



Projet Python :

Les arbres de Grenoble



L'emblématique arbre de Venon

Source : francebleu.fr

Objectifs du module

L'objectif de ce projet est de vous plonger, tête la première dans de l'analyse de données, sans aucun pré-requis technique. A l'issue de ce projet, vous aurez vu les bases de la **manipulation de données grâce à Python**.

Nous sommes conscients qu'il existe parmi vous une disparité de connaissances relatives au monde informatique, à la programmation et à l'analyse de données. Certains d'entre vous ont probablement déjà codé (régulièrement ou non) au cours de leur vie personnelle ou professionnelle (python, matlab, R ou autre) alors que d'autres découvrent ce monde aujourd'hui.

Il est donc tout à fait normal que le projet à suivre paraisse facile à certaines personnes et difficile pour d'autres. Tant pis, et tant mieux ! Nous vous encourageons autant que possible à vous entraider, à transmettre, et à collaborer. Prenez votre temps au Campus Numérique comme un **espace disponible vous permettant de vous auto-former et d'expérimenter à plusieurs**. N'hésitez pas à approfondir vos connaissances et à aller chercher plus loin que ce qui est strictement attendu de vous si vous en sentez le besoin ou le désir.

Le rôle des formateurs au cours de ce projet (et de la formation de manière plus générale) sera celui d'**accompagnateur méthodologique**. Nous n'aurons probablement pas la réponse à toutes vos questions.

Modalités

- Durée du projet : 5 jours.
- Ce projet sera réalisé en autonomie
- L'entraide est (très) fortement encouragée !

Description du projet

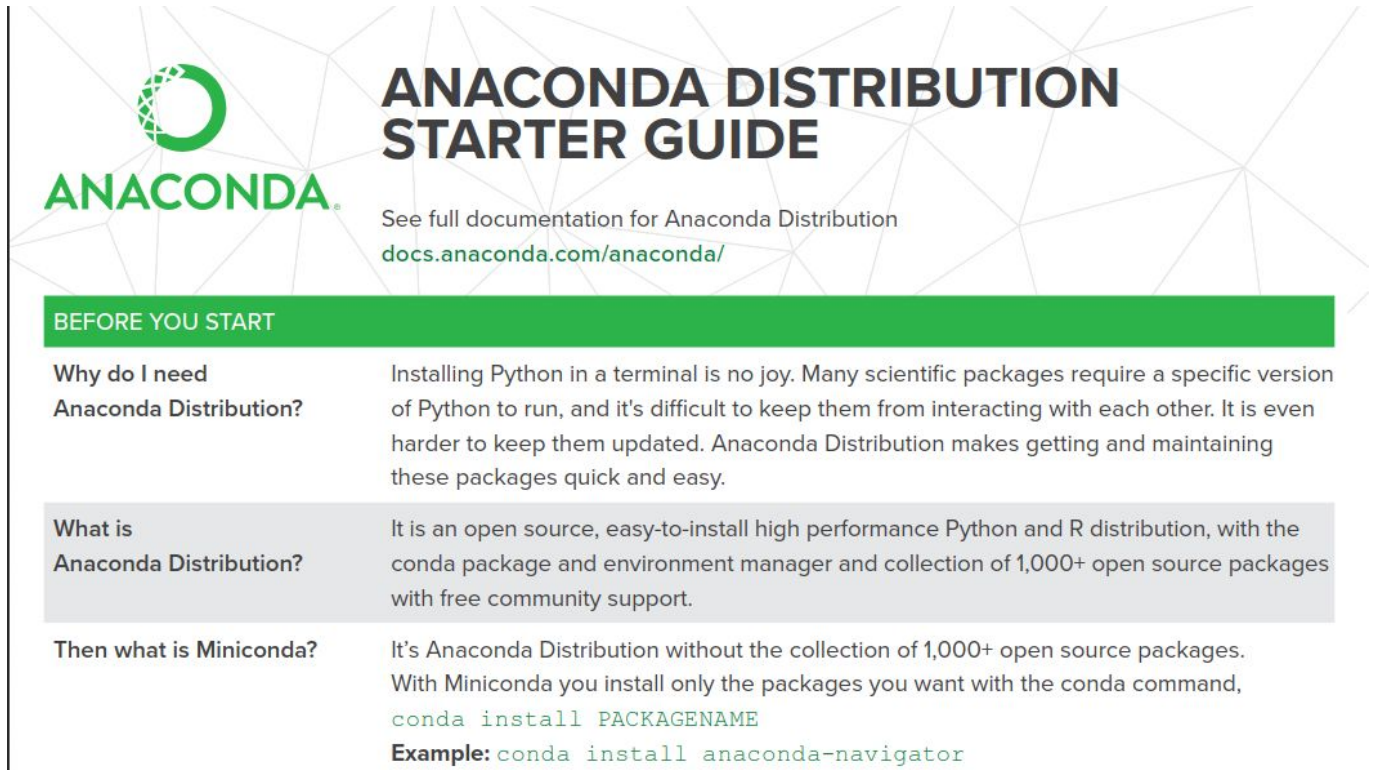
- Ce projet s'insère dans le cadre de l'**Open Data**, et en particulier des données libérées par la Ville de Grenoble. Ce jeu de données contient des informations relatives aux arbres situés sur le territoire de la ville : numéro d'identification, date de plantation, genre botanique, espèce, etc. Ce jeu de données est disponible à cette adresse : <http://data.metropolegrenoble.fr/ckan/dataset/les-arbres-de-grenoble>
- Vous explorerez les données via un **Jupyter notebook**, particulièrement adapté pour le calcul scientifique et la Data Science. Jupyter est un projet open-source et gratuit.



