

# Rapport de Projet : Analyse en Composantes Principales (ACP) et Régression Linéaire

Membres de l'équipe : Yuefan LIU, Mouzheng LI, Lianghong LI

Fichier de réponses : YuefanLIU\_MouzhengLI\_LianghongLI.csv

Langage utilisé : Python

**[q1a, q1b] De combien de villes proviennent les données météorologiques ? Combien de villes sont concernées par des mesures manquantes ?**

[q1a] Réponse : 25 villes.

[q1b] Réponse : 0 villes ont des valeurs manquantes et ont été supprimées de l'analyse.

**[q2a–q2h] Quelle est la ville associée à la valeur minimale et maximale de chaque variable météorologique ?  
(Températures, précipitations, ensoleillement)**

Numéro	Variable	Ville	Valeur
[q2a]	Température minimale minimale	Beauvais	7.7 °C
[q2b]	Température minimale maximale	Nice	14.2 °C
[q2c]	Température maximale minimale	Brest	15.5 °C
[q2d]	Température maximale maximale	Ajaccio	22.5 °C
[q2e]	Précipitations minimales	Ajaccio	515.8 mm
[q2f]	Précipitations maximales	Biarritz	1819.4 mm
[q2g]	Ensoleillement minimal	Brest	1363.2 h
[q2h]	Ensoleillement maximal	Ajaccio	2854.5 h

**[q2a] Température minimale minimale**

Réponse : Beauvais, 7.7°C

Commentaire : => Température très basse, climat plus continental avec nuits fraîches.

**[q2b] Température minimale maximale**

Réponse : Nice, 14.2°C

Commentaire : => Climat méditerranéen, hiver doux et faible amplitude thermique.

**[q2c] Température maximale minimale**

Réponse : Brest, 15.5°C

Commentaire : => Ciel souvent nuageux, climat océanique tempéré et humide.

**[q2d] Température maximale maximale**

Réponse : Ajaccio, 22.5°C

Commentaire : => Influence maritime et ensoleillement important.

**[q2e] Précipitations minimales**

Réponse : Ajaccio, 515.8 mm

Commentaire : => Climat sec et méditerranéen, précipitations faibles.

**[q2f] Précipitations maximales**

Réponse : Biarritz, 1819.4 mm

Commentaire : => Région très pluvieuse, influence atlantique et vents humides.

**[q2g] Ensoleillement minimal**

Réponse : Brest, 1363.2 h

Commentaire : => Zone nuageuse, influence océanique, faible luminosité.

**[q2h] Ensoleillement maximal**

Réponse : Ajaccio, 2854.5 h

Commentaire : => Région très ensoleillée, climat méditerranéen sec.

## **[q3a–q3d] Calculez la variance des variables météorologiques**

**[q3a] la variance de Température minimale :** 3.253599999999999

**[q3b] la variance de Température maximale :** 4.315833333333331

**[q3c] la variance de Hauteur des précipitations :** 82209.53310000002

**[q3d] la variance de Durée d'ensoleillement :** 198650.1257666667

**[q3a] Variance de la température minimale**

Réponse : 3.25

Commentaire : => Faible dispersion, températures nocturnes stables.

**[q3b] Variance de la température maximale**

Réponse : 4.32

Commentaire : => Légère fluctuation, stabilité globale du climat diurne.

**[q3c] Variance de la hauteur de précipitations**

Réponse : 82209.53

Commentaire : => Variabilité extrême, zones très humides et très sèches.

### [q3d] Variance de la durée d'ensoleillement

Réponse : 198650.13

Commentaire : => Très grande différence d'ensoleillement selon les villes.

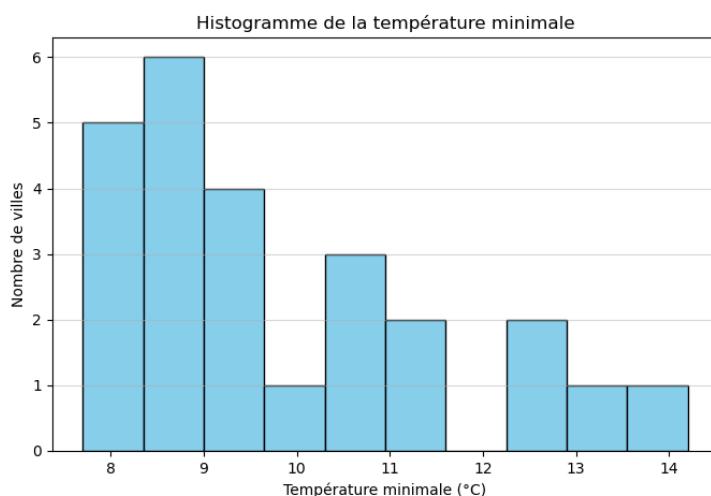
### [q4a–q4c] Moyenne, médiane et écart-type de la variable de variance minimale (température minimale)

