

<p>Enseignants</p> <p>Cours : A. NAJJAR-I. BEN OTHMEN-M. FARHAT</p> <p>TP : F. JENHANI- S.BIROUZA – I. HAMROUNI</p>	<p>TP1</p> <p>Machine Learning</p>	<p>Classe : 3ème GLSI</p>
--	--	----------------------------------

Partie 1- Rappel python

- 1- Ecrire un programme qui permet de résoudre une équation de second degré sous la forme $ax^2+bx+c=0$. Il faut alors saisir trois nombres flottants correspondants aux coefficients et les stocker dans des variables a, b et c.
- 2- Écrire une fonction nombreOcurrences (caractere, mot) qui prend en argument un caractère x et une chaîne de caractère mot et qui renvoie le nombre de fois où le caractère x est présent dans mot. Par exemple, si mot est le mot «java», nombreOcurrences('a', mot) vaut 2.

Partie 2- La bibliothèque numpy

- 1- Créer la liste suivante : revenus = [1800, 1500, 2200, 3000, 2172].
- 2- Ecrire un programme qui permet de calculer la moyenne.
- 3- Utiliser la structure array de la bibliothèque Numpy pour créer le tableau 'revenus'.
- 4- Recalculer la moyenne en utilisant la fonction mean.
- 5- Accéder à l'élément à l'indice de l'élément minimum puis maximum.
- 6- Ordonner le tableau par ordre croissant.
- 7- Afficher les 3 premiers éléments du tableau.
- 8- Afficher un élément sur 2.
- 9- Afficher les éléments dont la valeur est supérieure à 200.
- 10- Créer un tableau à deux dimensions qui contient des valeurs aléatoires entières comprises entre 1 et 10.
- 11- Accéder à la première ligne
- 12- Rajouter une ligne au tableau.
- 13- Supprimer la dernière ligne du tableau.

Partie 3 – La bibliothèque Pandas

1- En utilisant l'objet DataFrame de la bibliothèque Pandas, créer le tableau suivant :

	state	year	pop
0	Tunis	2000	1.5
1	Bizerte	2001	1.7
2	Sfax	2002	3.6
3	Sousse	2001	2.4
4	Gabes	2002	2.9

- 2- Afficher quelques info à propos du dataframe crée en utilisant la commande : `info()`.
- 3- Afficher le dataframe sous forme d'un tableau.
- 4- Afficher les noms des colonnes.
- 5- Afficher toute la colonne 'year'.
- 6- Changer la valeur de pop par 5.5.
- 7- Rajouter une variable 'zone'.
- 8- Renommer les colonnes comme suit : A, B, C, D.
- 9- Supprimer la variable 'A'.

Partie 4- Exploration des données

Plusieurs bases dédiées pour l'apprentissage automatique peuvent être trouvée sur le web. Consulter la page web suivante :

<https://archive.ics.uci.edu/ml/datasets.html>

Dans ce TP, on se basera sur la base "**pima-indians-diabetes.data.csv**".

Pima est un groupe d'indien de l'Amérique (**Native Americans**) vivant à Arizona. Ils sont sujets de plusieurs études car le changement dans leurs habitude alimentaire a favorisé qu'ils soient atteints de diabète de type 2.

La base inclus 767 femmes décrites par les 8 attributs suivants.

- 1- Nombre de grossesses (**NumTimesPrg**)
- 2- Concentration du plasma en glucose (**PIGlcConc**)
- 3- Tension artérielle (**BloodP**)
- 4- Épaisseur du pli cutané des triceps (**SkinThick**)
- 5- Taux d'insuline (**TwoHourSerIns**)
- 6- Indice de masse corporelle (**BMI**)
- 7- *Fonction pédigrée du diabète* : hérédité (**DiPedFunc**)
- 8- Age (**age**)

La dernière colonne "**HasDiabetes**" de la base (9^{ème} colonne) indique si la personne est diagnostiqué (1) de diabète ou pas (0).

On souhaite d'abord lire et inspecter les données dans la base. La base "**pima-indians-diabetes.data.csv**" est fournie avec le format **CSV (Comma Separated Values, valeurs séparées par des virgules)**. Il s'agit du format le plus commun dans l'import et l'export de feuilles de calculs et de bases de données.

Exercice

- 1- Lire le contenu du fichier "**pima-indians-diabetes.data.csv**".
- 2- Attribuer à chaque colonne le nom de l'attribut correspondant.
- 3- Afficher la structure de la base.
- 4- Afficher les 10 premières lignes de la base.
- 5- Afficher seulement les valeurs de la tension artérielle (attribut "BloodP") pour la totalité des individus.
- 6- Obtenir des statistiques sur les individus dans la base. Analyser certains résultats obtenus
- 7- Créer une matrice qui ne contient que les données des attributs (séparer les valeurs des attributs de leurs classes d'appartenances).