Participez à un concours sur la Smart City

Optimisation des tournées pour l'entretien des arbres de la ville de Paris

By Tyson JOHN

Content

Introduction

- Concour lancé par L'ONG "Data is for Good"
- Jue de données de la ville de Paris sur Opendata.paris.fr
- Aider Paris à devenir une smart-city !

Objective

- Données explorées à l’aide de Python et de ses librairies
- Méthodologique d’analyse de données
- Synthèse de l’analyse de données
- Une optimisation des tournées pour l’entretien des arbres de la ville

Initial Data - Presentation

Below is the initial Presentation of the data downloaded form Paris Data

id	type_emplacement	domanialite	arrondissement	complement_addresse	numero	lieu	id_emplacement	libelle_francais	genre	espece	variete	circonference_cm	hauteur_m	stade_developpement	remarquable	geo_point_2d_a	geo_point_2d
874	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	19	Marronnier	Aesculus	hippocastanum	NaN	20	5	NaN	0.0	48.857620	2.32096
875	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	20	If	Taxus	baccata	NaN	65	8	A	NaN	48.857656	2.32103
876	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	21	If	Taxus	baccata	NaN	90	10	A	NaN	48.857705	2.32106
877	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	22	Erable	Acer	negundo	NaN	60	8	A	NaN	48.857722	2.32100
