

TP2: Jouer avec Numpy et Pandas

Dataset

1. A partir du jeu de données original, créer:
 - un Dataframe comportant tous les modules en précisant l'effectif des apprenants, le nombre total de sessions, le score maximal, le score minimal, le taux de réussite (pour les apprenants des scores > 50)
 - un tableau 2D Numpy avec les données des 10 tops modules selon les taux de réussites (Chaque ligne représente un module)
2. Pour le top module et le dernier:
 - créer un Dataframe avec les données résumées de chaque apprenants: nombre total de sessions, nombre total de temps, score, resultat final (de valeur 0 si score < 50 et 1 sinon)

Restitution: Lundi prochain (07/11/2022)

Suite: Un clin d'oeil à Matplotlib

Pour les 10 top module:

- représenter par une courbe ces derniers dans un même repère sur une figure unique. A chaque apprenant correspond un point dont l'abscisse est le temps en minutes de celui-ci. L'ordonné est déterminé par son score maximum.
- Un diagramme en bâton avec le taux de réussite de chaque module.

Pour les 3 top module:

- représenter par une courbe chacun dans un même repère différent sur une même figure horizontalement subdivisée en trois. A chaque apprenant correspond un point dont l'abscisse est le rang de celui-ci. L'ordonné est déterminé par son score maximum.

Pour tous les modules:

- représenter l'ensemble par un nuage de points dans un même repère sur une même figure. A chaque ~~apprenant~~^{module} correspond un point dont l'abscisse est le rang de celui-ci. L'ordonné est déterminé par le temps global sur le module.
- Un camembert avec le taux des modules catégorisés en quatre classes:
 - Classe A: taux de réussite > = 80%
 - Classe B: taux de réussite > = 60%
 - Classe C: taux de réussite > = 50%
 - Classe D: taux de réussite < 50%