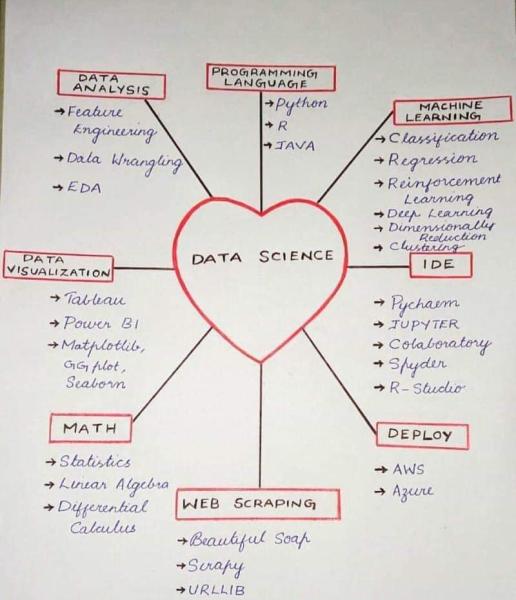
# **Programme Data Science**



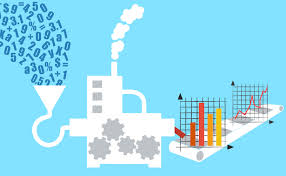
<https://pixelastic.github.io/pokemonorbigdata/>

# A lire – Important :

Modalités pédagogiques, livrables et évaluation : [Cliquer ici](https://1drv.ms/w/s!AmJGbSlW18YGq-YHEcsgT5hYDviImg?e=ym6EPO)