[q4a] Moyenne : 9.788°C

[q4b] Médiane : 9.2°C

[q4c] Écart-type : 1.8037738217415173°C

*l'histogramme de la température minimale :*



Commentaire :

=> Distribution asymétrique à gauche, forte concentration autour de 8-10°C, quelques villes avec températures minimales plus élevées.

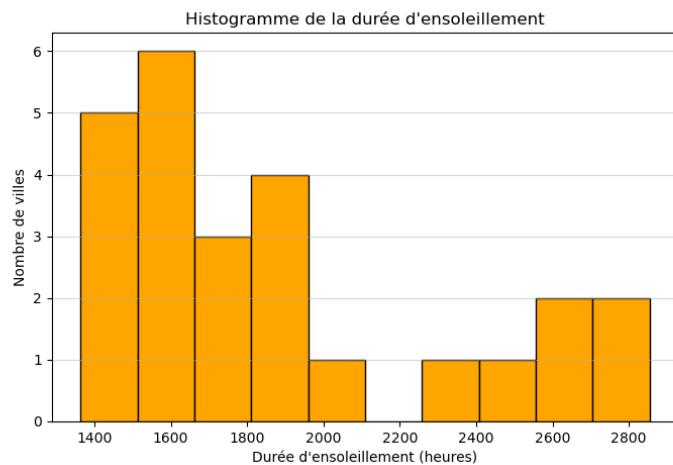
### [q5a–q5c] Moyenne, médiane et écart-type de la variable de variance maximale (durée d'ensoleillement)

[q5a] Moyenne : 1881.108 h

[q5b] Médiane : 1694.0 h

[q5c] Écart-type : 445.7018350496963 h

*l'histogramme de la durée d'ensoleillement :*



**Commentaire :**

=> Distribution bimodale, forte majorité autour de 1400-1900h, quelques villes très ensoleillées (2600-2800h).

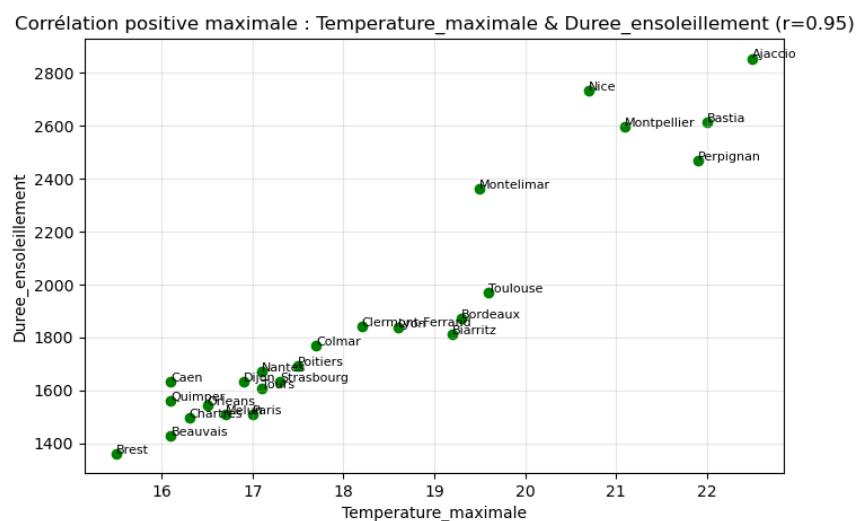
### [q6a-q6c] Quelles sont les deux variables les plus corrélées (positivement, négativement et faiblement) ?

