

Département	Informatique	Niveau	L3
Matière	LIFDATA		
Enseignants	Khalid Benabdeslem, Haytham Elghazel et Ichrak Ennaceur		
Intitulé du TP :	TPI : Analyse statistique de base		

### Remarque

**Ce TP est à rendre sur Moodle. Il faudra rendre le notebook jupyter commenté. Ne pas oublier de mentionner vos noms et prénoms dans le nom du fichier qui sera sous la forme : TP1\_Initiation\_A\_rendre\_Noms\_Prenoms.ipynb.**

### Préambule

Les données dans le fichier capital.xls sont extraites du magazine Capital qui a donné pour 100 villes françaises les valeurs du taux de la taxe d'habitation.

Les variables X : représente la ville, Y : le taux de la taxe d'habitation et Z : la région.

### Questions

1. Importer le fichier capital.xls.

**Dans la suite, il faut toujours donner un titre pour chaque graphique demandé. Les libellés des variables seront indiqués sur toutes les sorties.**

2. Donner une représentation graphique en histogramme du pourcentage de villes par région
3. Représenter graphiquement et de différentes manières la répartition de chaque région pour cet ensemble de données.
4. Quelle est la région la plus représentée dans cette étude ?
5. Calculer les différentes statistiques sur le taux de la taxe d'habitation. Regrouper ensuite ces statistiques par région.
6. Quelle est la région la plus intéressante pour le taux de la taxe d'habitation ?
7. Quelle est la région la plus représentative de la moyenne nationale ?
8. Donner en fonction du taux de la taxe d'habitation une typologie des villes françaises en 3 groupes.
9. En utilisant une fonction Python, donner les villes qui ont un taux de taxe d'habitation situé 30% au dessus ou en dessous de la moyenne.
10. Idem pour les villes qui ont un taux de taxe d'habitation situé 20% au dessus ou en dessous de la moyenne.