Installation de l'environnement Anaconda

Objectifs de l'activité

- Mettre en place l'environnement de développement scientifique avec Anaconda



The graphic is a 'Starter Guide' for Anaconda Distribution. It features the Anaconda logo (a green snake head) and the text 'ANA CONDA' in green. The main title is 'ANACONDA DISTRIBUTION STARTER GUIDE' in large, bold, black letters. Below the title, it says 'See full documentation for Anaconda Distribution' and provides the URL 'docs.anaconda.com/anaconda/'. A green horizontal bar contains the text 'BEFORE YOU START'. Below this bar, there are three sections with questions and answers. The first section asks 'Why do I need Anaconda Distribution?' and answers that installing Python in a terminal is difficult and Anaconda makes it easier. The second section asks 'What is Anaconda Distribution?' and answers that it is an open source, easy-to-install high performance Python and R distribution. The third section asks 'Then what is Miniconda?' and answers that it is Anaconda Distribution without the collection of 1,000+ open source packages. It also provides a command to install a package: `conda install PACKAGENAME` and an example: `conda install anaconda-navigator`.

ANA CONDA

ANACONDA DISTRIBUTION STARTER GUIDE

See full documentation for Anaconda Distribution
docs.anaconda.com/anaconda/

BEFORE YOU START

Why do I need Anaconda Distribution? Installing Python in a terminal is no joy. Many scientific packages require a specific version of Python to run, and it's difficult to keep them from interacting with each other. It is even harder to keep them updated. Anaconda Distribution makes getting and maintaining these packages quick and easy.

What is Anaconda Distribution? It is an open source, easy-to-install high performance Python and R distribution, with the conda package and environment manager and collection of 1,000+ open source packages with free community support.

Then what is Miniconda? It's Anaconda Distribution without the collection of 1,000+ open source packages. With Miniconda you install only the packages you want with the conda command,
`conda install PACKAGENAME`
Example: `conda install anaconda-navigator`

Consignes

- Pour Linux :
 - Télécharger <https://www.anaconda.com/distribution/#linux>
 - Choisir Python 3.7 => 64-bit (x86) installer (506MB)
 - Consignes d'installation : <https://docs.anaconda.com/anaconda/install/linux/>
 - Pour ouvrir un terminal à l'endroit où vous avez téléchargé Anaconda
 - Click droit => Open Terminal Here
 - Vous pouvez alors exécuter les instructions de la doc
- Pour Windows
 - Idem mais avec l'interface graphique !

Itération 1

Exploration préliminaire du dataset “Arbres”

Objectifs de l'activité

- Installer Anaconda
- Nous nous Intéresserons nous dans un premier temps à quelques **structures de données internes à Python** - *built-in data structures* - sans ajout de librairie spécialisée (type *pandas*) et aux opérations de bases (**boucles et conditions**)
- Ouverture et lecture du contenu d'un fichier “.csv”
- Exploration préliminaire des données.

