

NOM : ESSOU

PRÉNOMS : Jean-Jéliel Yao

PROJET : Optimisation des tournées d'entretien des arbres de Paris



Contexte : La ville de Paris souhaite optimiser l'entretien de ses arbres en améliorant les tournées de ses équipes. Une meilleure organisation permettrait de réduire les coûts et l'impact environnemental, tout en assurant un entretien régulier et efficace de ce patrimoine végétal précieux.

Objectif : Vous êtes chargé d'analyser les données des arbres de Paris pour identifier des pistes d'optimisation des tournées d'entretien. Vous devrez explorer les données, réaliser des visualisations pertinentes, et formuler des recommandations concrètes basées sur vos analyses.

1. Chargement et exploration des données :

Présentation des données :

Le jeu de données fourni contient des informations sur les arbres de la ville de Paris.

Data columns (total 18 columns):

#	Column	Non-Null Count	Dtype
0	id	200137 non-null	int64
1	type_emplacement	200137 non-null	object
2	domanialite	200136 non-null	object
3	arrondissement	200137 non-null	object
4	complement_adresse	30902 non-null	object
5	numero	0 non-null	float64
6	lieu	200137 non-null	object
7	id_emplacement	200137 non-null	object
8	libelle_francais	198640 non-null	object
9	genre	200121 non-null	object
10	espece	198385 non-null	object
11	variete	36777 non-null	object
12	circonference_cm	200137 non-null	int64
13	hauteur_m	200137 non-null	int64
14	stade_developpement	132932 non-null	object
15	remarquable	137039 non-null	float64
16	geo_point_2d_a	200137 non-null	float64
17	geo_point_2d_b	200137 non-null	float64

Résumé des observations

1. Colonnes importantes pour l'analyse :

- type_emplacement, arrondissement, espece, circonference_cm, hauteur_m, stade_developpement, remarquable, geo_point_2d_a, geo_point_2d_b.

2. Colonnes à nettoyer ou compléter :

- libelle_francais, genre, espece : Compléter les valeurs manquantes si possible.
- stade_developpement et remarquable : Prendre en compte les données manquantes pour éviter des biais.

3. Colonnes à ignorer ou supprimer :

- numero (vide).
- complement_adresse et variete : Utiles uniquement pour des analyses très détaillées.

2. Analyse descriptive univariée :

	circonference_cm	hauteur_m
count	197866.00000	197866.000000
mean	80.25867	8.356706
std	64.54718	6.192783
min	0.00000	0.000000
25%	30.00000	5.000000
50%	70.00000	8.000000
75%	115.00000	12.000000
max	1220.00000	42.000000

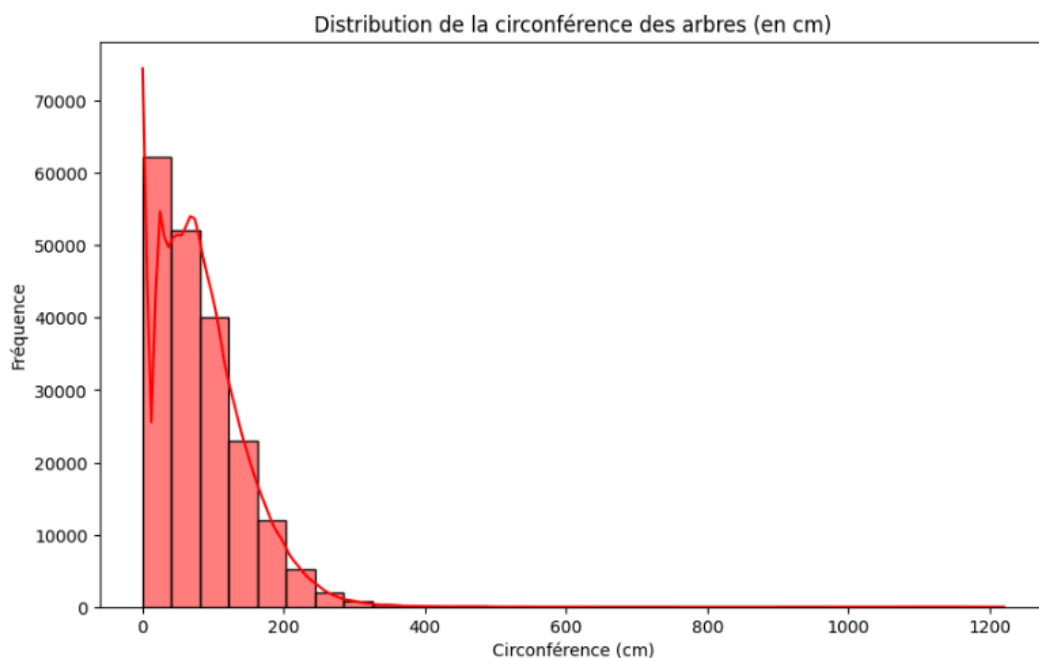
Interprétation :