**[q6a]** Les plus positivement corrélées : Température minimale & Température maximale ( $r = 0.9490537451909286$ )

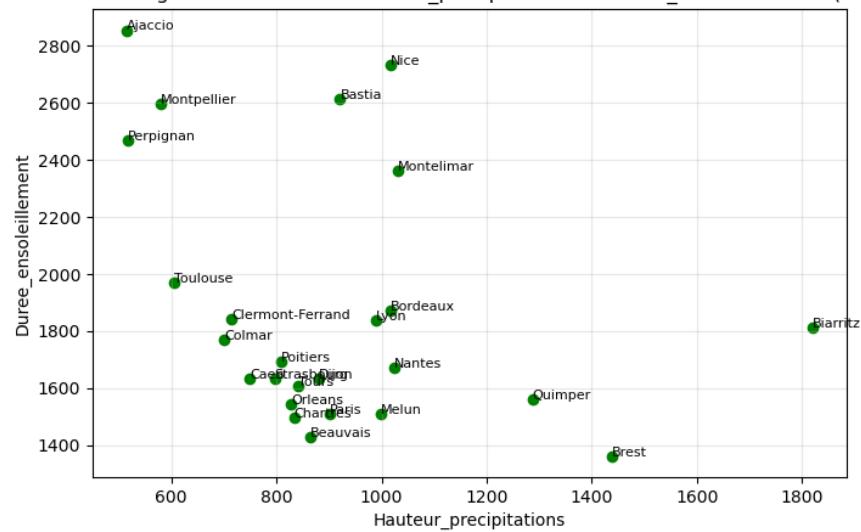
**[q6b]** Les plus négativement corrélées : Température minimale & Hauteur des précipitations ( $r = -0.3298125874294128$ )

**[q6c]** Les moins corrélées : Hauteur des précipitations & Ensoleillement ( $r = -0.005084415280884977$ )

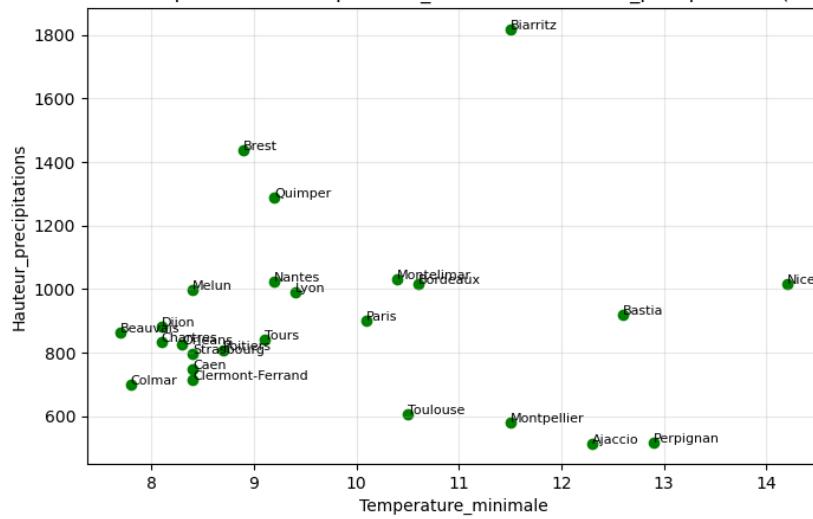
*les trois nuages de points :*



Corrélation négative maximale : Hauteur\_precipitations & Duree\_ensoleillement ( $r=-0.33$ )



Corrélation la plus faible : Temperature\_minimale & Hauteur\_precipitations ( $r=-0.01$ )



=> Corrélation positive forte (Température\_maximale & Durée\_ensoleillement) :

Les villes avec plus de soleil tendent à avoir des températures maximales plus élevées ; tendance linéaire nette et forte dispersion.

