

# Participez à un concours sur la Smart City

Projet 2

A large, dark blue, diagonal shape that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

# Sommaire

- Introduction du sujet et de l'objectif
- Présentation du dataframe
- Nettoyage du dataset
- Visualisation du dataset

# Introduction du sujet et de l'objectif

- Faire une analyse exploratoire avec un jeu de données portant sur les arbres de la ville de Paris afin d'optimiser l'arrosage des arbres

## Objectif :

- Analyse univariée de toutes les données disponibles.
- Analyse et nettoyage de certaines données.
- Identifier le nombre d'arbres nécessitant un arrosage spécifique ou pas par arrondissement pour aider à l'organisation de l'arrosage
- Identifier le nombre d'arbres remarquables par arrondissement afin d'apporter l'attention nécessaire et l'arrosage adéquat.

# Analyse et nettoyage de données.

- J'ai commencé par importer le csv et avoir une première analyse du dataset.

```
P2ArbresFr =  
pd.read_csv("p2-arbres-fr.csv", sep=';')  
P2ArbresFr.head(10)  
P2ArbresFr.info()
```

- Puis, j'ai analysé le taux de valeurs et supprimer les colonnes qui ne seront pas utilisées et les lignes dupliquées

```
len(P2ArbresFr)  
len(P2ArbresFr.drop_duplicates())  
P2ArbresFr=P2ArbresFr.drop(columns='numero')
```

# Analyse et nettoyage de données.

- J'ai fait d'abord une analyse univariée de toutes les données disponibles.
- Après une visualisation des différents graphes, on remarque que plusieurs colonnes censées être une répartition un peu bizarre, on va s'intéresser dans le cadre de cette analyse, à 3 colonnes en particulier :
  1. La colonne **Remarquable** qui doit être un booléen et non une valeur numérique -> On remplacera ces valeurs numériques par un booléen
  2. Les colonnes **circonférence** et **hauteur** dont l'analyse univariée en utilisant la méthode interquantile nous démontre qu'il y a plusieurs valeurs considérées "aberrantes" et aussi qui paraissent illogiques.

# Analyse et nettoyage de données.

- Changer le type de la colonne “remarquable” de numérique à un booléen

Lignes : 679 à 680

- Identification des valeurs normales et atypiques pour la circonférence / Hauteur en utilisant la médiane, l'écart-type, les valeurs max et min

