

Traitement de données en table

Liste des prénoms

I Le fichier des prénoms de l'INSEE

L'INSEE (Institut National de la Statistique et des Études Économiques) est une source de données en « open data ». Il propose en particulier :

« Le fichier des prénoms contient des données sur les prénoms attribués aux enfants nés en France entre 1900 et 2019. »

À faire soi-même 1

Ouvrir l'adresse : <https://www.insee.fr/fr/statistiques/2540004#consulter>

Télécharger le fichier CSV correspondant à la France hors Mayotte.

L'onglet « Dictionnaire des variables » nous dit que :

Le premier **fichier national** comporte 652 056 enregistrements et quatre variables décrites ci-après.

Ce fichier est trié selon les variables SEXE, PREUSUEL, ANNAIS.

- Nom : SEXE - intitulé : sexe - Type : caractère - Longueur : 1 - Modalité : 1 pour masculin, 2 pour féminin
- Nom : PREUSUEL - intitulé : premier prénom - Type : caractère - Longueur : 25
- Nom : ANNAIS - intitulé : année de naissance - Type : caractère - Longueur : 4 - Modalité : 1900 à 2019, XXXX
- Nom : NOMBRE - intitulé : fréquence - Type : numérique - Longueur : 8

Il y a donc quatre descripteurs et pour chacun d'entre eux le type est décrit.

À faire soi-même 2

Ouvrir le fichier téléchargé à l'aide d'un tableur. Les premières lignes doivent ressembler à ceci :

	A	B	C	D
1	sexe	preusuel	annais	nombre
2	1	A	1980	3
3	1	A	1998	3
4	1	A	XXXX	21
5	1	AADAM	2009	4
6	1	AADAM	2014	3
7	1	AADAM	2016	4

À faire soi-même 3

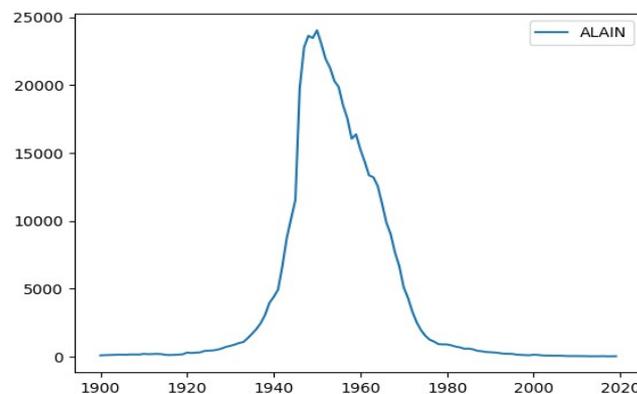
Aller dans l'onglet « Documentation » et expliquer à quoi correspond la valeur XXXX pour le descripteur ANNAIS.

Expliquer à quoi correspond la valeur « _PRENOMS_RARES_ » pour le descripteur PREUSUEL.

II Attribution d'un prénom en fonction du temps

Vous allez écrire une fonction qui affiche un graphique du nombre de fois où un prénom a été donné en fonction de l'année de naissance.

Pour le prénom de sexe masculin 'ALAIN' nous devons obtenir :



À faire soi-même 4

Recopier le début de programme ci-dessous qui importe les modules csv pour les fichiers .csv et matplotlib.pyplot pour créer le graphique. On suppose que le fichier nat2019.csv est dans le même répertoire que votre programme.

```
import csv
import matplotlib.pyplot as plt

def charger_fichier(nom_fich):
    """ string -> List[dict]
    Charge le fichier nom_fich et retourne une liste de dictionnaires contenant
    les valeurs du fichier
    """
    liste = []
    with open(nom_fich, "r", encoding="utf-8") as csvfile :
        lecteur = csv.DictReader(csvfile , delimiter=";")
        for ligne in lecteur :
            liste.append(ligne)
    return liste

prenoms = charger_fichier("nat2019.csv")
print(prenoms[96164])
```

Ce programme doit afficher ;

```
{'sexe': '1', 'preusuel': 'GABI', 'annais': '2009', 'nombre': '12'}
```

Le prénom masculin 'GABI' a été donné 12 fois en 2009.

À faire soi-même 5

Vous devez recopier puis compléter la fonction suivante qui prend en argument la liste des données sur les prénoms, le sexe et le prénom étudié. L'algorithme est le suivant :

Lire les données.

Si le prénom et le sexe du dictionnaire lu sont égaux à sexe et prenom en majuscule, et si annee est différent de 'XXXX' alors on ajoute à la liste annees l'année de ce dictionnaire et à la liste effectifs le nombre de fois où ce prénom a été attribué.

