Projet final

Comprendre les comportements de réservation des leçons de conduite, pour nous permettre de comprendre ce qu'on appelle la liquidité de la plateforme (à quel point les élèves arrivent à trouver des leçons et à quel point les enseignants arrivent à remplir leur planning).

Donc quels sont les patterns de réservation, quelle est la compétition pour certains types de créneaux (est-ce que les élèves préfèrent les leçons de telle heure et/ou telle jour et est-ce qu'ils trouvent facilement leur bonheur) ?

Et on a besoin de comprendre ça à plusieurs échelles géographiques : au niveau national pour avoir une tendance moyenne, au niveau de chaque territoire, au niveau du département et au niveau de la zone (au sens d'Ornikar, voir plus bas la définition).

Comprendre les comportements de réservation :

3 axes d’étude :

* à l’échelle d’un individu (comment et pk il réserver ? mais il faudrait avoir plus d’informations sur l’individu. Ce que je n’ai pas)
* à l’échelle géographique (selon la zone, le département, le territoire, il semblerait qu’il y ait trop peu de données)
* à l’échelle du temps (horaire, jours de la semaine, jour de l’année (vacances scolaires, jours fériés))

Comment je peux savoir quels sont les créneaux préférés ?

Par learner\_id, les heures sont noté HCR alors qu’elles n’ont pas encore été effectuées.

Comment prédire les comportements de réservation sur le dataset 2020 ?

Les comportements sont biaisés par le fait que les créneaux de réservation ont été bloqués à la réservation pendant 2 mois. Les comportements de réservation post confinement peuvent être biaisés par le fait que tout le monde se soit rué sur les dates ouvertes. Les gens s’adaptent en fonction des heures ouvertes ?

De quelles features je dispose pour pouvoir faire des prédictions ?

Faire du feature engineering mais à partir de quelles features ?

* date de réservation de la leçon
* remaining tokens
* date d’inscription sur Ornikar