Compétences

- Ouvrir et lire le contenu d'un fichier .csv
- Utiliser des boucles for
- Utiliser des structures conditionnelles
- Manipuler des listes et tuples

Consignes

1. Avant d'explorer les données pensez à **ranger votre arborescence de fichier**. Vous allez très rapidement accumuler de nombreux dossiers et fichiers, pas question de tout mettre en b@r#*l sur le bureau ! Vous y perdriez beaucoup de temps et d'énergie assez rapidement. Au sein de votre répertoire de travail, créez un dossier spécifique à ce projet ayant une structure :

```
Projet_arbre
|_ data
    |_ arbres.csv
|_ notebooks
    |_ arbres.nbpy
```

2. Avant même de commencer à coder, il est très important d'observer les données. Ouvrez donc vos données dans un éditeur de texte (vscode, sublime text, etc.) ou même dans le *bash* (*cat* ou *nano* par exemple). Ou même dans Excel 2000 ! Regardez les données à traiter droit dans les yeux !
3. A l'aide du notebook, vous écrirez maintenant du code (uniquement à l'aide du module csv de la librairie standard dans un premier temps) permettant de :
 - ❑ Lire les deux premières ligne du csv des arbres de Grenoble (module csv, voir doc officielle). Que représentent elles ?
 - ❑ Extraire l'information concernant l'année de plantation pour l'ensemble des lignes (*rows*) et stocker ces informations dans une variable de type *List*
 - ❑ Afficher les 50 premières lignes et les 50 dernières lignes de cette liste.
 - ❑ Pour combien d'arbres manque-t-il l'information concernant la date de plantation ?

- ❑ Extraire l'ensemble des lignes (*rows*) et stocker ces informations dans une variable de type *List*
- ❑ Combien d'arbres sont recensés dans ce jeu de données ?
- ❑ Combien d'arbres ont été plantés en moyenne chaque année ?
- ❑ Afficher tous les arbres ayant été plantés l'année de votre naissance.
- ❑ Combien d'arbres ont été plantés depuis votre naissance ?
- ❑ Quel Maire a planté le plus d'arbre à Grenoble ?
- ❑ Utilisez Matplotlib pour tracer l'histogramme représentant le nombre d'arbres plantés par année.
- ❑ Récupérez maintenant l'information concernant le genre botanique et la stocker conjointement avec l'année de plantation dans un tuple. Les tuples seront stockés dans une liste.
- ❑ Pour combien d'arbre manque-t-il l'information concernant le genre botanique ?
- ❑ Triez les tuples en fonction de l'année de plantation par ordre décroissant.

Note sur les ressources :

Les ressources que vous trouverez sur Internet concernant Python (et plus largement sur les langages de programmation) sont de plusieurs types : **tutoriels** (écrits, vidéos, livres), **how-to** (orientés sur un but précis) et **forums** (stackoverflow par exemple) et **documentation officielle**. Vous serez régulièrement amenés à jongler entre toutes ces sources d'information. Ne négligez pas la documentation officielle! Elle est souvent aride, mais elle est exhaustive. Il est important de s'y confronter régulièrement pour finir par la dompter.

Ressources

- Documentation officielle Python 3
 - <https://docs.python.org/3/library/csv.html>
 - <https://docs.python.org/3/howto/sorting.html>
 - <https://docs.python.org/3/tutorial/introduction.html#lists>
- Documentation W3C
 - https://www.w3schools.com/python/python_datatypes.asp
- Tutoriel par Google :
 - <https://developers.google.com/edu/python>
- Documentation Matplotlib :
 - <https://matplotlib.org/tutorials/index.html>
- Un peu de biologie (désolé pour les vrais biologistes...) :
 - [https://fr.wikipedia.org/wiki/Genre_\(biologie\)](https://fr.wikipedia.org/wiki/Genre_(biologie))

Interlude

The Weird Trees Game

Un second fichier recensant les mêmes arbres a été découvert. Celui-ci s'appelle "weird_trees.csv". Ce fichier a été corrompu de manière aléatoire.

A ↔ あ

Pour des raisons obscures qui nous échappent - probablement politiques - nous sommes missionnés pour nous assurer que ce fichier redevienne strictement identique au fichier original. Nous ne pouvons pas simplement l'effacer et faire une copie du fichier original.

Repérez les corruptions et réparez !

Ressources

- W3C
 - https://www.w3schools.com/python/python_ref_string.asp

Avertissement préliminaire



Les métiers de Data Analyst et Data Scientist visent à recueillir des données, les traiter et s'en servir pour extraire une information "métier" pertinente.

Le projet sur lequel nous travaillons a un intérêt en terme pédagogique (python) et également un potentiel intérêt réel à l'échelle du territoire de la métropole grenobloise. Cependant, il est important de garder en tête que sans les compétences métiers associées au projet - dans notre cas celles des urbanistes, pépiniéristes, pédologues et biologistes - les analyses effectuées n'auront que peu d'intérêt réel.

