

ET SI ON PARLAIT DU PROJET?

# LE CONCEPT

Click to add subtitle or delete text box if not required

---

- **Groupes au hasard (pour refléter la réalité du monde du travail ou vous ne choisirez pas vos collègues)**
- **Sur un repo github, où vous m'inviterez**
- **Sur le jeu de donnée que vous voulez**
- **Vous répondrez à la problématique que vous choisissiez**
- **Ex: Sur Kaggle il existe un dataset représentant l'ensemble des matchs professionnels de tennis féminins (WTA) de 2000 à 2016, et j'aimerais les utiliser pour prouver la pertinence d'un algorithme capable de prédire la victoire d'une joueuse face à une autre**
- **Vous implementerez au moins plusieurs algos de machine learning supervisés afin de les comparer**
- **Vous présenterez, en groupe, pendant une soutenance de 30 minutes vos résultats**
- **Tout travail supplémentaire, scrapping, build, tests, etc... sera fortement valorisé**

# LA BASE C'EST DE SE POSER LA BONNE QUESTION

Click to add subtitle or delete text box if not required

---

- **Kaggle (ou autre) pour explorer de la donnée et en déduire une problématique**
- **En explorant la donnée, essayer de se poser des questions pertinentes, est-ce qu'il existe un label sur lequel on peut émettre une prediction, est-ce que les colonnes/features qui constituent la donnée peuvent se combiner pour extraire de l'intelligence ?**
- **Commencez à explorer la donnée, mapper le dataset, repérer les features interessantes et les autres**
- **Discutez entre vous pour trouver une problématique qui vous interesse, profitez car dans le monde du travail vous n'aurez pas tout le temps cette chance**
- **N'hésitez pas à choisir plusieurs dataset si vous arrivez à trouver des liens entre eux, plus de données c'est toujours mieux**
- **Les TP seront dédiés à l'avancement du projet, ils ne sont pas néanmoins facultatifs, un point sera fait avec chaque équipe pour constater l'avancement, étudier les points de blocage etc...**