878	Arbre	Jardin	PARIS 17E ARRDT	NaN	NaN	PARC CLICHY-BATIGNOLLES-MARTIN LUTHER KING	000G0037	Arbre à miel	Tetradium	daniellii	NaN	38	0	NaN	NaN	48.890435	2.31528

Initial Data - Presentation

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200137 entries, 0 to 200136
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_addresse   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
dtypes: float64(4), int64(3), object(11)
memory usage: 27.5+ MB
```

A quickview about the data

- 200137 Lines and 18 columns
- each row is details on a tree
- each column is different characteristic of a tree
- Datatype of each column and notnull values mentioned

Data Understanding

From 18 columns we keep only 7 columns for Data analysis

id	arrondissement	circonference (cm)	hauteur (m)	stade	geoLoc a	geoLoc b
99874	PARIS 7E ARRDT	20	5	NaN	48.857620	2.320962
99875	PARIS 7E ARRDT	65	8	A	48.857656	2.321031
99876	PARIS 7E ARRDT	90	10	A	48.857705	2.321061
99877	PARIS 7E ARRDT	60	8	A	48.857722	2.321006
99878	PARIS 17E ARRDT	38	0	NaN	48.890435	2.315289

Arrondissement & GeoLoc for location

Stade to quantify the care

Hauteur and circonference for equipments

Data Preparation

Descriptive statistics on the Data

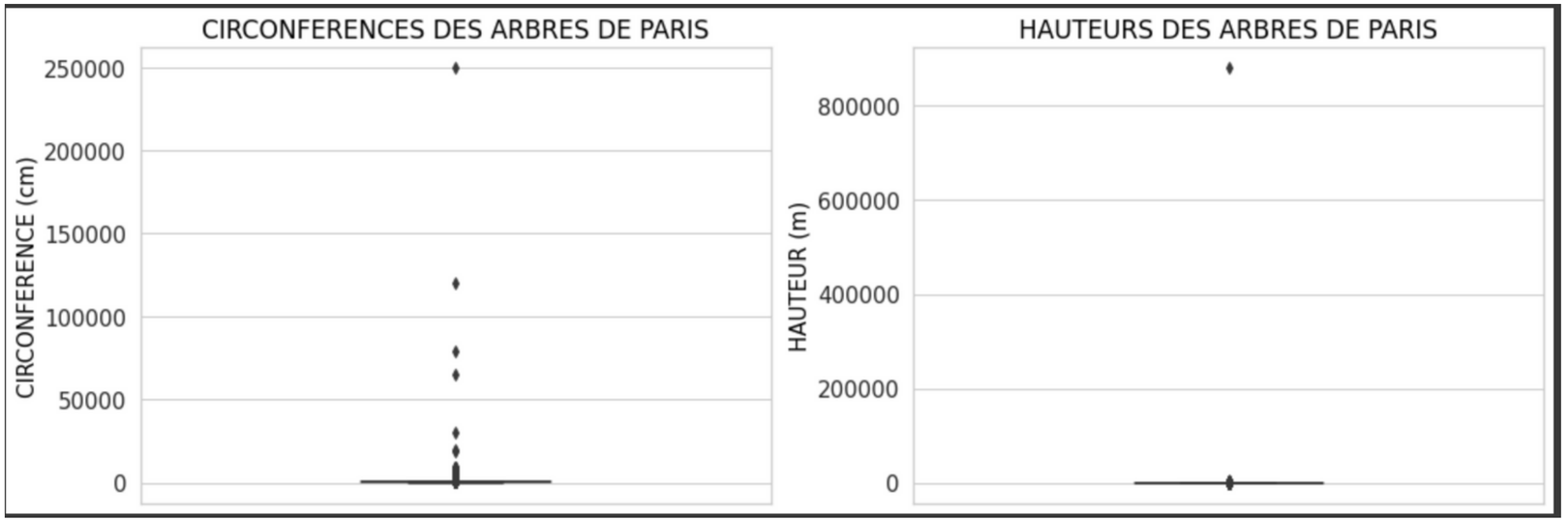
Clean the unrealist data like