=> Corrélation négative modérée (Hauteur\_précipitations &

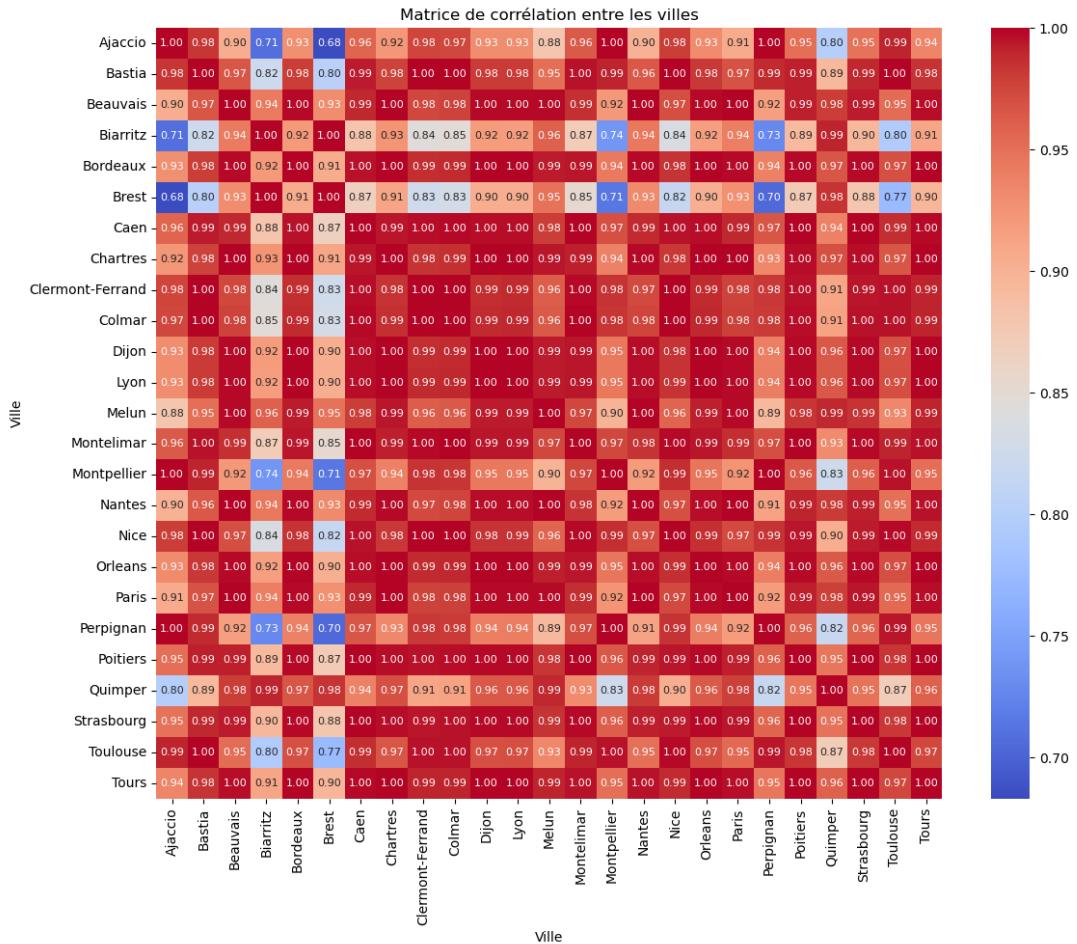
Durée\_ensoleillement) : Les villes très ensoleillées reçoivent en général moins de précipitations, mais la relation n'est pas parfaitement linéaire.

=> Corrélation quasi nulle (Température\_minimale & Hauteur\_précipitations) :

Aucune tendance claire entre ces variables ; points dispersés sans alignement.

## [q7] Matrice de corrélation entre les villes

*la heatmap de corrélation :*



### Commentaire :

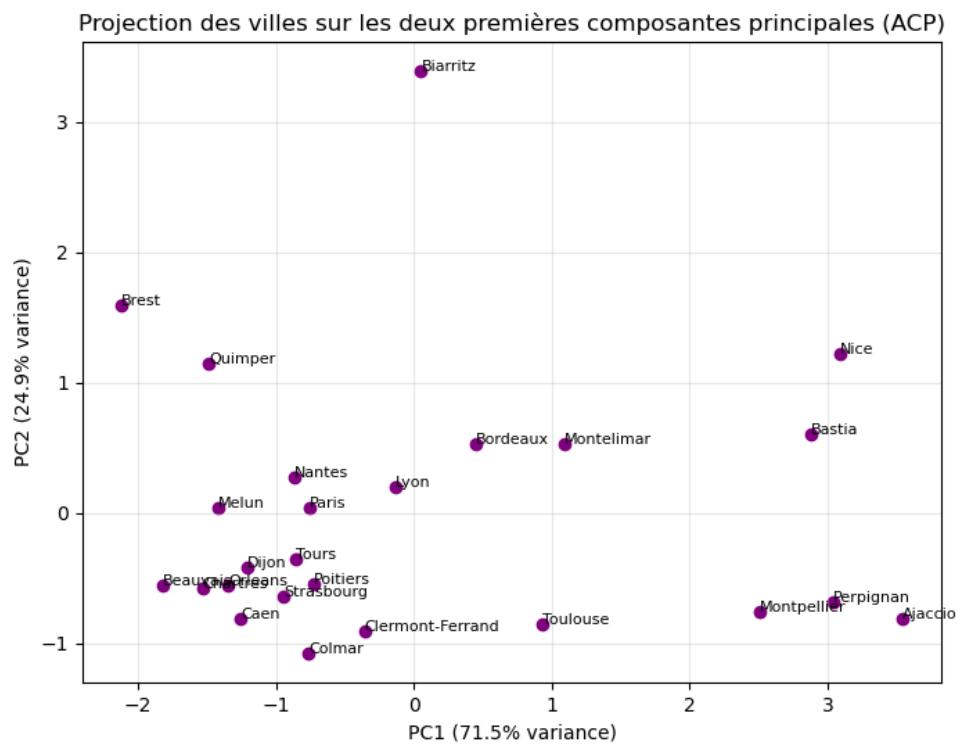
=> Corrélations globalement très fortes (>0.95) entre les villes, climat homogène sur la majorité du territoire ; quelques exceptions (ex. Biarritz, Brest) montrent des variations locales plus marquées.

