

# **Soutenance de stage d'initiation**

Prévision et analyse de données météorologiques

*Réalisé par : Louise Vannetzel*

*Encadré par : M.Modeste*

*Année Universitaire 2023-2024*

# Sommaire

## I. Modèle de prédiction de températures simple

- A. Estimation d'une station grâce à une autre

- B. Estimation d'une station grâce aux onze autres

## II. Clustering : méthode des k-means

- A. Sur la position

- B. Sur la température

## III. Conclusion

# Premier modèle de prédiction

- Nettoyage de la base de données

	annee	jour	Cap-De-La-Heve	Rennes	Le Mans	Paris	Beauvais	\
186	1956	1	8.2	6.8	4.0	6.2	4.4	
187	1956	2	6.3	7.7	6.6	5.2	4.7	
188	1956	3	5.4	4.3	4.0	2.2	2.0	
189	1956	4	2.2	1.4	0.6	1.8	-0.2	
190	1956	5	3.6	1.6	2.9	3.9	1.9	
...	...	...	...	...	...	...	...	
1948	2012	27	6.9	5.0	4.3	6.6	3.4	
1949	2012	28	5.2	5.1	4.1	5.4	3.3	
1950	2012	29	4.6	4.7	4.9	3.0	1.6	
1951	2012	30	1.0	2.5	1.0	0.7	-0.2	
1952	2012	31	-0.4	1.3	0.0	0.1	-0.6	
	Strasbourg	Vichy	Biarritz	Toulouse	Carcassone	Nimes	Mont-Aigoual	
186	3.1	4.0	8.8	6.9	7.5	6.4	-4.3	
187	5.0	5.4	9.0	7.0	7.6	7.5	-3.8	
188	2.7	1.6	4.0	4.9	5.8	7.1	-2.3	
189	-0.8	1.4	4.0	0.1	3.1	6.2	-0.1	
190	1.5	-2.9	3.4	-0.8	-0.8	4.8	3.2	
...	...	...	...	...	...	...	...	
1948	4.2	5.7	9.1	8.6	9.3	8.8	-0.9	
1949	3.5	2.3	7.3	5.7	6.7	7.8	-4.1	
1950	1.6	1.8	4.1	3.0	3.1	6.0	-6.0	
1951	0.0	0.7	4.1	2.2	3.7	5.8	-6.5	
1952	-1.2	-0.5	7.3	4.3	3.8	1.1	-7.1	

[1705 rows x 14 columns]

# Estimation d'une station grâce à une autre

- $Y = a + bX$
- $Y = \mathbb{E}[X]$

avec  $a = \mathbb{E}[Y - bX]$  et  $b = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$



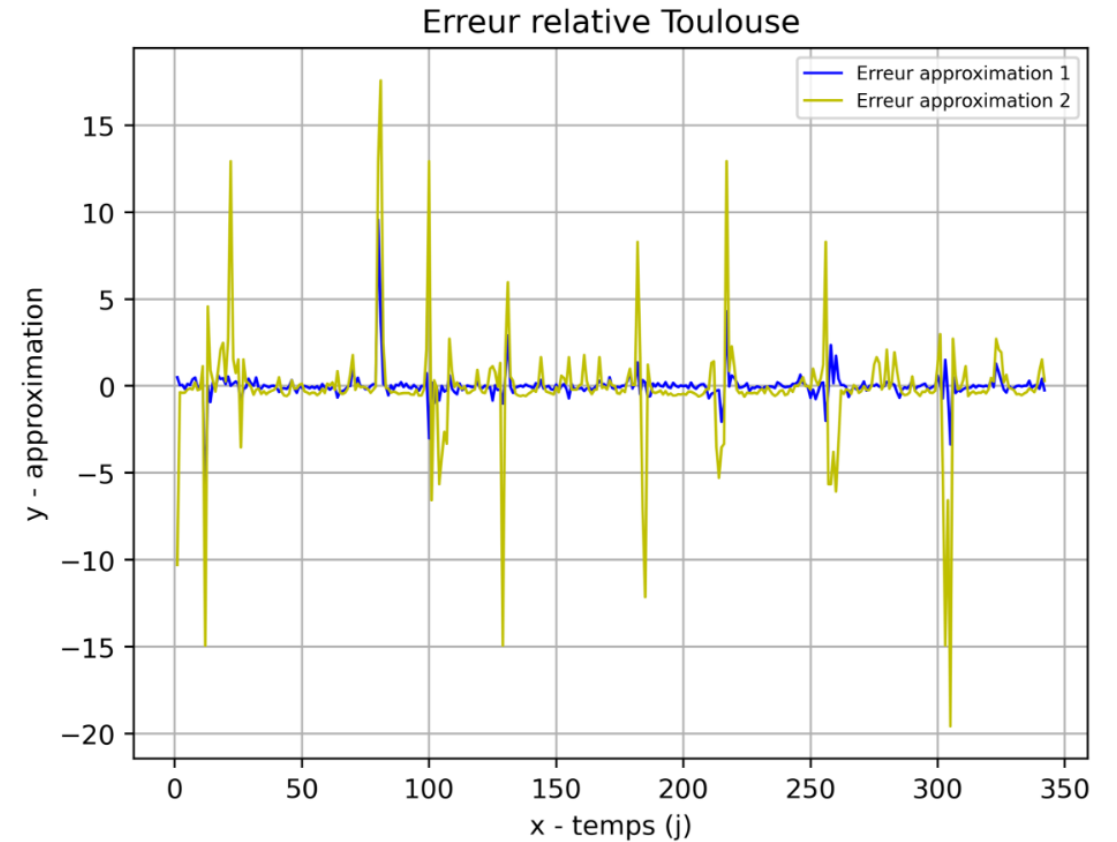
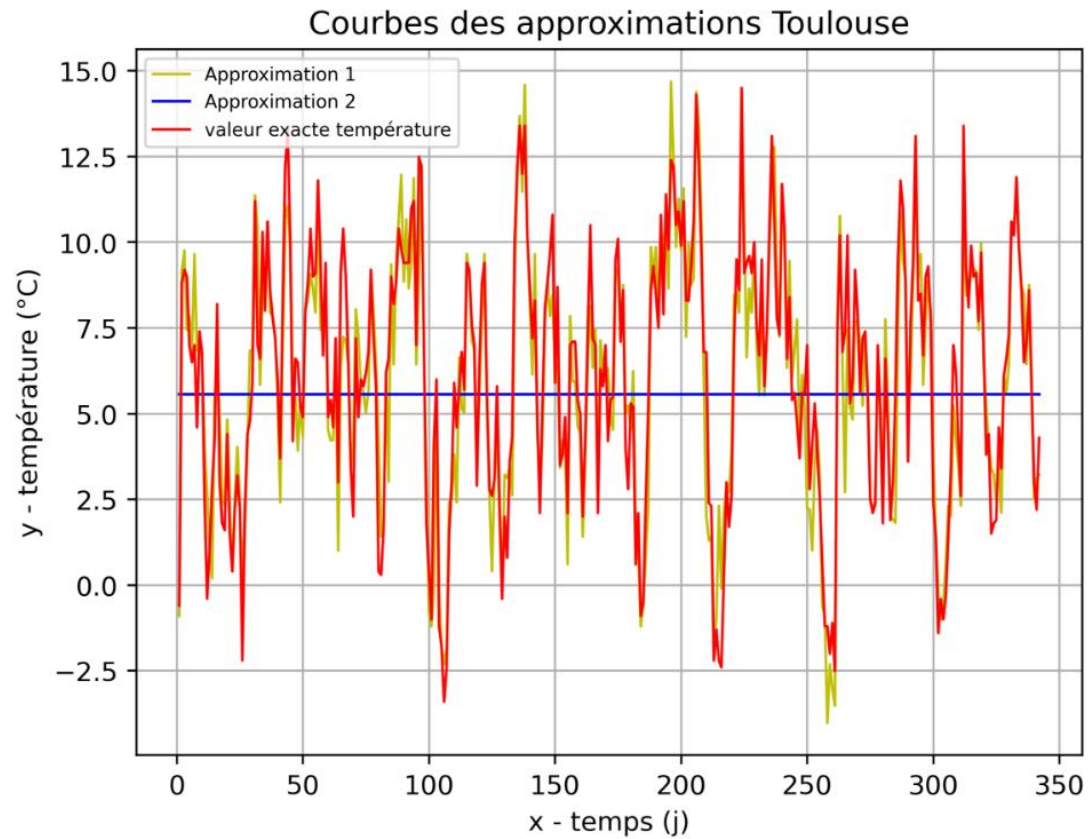
Y variable aléatoire  
températures ville  
cible