Données aberrantes : Les valeurs minimales (0 pour `circonference_cm` et `hauteur_m`) sont clairement incorrectes. Ces enregistrements doivent être exclus ou corrigés.

Dispersion des tailles : Les arbres parisiens présentent une grande diversité, comme en témoignent les écart-types élevés pour les deux variables.

Extrêmes élevés : La circonférence maximale (1 220 cm) et la hauteur maximale (42 m) doivent être vérifiées, car elles peuvent refléter des cas uniques ou des erreurs.

Quartiles informatifs : Les quartiles montrent que la majorité des arbres sont de taille moyenne à petite, ce qui est attendu dans un contexte urbain.



Interprétation :

Majorité des arbres de petite taille :

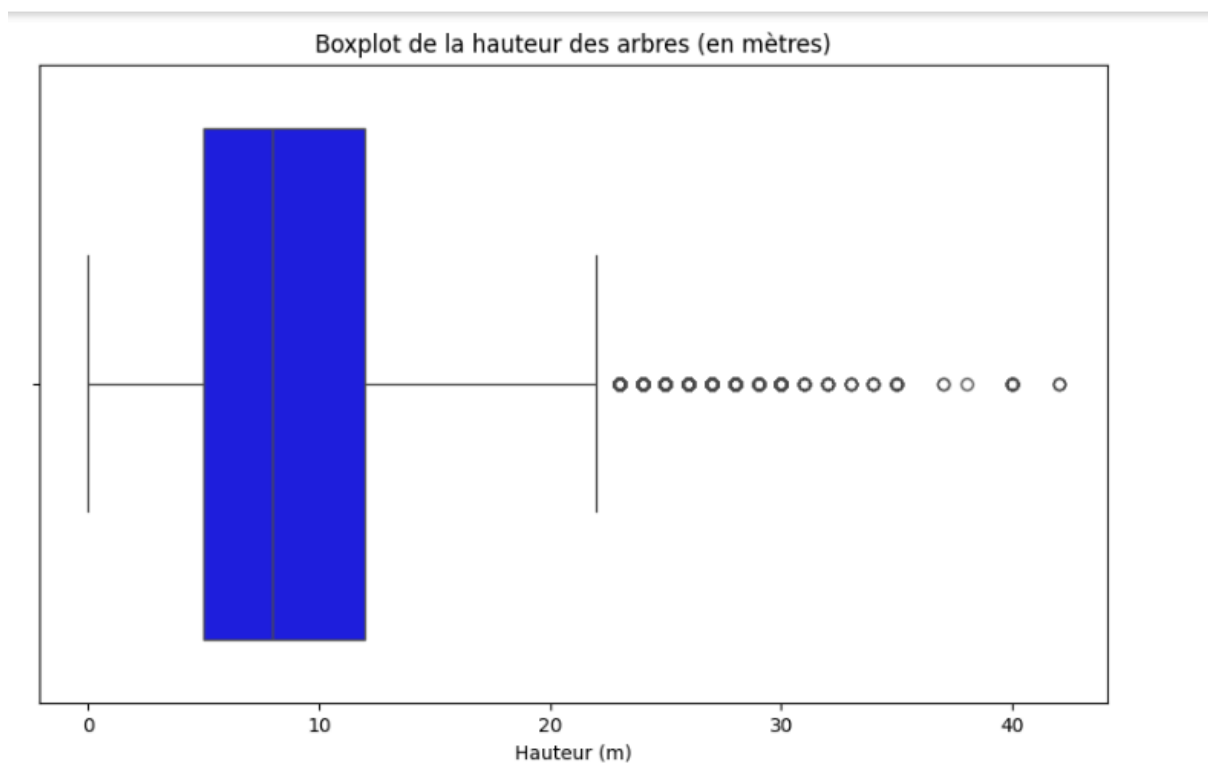
La prévalence d'arbres ayant une circonférence entre 0 et 200 cm reflète probablement des plantations récentes ou des espèces ayant des troncs naturellement fins.

