

# Tables des matières

- [1 Pratique visualisation avec Seaborn](#)
  - [1.1 Importation des modules](#)
  - [1.2 Chargement du dataset](#)
  - [1.3 Statistiques sur le dataframe](#)
    - [1.3.1 Vérification des NaN](#)
    - [1.3.2 Statistiques de base](#)
- [2 Matrice de corrélation](#)
  - [2.1 Heatmap](#)
  - [2.2 Détermination de linéarité](#)

## Pratique visualisation avec Seaborn

### Importation des modules

```
Entrée [8]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.datasets import fetch_california_housing
dataset = fetch_california_housing()

#Mettre dans un dataframe
# dataset = load_boston()
housing = pd.DataFrame(dataset.data, columns = dataset.feature_names)
housing['MEDV'] = dataset.target
```

```
[4.526 3.585 3.521 ... 0.923 0.847 0.894]
```

## Chargement du dataset

Entrée [3]: *#afficher une partie du dataframe*  
housing.head()

Out[3]:

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MEDV
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.52
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.58
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.52
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.41
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.42

## Statistiques sur le dataframe

### Vérification des NaN

Entrée [4]: *#verifier la présence de NaN*  
housing.isnull().sum()

Out[4]:

MedInc	0
HouseAge	0
AveRooms	0
AveBedrms	0
Population	0
AveOccup	0
Latitude	0
Longitude	0
MEDV	0

dtype: int64

### Statistiques de base

Entrée [5]: *#description stats sur la cible target*  
housing['MEDV'].describe()

Out[5]:

count	20640.000000
mean	2.068558
std	1.153956
min	0.149990
25%	1.196000
50%	1.797000
75%	2.647250
max	5.000010

Name: MEDV, dtype: float64

# Matrice de corrélation

## Heatmap

Entrée [6]: *#correlation entre les différents éléments du dataset*

```
sns.set(style="white", font_scale = 0.9)
sns.heatmap(data= housing.corr().round(2), cmap='coolwarm', annot=True,
            annot_kws={"size":8} )
plt.tight_layout()
plt.show()
```