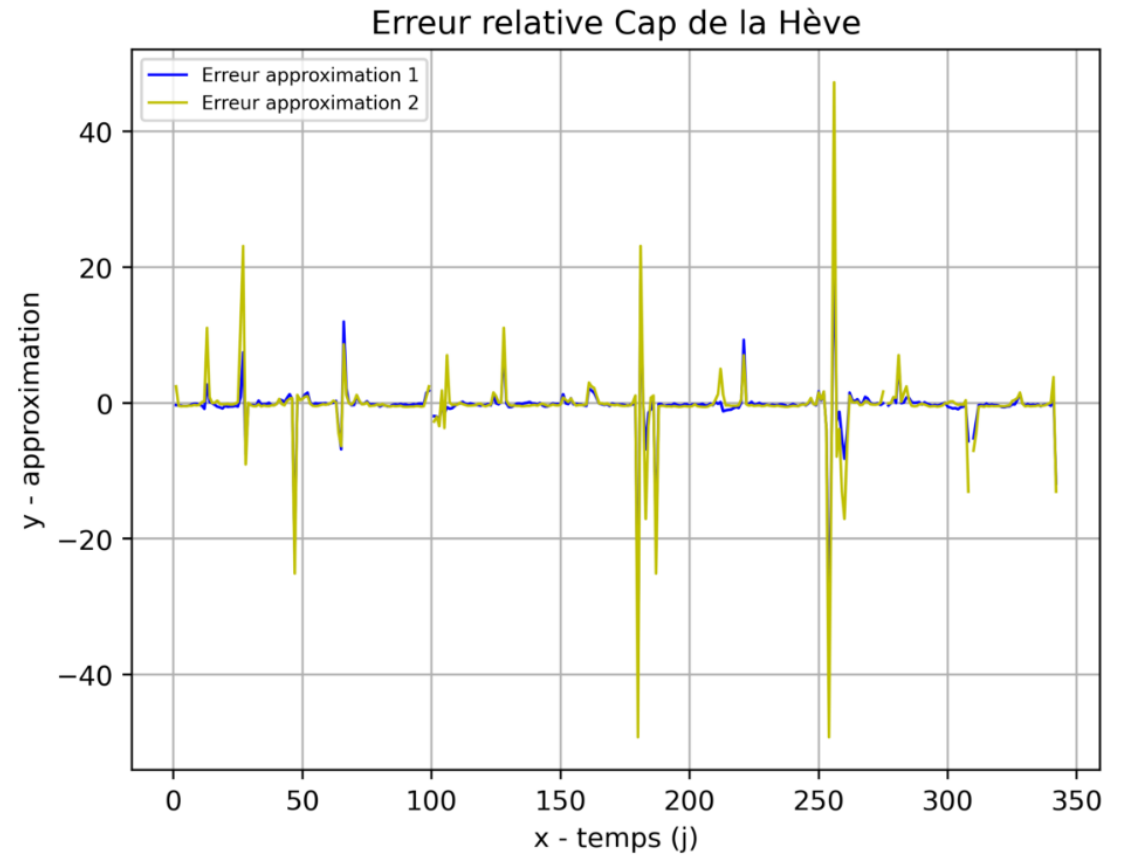
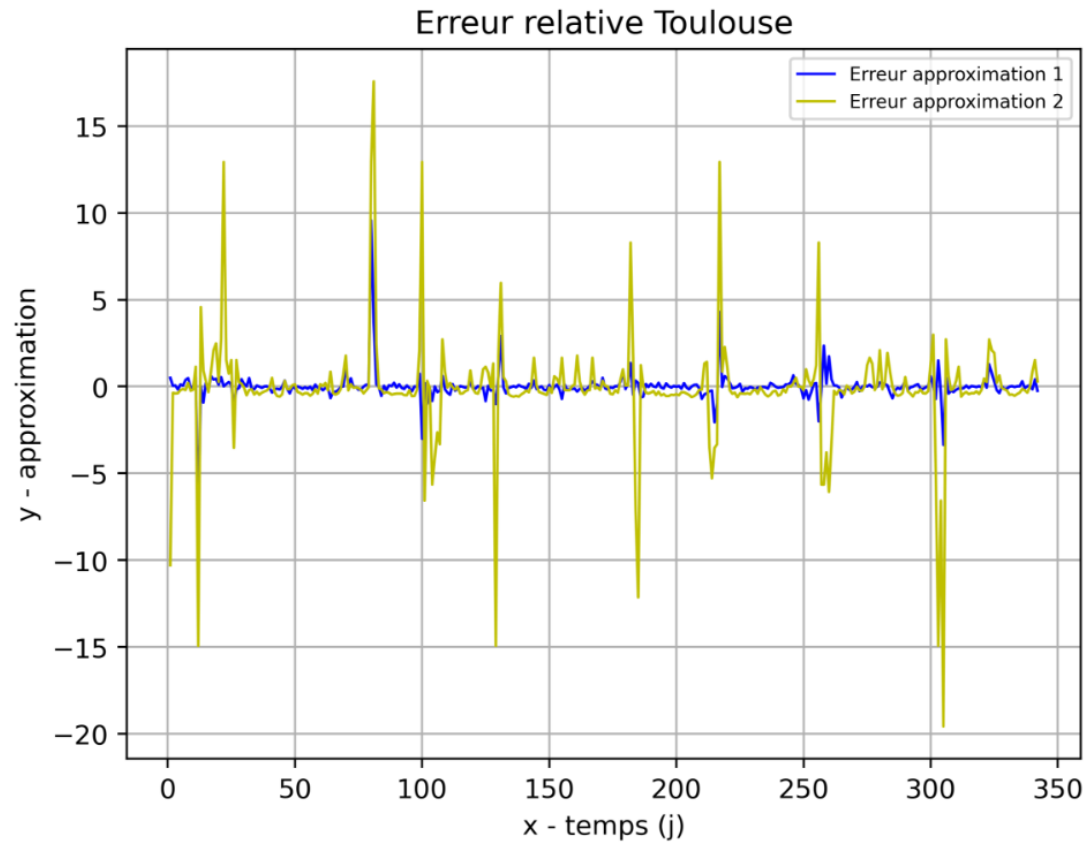
X ville utilisée

