NameError                                Traceback (most recent call last)
/tmp/ipython-input-1971844149.py in <cell line: 0>()
----> 1 df[(df['magasin'] == "JD Créteil Soleil") & (df['couleur'] == "Blanc")]

NameError: name 'df' is not defined
```

```
In [ ]: df[(df['ville']=='Paris')&(df['couleur']=='Blanc')]
```

```
In [ ]: df[(df['ville']=="Paris")]
```

iteration = ordre execution cellule se retranscrit par nb entre crochet

[texte du lien](#)