- The tallest tree in the world is at 115.92 M
- The Circumference cannot be more than 4000 cm

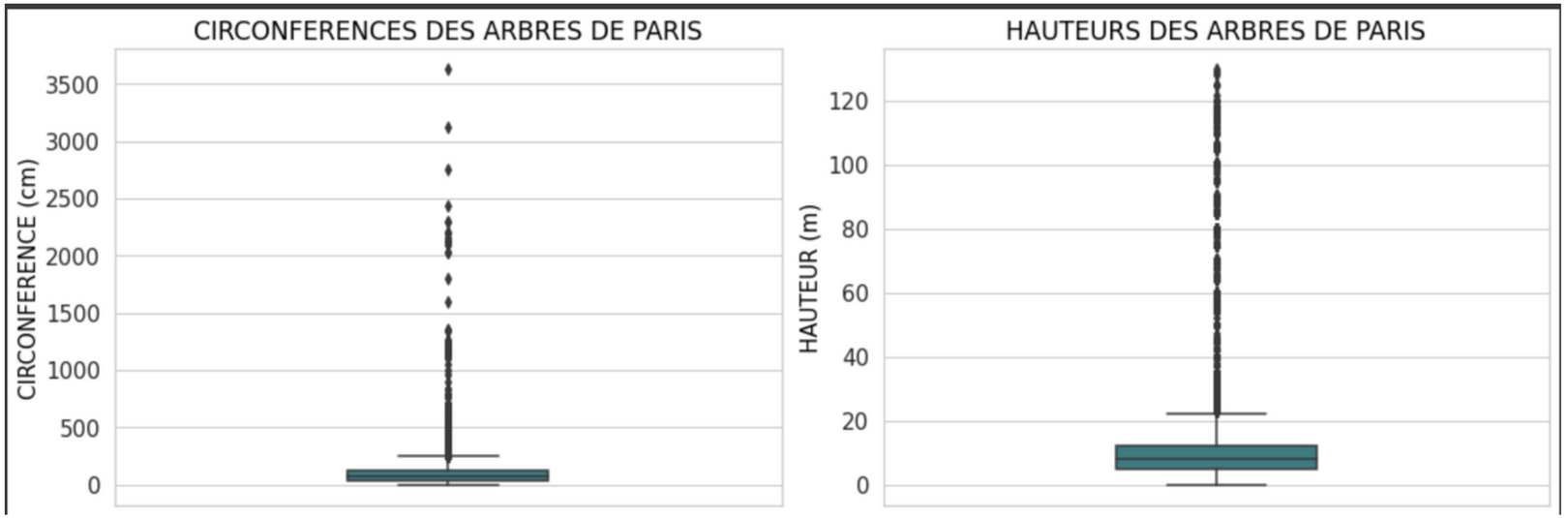
	id	arrondissement	libelle	circonference (cm)	hauteur (m)	stade	geoLoc a	geoLoc b
count	2.001370e+05	200137	198640	200137.000000	200137.000000	132932	200137.000000	200137.000000
unique	NaN	25	192	NaN	NaN	4	NaN	NaN
top	NaN	PARIS 15E ARRDT	Platane	NaN	NaN	A	NaN	NaN
freq	NaN	17151	42508	NaN	NaN	64438	NaN	NaN
mean	3.872027e+05	NaN	NaN	83.380479	13.110509	NaN	48.854491	2.348208
std	5.456032e+05	NaN	NaN	673.190213	1971.217387	NaN	0.030234	0.051220
min	9.987400e+04	NaN	NaN	0.000000	0.000000	NaN	48.742290	2.210241