La première ligne commençant par plt prend en argument la liste des années qui sera en abscisse et la liste des effectifs qui sera en ordonnées, label=prenom permet d'afficher le prénom étudié sur le graphique. La ligne suivante ajoute effectivement la légende sur le graphique et celle qui suit demande l'affichage du graphique.

```
def affiche_prenom(liste, sexe, prenom):
    """ list[dict] * string * string -> list * list
    Affiche le nombre de fois où un prénom est donnée en fonction
    de l'année de naissance
    Retourne la liste des années et des effectifs correspondants
    """
    annees=[]
    effectifs=[]
    prenom = prenom.upper() #Assure que le prénom est en majuscules
    for dico in liste:
        if # À compléter
            if dico['annais'] != 'XXXX':
                annees.append(int(dico['annais']))
                effectifs.append(int(dico['nombre']))
    plt.plot(annees, effectifs, label=prenom)
    plt.legend()
    plt.show()
    return annees, effectifs
```

```
affiche_prenom(prenoms, '1', 'ALAIN')
```

Ce programme doit afficher le graphique du début de la page précédente.

Vous pouvez le tester avec d'autres prénoms, le vôtre par exemple, et comparer leur le nombre de fois où ils ont été choisis en fonction du temps.

À faire soi-même 6

Vous allez compléter une fonction qui permet d'obtenir les 10 prénoms les plus attribués en fonction de l'année.

Ce programme prend en argument la liste des prénoms et l'année que l'on veut étudier. Elle retourne une liste de dictionnaires correspondant aux 10 prénoms les plus choisis classés du plus choisi au moins choisi parmi les 10.

Cette fonction créer une liste de dictionnaires nommée `liste_annee` qui contient tous les dictionnaires correspondant aux prénoms de l'année `annee`. Puis `liste_annee_triee` contient cette même liste triée dans l'ordre décroissant du nombre de fois où un prénom a été donnée.

La fonction `sorted` (sort veut dire trier en anglais) retourne une liste triée selon la fonction donnée après « `key=` ». « `reverse = True` » signifie que l'on ordonne du plus grand au plus petit.

```
def top10(liste,annee):
    """ list[dict] * int -> list[dict]
    Retourne les dictionnaires correspondant aux 10 prénoms les plus données
    l'année annee du plus donné au moins donné
    """
    liste_annee = []
    for ligne in liste:
        # Si le dictionnaire ligne contient l'année annee.
        if # À compléter
            ligne['nombre'] = int(ligne['nombre'])
            liste_annee.append(ligne)
    # liste_annee_triee contient tous les prénoms de l'année annee du plus attribué au
    moins attribué
    def critere(dico):
        return dico['nombre']
    liste_annee_triee = sorted(liste_annee,key =critere, reverse = True )
    # Retourne les 10 premiers éléments de la liste liste_annee_triee
    return #À compléter

print(top10(prenoms,1982))
```

Doit afficher (ici sur deux colonnes pour des questions de place) :

```
[OrderedDict([('sexe', '1'), ('preusuel', 'NICOLAS'), ('annais', '1982'), ('nombre', 17801)]),
OrderedDict([('sexe', '1'), ('preusuel', 'JULIEN'), ('annais', '1982'), ('nombre', 16446)]),
OrderedDict([('sexe', '2'), ('preusuel', 'AURÉLIE'), ('annais', '1982'), ('nombre', 13679)]),
OrderedDict([('sexe', '1'), ('preusuel', 'SÉBASTIEN'), ('annais', '1982'), ('nombre', 13551)]),
OrderedDict([('sexe', '2'), ('preusuel', 'CÉLINE'), ('annais', '1982'), ('nombre', 11215)]),
OrderedDict([('sexe', '2'), ('preusuel', 'EMILIE'), ('annais', '1982'), ('nombre', 10361)]),
OrderedDict([('sexe', '1'), ('preusuel', 'DAVID'), ('annais', '1982'), ('nombre', 9265)]),
OrderedDict([('sexe', '2'), ('preusuel', 'LAETITIA'), ('annais', '1982'), ('nombre', 8387)]),
OrderedDict([('sexe', '1'), ('preusuel', 'GUILLAUME'), ('annais', '1982'), ('nombre', 8258)]),
OrderedDict([('sexe', '2'), ('preusuel', 'STÉPHANIE'), ('annais', '1982'), ('nombre', 8037)])]
```

Déterminer le top 10 pour l'année 1992 puis pour l'année 2002. Que constatez-vous ?