Diversité limitée :

La forte concentration dans les petites circonférences suggère une homogénéité dans les tailles des arbres parisiens, probablement due à des politiques de plantation spécifiques.

Cas particuliers :

Les très grandes circonférences pourraient indiquer des espèces spécifiques ou des arbres historiques/exceptionnels (comme des platanes remarquables). Une analyse complémentaire pourrait être menée sur ces valeurs extrêmes.



Interprétation :

Urbanisme et gestion :

Les arbres de taille modérée (5-12 m) dominent, ce qui peut indiquer un effort pour limiter les impacts environnementaux liés à des arbres plus grands (taille, entretien, sécurité).

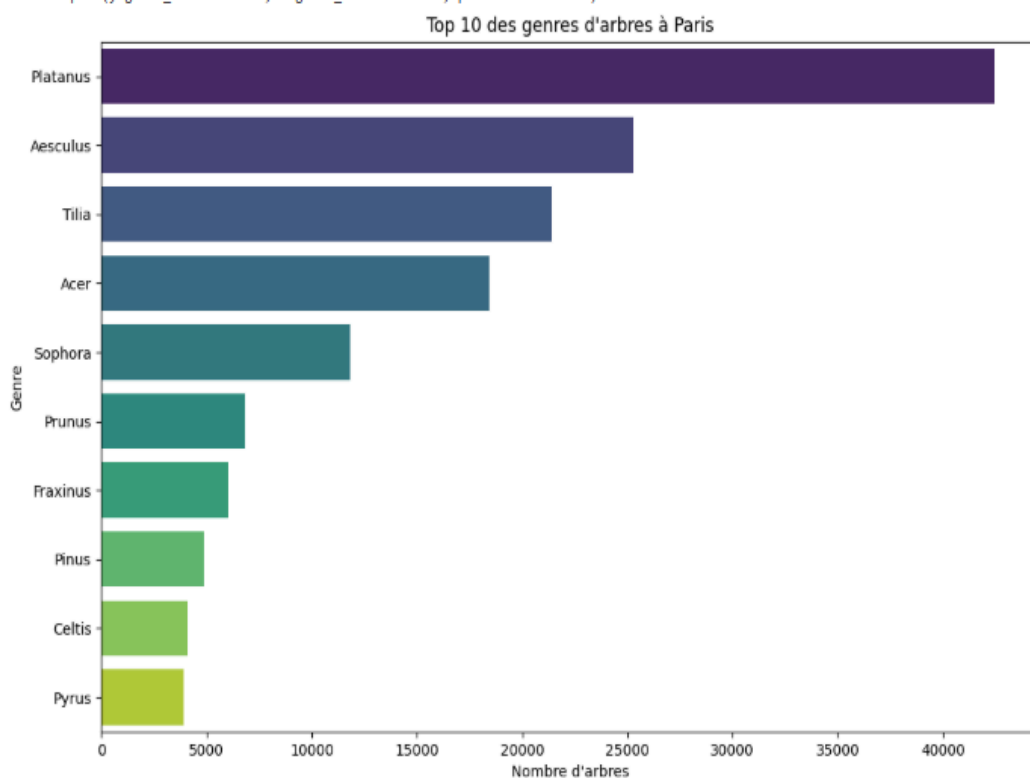
Arbres remarquables :

Les outliers (arbres de plus de 20 mètres) mériteraient une analyse supplémentaire pour déterminer:

- Les espèces associées.
- Leur localisation (parcs, grands boulevards, etc.).
- Leur rôle dans l'écosystème urbain.