# Grands Projets (pour la présentation finale)

**Accéder à la liste des** [**grands projets**](https://1drv.ms/w/s!AmJGbSlW18YGq-dbhQeOJHtujQ9s_w?e=wf4R8f)

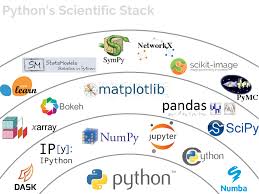
** **

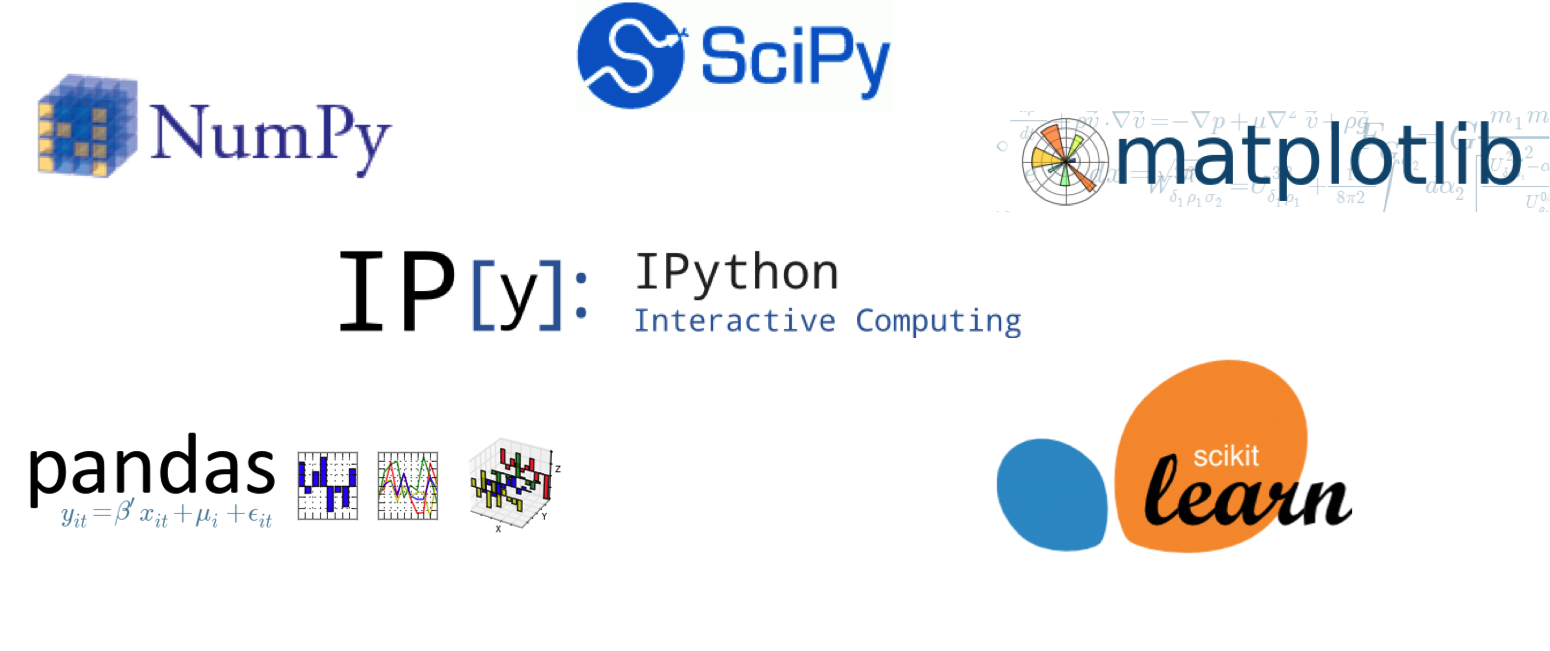
# Antisèches

* + - Python [[memonto complet](https://1drv.ms/b/s!AmJGbSlW18YGq-kIg9Xr_JQqg0Kzmg?e=q5gvoY)] [[niveau de base](https://1drv.ms/b/s!AmJGbSlW18YGp-gKOxKUvSyS87ZlBw?e=mdKtGo)] [[intermédiaire](https://1drv.ms/b/s!AmJGbSlW18YGp-gJ1Gei5cYMJRUtcw?e=5ysauh)]
    - Numpy [[datacamp](https://1drv.ms/b/s!AmJGbSlW18YGq-YKtNXAWLTgaRDpvw?e=RazTgz)] [[dataquest](https://1drv.ms/b/s!AmJGbSlW18YGq-YJAI7r6ItLkKL5WA?e=NnadQb)]
    - Data Wrangling avec Pandas [[cliquer ici](https://1drv.ms/b/s!AmJGbSlW18YGp-gIt2sV8pNk4bmsQQ?e=bRdzfd)]
    - Data viz avec Matplotlib [[cliquer ici](https://1drv.ms/b/s!AmJGbSlW18YGqf1D0JjGuBwGFqsNqQ?e=76Q9li)]

# Python

* + **Projet Casino** : [[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGqNUpJDCx_3SIhjeeKA?e=9RnnMV)],  ***correction***
  + **Python les fondamentaux** :
    - Devenir un développeur pythonique [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-cj2rZ7XIfi1nAisw?e=2expxg)],  ***correction***
    - Flow control [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-chf9EFGizsJ4YL4g?e=l1QZZj)],  ***correction***
    - Gestion des exceptions [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-ce4Jcf_3vy0nbZDw?e=d7mesM)],  ***correction***
    - Les fonctions [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-cfMUQsaMIT7wMftQ?e=bGTlvo)] ,  ***correction***
    - Maîtriser l’encoding [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-cgrM-7JBUUPqiOXQ?e=nMCeKJ)] ,  ***correction***
    - Gestion des fichiers [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-ciLR4q8VCYRsB1aQ?e=I3hqAd)] ,  ***correction***
    - Listes, Tuples et Dico [accéder au [notebook](https://1drv.ms/u/s!AmJGbSlW18YGq-cqWpqHWmX8tWKAZQ?e=LxHf9f), aux données : [result\_foot.txt](https://1drv.ms/t/s!AmJGbSlW18YGm6oGtUOR08fXm1EcWg?e=e00Qfz), [cities\_rates.txt](https://1drv.ms/t/s!AmJGbSlW18YGm6ool5kunY3vWMZDpg?e=WiMLEN)] ,  ***correction***

******



# Numpy

* + **Manipuler efficacement des objets numpy ?** [[Cours](https://1drv.ms/t/s!AmJGbSlW18YGq-dp8Rq_kSOhpHohRg?e=ElrddR)], [[TP](https://1drv.ms/t/s!AmJGbSlW18YGq-dqNcBMXNBrIfcFyA?e=LCebe9)]**, *correction***

# Pandas

* + **Projet complet : De l'agilité avec les données vers la data viz ! Secteur Life Science** [Projet : [notebook](https://1drv.ms/u/s!AmJGbSlW18YGr5BKqz0g1_KBQT64AA?e=pJo8NP)] [[data](https://1drv.ms/t/s!AmJGbSlW18YGr5BJqzzGQapf_G2v-A?e=lfedRN) : fichier *sante.txt*]
  + **Lire efficacement des fichiers plats ?** [[TP](https://1drv.ms/w/s!AmJGbSlW18YGq-dEazH5y_IVQkf_cQ?e=eV6Y4v)]**, *correction***
    - * + ***Pour vous aider* :** article datacamp avec des exercices en exécution interactive du code ; [[Cliquer ici](https://www.datacamp.com/community/tutorials/pandas-read-csv)]
  + **Service facturation** - *Billing Dispute files* [[TP](https://1drv.ms/w/s!AmJGbSlW18YGq-lPonxDwncUwvKMUQ?e=dbB0Zy)]**, *correction***
  + **Brief Project : Manipuler efficacement les vars de type categorical ?**