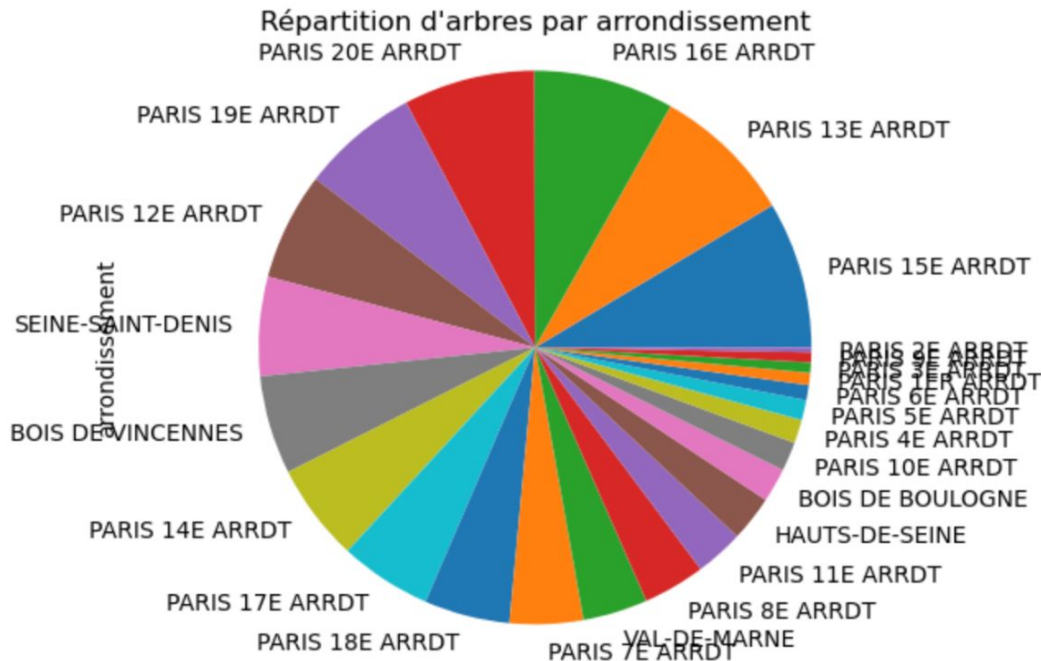
Lignes : 680 à 699

- Puis, on remplace toutes les valeurs aberrantes par un NaN

Lignes : 688, 689 et 695

- Nous avons par la suite, un dataframe avec des valeurs qui semblent pertinentes
  - Nom du dataframe : `P2ArbresFrClean`

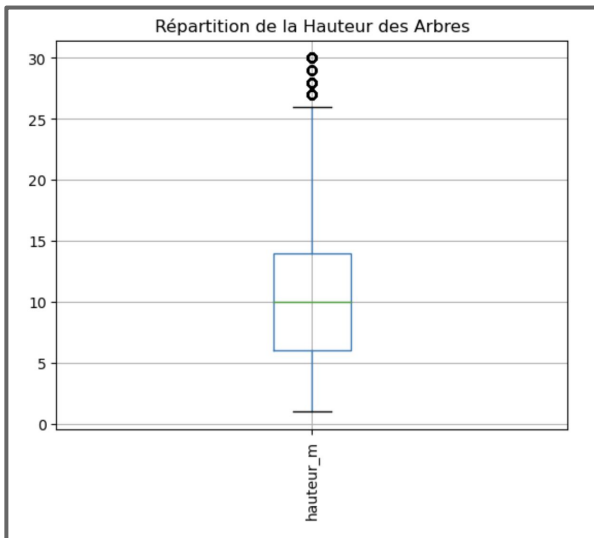
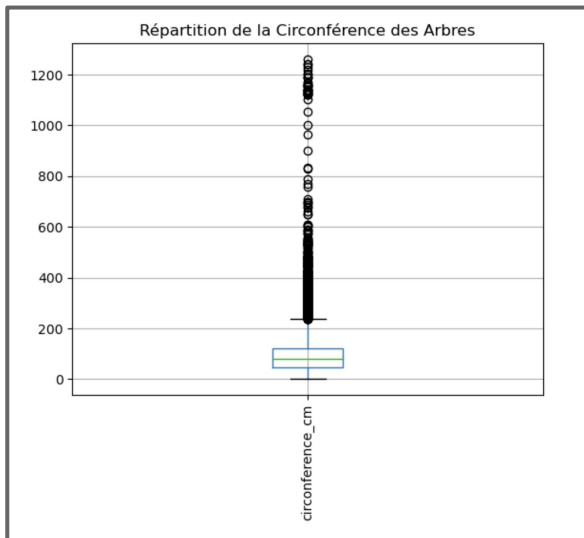
# Visualisation et répartition des arbres par arrondissement



# Répartition de la circonférence et de la hauteur en utilisant la méthode interquantile

## Hypothèse :

N'ayant aucune info métier sur le type d'arrosage, j'ai fait appel à ChatGBT pour identifier à partir de quelle circonférence et hauteur on devra faire appel à de nouvelles techniques d'arrosage. Une estimation à 35 cm pour la circonférence et 7 m a été communiquée. Cette hypothèse n'est pas une information à challenger. j'ai aussi considéré que les arbres remarquables nécessiteront une attention particulière car faisant partie du patrimoine de Paris, donc cela nécessiterait potentiellement une équipe plus expérimentée.