25%	1.559270e+05	NaN	NaN	30.000000	5.000000	NaN	48.835021	2.307530
50%	2.210780e+05	NaN	NaN	70.000000	8.000000	NaN	48.854162	2.351095
75%	2.741020e+05	NaN	NaN	115.000000	12.000000	NaN	48.876447	2.386838
max	2.024745e+06	NaN	NaN	250255.000000	881818.000000	NaN	48.911485	2.469759

Data Preparation

BoxPlot Before Cleaning

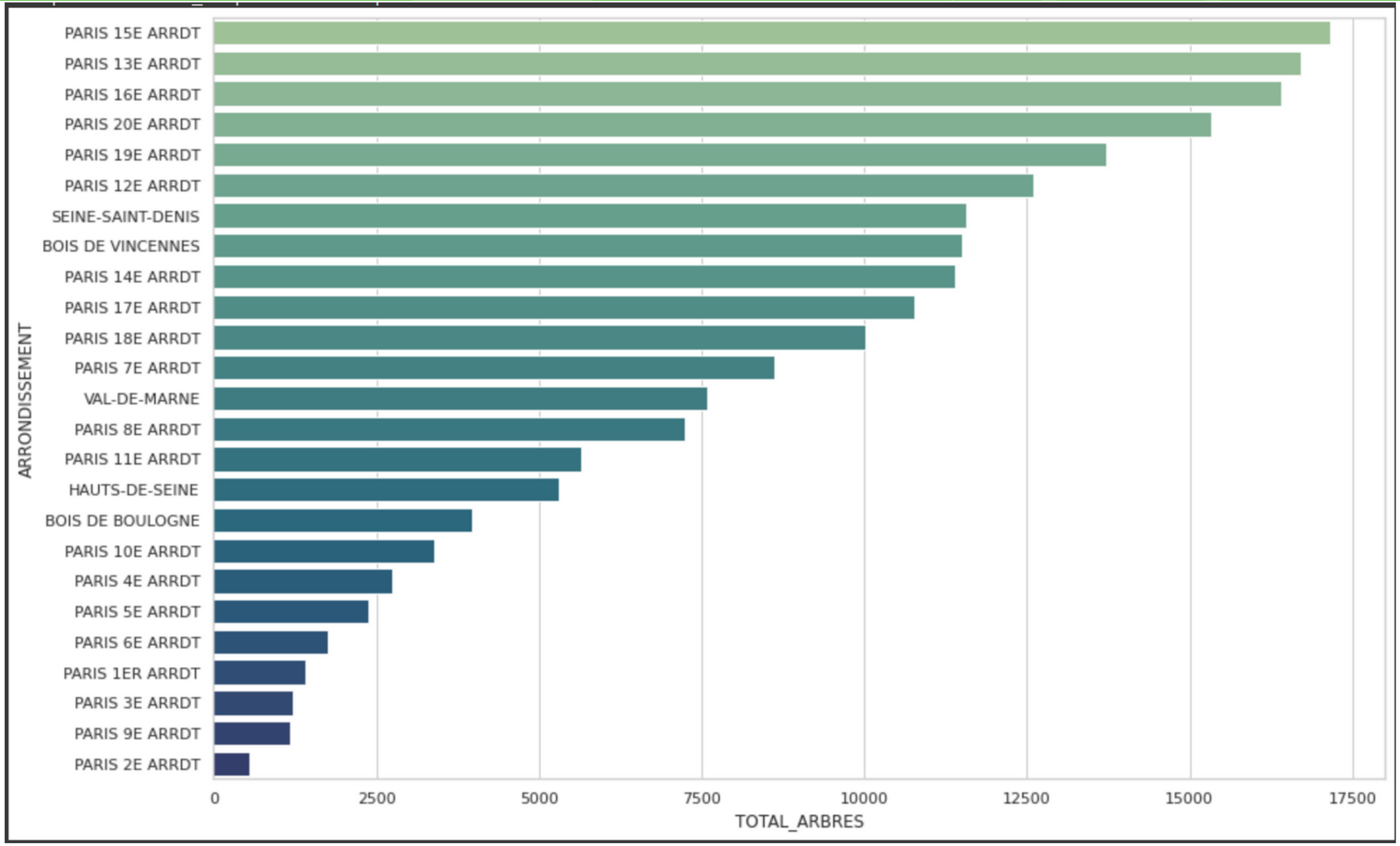


After Cleaning



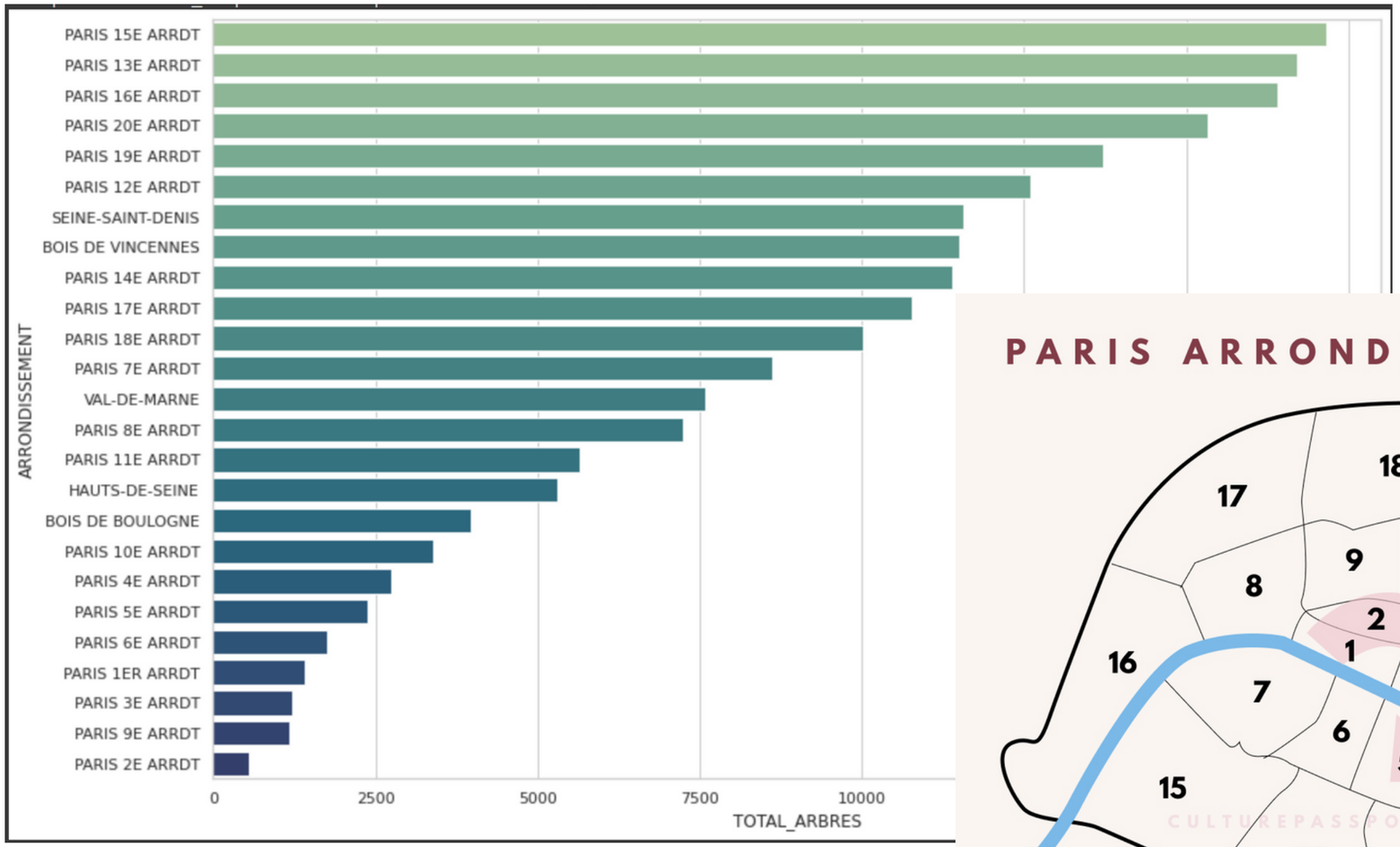
Data Visualization

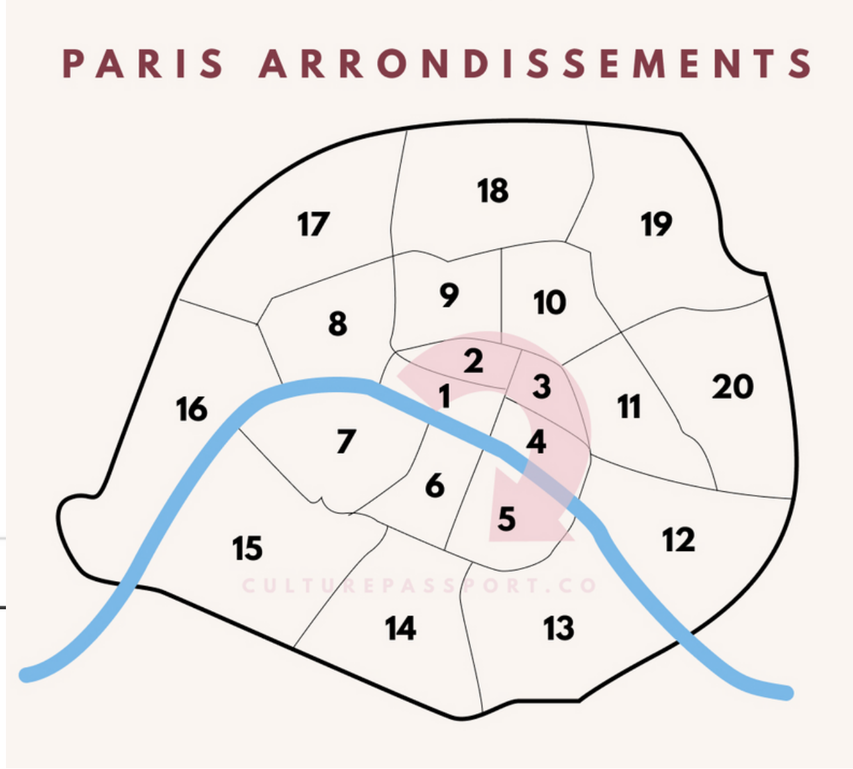
Tree distribution by districts in Paris



Data Visualization

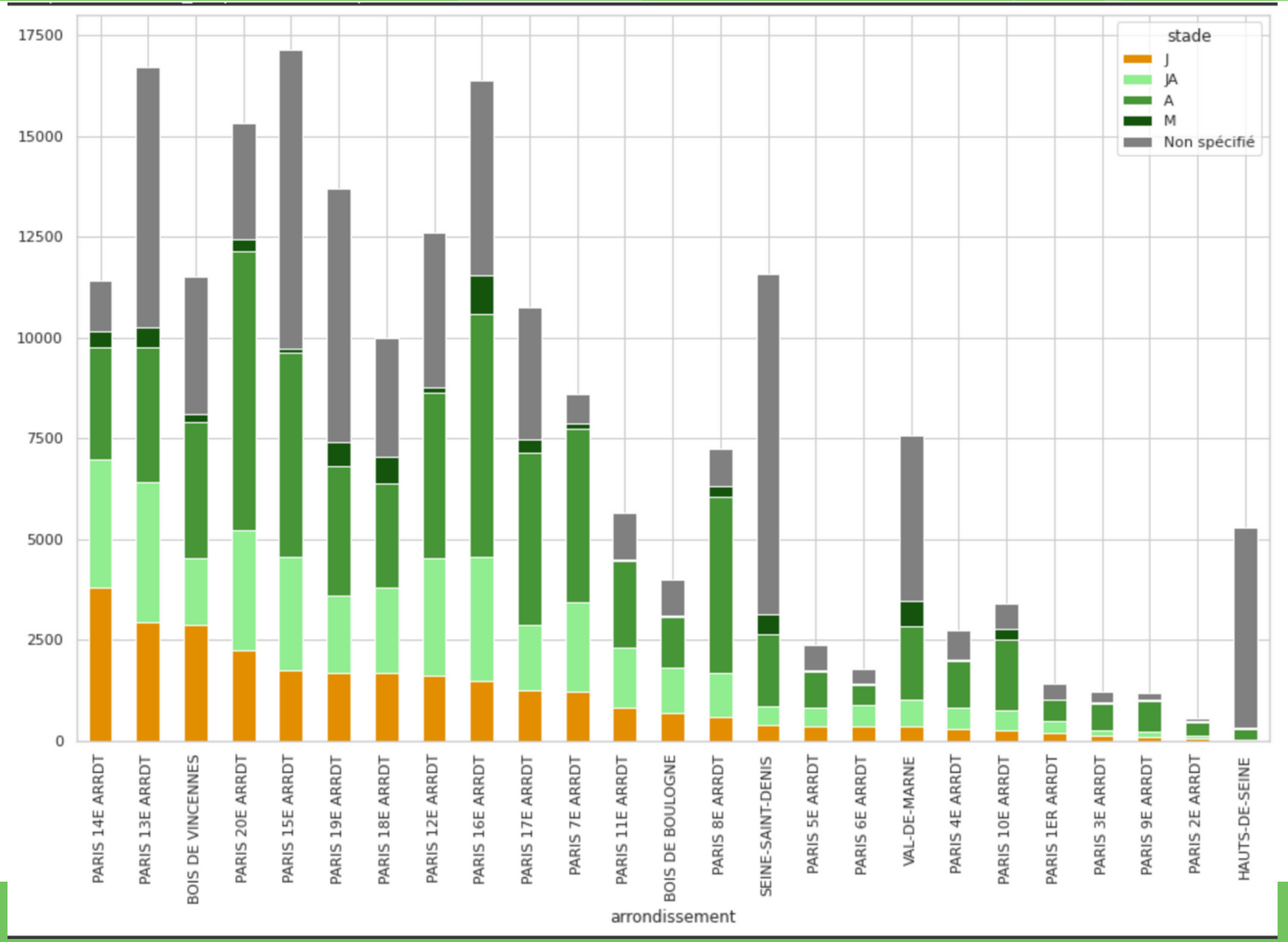
The trees are concentrated in the outer districts.