- $\bar{X}_d = \frac{1}{d} \sum_{i=1}^d X_i^p$
- $\bar{Y}_d = \frac{1}{d} \sum_{i=1}^d Y_i^p$
- $\bar{S}_d^2 = \frac{1}{d-1} \sum_{i=1}^d (X_i - \bar{X}_d)^2$
- $\frac{1}{d-1} \sum_{i=1}^d (X_i - \bar{X}_d) \times (Y_i - \bar{Y}_d)$

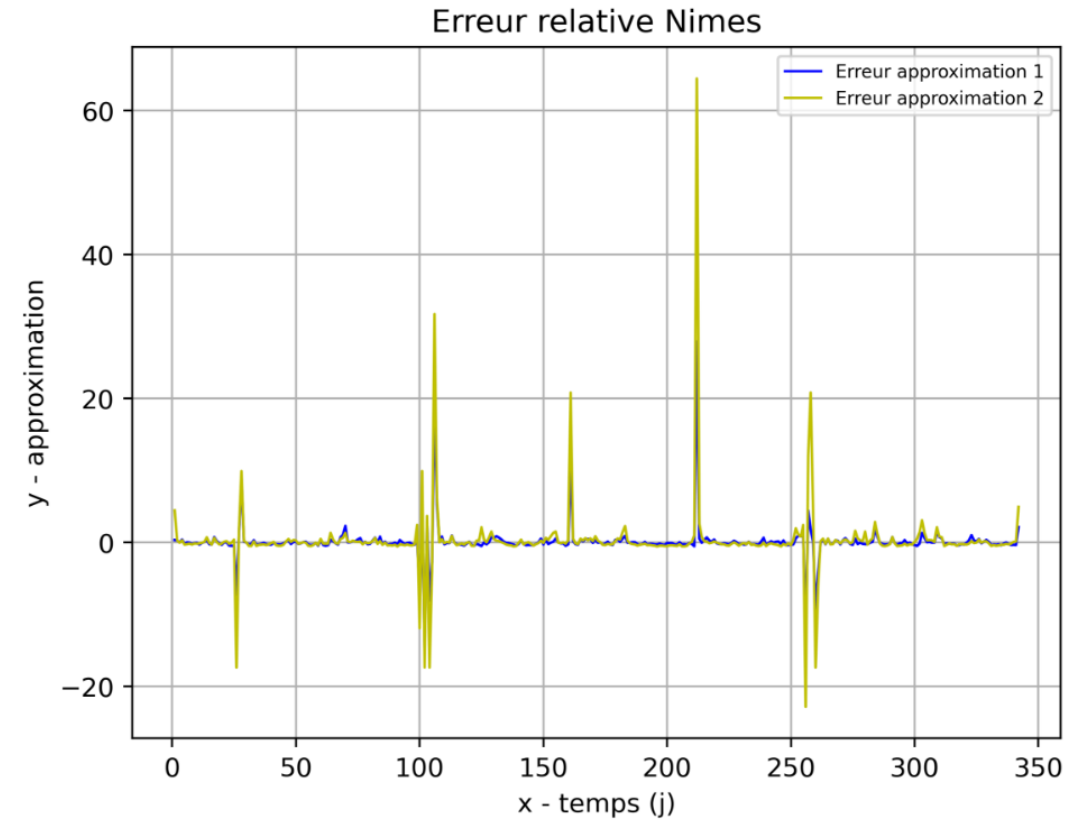
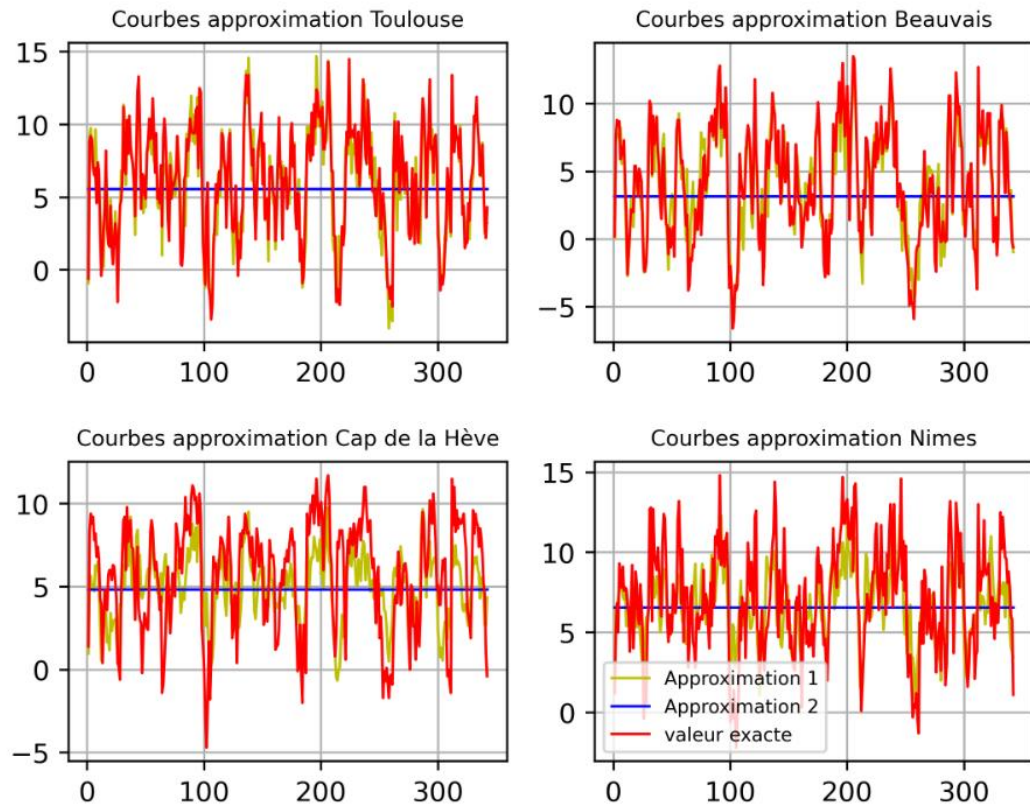
# Résultats obtenus : Toulouse