## [q8a, q8b] Quel est le pourcentage de variance expliquée par les deux premières composantes principales ?

[q8a] PC1 : 71.54%

[q8b] PC2 : 24.93%

le nuage de points ACP des villes :

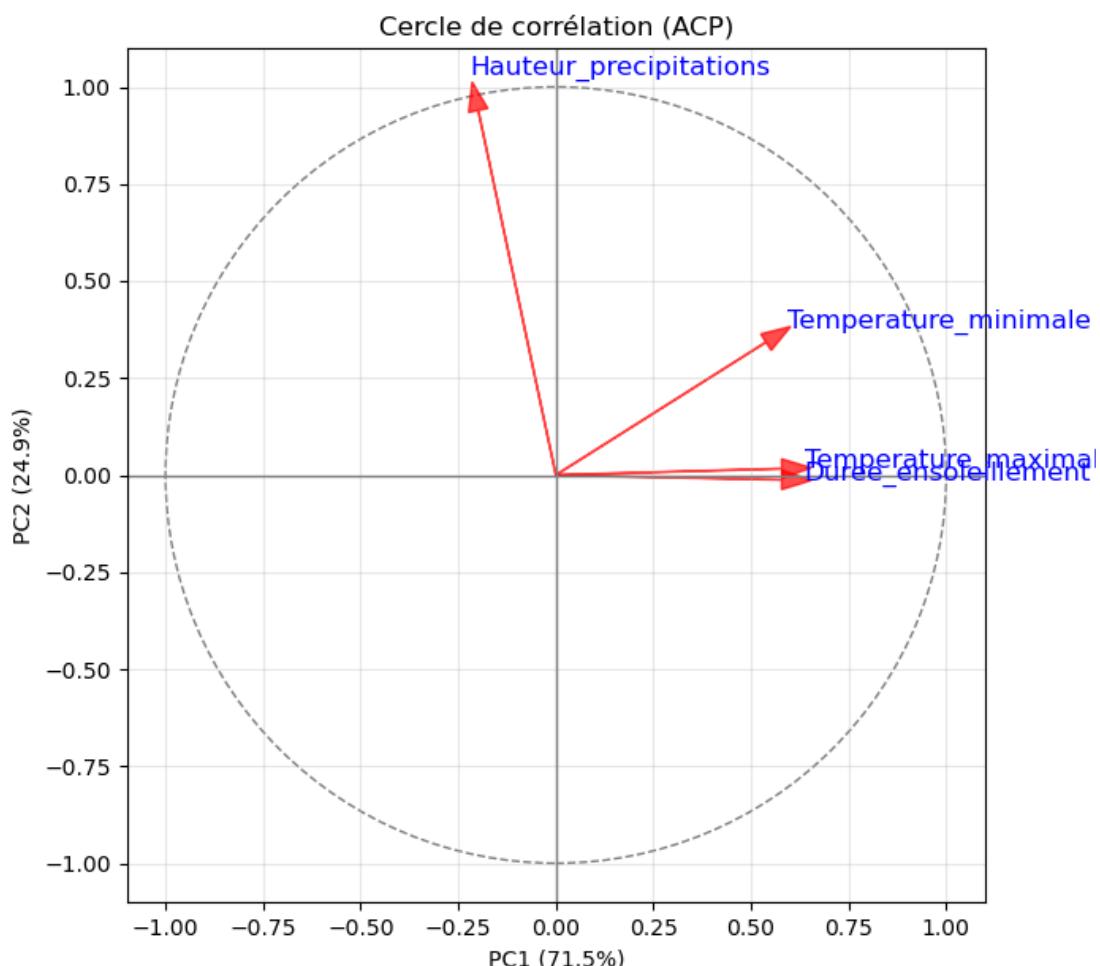


### Commentaire :

=> PC1 (71.5%) sépare les villes selon l'ensoleillement et la température maximale ;  
