

Participez à un concours sur la Smart City

Présentation "Projet 2" chez "OPENCLASSROOM"
Jaoid KRAIRI
(Mai 2021)

SOMMAIRE



Introduction



Présentation générale du jeu de données



Démarche méthodologique d'analyse de données



Synthèse de mon analyse de données



Conclusion



Remerciements

INTRODUCTION : Végétaliser la ville de Paris



PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement



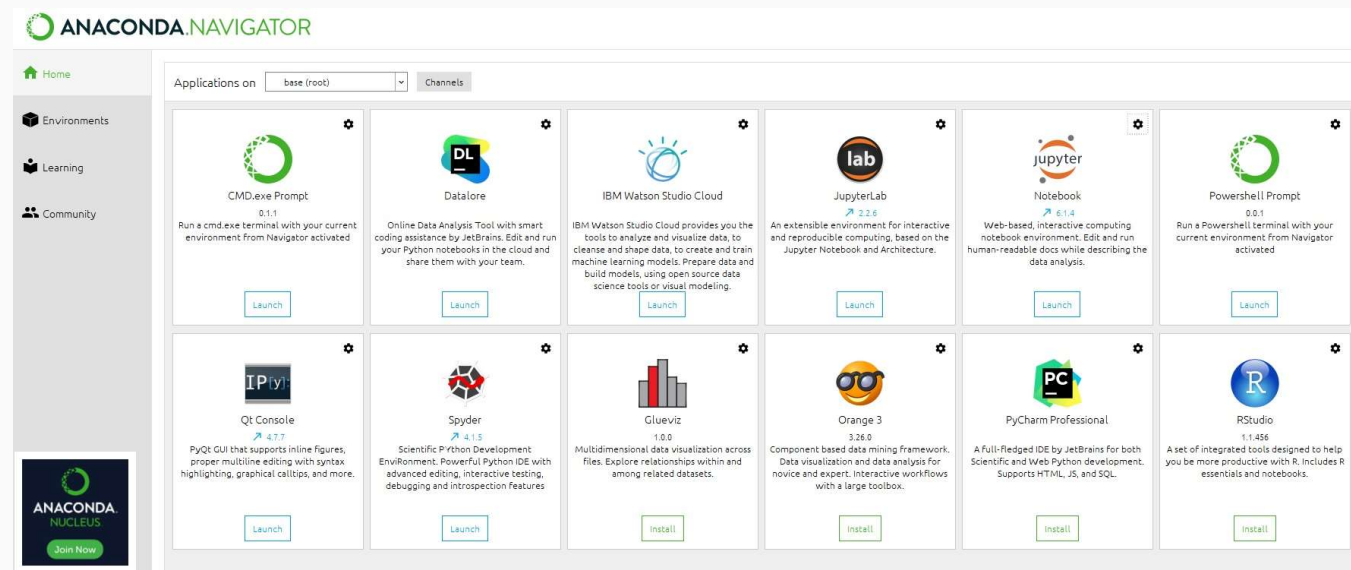
Ce PC > Téléchargements				
	Nom	Modifié le	Type	Taille
	p2-arbres-fr 1	17/04/2021 05:05	Fichier CSV Micro...	28 301 Ko

Ce PC > Téléchargements				
	Nom	Modifié le	Type	Taille
	Anaconda3-2020.11-Windows-x86_64	05/04/2021 08:55	Application	468 161 Ko

PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement

Environnement
de
développement



PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement

Installation des
bibliothèques Python



Création

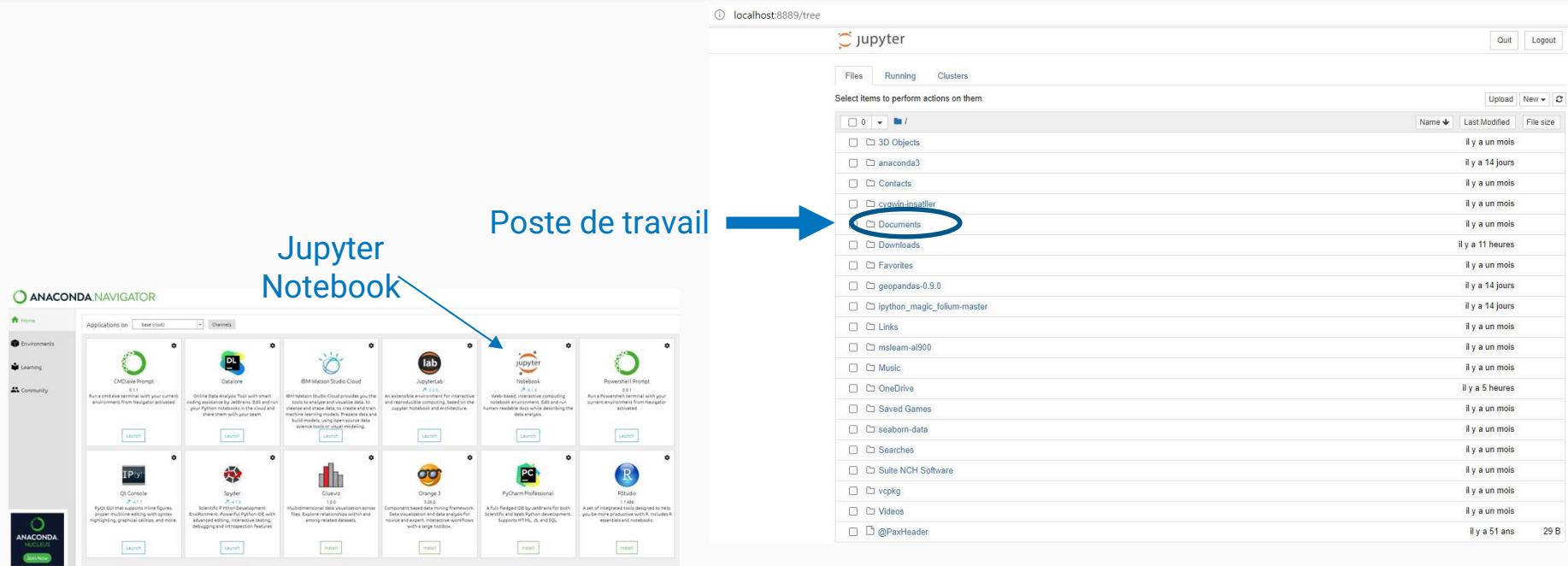


Name	Description	Version
<input checked="" type="checkbox"/> _ipyw_jlab_nb_ex...	A configuration metapackage for enabling anaconda-bundled jupyter extensions	0.1.0
<input checked="" type="checkbox"/> alabaster	Configurable, python 2+3 compatible sphinx theme.	0.7.12
<input checked="" type="checkbox"/> anaconda	Simplifies package management and deployment of anaconda	2020.11
<input checked="" type="checkbox"/> anaconda-client	Anaconda.org command line client library	1.7.2
<input checked="" type="checkbox"/> anaconda-project	Tool for encapsulating, running, and reproducing data science projects	0.8.4
<input checked="" type="checkbox"/> argh	The natural cli.	0.26.2
<input checked="" type="checkbox"/> argon2-cffi	The secure argon2 password hashing algorithm.	20.1.0
<input checked="" type="checkbox"/> asn1crypto	Python asn.1 library with a focus on performance and a pythonic api	1.4.0
<input checked="" type="checkbox"/> astroid	A abstract syntax tree for python with inference support.	2.4.2
<input checked="" type="checkbox"/> astropy	Community-developed python library for astronomy	4.0.2
<input checked="" type="checkbox"/> async-generator		1.10
<input checked="" type="checkbox"/> async_generator	Async generators and context managers for python 3.5+	1.10
<input checked="" type="checkbox"/> atomicwrites	Atomic file writes	1.4.0
<input checked="" type="checkbox"/> attrs	Attrs is the python package that will bring back the joy of writing classes by relieving you from the drudgery of implementing object protocols (aka dunder methods).	20.3.0
<input checked="" type="checkbox"/> autopep8	A tool that automatically formats python code to conform to the pep 8 style guide	1.5.4
<input checked="" type="checkbox"/> babel	Utilities to internationalize and localize python applications	2.8.1
<input checked="" type="checkbox"/> backcall	Specifications for callback functions passed in to an api	0.2.0
<input checked="" type="checkbox"/> backports		1.0
<input checked="" type="checkbox"/> backports.function...		1.6.1

PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement

Jupyter Notebook → **Poste de travail** →



The image displays two screenshots of data science environments. The left screenshot shows the Anaconda Navigator interface with a grid of application tiles. The 'Jupyter Notebook' tile is highlighted with a blue arrow pointing to the 'Poste de travail' (Workspace) label. The right screenshot shows the JupyterLab interface, which includes a file explorer on the left. The 'Documents' folder is highlighted with a blue circle, and a blue arrow points from the 'Poste de travail' label to it.

Left Screenshot (Anaconda Navigator):

- Applications on: Channels
- Grid of application tiles: CML, DataLab, IBM Watson Studio Cloud, JupyterLab, Jupyter Notebook, PowerShell Prompt, Jupyter Notebook, Jupyter Notebook, Jupyter Notebook, Jupyter Notebook, Jupyter Notebook, Jupyter Notebook.

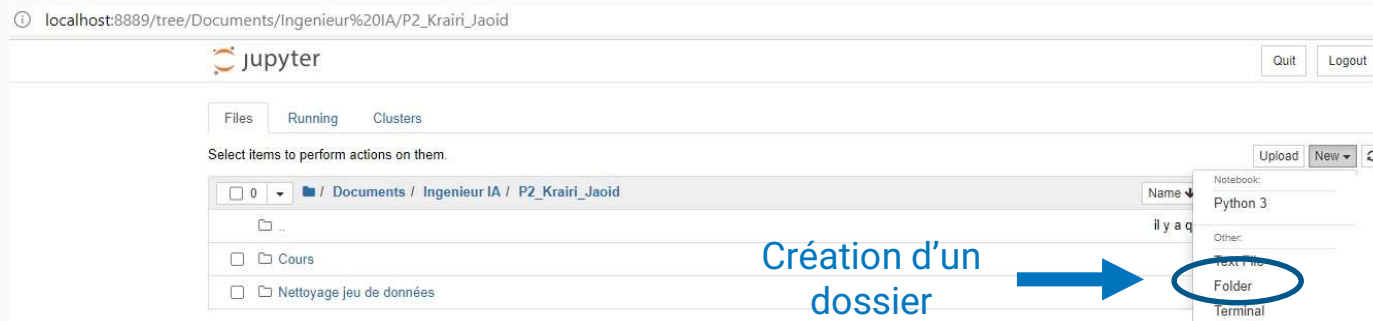
Right Screenshot (JupyterLab):

- URL: localhost:8889/tree
- Buttons: Quit, Logout
- Tabs: Files, Running, Clusters
- Select items to perform actions on them. Upload New
- File Explorer Table:

	Name	Last Modified	File size
<input type="checkbox"/>	0		
<input type="checkbox"/>	3D Objects	il y a un mois	
<input type="checkbox"/>	anaconda3	il y a 14 jours	
<input type="checkbox"/>	Contacts	il y a un mois	
<input type="checkbox"/>	cywin-installer	il y a un mois	
<input type="checkbox"/>	Documents	il y a un mois	
<input type="checkbox"/>	Downloads	il y a 11 heures	
<input type="checkbox"/>	Favorites	il y a un mois	
<input type="checkbox"/>	geopandas-0.9.0	il y a 14 jours	
<input type="checkbox"/>	ipython_magic_folium-master	il y a 14 jours	
<input type="checkbox"/>	Links	il y a un mois	
<input type="checkbox"/>	mslearn-ai900	il y a un mois	
<input type="checkbox"/>	Music	il y a un mois	
<input type="checkbox"/>	OneDrive	il y a 5 heures	
<input type="checkbox"/>	Saved Games	il y a un mois	
<input type="checkbox"/>	seaborn-data	il y a un mois	
<input type="checkbox"/>	Searches	il y a un mois	
<input type="checkbox"/>	Suite NCH Software	il y a un mois	
<input type="checkbox"/>	vcpkg	il y a un mois	
<input type="checkbox"/>	Videos	il y a un mois	
<input type="checkbox"/>	@PaxHeader	il y a 51 ans	29 B

PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement



PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement

The image displays two screenshots of the JupyterLab web interface. The top screenshot shows the 'Files' tab with a file browser view of the directory `/ Documents / Ingenieur IA / P2_Krairi_Jaoid / p2`. A blue arrow points from the text 'Créer un fichier Python' to the 'New' button in the top right corner. A dropdown menu is open, showing options: 'Notebook: Python 3', 'Other: Text File', 'Folder', and 'Terminal'. The bottom screenshot shows the same directory view, but now a file named `P2_01_notebook.ipynb` has been created and is circled in blue. A blue arrow points from the text 'Fichier créé' to this circled file. The interface includes a top bar with the Jupyter logo, 'Quit', and 'Logout' buttons, and a middle bar with 'Files', 'Running', and 'Clusters' tabs.