# Comparaison : Cap de la Hève



# Le Mont Aigoual



# Estimation d'une station grâce aux onze autres



Y variable aléatoire  
températures ville  
cible

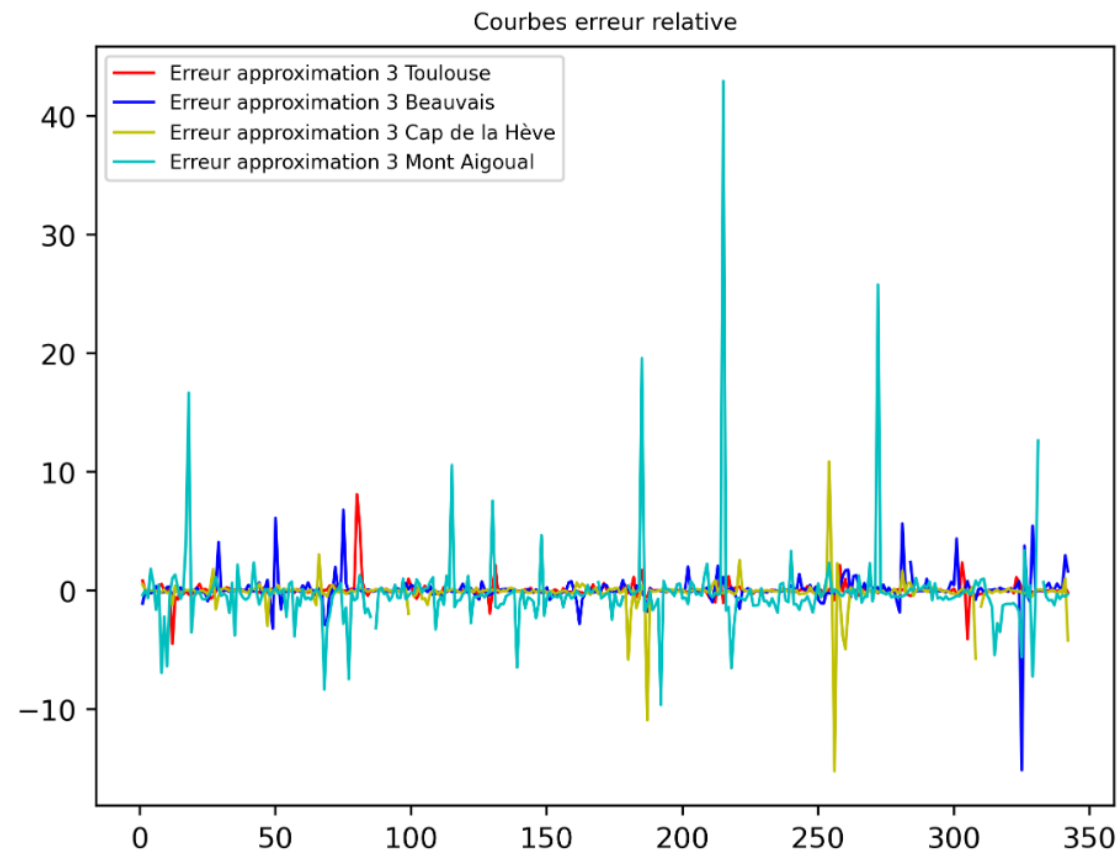
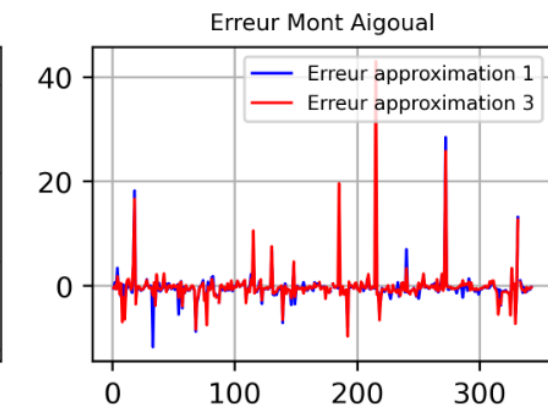
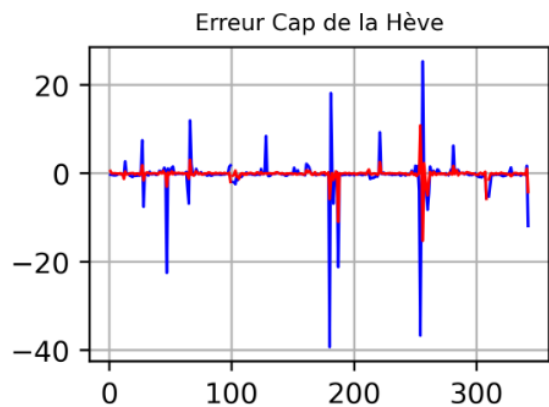
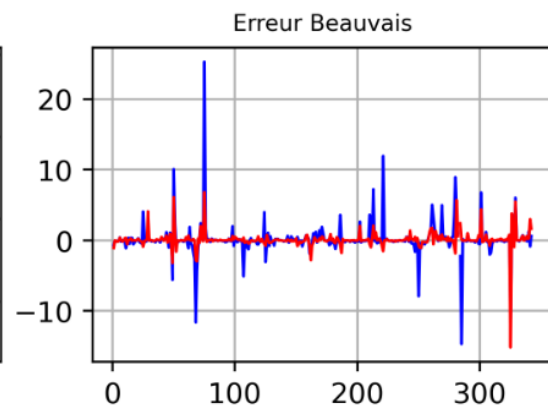
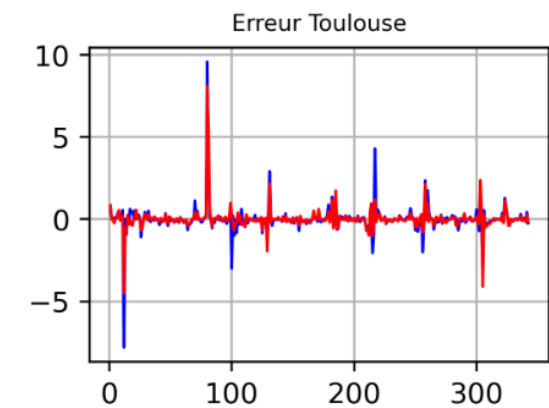