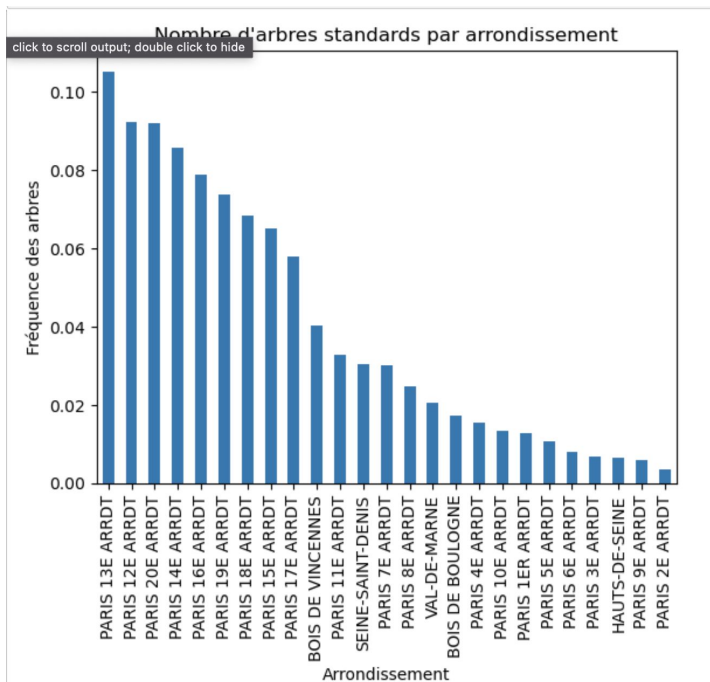


En étudiant l'ensemble du parc de Paris, on remarque que + du  $\frac{3}{4}$  du parc nécessite une technique d'arrosage simple. Donc, cela nous démontre qu'on peut diviser le parc en 3 catégories différentes de types d'arbres :

- Les arbres standards qui nécessite une équipe standard.
- Les arbres spécifiques avec des dimensions qui obligent la ville de mobiliser du matériel spécifique.
- Et finalement, les arbres remarquables qui nécessitent une attention particulière puisqu'ils nécessitent un suivi plus avancé.



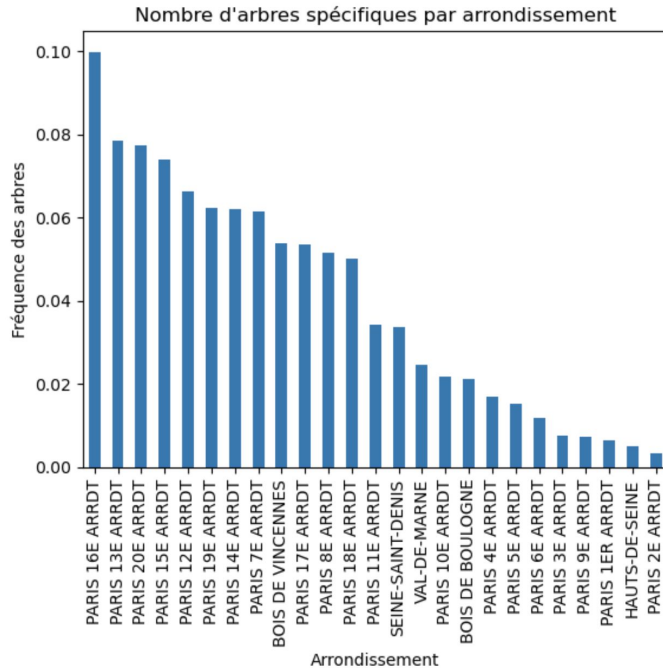
# Analyse univarié du nombre d'arbres standards par arrondissement



Cette analyse univariée nous permet d'identifier les arrondissements avec le plus d'arbres standards, cela permettra aux équipes d'organiser leur tournée d'arrosage.