Data Visualization

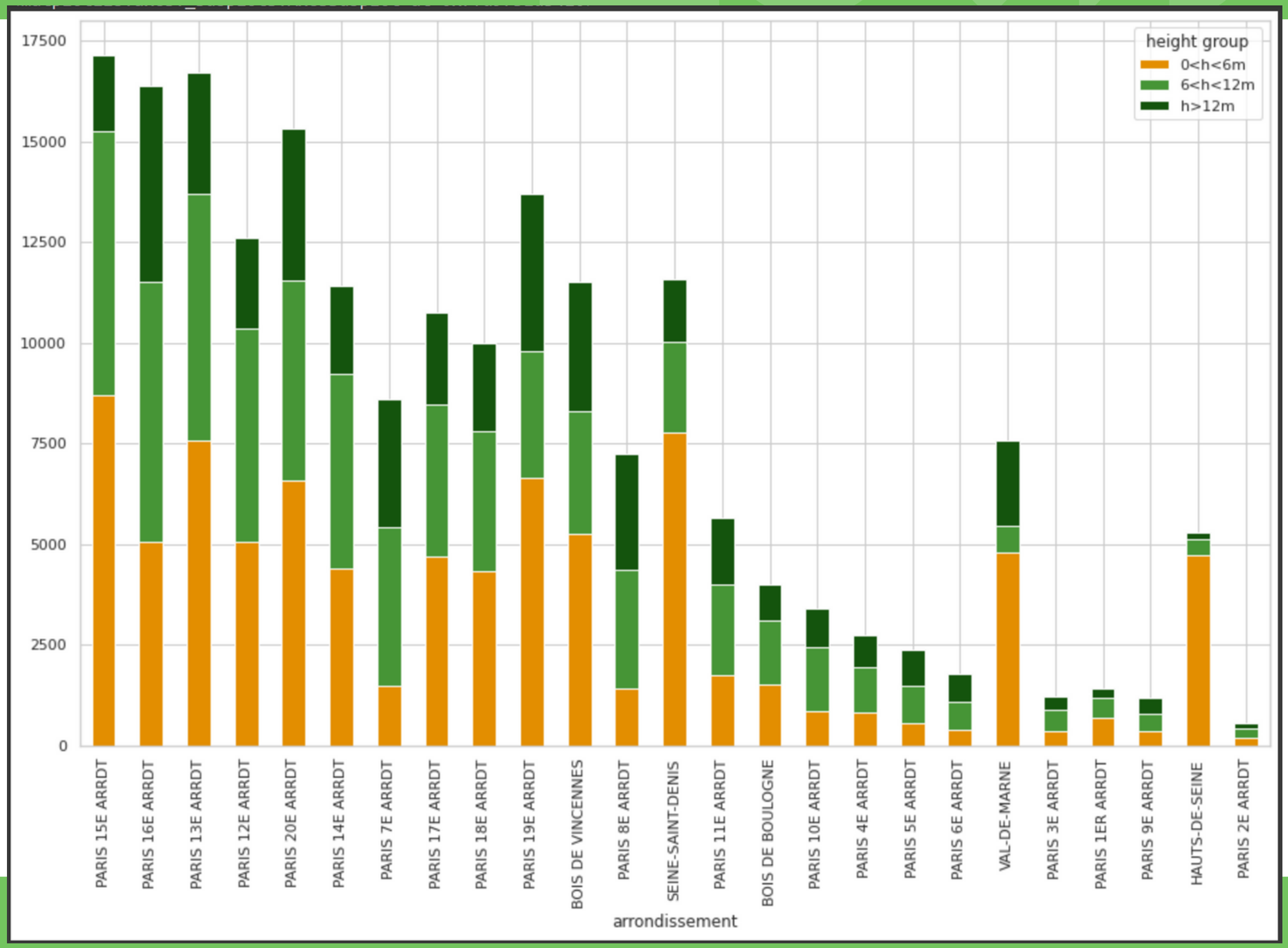
Tree distribution by stage of development and districts in Paris



Data Visualization

Trees distribution by height and district

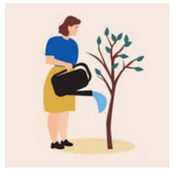
To decide on the equipment's required



Conclusion

Concentration of Trees

Paris peripheric districts are having
most of the tree concentration



High work Zones

Paris peripheric districts are demanding high maintanance
due to the high concentration of young trees

Heavy duty equipments zone

District 7 and all peripheric districts demands boom lift like
heavy equipments for its high concentration of tall trees