X vecteur températures  
des 11 autres stations

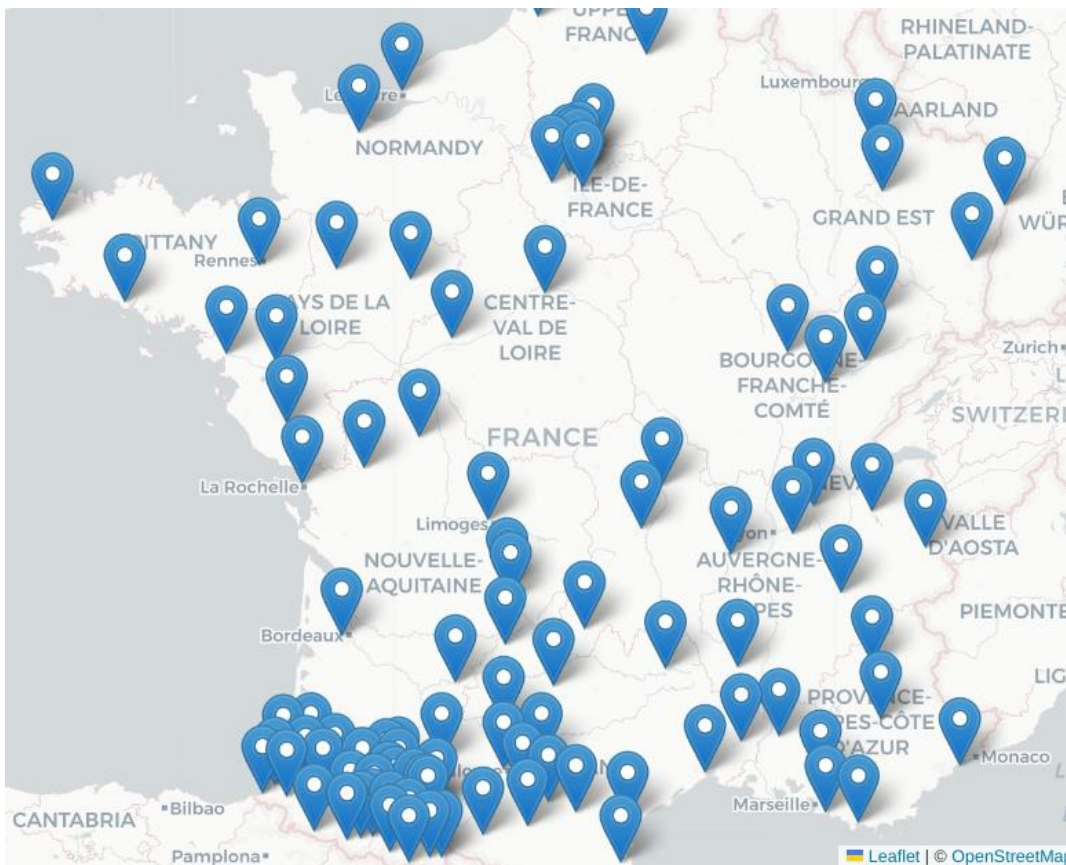
$$\mathbf{Y} = \mathbb{E}[Y] + \mathbf{C}^t \mathbf{K}^{-1} (\mathbf{X} - \mathbb{E}[\mathbf{X}]) , \text{ avec } \mathbf{K} = \text{Cov}[X_i, X_j]_{i,j} \text{ et } \mathbf{C} = \text{Cov}[X_i, Y]_i$$



# Résultats obtenus



# Deuxième modèle de prédiction



```
[5]: localisation
```

[5]:

	indic	lat	lon	alti	nom_usuel.commune
0	1089001	45.98	5.33	250	AMBERIEU
1	2320001	49.82	3.21	98	ST QUENTIN
2	3060001	46.17	3.40	249	VICHY-CHARMEIL
3	4070009	44.07	6.19	554	DIGNE LES BAINS
4	5061009	44.58	6.08	885	GAP
...	...	...	...	...	...
99	86027001	46.59	0.31	125	POITIERS-BIARD
100	87085006	45.86	1.18	402	LIMOGES-BELLEGARDE
101	91027002	48.72	2.40	86	ORLY
102	92040002	48.83	2.28	32	ISSY-HELIPORT
103	95527001	49.02	2.53	108	ROISSY