Créer un fichier Python

Fichier créé

PRESENTATION GENERALE DU JEU DE DONNEES :

1/ L'environnement de développement

Jeu de données
des arbres de la
ville de Paris



Ce PC > Téléchargements				
	Nom	Modifié le	Type	Taille
	p2-arbres-fr 1	17/04/2021 05:05	Fichier CSV Micro...	28 301 Ko

Importation du jeu
de données dans le
dossier source



localhost:8889/tree/Documents/Ingenieur%20IA/P2_Krairi_Jaoid/p2

jupyter Quit Logout

Files Running Clusters

Select items to perform actions on them. Upload New ↻

	Documents / Ingenieur IA / P2_Krairi_Jaoid / p2	Name	Last Modified	File size
	..		il y a quelques secondes	
	P2_01_notebook.ipynb		il y a 4 jours	236 kB
	p2-arbres-frjk.xlsx		il y a 21 jours	17.4 MB

PRESENTATION GENERALE DU JEU DE DONNEES :

2/ Importer les librairies Python et le jeu de données

Ouvrir le fichier
Python



localhost:8889/tree/Documents/Ingenieur%20IA/P2_Krairi_Jaoid/p2

jupyter

Files Running Clusters

Select items to perform actions on them.

Upload New

	Name	Last Modified	File size
<input type="checkbox"/>	..	il y a quelques secondes	
<input checked="" type="checkbox"/>	P2_01_notebook.ipynb	il y a 4 jours	236 kB
<input type="checkbox"/>	p2-arbres-frjk.xlsx	il y a 21 jours	17.4 MB

Importer les librairies

```
Entrée [1]: 1 import pandas as pd
            2 import numpy as np
            3 import seaborn as sb
            4 import matplotlib.pyplot as plt
```

Charger le fichier dans un dataframe

```
Entrée [2]: 1 df = pd.read_excel('p2-arbres-frjk.xlsx')
```

PRESENTATION GENERALE DU JEU DE DONNEES :

3/ Le résultat

Afficher les 5 première lignes de ma dataframe

Entrée [3]:

```
1 df.head(5)
```

Out[3]:

	id	type_emplacement	domanialite	arrondissement	complement_adresse	numero	lieu	id_emplacement	libelle_francais	genre	e
0	99874	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	19	Marronnier	Aesculus hippoc	
1	99875	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	20	If	Taxus	b
2	99876	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	21	If	Taxus	b
3	99877	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	22	Erable	Acer	ne
4	99878	Arbre	Jardin	PARIS 17E ARRDT	NaN	NaN	PARC CLICHY- BATIGNOLLES- MARTIN LUTHER KING	000G0037	Arbre À miel	Tetradium	c

Afficher les 5 dernières lignes de ma dataframe

Entrée [4]:

```
1 df.tail(5)
```

Out[4]:

	id	type_emplacement	domanialite	arrondissement	complement_adresse	numero	lieu	id_emplacement	libelle_francais	genre	e
200132	2024741	Arbre	Jardin	BOIS DE VINCENNES	NaN	NaN	ARBORETUM DE L'ECOLE DU BREUIL - ROUTE DE LAF...	720170151	NaN	Castanea	
200133	2024742	Arbre	Jardin	BOIS DE VINCENNES	NaN	NaN	ARBORETUM DE L'ECOLE DU BREUIL - ROUTE DE LAF...	720170152	Arbre aux mouchors	Davidia in	
200134	2024743	Arbre	Jardin	BOIS DE VINCENNES	NaN	NaN	ARBORETUM DE L'ECOLE DU BREUIL - ROUTE DE LAF...	720170153	Peuplier	Populus	
200135	2024744	Arbre	Jardin	BOIS DE VINCENNES	NaN	NaN	ARBORETUM DE L'ECOLE DU BREUIL - ROUTE DE LAF...	720170154	Chêne	Quercus	
200136	2024745	Arbre	Jardin	BOIS DE VINCENNES	NaN	NaN	ARBORETUM DE L'ECOLE DU BREUIL - ROUTE DE LAF...	720170155	Raisiner	Hovenia	

Nombre de lignes et de colonnes

Entrée [5]:

```
1 df.shape
```

Out[5]: (200137, 18)

PRESENTATION GENERALE DU JEU DE DONNEES :

3/ Le résultat

Verifier les différents types de nos variable

Entrée [6]:

```
1 df.info()
```

```
<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_adresse     30901 non-null   object  
5   numero                 0 non-null       float64 
6   lieu                   200137 non-null  object  
7   id_emplacement         199892 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            184 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
```

Entrée [7]:

```
1 df.dtypes.value_counts()
```

```
Out[7]: object    11
float64    4
int64      3
dtype: int64
```

PRESENTATION GENERALE DU JEU DE DONNEES :

3/ Le résultat

Nombre de valeurs manquantes

Entrée [8]: 1 (df.isna().sum()/df.shape[0]).sort_values(ascending=True)

```
Out[8]: id                0.000000
hauteur_m              0.000000
circonference_cm        0.000000
geo_point_2d_a          0.000000
id_emplacement          0.000000
lieu                   0.000000
geo_point_2d_b          0.000000
arrondissement          0.000000
type_emplacement        0.000000
domanialite             0.000005
genre                  0.000080
libelle_francais        0.007480
espece                 0.008754
remarquable            0.315274
stade_developpement     0.335795
variete                0.816241
complement_adresse      0.845596
numero                 1.000000
dtype: float64
```