- On peut imaginer par exemple que le 13ème et le 12ème arrondissement nécessitera peut être plusieurs jours d'arrosage et donc cela nous oblige à **faire encore une analyse plus poussée pour cette arrondissement**
- D'un autre coté, le 2ème, 3ème et 9ème arrondissement ne nécessiteront pas autant de temps d'arrosage, et on peut donc imaginer par exemple de **les grouper puisqu'ils sont limitrophes pour optimiser les journées des équipes d'arrosage.**

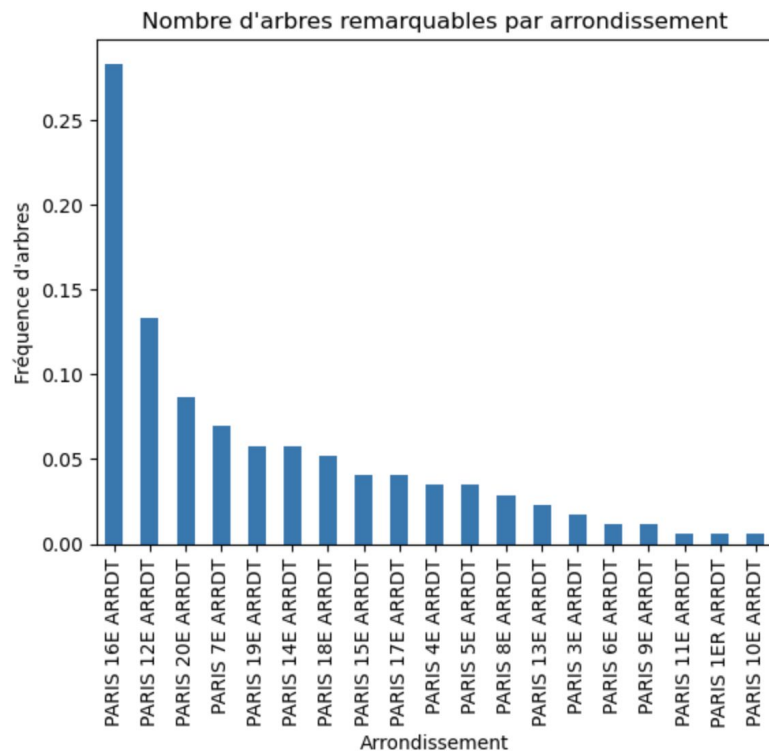
# Analyse univarié du nombre d'arbres spécifiques par arrondissement



De la même façon, cette analyse univariée nous permet d'identifier les arrondissements avec le plus d'arbres spécifiques, cela permettra aux équipes d'organiser leur tournée d'arrosage.

- On peut imaginer par exemple que le 16ème arrondissement nécessitera peut être plusieurs jours d'arrosage et donc cela nous oblige à faire **une analyse plus poussée pour cet arrondissement**
- D'un autre côté, le 2ème, 3ème, 9ème et 1er arrondissement ne nécessiteront pas autant de temps d'arrosage, et on peut donc imaginer par exemple de les grouper puisqu'ils sont limitrophes pour optimiser les journées des équipes d'arrosage.

# Analyse univarié du nombre d'arbres remarquables par arrondissement



De la même façon, cette analyse univariée nous permet d'identifier les arrondissements avec le plus d'arbres remarquables, cela permettra aux équipes d'organiser leur tournée d'arrosage.

- On peut imaginer par exemple que le 16ème arrondissement nécessitera peut être plusieurs jours d'arrosage et donc cela nous oblige à faire encore une analyse plus poussée pour cette arrondissement
- D'un autre coté, le 9, 10 et 11ème arrondissement ne nécessiteront pas autant de temps d'arrosage, et on peut donc imaginer par exemple de les grouper puisqu'ils sont limitrophes pour optimiser les journées des équipes d'arrosage.