104 rows × 5 columns

```
[5]: date
```

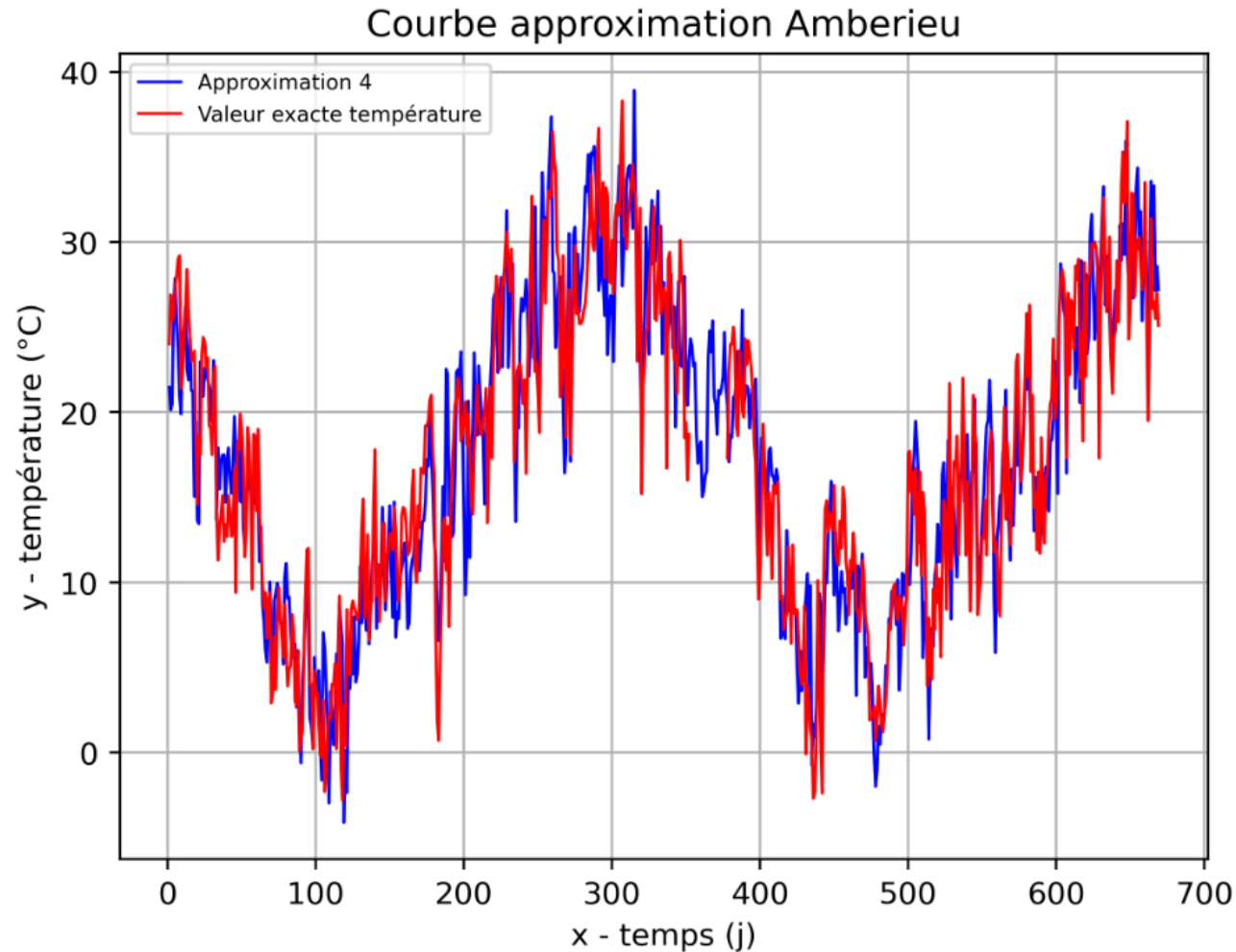
[5]:

**dates**

0	2021-08-01
1	2021-08-02
2	2021-08-03
3	2021-08-04
4	2021-08-05
...	...
694	2023-07-27
695	2023-07-28
696	2023-07-29
697	2023-07-30
698	2023-07-31

699 rows x 1 columns

# Test sur Amberieu



```
[3]: temp
```

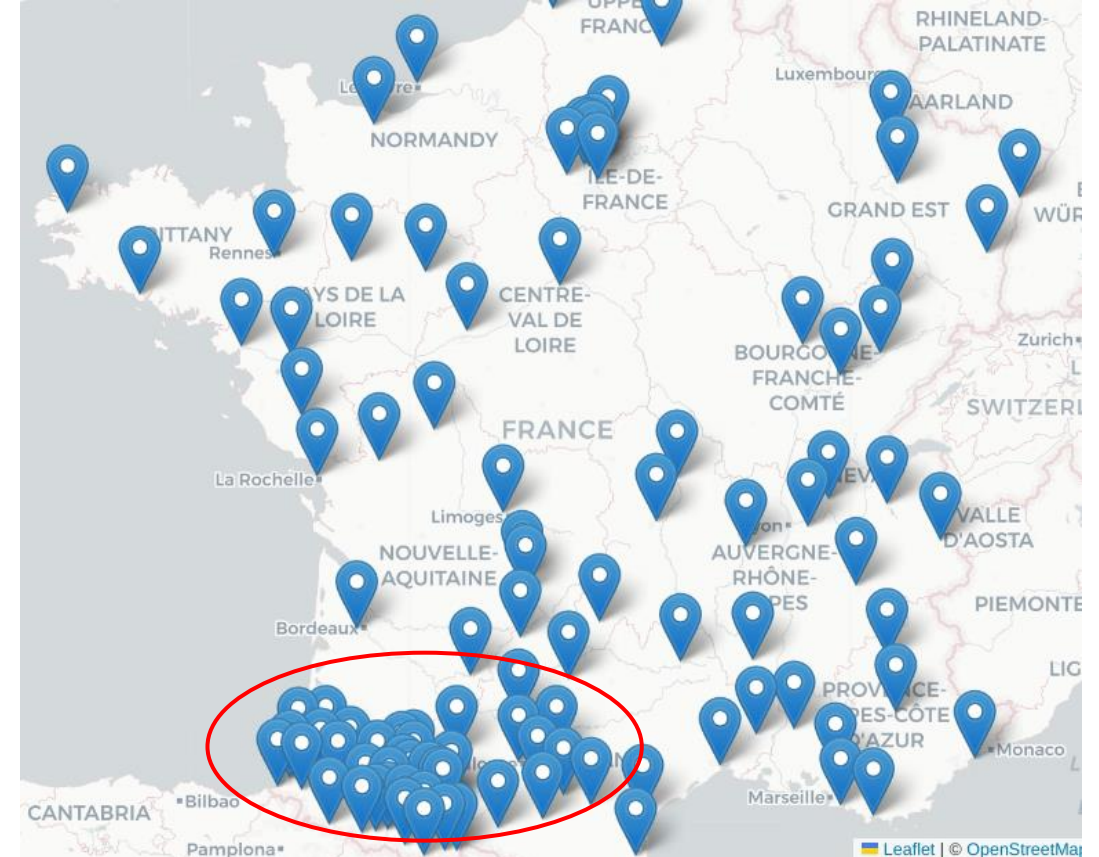
```
[3]: xarray.DataArray (dim_0: 104, dim_1: 699, dim_2: 1)
```

```
array([[22.60000038],  
       [15.89999962],  
       [16.79999924],  
       ...,  
       [25.5         ],  
       [27.         ],  
       [25.10000038]],  
  
       [[19.10000038],  
       [18.60000038],  
       [20.20000076],  
       ...,  
       [20.70000076],  
       [17.60000038],  
       [21.5         ]],  
  
       [[22.20000076],  
       [16.79999924],  
       [20.79999924],  
       ...,  
       ...,  
       ...,  
       ...,  
       [23.60000038],  
       [21.70000076],  
       [22.60000038]],
```