PC2 (24.9%) distingue les variations de précipitations et de température minimale ;  
dispersion modérée, villes bien regroupées et quelques exceptions.

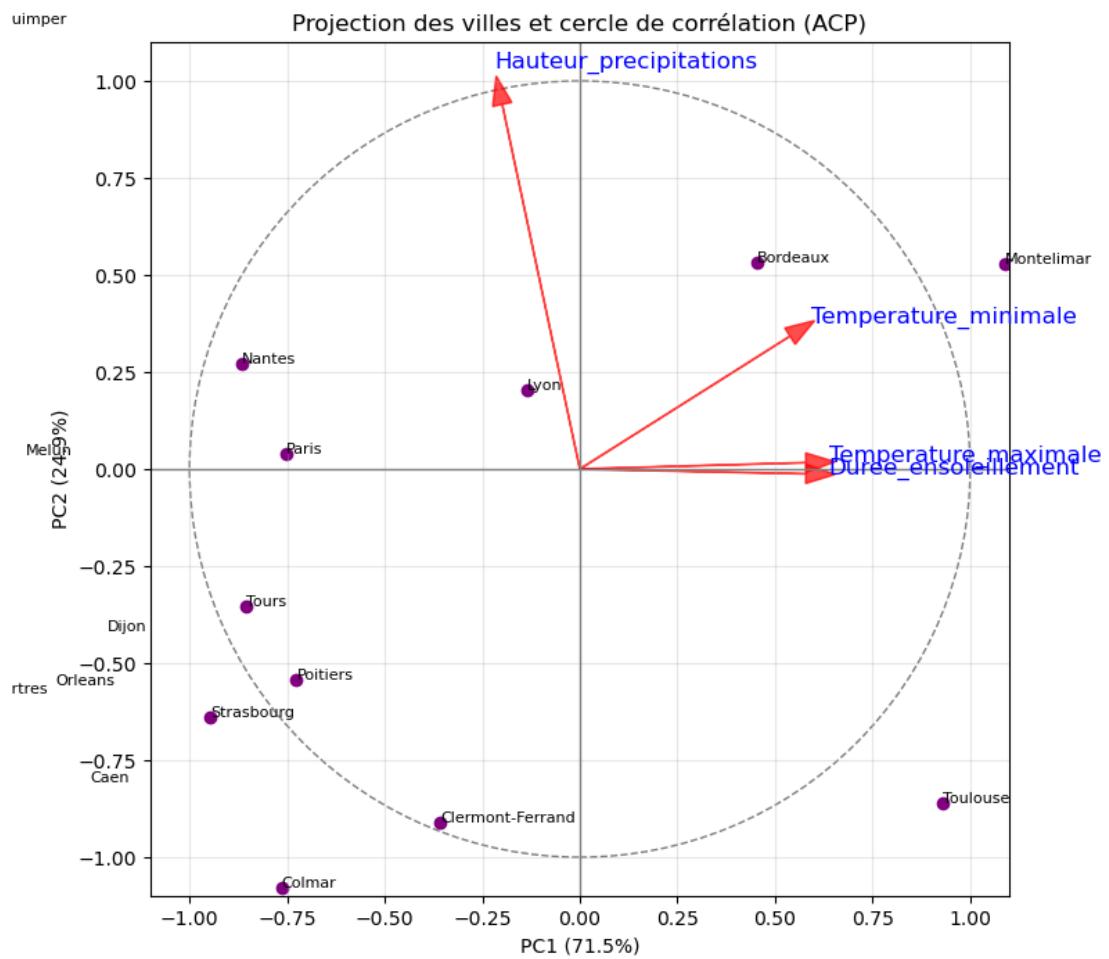
## [q9] Affichez et commentez le cercle de corrélation

le cercle de corrélation :