Entrée [9]: 1 df.isnull().sum().sum()

```
Out[9]: 666301
```

DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

1/ Supprimer des valeurs manquantes

```
Entrée [10]: 1 df = df.drop(columns=['id', 'type_emplacement', 'complement_adresse', 'numero', 'id_emplacement', 'variete'])
```

DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

2/ Couper notre data frame en 2 parties ((Catégorique : Qualitative) et (Numérique : Quantitative))

```
Entrée [11]: 1 cat_data=[]
2 num_data=[]
3
4 for i,c in enumerate(df.dtypes):
5     if c=="object":
6         cat_data.append(df.iloc[:,i])
7     else:
8         num_data.append(df.iloc[:,i])
9 cat_data=pd.DataFrame(cat_data).transpose()
10 num_data=pd.DataFrame(num_data).transpose()
```


DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

3/ Remplacer les valeurs manquantes Catégorique

```
Entrée [12]: 1 cat_data=cat_data.apply(lambda x:x.fillna(x.value_counts().index[0]))  
2 cat_data.isnull().sum()
```

```
Out[12]: domanialite      0  
arrondissement    0  
lieu              0  
libelle_francais  0  
genre            0  
espece           0  
stade_developpement  0  
dtype: int64
```

DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

4/ Vérifier les valeurs manquantes Numérique

```
Entrée [13]: 1 num_data = num_data.fillna(num_data.median())  
            2 num_data.isnull().sum()
```

```
Out[13]: circonference_cm    0  
         hauteur_m          0  
         remarquable         0  
         geo_point_2d_a       0  
         geo_point_2d_b       0  
         dtype: int64
```

DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

5/ Résumer statistique des variables Catégorique de manière rapide

Entrée [14]:

```
1 cat_data.describe()
```

Out[14]:

	domanialite	arrondissement	lieu	libelle_francais	genre	espece	stade_developpement
count	200137	200137	200137	200137	200137	200137	200137
unique	9	25	6921	192	175	539	4
top	Alignement	PARIS 15E ARRD	PARC FLORAL DE PARIS / ROUTE DE LA PYRAMIDE	Platane	Platanus	x hispanica	A
freq	104950	17151	2995	44005	42607	38161	131643

DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

6/ Résumer statistique des variables numérique de manière rapide

Entrée [15]:

```
1 num_data.describe()
```

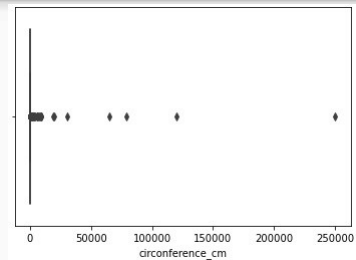
Out[15]:

	circonference_cm	hauteur_m	remarquable	geo_point_2d_a	geo_point_2d_b
count	200137.000000	200137.000000	200137.000000	200137.000000	200137.000000
mean	83.380479	13.110509	0.000919	48.854491	2.348208
std	673.190213	1971.217387	0.030307	0.030234	0.051220
min	0.000000	0.000000	0.000000	48.742290	2.210241
25%	30.000000	5.000000	0.000000	48.835021	2.307530
50%	70.000000	8.000000	0.000000	48.854162	2.351095
75%	115.000000	12.000000	0.000000	48.876447	2.386838
max	250255.000000	881818.000000	1.000000	48.911485	2.469759

DEMARCHE METHODOLOGIQUE DE L'ANALYSE DE DONNEES :

7/ Les valeurs aberrantes

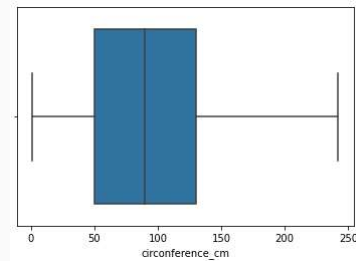
AVANT



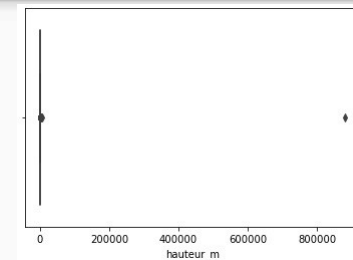
METHODE

```
Entrée [20]: 1 mask = (df['circonference_cm']>0) & (df['circonference_cm']<Upper_Whisker)
2 df['circonference_cm'] = df['circonference_cm'].where(mask, df.loc[mask, 'circonference_cm'].quantile(0.80))
```

APRES



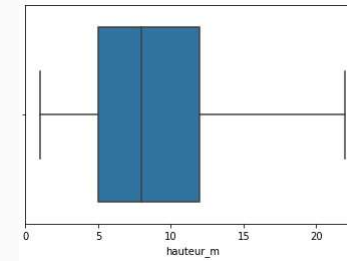
AVANT



METHODE

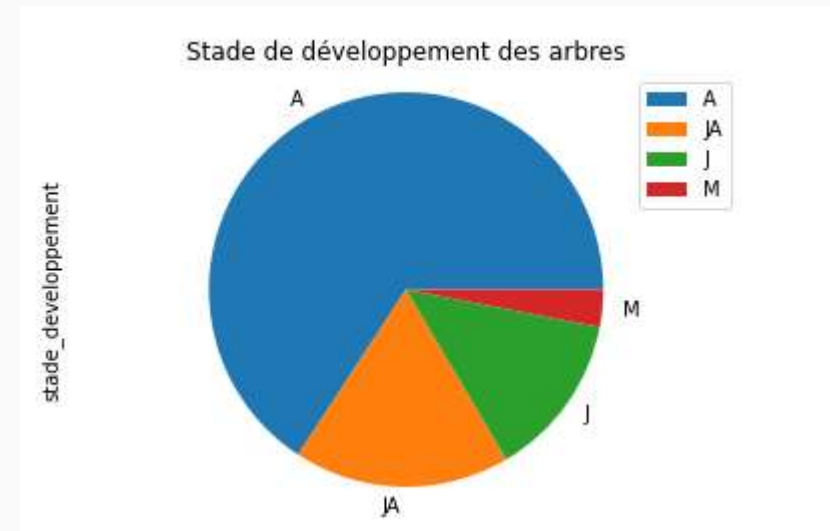
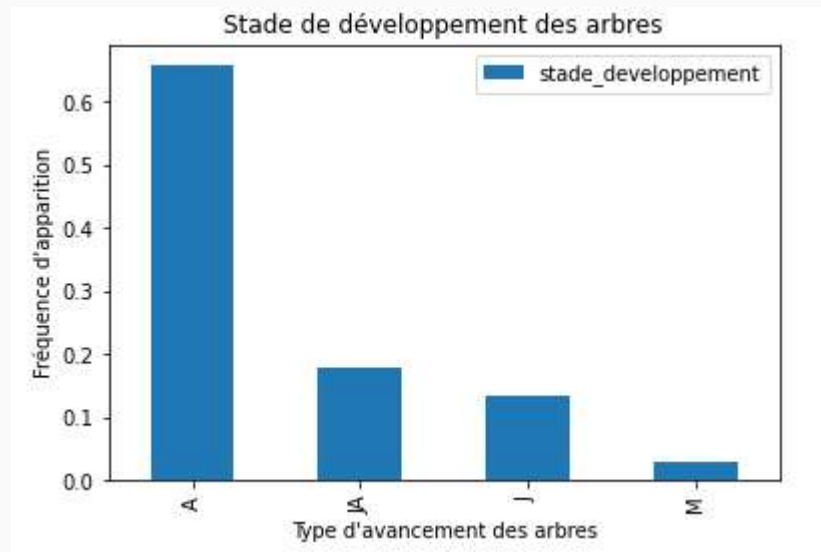
```
Entrée [24]: 1 mask = (df['hauteur_m']>0) & (df['hauteur_m']<Upper_Whisker)
2 df['hauteur_m'] = df['hauteur_m'].where(mask, df.loc[mask, 'hauteur_m'].quantile(0.20))
```