Limite basse :

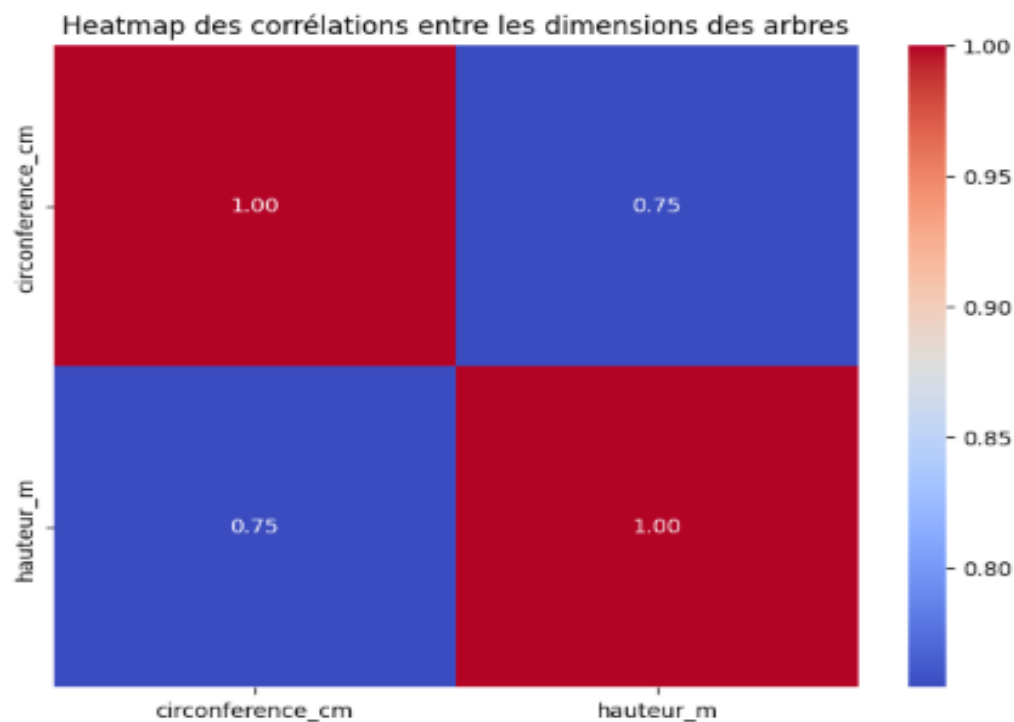
Aucun arbre ne semble avoir une hauteur proche de 0 m, ce qui est cohérent avec des données fiables.



Interprétation :

Le graphique montre que le **Platanus** (platane) est largement dominant parmi les arbres de Paris, suivi des genres **Aesculus** (marronnier), **Tilia** (tilleul), et **Acer** (érable). Ces espèces sont choisies pour leur résistance et leur esthétique dans un environnement urbain. Cependant, cette homogénéité peut poser des risques écologiques en cas de maladies ou parasites. Une diversification des genres renforcerait la résilience écologique et le patrimoine arboré de la ville.

3. Analyse bivariée et identification de relations :



Interprétation :

Corrélation positive entre circonférence et hauteur :

- La corrélation entre la **circonférence** des arbres (en cm) et leur **hauteur** (en m) est de **0,75**, indiquant une relation fortement positive. Cela signifie que, généralement, les arbres plus grands tendent à avoir une circonférence plus importante.