[[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-knEyKBNkZB8BuRgg?e=ClTVT2)]**,** [**data**](https://1drv.ms/t/s!AmJGbSlW18YGm6onPI6YRfBrs9nIJg?e=aqMURc)**,** [***correction***](https://1drv.ms/u/s!AmJGbSlW18YGq-koxljX62lPHtDwCA?e=s6HwxH)

* + **Brief Project : Traiter efficacement les données manquantes ?**

[[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-krBHRFeOs4hok9gw?e=dUZjUe)]**,** [**data**](https://1drv.ms/t/s!AmJGbSlW18YGm6oWmUD8oD43Sr3Wwg?e=rBoZF5)**,** [***correction***](https://1drv.ms/u/s!AmJGbSlW18YGq-koxljX62lPHtDwCA?e=s6HwxH)

* + **Brief Project : Manipuler efficacement les vars de type string ?**

[[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-kyQnJPeY4k1fkblA?e=GCD4Sd)]**,** [**data**](https://1drv.ms/u/s!AmJGbSlW18YGq-k0P36aRoKQIWSsqw?e=IjvtGz)

* + **Brief Project : Travailler avec les vars de type datetime ?**

[[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-k4kklKpLeQk4fhHQ?e=KjMg0C)]**,** [**data**](https://1drv.ms/t/s!AmJGbSlW18YGm6oaFqDuksOX8jIO0Q?e=7I6DGa)

* + **Brief Project : Travailler avec les axes d’un dataframe ?**

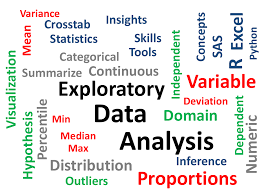
[[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-k8kRocMCdrvBjrhw?e=RrF3Th)]**,** [**data**](https://1drv.ms/t/s!AmJGbSlW18YGm6oUhjviSaH3CW62FQ?e=BByqXw)

* + **Brief Project : Travailler avec les axes d’un dataframe ?**

[[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-lBbc1JfC_ctaajww?e=rjLgMh)]**,** [**data**](https://1drv.ms/t/s!AmJGbSlW18YGm6ocfH6S-3gG36afCw?e=jq2naN)

* + **Brief Project : *Some utilities for df* ?** [[Enoncé](https://1drv.ms/u/s!AmJGbSlW18YGq-lBbc1JfC_ctaajww?e=hMABVV)]**,** [**data**](https://1drv.ms/t/s!AmJGbSlW18YGm6ocfH6S-3gG36afCw?e=XB71wh)

# Analyse de données :



# Ressources classe inversée**:**

* Exploratory Data Analysis EDA ? [[Article Kaggle](https://towardsdatascience.com/a-gentle-introduction-to-exploratory-data-analysis-f11d843b8184)]
* Analyse bivariée [Créer des [correlation heatmap](https://towardsdatascience.com/formatting-tips-for-correlation-heatmaps-in-seaborn-4478ef15d87f)]

# Cours

* + - Ebay voiture d’occasion : Nettoyer un dataframe après avoir scrapper le site d’Ebay [[Notebook corrigé](https://1drv.ms/u/s!AmJGbSlW18YGqf4IeKa4LX37pAq7uw?e=mjkfws)] [[dataset-brut](https://1drv.ms/u/s!AmJGbSlW18YGqf4KmWGPPuN69hnh5Q?e=51Plwn)] [[dataset-clean](https://1drv.ms/u/s!AmJGbSlW18YGqf4J5yShXs0khyrxOw?e=arDJHB)]
    - Viz avec matplotlib[[cliquer ici](https://1drv.ms/u/s!AmJGbSlW18YGqf1J4iYt_O02oo3Mow?e=ZD46h6)]

# TP :

* + *Service Support* d’une Entreprise qui retrace l’ensemble des incidents informatique : [[accéder aux données ici](https://1drv.ms/x/s!AmJGbSlW18YGqNUqLMiWa5Y-pagqvA?e=F6nNja)] données réelles
* *Q principale à traiter : scinder la colonne* "*Sujet (complet)*" *suivant les* "*/*" *en plusieurs colonnes* ?

*Correction :*

* + Data : données réelles et récentes du [Google Store](https://store.google.com/) où on retrace toutes les activités des visiteurs du site lors de leur session de navigation : nombre de clics, liens sur lesquels les visiteurs ont cliqué, durée de la session, composition du panier, achat ou non, ... .
* *Q : Dans certaines colonnes de cet énorme jeu de données, on y trouve du json stocké* (*souvent très imbriqués*)*, Splitter ces colonnes* ?
* *Ressources pour vous aider sur le Data Wrangling avec JSON* : [[cours](https://1drv.ms/u/s!AmJGbSlW18YGp69YZekTsF7WHw9j4Q?e=ydSgrv)], [[TP](https://1drv.ms/u/s!AmJGbSlW18YGp-M95XbuREMej46X4A?e=Wpam0Z)]

*Correction ;*