APRES



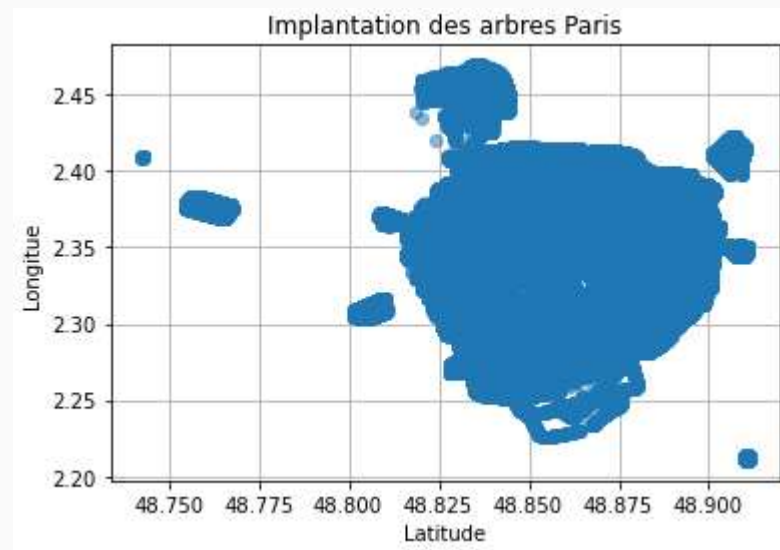
SYNTHESE DE MON ANALYSE DE DONNEES:

1/ Représentation d'une variable sous forme graphique



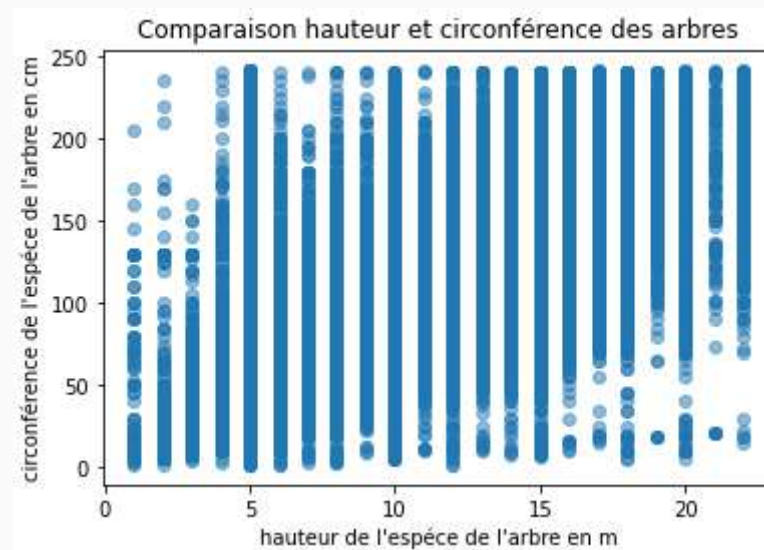
SYNTHESE DE MON ANALYSE DE DONNEES:

2/ Carte géographique des implantations d'arbres



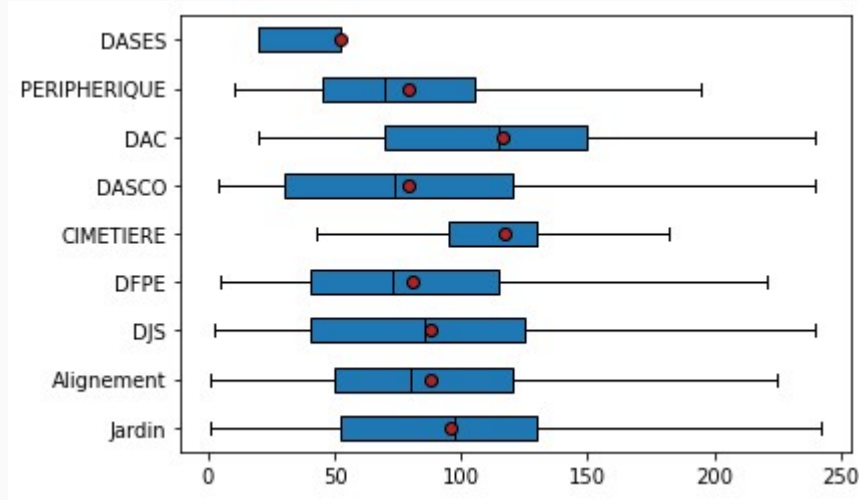
SYNTHESE DE MON ANALYSE DE DONNEES:

3/ Analyser la corrélation entre 2 variables quantitative



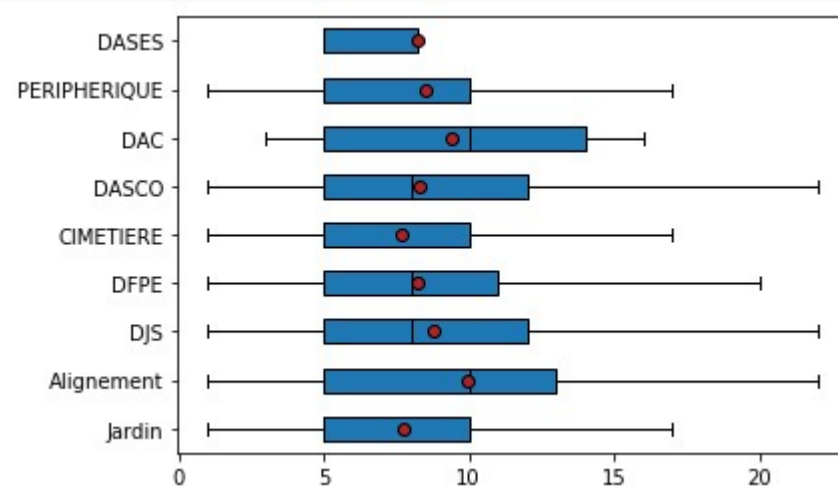
SYNTHESE DE MON ANALYSE DE DONNEES:

4/ Analysez la variable quantitative "circonference_cm" et une qualitative "domanialite"



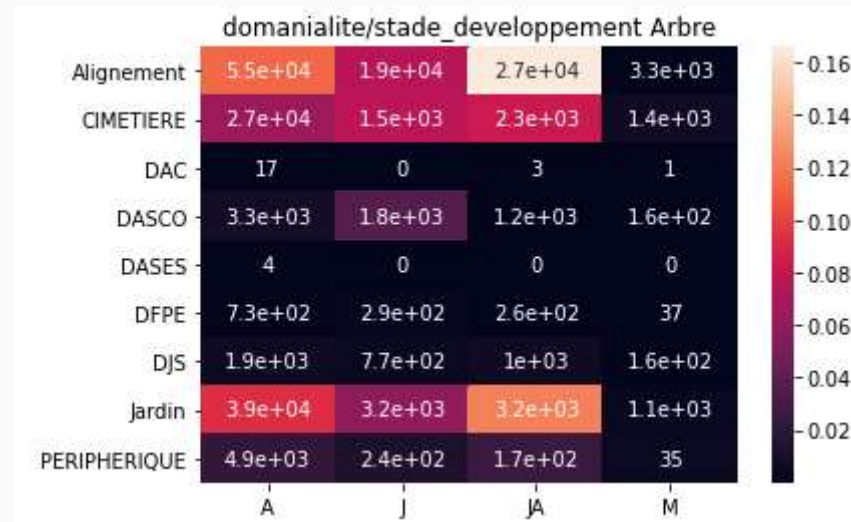
SYNTHESE DE MON ANALYSE DE DONNEES:

5/ Analysez la variable quantitative "hauteur_m" et une qualitative "domanialite"



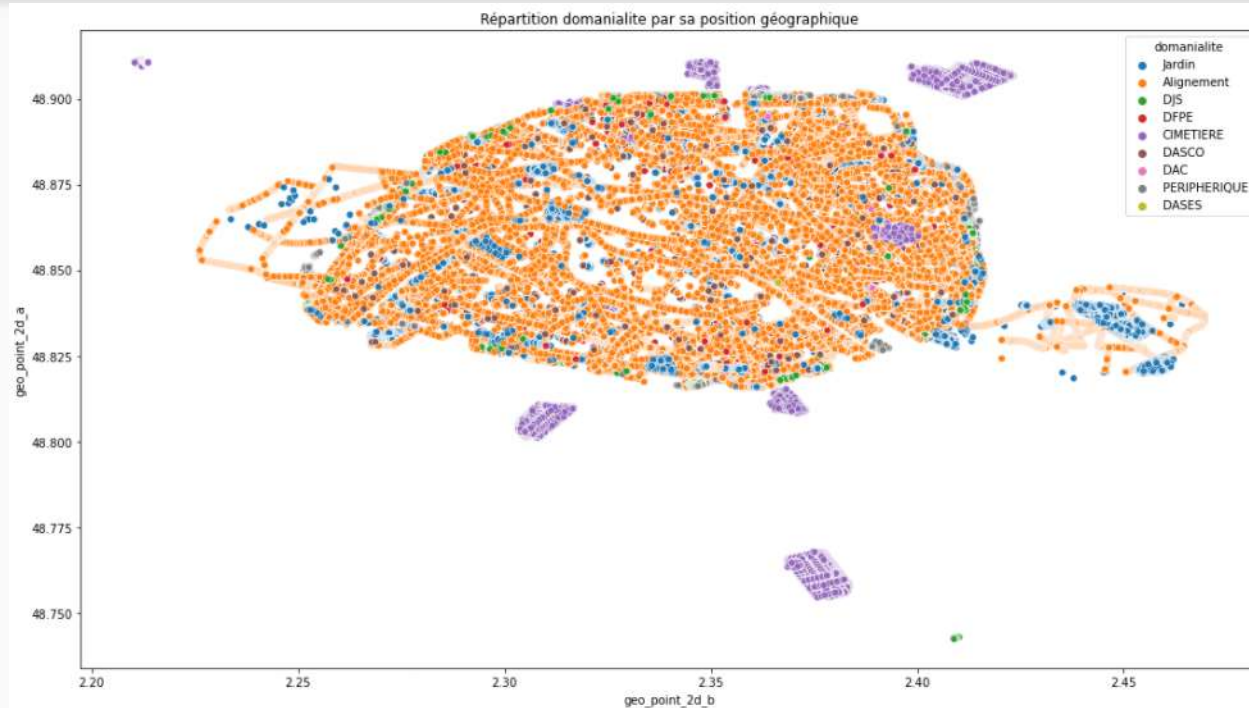
SYNTHESE DE MON ANALYSE DE DONNEES:

6/ Analyse de 2 variables qualitatives "domanialite" et "stade_developpement"