Relation non parfaite :

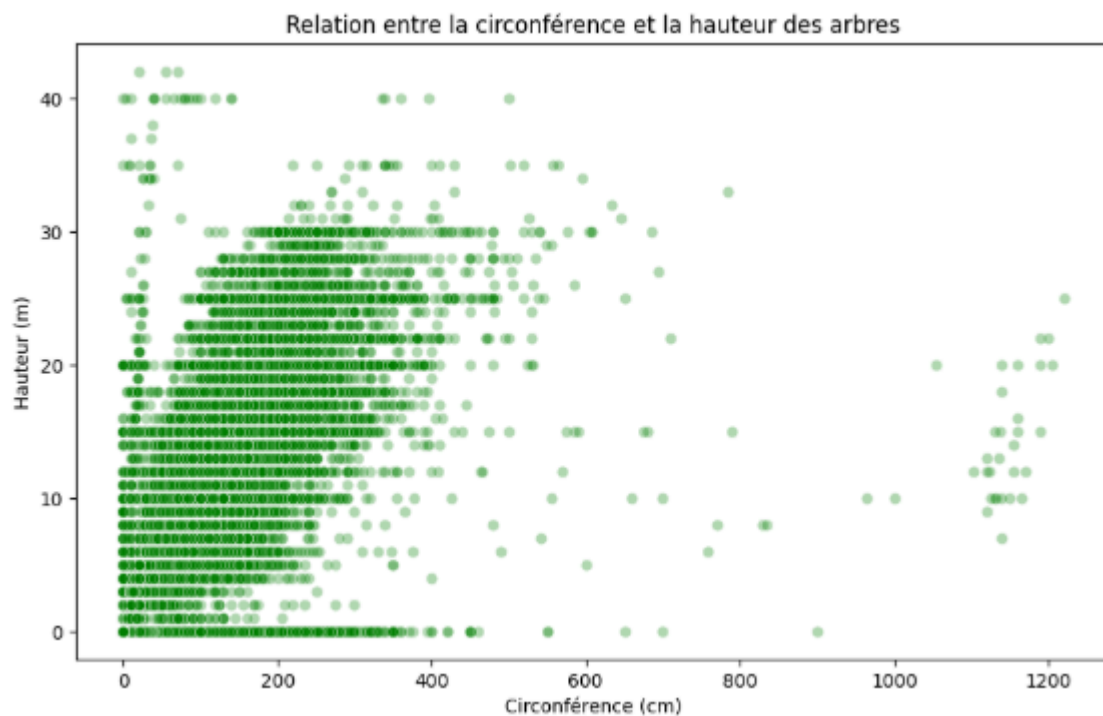
- Bien que la corrélation soit élevée, elle n'est pas de 1, ce qui suggère que d'autres facteurs (comme l'espèce, l'environnement, ou l'âge de l'arbre) peuvent influencer ces dimensions.

Implications pratiques :

- Cette corrélation forte mais non parfaite peut être utilisée pour estimer une dimension à partir de l'autre. Par exemple, la hauteur pourrait être approximée si seule la circonférence est mesurée.

Uniformité des données :

- Les valeurs de 1 sur la diagonale confirment que chaque variable est parfaitement corrélée avec elle-même, ce qui est attendu.