Itération 2

Analyse de l'état de la diversité des arbres

Objectifs de l'activité

- Analyser la répartition des espèces d'arbres à Grenoble, l'évolution de cette diversité au cours du temps, comparaison avec les objectifs de la Métro en terme de diversification des espèces.
- Utilisation d'autres structures de données : Sets, Dict

Contexte

Diversifier les essences d'arbres en ville est crucial pour de nombreuses raisons : résilience des écosystèmes urbains, lutte contre les îlots de chaleur, renouvellement de l'air, etc. Or, le constat fait par la Métropole de Grenoble est que la diversité des arbres est faible sur le territoire (voir ci-dessous).



Constat : Une faible diversité des arbres en ville

Une faible diversité du patrimoine arboré dans les villes est observée, c'est aussi le cas au sein du territoire de Grenoble-Alpes Métropole. Aussi, des efforts pour diversifier les plantations sont faits depuis plusieurs années.

La diversité des arbres en ville

La diversité des arbres en ville est considérée comme bonne, d'après la règle du 10-20-30.

- pas plus de 10% d'une espèce;
- pas plus de 20% d'un genre;
- pas plus de 30% d'une famille (règle respectée sur le territoire).

Cependant, planter le bon arbre, au bon endroit, et s'assurer de sa croissance par des études climatiques, pédologiques, des contraintes du lieu de plantation est aussi un moyen de s'assurer d'un maintien de la diversité arborée.

Source : <https://arbres.lametro.fr/811-le-choix-d-un-arbre-en-pepiniere.htm>

Un parc paysager plus varié

Sur l'ensemble du territoire, la Ville de Grenoble a choisi **d'accentuer la diversité des espèces**. En 2006, Grenoble en comptait 292 contre plus de **500 espèces en 2017**.

Source : dossier de presse de la ville de Grenoble (2017 ?)

Consignes

- Pour commencer identifiez tous les genres botaniques représentés à Grenoble. Combien y en a-t-il ? Plusieurs manières de résoudre ce problème :
 - Essayez différentes méthodes et chronométrez les temps d'exécution de vos instructions. Comparez avec les autres membres de votre groupe et essayez de comprendre les différences.
- Identifiez maintenant toutes les espèces différentes présentes à Grenoble. N'oubliez pas que les espèces sont une sous catégorie de genre botanique ! (l'espèce est 'Platanus Acerifolia' et pas uniquement 'Acerifolia')
- En utilisant la structure de données *Dict* compter le nombre de représentants de chaque genre botanique :
 - Quels sont les genres botaniques les plus représentés ? Quelle est la proportion par rapport au total ? Les objectifs de la Métro sont-ils respectés en terme de diversité ?
 - Listez tous les représentants uniques de leurs genre botanique
- Idem précédemment mais en comptant les espèces.

- Comment la pratique de plantation a-t-elle évolué au cours du temps ?
Diversifie-t-on plus maintenant qu'avant ? Vos analyses sont-elles en accord avec le communiqué de presse ? (plus dur)

Ressources

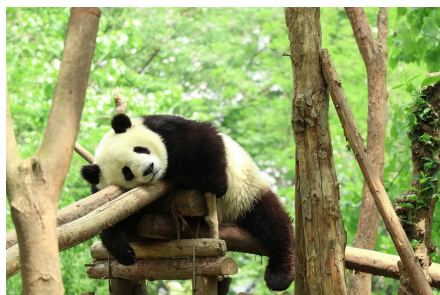
- Doc officielle :
 - Timer : <https://ipython.readthedocs.io/en/stable/interactive/magics.html>
 - Collections : <https://docs.python.org/2/library/collections.html>
- Lien vers le dossier de presse :
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&cad=rja&uact=8&ved=2ahUKEwiPr-TSII3mAhWPHhQKH5qLDGoQFjAEegQIChAC&url=https%3A%2F%2Fwww.grenoble.fr%2Fuploads%2FExterne%2F96%2F470_610_2-700-arbres-plantes-en-3-saisons.pdf&usq=AOvVaw1P0stH-XP1wSivUxmsa6cN

Itération 3

Analyse et stratégie

Objectifs de l'activité

- Analyser la répartition des espèces d'arbres à Grenoble, l'évolution de cette diversité au cours du temps, comparaison avec les objectifs de la Métro en terme de diversification des espèces et proposition de stratégie de plantation.
- Utilisation de pandas



La vie est plus simple avec un panda

@ [Needpix.com](https://www.needpix.com)

Consignes

- Commencez par vous familiariser avec pandas en manipulant un peu les données. Vous pouvez même reprendre certaines des consignes précédentes.
 - Essayez de reprendre certains des problèmes précédents **sans utiliser de boucle for**.
- Plusieurs sources évoquent une espérance de vie moyenne des arbres en ville de 30 à 80 ans¹². Les données de terrain disponibles pour Grenoble corroborent-elles ces sources ?
- En vous servant de l'annexe A, identifiez toutes les essences "non désirées" à Grenoble. Quelle est la proportion "essences désirées" vs "non désirées" ?
- Essayer de proposer une liste d'espèce à planter qui permet de remplir les objectifs annoncés en terme de diversification.