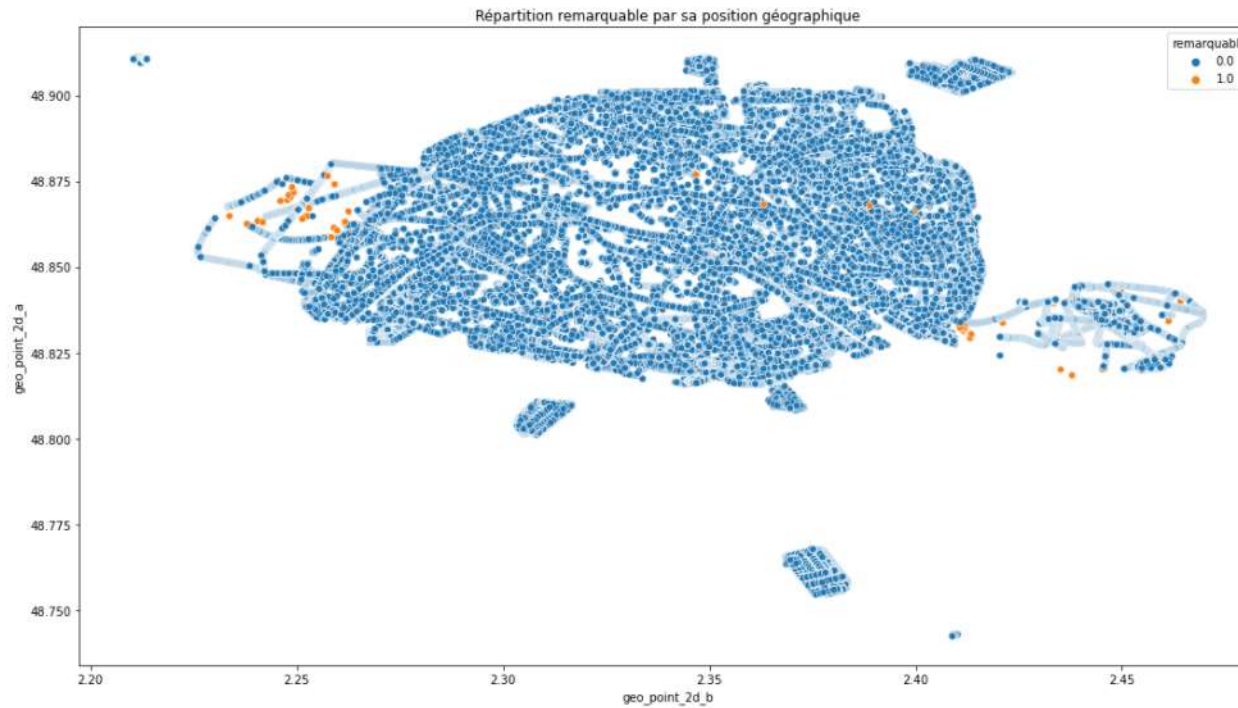
SYNTHESE DE MON ANALYSE DE DONNEES:

7/ Analyse d'une variable qualitatives « domanialite » et sa position géographique



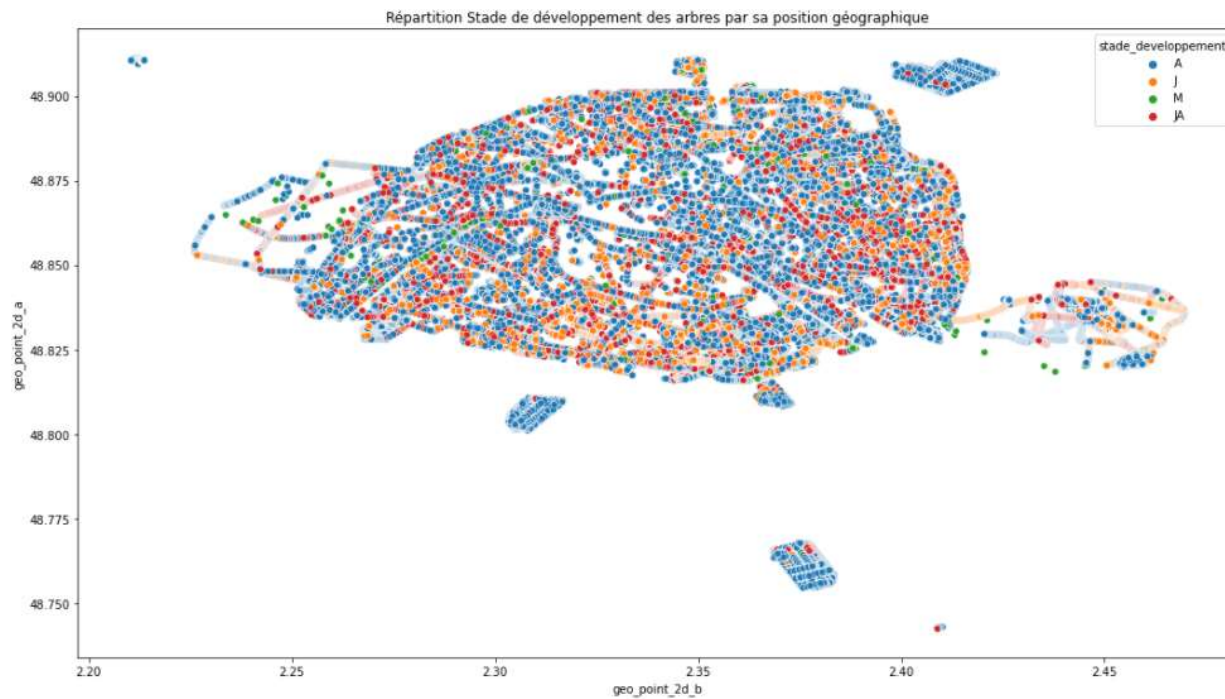
SYNTHESE DE MON ANALYSE DE DONNEES:

8/ Analyse d'une variable quantitative « remarquable » et sa position géographique



SYNTHESE DE MON ANALYSE DE DONNEES:

9/ Analyse d'une variable qualitative « stade_développement » et sa position géographique



SYNTHESE DE MON ANALYSE DE DONNEES:

10/Analyse statistique par stade de développement des arbres de Paris

1/ Arbre de petite taille

stade_developpement	domanialite	arrondissement	lieu	libelle_francais	genre	espece	circonference_cm	hauteur_m	remarquable	geo_point_2d_a	geo_point_2d_b
A	Alignement	BOIS DE BOULOGNE	5 RUE CLEMENT MYIONNET = 22 SEBASTIEN MERCIER	Abelia	Abelia	Fleur n. sp.	1.0	1.0	0.0	48.742371	2.210241
J	Alignement	BOIS DE BOULOGNE	28 BOULEVARD DE DOUAUMONT	Abricotier	Abies	Fleur n. sp.	1.0	1.0	0.0	48.759069	2.212541
JA	Alignement	BOIS DE BOULOGNE	ALAIN FOURNIER (5-7) TEP	Abricotier	Abies	Fleur n. sp.	4.0	1.0	0.0	48.742290	2.212229
M	Alignement	BOIS DE BOULOGNE	ALLEE DES CYGNES	Ailante	Acer	alba	3.0	2.0	0.0	48.742421	2.211263

SYNTHESE DE MON ANALYSE DE DONNEES:

10/Analyse statistique par stade de développement des arbres de Paris

2/ Arbre de grande taille

	domanialite	arrondissement	lieu	libelle_francais	genre	espece	circonference_cm	hauteur_m	remarquable	geo_point_2d_a	geo_point_2d_b
stade_developpement											
A	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Xanthoceras	x Sorbopyrus	yunnanensis	242.0	22.0	1.0	48.911485	2.469752
J	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Virgilier	Cupressocyparis ^x	xerubescens	240.0	20.0	1.0	48.910713	2.469759
JA	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Virgilier	Cupressocyparis ^x	xerubescens	240.0	22.0	1.0	48.910689	2.469718
M	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Tulipier	Cupressocyparis ^x	x hollandica	242.0	22.0	1.0	48.911003	2.465116

SYNTHESE DE MON ANALYSE DE DONNEES:
10/Analyse statistique par stade de développement des arbres de Paris
3/ Arbre de taille moyenne

	circonference_cm	hauteur_m
stade_developpement		
A	107.496965	9.350547
J	39.117534	5.741359
JA	70.185730	8.932146
M	175.868968	14.309341

SYNTHESE DE MON ANALYSE DE DONNEES:
10/Analyse statistique par stade de développement des arbres de Paris
4/ Ecart-type

	circonference_cm	hauteur_m
stade_developpement		
A	47.055011	4.836296
J	19.651238	2.032855
JA	25.797135	3.202862
M	42.514226	6.077166

SYNTHESE DE MON ANALYSE DE DONNEES:

11/Analyse de jeu de données 'entretiens des arbres de Paris'

1/ Arbre de petite taille



Arbre de petite
taille

stade_developpement	domanialite	arrondissement	lieu	libelle_francais	genre	espece	circonference_cm	hauteur_m	remarquable
A	Alignement	BOIS DE BOULOGNE	5 RUE CLEMENT MYIONNET = 22 SEBASTIEN MERCIER	Abelia	Abelia	Fleur n. sp.	1.0	1.0	0.0
J	Alignement	BOIS DE BOULOGNE	28 BOULEVARD DE DOUAUMONT	Abricotier	Abies	Fleur n. sp.	1.0	1.0	0.0
JA	Alignement	BOIS DE BOULOGNE	ALAIN FOURNIER (5-7) TEP	Abricotier	Abies	Fleur n. sp.	4.0	1.0	0.0
M	Alignement	BOIS DE BOULOGNE	ALLEE DES CYGNES	Ailante	Acer	alba	3.0	2.0	0.0

SYNTHESE DE MON ANALYSE DE DONNEES:

11/Analyse de jeu de données 'entretiens des arbres de Paris'

2/ Arbre de grande taille





Arbre de grande
taille

	domanialite	arrondissement	lieu	libelle_francais	genre	espece	circonference_cm	hauteur_m	remarquable
stade_developpement									
A	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Xanthoceras	x Sorbopyrus	yunnanensis	242.0	22.0	1.0
J	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Virgilier	Cupressocyparis ^x	xerubescens	240.0	20.0	1.0
JA	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Virgilier	Cupressocyparis ^x	xerubescens	240.0	22.0	1.0
M	PERIPHERIQUE	VAL-DE-MARNE	YEO THOMAS (5) MAT	Tulipier	Cupressocyparis ^x	x hollandica	242.0	22.0	1.0

SYNTHESE DE MON ANALYSE DE DONNEES:
11/Analyse de jeu de données 'entretiens des arbres de Paris'
3/ Conclusion

Arbre de petite taille	Arbre de grande taille
Adulte	Adulte
Alignement	Périphérique
BOIS DE BOULOGNE	VAL-DE-MARNE
Abelia	Xanthoceras
Taille 1cm X 1m	Taille 242cm X 22m

 Non entretenue
 Entretienue

CONCLUSION

- ✓ L'environnement de développement est installé et fonctionnel sur mon poste de travail (Python et Jupyter),
- ✓ Un environnement virtuel a été créé pour assurer l'isolement de mon projet et la gestion des dépendances,
- ✓ Les librairies python spécialisées ont été importées dans le Jupyter Notebook,
- ✓ Le jeu de données a été décrit brièvement en ayant chargé le fichier dans un data frame,
- ✓ Des indicateurs statistiques basiques (moyenne et écart-type) ont été calculés pour les différentes colonnes figurant dans mon fichier 'P2_01_notebook',
- ✓ Les ordres de grandeur, des grandeurs statistiques des différentes colonnes ont été comparés figurant dans mon fichier 'P2_01_notebook',
- ✓ Les fonctionnalités d'édition de cellule Markdown du Jupyter Notebook sont utilisées dans au moins trois cellules pour commenter l'analyse et la mettre en forme dans mon fichier 'P2_01_notebook',
- ✓ Les titres des trois parties sont visuellement en évidence dans mon fichier 'P2_01_notebook',
- ✓ Les moyennes, médianes et quantiles des distributions sont calculées pour les valeurs quantitative,
- ✓ Au moins une représentation graphique d'une distribution statistique a été tracée,
- ✓ Les éventuelles valeurs aberrantes ont été identifiées,
- ✓ Les graphiques sont lisibles

REMERCIEMENT

- ❖ Remercier mon Mentor Mr Moussa CAMARA
- ❖ Remercier OPENCLASSROOM
- ❖ Merci de m'avoir écouté

REPONDRE AUX QUESTIONS