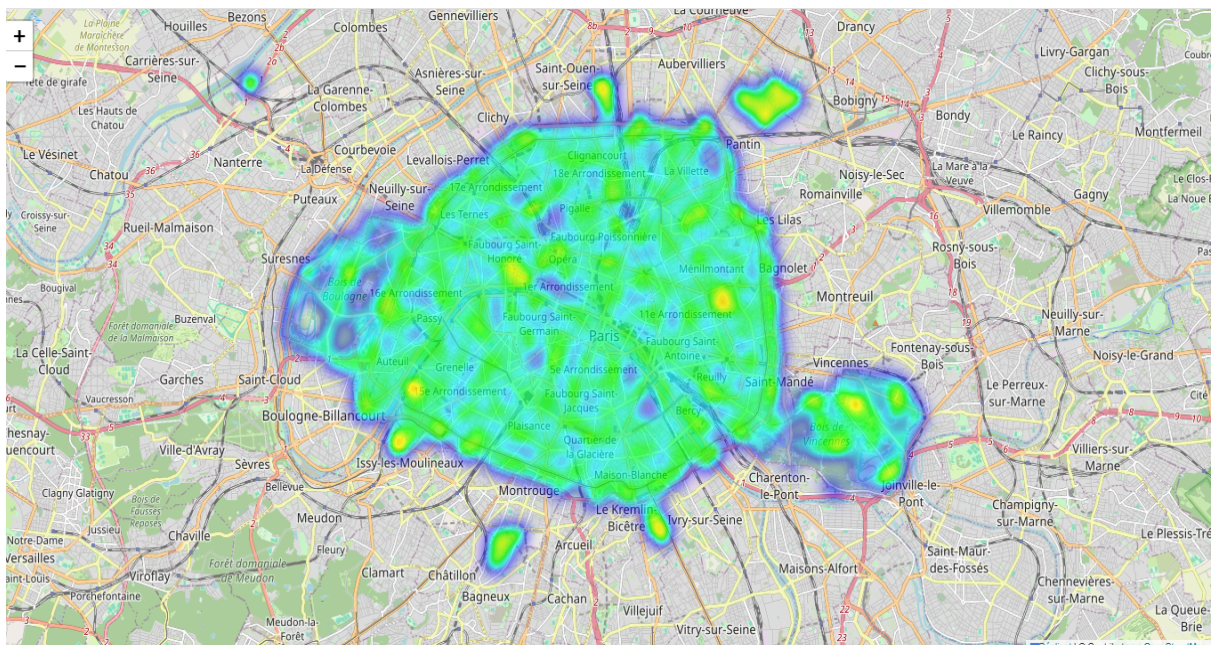


Interprétation :

- **Lien entre circonférence et hauteur :**
 - Une forte corrélation entre la circonférence et la hauteur signifie que les deux dimensions sont étroitement liées. Par conséquent, mesurer l'une (par exemple, la circonférence) pourrait fournir une estimation fiable de l'autre (la hauteur). Cela simplifie certaines décisions d'entretien, car il devient possible de prédire les caractéristiques globales d'un arbre à partir d'une seule mesure.
- **Optimisation de l'entretien :**
 - Les interventions telles que la coupe, la fertilisation ou le traitement des maladies peuvent être planifiées plus efficacement, car une forte corrélation permet d'estimer avec précision les dimensions globales d'un arbre en se basant sur une seule caractéristique.
- **Études supplémentaires :**

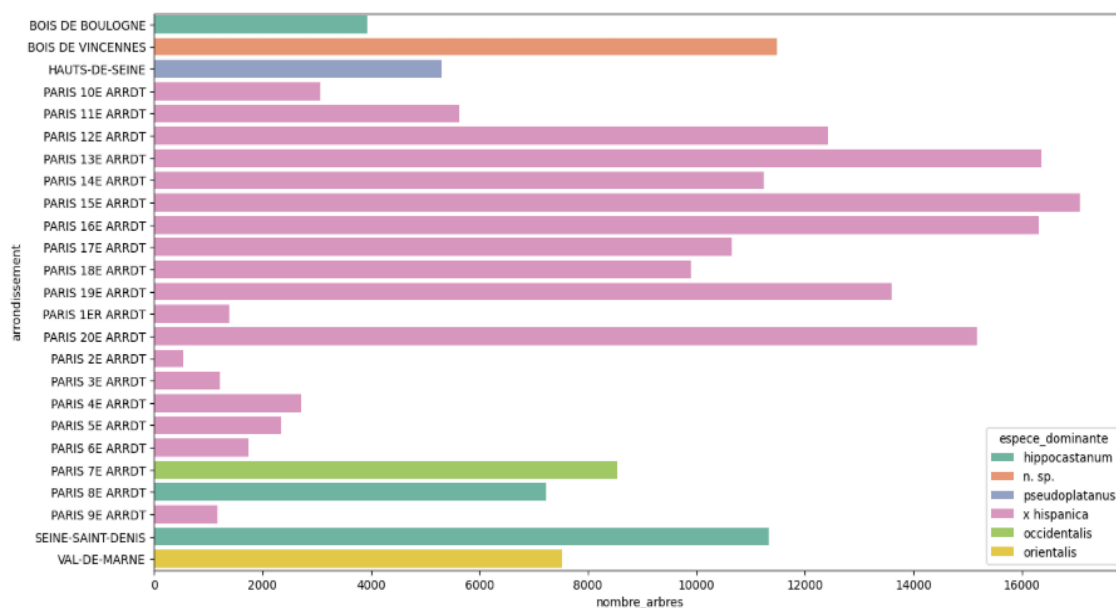
- Bien que la relation soit forte, il reste pertinent d'examiner d'autres facteurs influençant les dimensions des arbres, comme le genre, le stade de développement ou la localisation. Cela pourrait affiner encore davantage les stratégies d'entretien.