# Clustering : algorithme des k-means



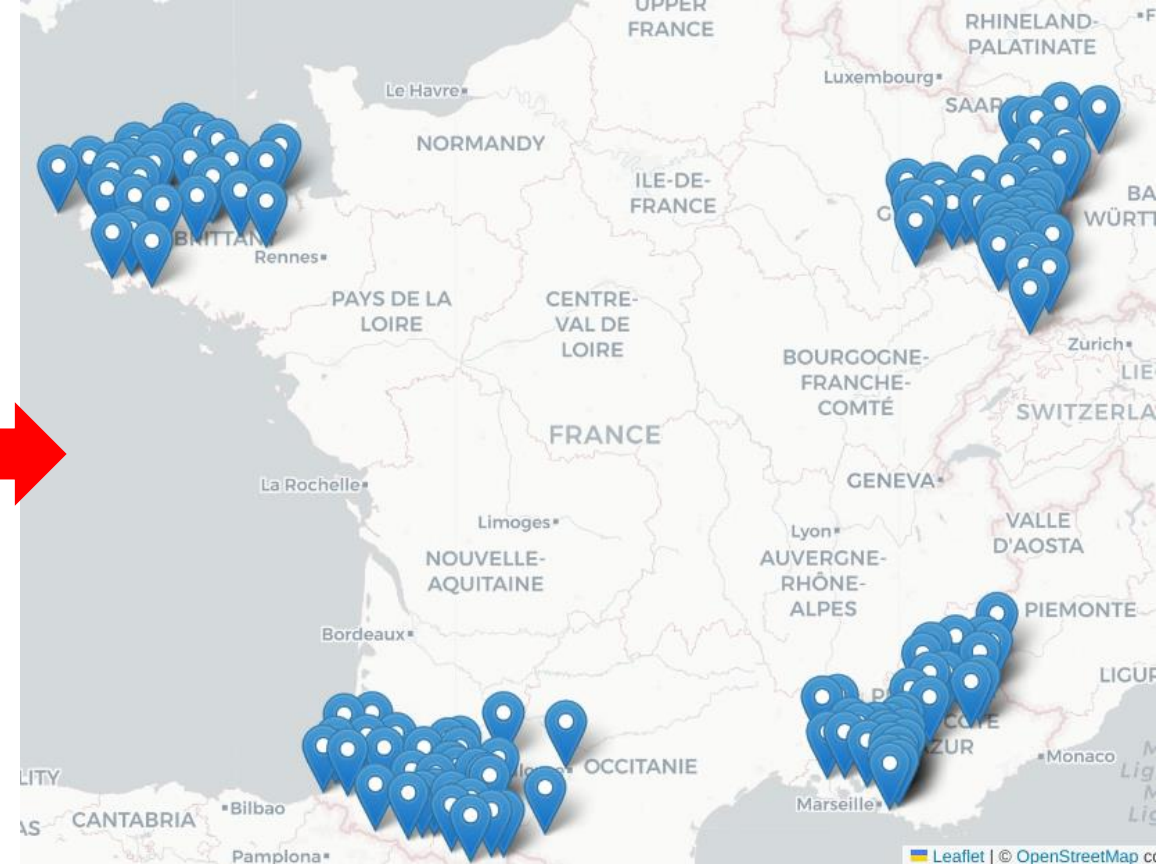
3ème jeu de données



2ème jeu de données

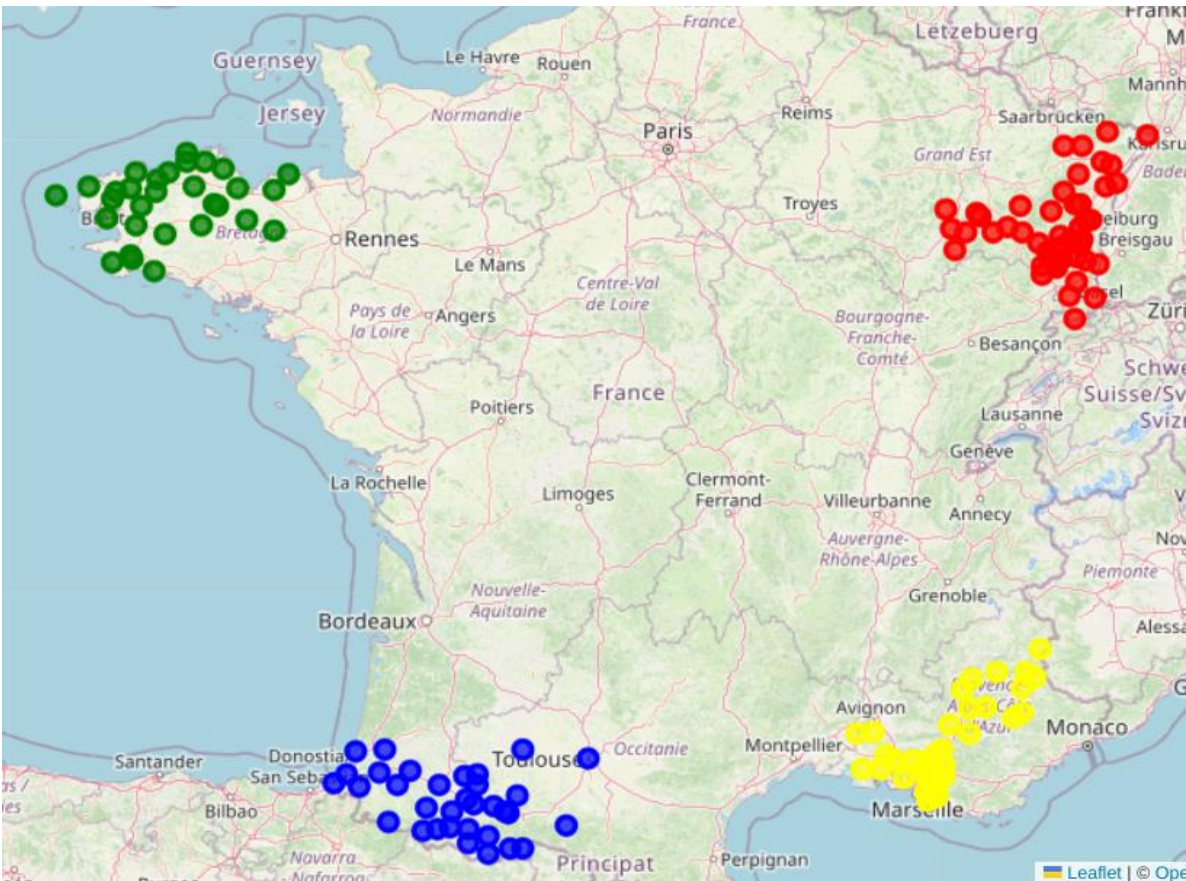


# Préparation du jeu de données

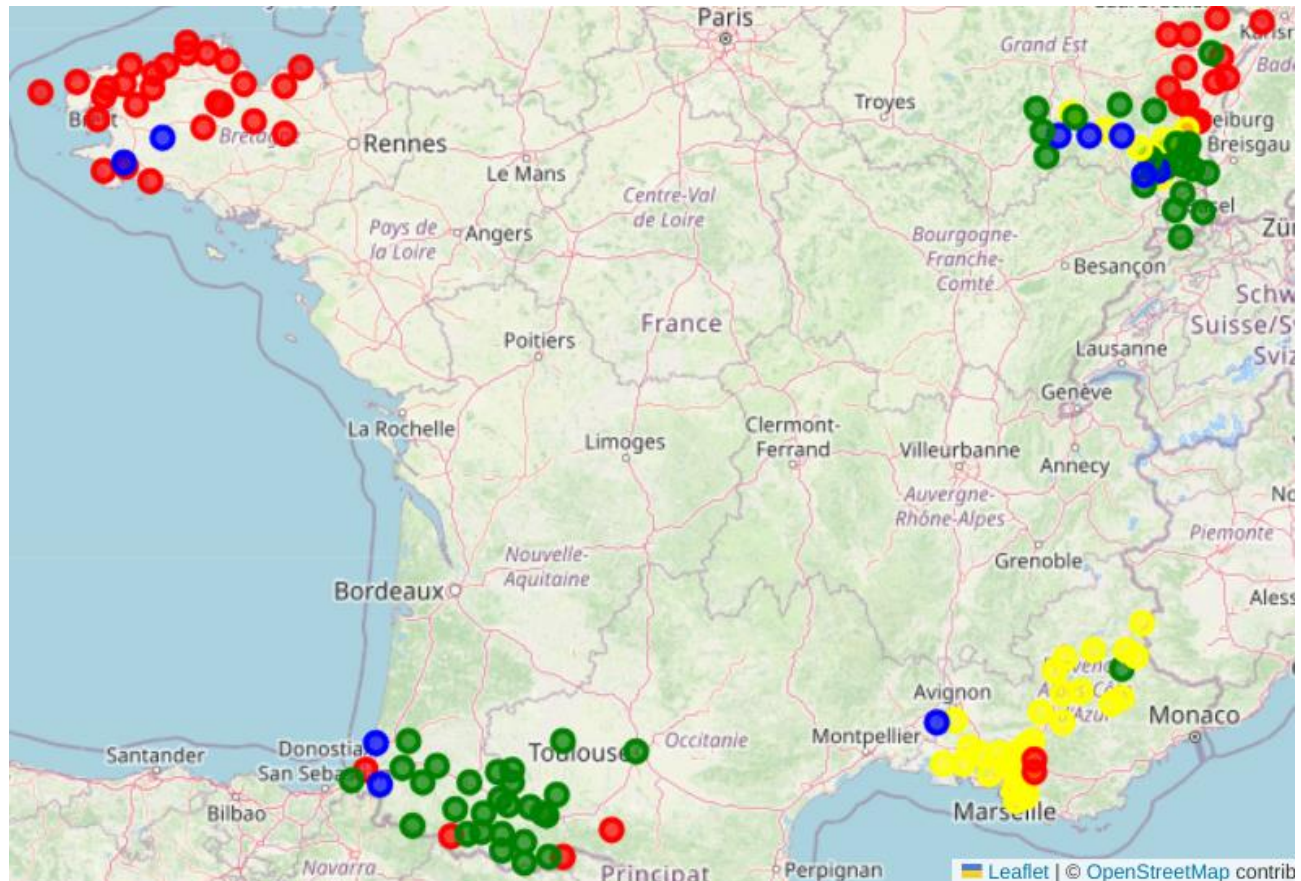


3ème jeu de données

# Résultats obtenus



## K-means sur la position



## K-means sur la température

# Conclusion

- Construction d'approximation en utilisant la méthode des moments, ou l'espérance conditionnelle
- Clustering : méthode des k-means
- Compétences acquises : traitement et analyse des données