

[Réponse à Q1]

Les données couvrent le parcours apprenant de l'inscription à la correction des travaux pratiques. Le tableau `training_sales` décrit les ventes, avec une ligne par apprenant inscrit et le montant payé. `training_user_progress_snapshot` fournit une vue historisée de la progression, avec un score d'avancement entre 0 et 1 et des dates de validité qui permettent de suivre l'évolution dans le temps. `user_homework_submission` liste les travaux pratiques soumis, avec leur date de soumission et éventuellement un score lorsqu'ils ont été corrigés. Ces trois tables permettent de construire des indicateurs business (CA, marge), de progression (avancement moyen, stagnation) et pédagogiques (volume et qualité des TP), sous réserve de bien gérer les valeurs manquantes, la qualité de l'historisation et l'alignement entre ventes, progression et travaux pratiques.

[Réponse à Q5]

Avec davantage de données, l'objectif serait de passer d'un simple constat (taux de complétion par formation) à une compréhension fine des mécanismes de décrochage, puis à l'identification de leviers concrets pour le réduire.

A/ Une première piste serait d'analyser les parcours par cohorte d'inscription : comparer, mois par mois, l'évolution de la progression et des travaux pratiques corrigés pour voir à quel moment les apprenants décrochent le plus, et si certaines formations, périodes ou profils sont plus exposés que d'autres. On pourrait, par exemple, suivre la proportion d'apprenants qui atteignent 25 %, 50 % ou 75 % d'avancement à J+30, J+60, J+90, et voir où les courbes se "casse" selon la formation ou la période d'inscription.

B/ Ensuite, il serait intéressant de reconstituer de véritables trajectoires d'usage : non seulement le niveau de progression atteint, mais la façon dont il est atteint dans le temps. En suivant la courbe d'avancement jour après jour (ou semaine par semaine) depuis la date de début, on peut repérer des phases de stagnation prolongée, des profils "binge learning" qui progressent très vite puis s'arrêtent net, ou au contraire des profils réguliers. Ces trajectoires peuvent être segmentées par formation, par typologie d'apprenant (âge, situation pro, temps disponible, si ces données existent) et croisées avec des événements clés (premier TP soumis, première correction, premières difficultés).

C/ Un autre axe essentiel concerne la qualité et la rapidité des corrections. Avec des données plus complètes sur les délais entre la soumission d'un TP et sa correction, on pourrait mesurer l'impact du feedback sur la poursuite de la formation : typiquement, vérifier si un délai de correction trop long augmente la probabilité de décrochage, ou si un certain volume de TP corrigés (par exemple au moins trois TP validés dans les deux premiers mois) est un bon signal de réussite future. Cela permettrait d'identifier des goulets d'étranglement opérationnels côté équipe pédagogique et de dimensionner les ressources de correction en conséquence.

D/ Utilisation de ces analyses

Enfin, ces analyses gagneraient à être reliées à des tests concrets de leviers. Une fois les facteurs de risque identifiés (par exemple : délai de correction trop long, manque de TP dans le premier mois, absence de jalons intermédiaires), on pourrait mettre en place des expérimentations contrôlées, c'est-à-dire raccourcir les délais de correction sur un échantillon, ajouter un TP obligatoire très tôt dans la formation, proposer des relances automatiques après X jours sans activité... et mesurer l'impact sur la progression et la complétion. À terme, cela permettrait de passer d'un diagnostic statistique à une feuille de route opérationnelle pour réduire le décrochage formation par formation.