4. Analyse spatiale :



Carte interactive (Folium)

- **Points individuels** : Chaque arbre est représenté par un point sur la carte. Ces points permettent de visualiser la localisation exacte de chaque arbre.
 - Observation possible : Certaines zones, comme les grands parcs ou avenues boisées, ont une concentration visible de points.
- **Heatmap (carte de densité)** : Les zones où plusieurs arbres sont proches les uns des autres apparaîtront comme des « points chauds » sur la carte.
 - Zones à forte densité : Des parcs comme le Bois de Boulogne ou le Bois de Vincennes ont des densités élevées.
 - Impact : Les zones à forte densité nécessitent une planification spécifique pour l'entretien collectif (élagage, arrosage, etc.)



Principales conclusions :

1. **Répartition des arbres :** Le genre **Platanus** est dominant, suivi par quelques autres genres comme **Aesculus**, **Tilia**, et **Acer**, ce qui illustre une homogénéité dans les plantations.
2. **Caractéristiques physiques :** La circonférence et la hauteur des arbres présentent une forte corrélation (0,75), suggérant une relation naturelle entre la taille et l'âge probable des arbres.
3. **Valeurs atypiques :** Certains arbres atteignent des dimensions extrêmes, ce qui pourrait indiquer des spécimens remarquables ou des erreurs de saisie dans les données.
4. **Données manquantes et qualité des données :** Des colonnes clés, comme le stade de développement ou certaines informations d'identification, présentent un pourcentage significatif de valeurs manquantes, limitant l'analyse.

Recommandations pour optimiser les tournées d'entretien :

1. **Priorisation par taille et état :**
 - Identifier les arbres ayant une hauteur ou une circonférence importante pour les inclure dans les priorités d'entretien, car ils sont plus sujets à des risques (branches cassées, stabilité).
 - Regrouper les arbres par stade de développement pour cibler les jeunes arbres nécessitant des soins ou les spécimens matures à surveiller.
2. **Optimisation géographique :**
 - Utiliser les coordonnées géographiques des arbres pour planifier des tournées basées sur la proximité, en minimisant les déplacements.
 - Donner la priorité aux arrondissements ayant une forte densité d'arbres à surveiller.
3. **Gestion des arbres remarquables :**

- Établir une tournée spéciale pour les spécimens remarquables identifiés par leur taille ou leur rareté.

Pistes d'amélioration pour la collecte et la gestion des données :

1. Standardisation et qualité des données :

- Mettre en place des processus réguliers de validation et de nettoyage des données pour corriger les incohérences et supprimer les doublons.
- Remplir systématiquement les colonnes manquantes critiques, comme le **stade de développement**, lors des inspections de terrain.

2. Technologie et suivi :

- Utiliser des outils numériques (applications mobiles, GPS) pour collecter directement les données lors des interventions.
- Intégrer des capteurs pour surveiller automatiquement l'état de santé des arbres (humidité du sol, inclinaison, etc.).

3. Diversité et résilience :

- Favoriser la plantation d'espèces variées pour réduire les risques liés aux maladies ou parasites qui affectent les genres dominants.
- Inclure des données sur les espèces introduites pour suivre leur évolution dans le milieu urbain.

Structuration selon les contraintes :

1. Opérationnelles :

- Former les équipes à l'utilisation d'outils numériques pour réduire les erreurs humaines.
- Planifier les tournées d'entretien selon des horaires adaptés aux contraintes de circulation urbaine.

2. Budgétaires :

- Prioriser les investissements dans la maintenance des arbres remarquables ou les plus à risque.
- Étaler les efforts d'amélioration des données sur plusieurs années pour limiter les coûts initiaux.

3. Environnementales :

- Intégrer les arbres dans une stratégie plus large de lutte contre les îlots de chaleur urbains, en plantant dans des zones déficitaires en végétation.