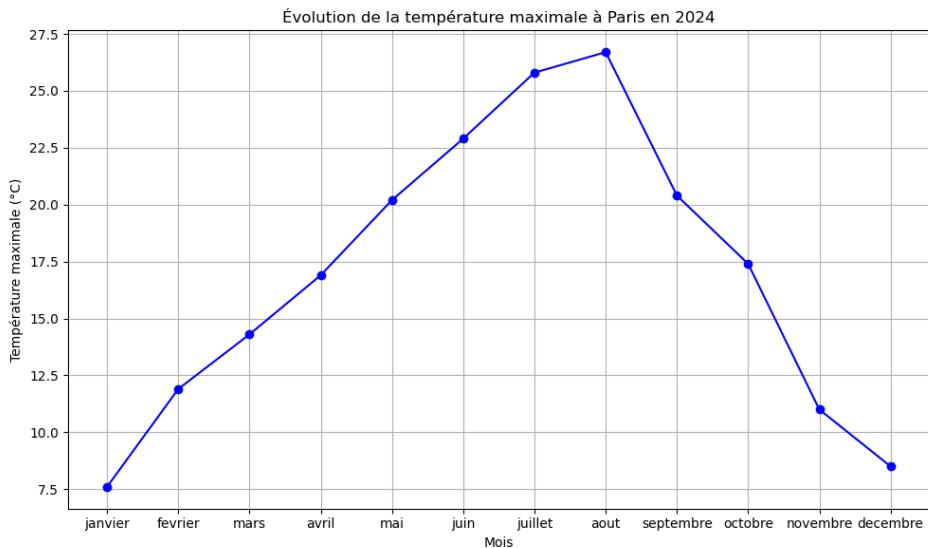
## [q10] Superposez le cercle de corrélation et la projection des villes

la figure combinée ACP + cercle :



## [q11] Évolution de la température maximale à Paris en 2024

*la courbe température 2024 :*



**Commentaire :**

=> Hausse progressive de janvier à août (pic estival), chute rapide dès août, tendance saisonnière nette.

## [q12a–q12e] Quel est le meilleur modèle de régression linéaire sur les mois de 2024 ?

[q12a] Valeur optimale de n : 5

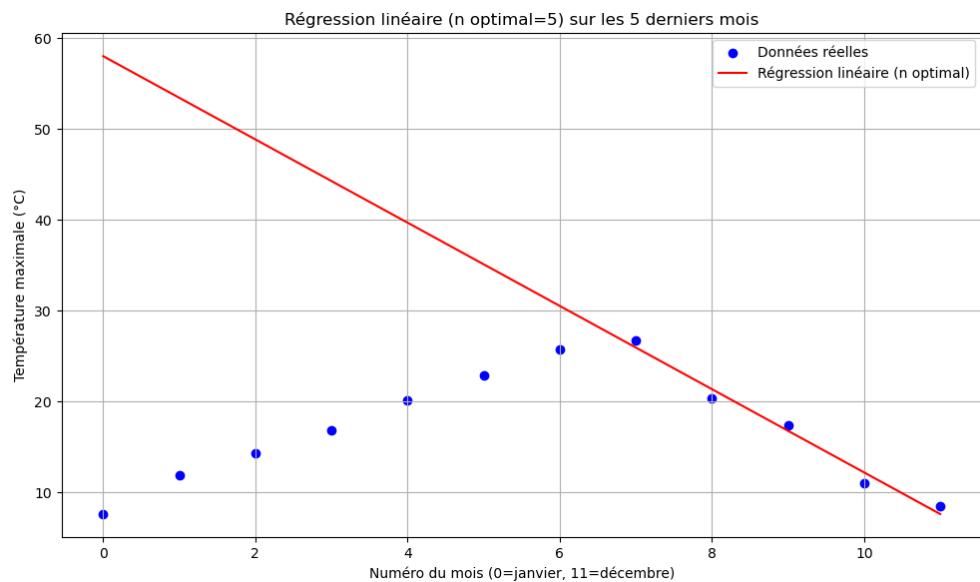
[q12b] R<sup>2</sup> ajusté optimal : 0.9745

[q12c] R<sup>2</sup> optimal : 0.9808

[q12d] Coefficient  $\beta_0$  : 58.0200

[q12e] Coefficient  $\beta_1$  : -4.5800

*la régression linéaire :*



[q13a, q13b] Quelle est la température prédictive pour janvier 2025 ? Quel est l'écart avec la température réelle ?