Ressources

- Towards Data Science :
 - <https://towardsdatascience.com/21-pandas-operations-for-absolute-beginners-5653e54f4cda>

¹ https://fr.wikipedia.org/wiki/Arbre_urbain

² <https://www.natural-solutions.eu/blog/arbre-en-milieu-urbain>

Itération 4

Distribution géographique des arbres

Objectifs de l'activité

- Analyser et représenter la distribution géographique des arbres.

Consignes

- Mettez les arbres sur une carte en utilisant Folium.
- Calculer la distance entre deux arbres de deux manières différentes
 - Grâce au module geopy
 - En implémentant vous même la formule de Haversine
- Quels sont les 10 arbres les plus proches de chez vous ?
- Quelle est la distance moyenne entre les arbres ?
- Quel est l'arbre le plus entouré ?
- Quel est l'arbre le plus solitaire ?

Annexe A : Essences non désirées à Grenoble

LISTE DES ESPECES A EVITER AUX ABORDS DES ESPACES CIRCULES

Genre, espèce*	Argumentaire
Acer saccharinum	Supporte mal la taille (danger de chute de branches dans un milieu très fréquenté), désordre racinaire.
Albizzia julibrissin	Remontée de couronne difficile sur des grandes tiges (petit arbre).
Catalpa bignonioides	Bois cassant.
Ginkgo biloba	Fruits nauséabonds (pieds femelle).
Gleditsia triacanthos	Remontée de couronne difficile sur des grandes tiges (petit arbre).
Koelreuteria paniculata	Remontée de couronne difficile sur des grandes tiges (petit arbre).
Laburnum anagyroides	Remontée de couronne difficile sur des grandes tiges (petit arbre).
Paulownia tomentosa	Remontée de couronne difficile sur des grandes tiges.
Prunus pissardi	Maladie : phellin, faible résistance à la taille.
Sophora japonica	Production de bois mort, désordre racinaire pouvant altérer la chaussée.

LISTE DES ESPECES SUPPORTANT MAL LE CLIMAT URBAIN DE LA METROPOLE

Genre, espèce	Argumentaire
Acer platanoides	Supporte mal la sécheresse.
Acer pseudoplatanus	Supporte mal la sécheresse.
Betula costata, davurica, jacquemontii, lenta, pendula,maximowicziana, pubescens, utilis	Supporte mal la sécheresse.
Cryptomeria japonica	Supporte mal la sécheresse.
Cupressus cashmeriana	Supporte mal la sécheresse.
Davidia involucrata	Supporte mal la sécheresse.
Fagus sylvatica	Supporte mal la sécheresse (sensible aux échaudures en tige).
Juglans nigra, ailantifolia	Supporte mal la sécheresse.
Liriodendron tulipifera	Supporte mal la sécheresse.
Magnolia grandiflora	Supporte mal la sécheresse.
Nerium oleander	Supporte mal le gèle et la sécheresse.
Picea abies	Supporte mal le gèle et la sécheresse.
Populus alba	Désordre racinaire, faible longévité.
Quercus bicolor, macrocarpa, nigra, rubra, alba	Supporte mal la sécheresse.
Salix gracolistyla,daphnoides, cinerea, arenaria, alba, acutifolia	Supporte mal la sécheresse.
Sorbus aucuparia	Supporte mal la sécheresse.

LISTE DES ESPECES SUJETTES AUX PATHOGENES

Genre, espèce	Argumentaire
Aesculus	Maladies : chancre bactérien, mineuse
Fraxinus	Maladie : la chalarose
Platanus	Maladie : chancre coloré

LISTE DES ESPECES INVASIVES

A éviter absolument

Genre, espèce
Ailanthus altissima
Robinia pseudoacacia
Acer negundo

LISTE DES ESPECES SUR REPRESENTEES A L'ECHELLE DU TERRITOIRE

Non souhaités par la Métropole dans un soucis de diversification des espèces

Genre, espèce	Argumentaire supplémentaire
Platanus x acerifolia	Chancre coloré (toutes espèces confondues)
Acer platanoides et pseudoplatanus	Sensible aux modifications climatique à l'échelle de la métropole

LISTE DES GENRES ALLERGENES PRINCIPAUX

Espèces anémophiles à éviter de planter à proximité de zones très fréquentées

Genre
Betula
Castanea
Cupressus
Fraxinus
Olea
Platanus