\*Round 1 : pourcentage de chances d’avoir un second rdv

Round 2 : pourcentage de chances d’avoir un second rdv

Itérer jusqu’à la fin et faire une moyenne

Problématiques :

**Etape 1 : approche par la moyenne de wave**

Quelle est la chance de décrocher une bonne note pour chaque wave ?

Quelle est la moyenne de chance de décrocher un rdv sur tous les rounds ?

Quel round est le plus propice pour décrocher un rendez-vous (zone avec scatter) ? => Voir la densité autour de la moyenne pour délimiter une zone de win

Uniformiser la notation pour la pertinence. (multiplier par 10)

Etablir une corrélation entre le nombre d’individus rencontrés et le wave ( nombre total d’individus présents par wave => établir des corrélations par tranche. ex: de 0>5, 5>10, 10>20) => créer des groupes de wave par nb d’individus et les comparer entre eux

**Etape 2 : critères déterminants de notation**

Quelles caractéristiques ont été les plus votées ? :

* Par les femmes
* Par les hommes

**Etape 3 :**

-samerace

-imprace:

How important is it to you (on a scale of 1-10) that a person you date be of the same racial/ethnic background?

Comparer les deux variables pour déterminer :

1-la véracité du propos des gens

2-les gens sont-ils plus influencés par le critère du samerace au final (par rapport à la moyenne globale à l’étape 2 ou « samerace = no ou yes ») ?

Aller voir notebook d’autres personnes sur kaggle

Commenter nos notebook avec le markdown pour structurer notre texte

Creuser les différentes variables

2 semaines pour pusher le projet dans git : rendre avant 2 semaines (avant le 15)

Package python « missingno» (quelles sont les colonnes qui contiennent le plus de nuls?)