[q13a] Température maximale prédictive pour janvier 2025 : 3.06°C

[q13b] Écart entre la température prédictive et réelle : -4.44°C

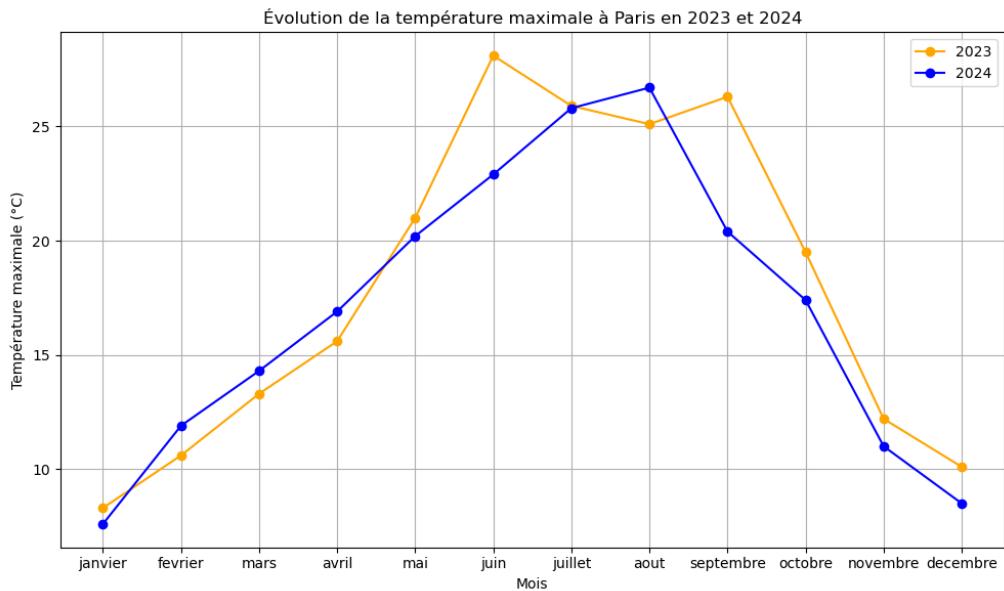
[q14a, q14b] Test d'hypothèse sur la pente  $\beta_1$  : p-valeur et conclusion ?

[q14a] p-value pour le test de  $\beta_1$  : 0.001131

[q14b] Conclusion ( $\alpha=5\%$ ) : Oui, la pente est significative : il existe une relation linéaire.

[q15] Évolution des températures à Paris en 2023 et 2024

la double courbe Paris 2023–2024 :



#### Commentaire :

=> Tendances similaires pour les deux années, été 2023 plus chaud avec pics marqués, 2024 plus stable et moins extrême.

### [q16a, q16b] Combien de combinaisons possibles et quel est le meilleur modèle multivarié ?

[q16a] Nombre total de combinaisons : 4095

[q16b] Meilleur modèle :

- Variables sélectionnées : ['Temp\_1', 'Temp\_2', 'Temp\_5', 'Temp\_7', 'Temp\_8', 'Temp\_9', 'Temp\_10', 'Temp\_11', 'Temp\_12']
- Nombre de variables : 9
- R<sup>2</sup> ajusté : 0.99967
- Significativité : Oui, toutes les variables présentent une significativité (p < 0,05), indiquant l'existence d'une relation linéaire significative ( $\alpha=5\%$ ).

#### Commentaire :

=> Modèle optimal très performant (R<sup>2</sup> ajusté proche de 1), forte dépendance linéaire des températures mensuelles ; relation linéaire confirmée pour toutes les variables à  $\alpha=5\%$ .

### [q17] Écart entre les prédictions pour janvier–avril 2025 et les valeurs réelles

Mois	[q17] Température prédictive	Température réelle	Écart
Janvier	7.95°C	7.5°C	0.45°C

Mois	[q17] Température prédictive	Température réelle	Écart
Février	8.65°C	8.6°C	0.05°C
Mars	11.96°C	14.6°C	-2.64°C
Avril	18.91°C	20.0°C	-1.09°C

**Commentaire :**

=> Bonne précision pour janvier et février (écart faible), prédictions moins fiables à partir de mars, modèle à affiner pour la saison printanière.