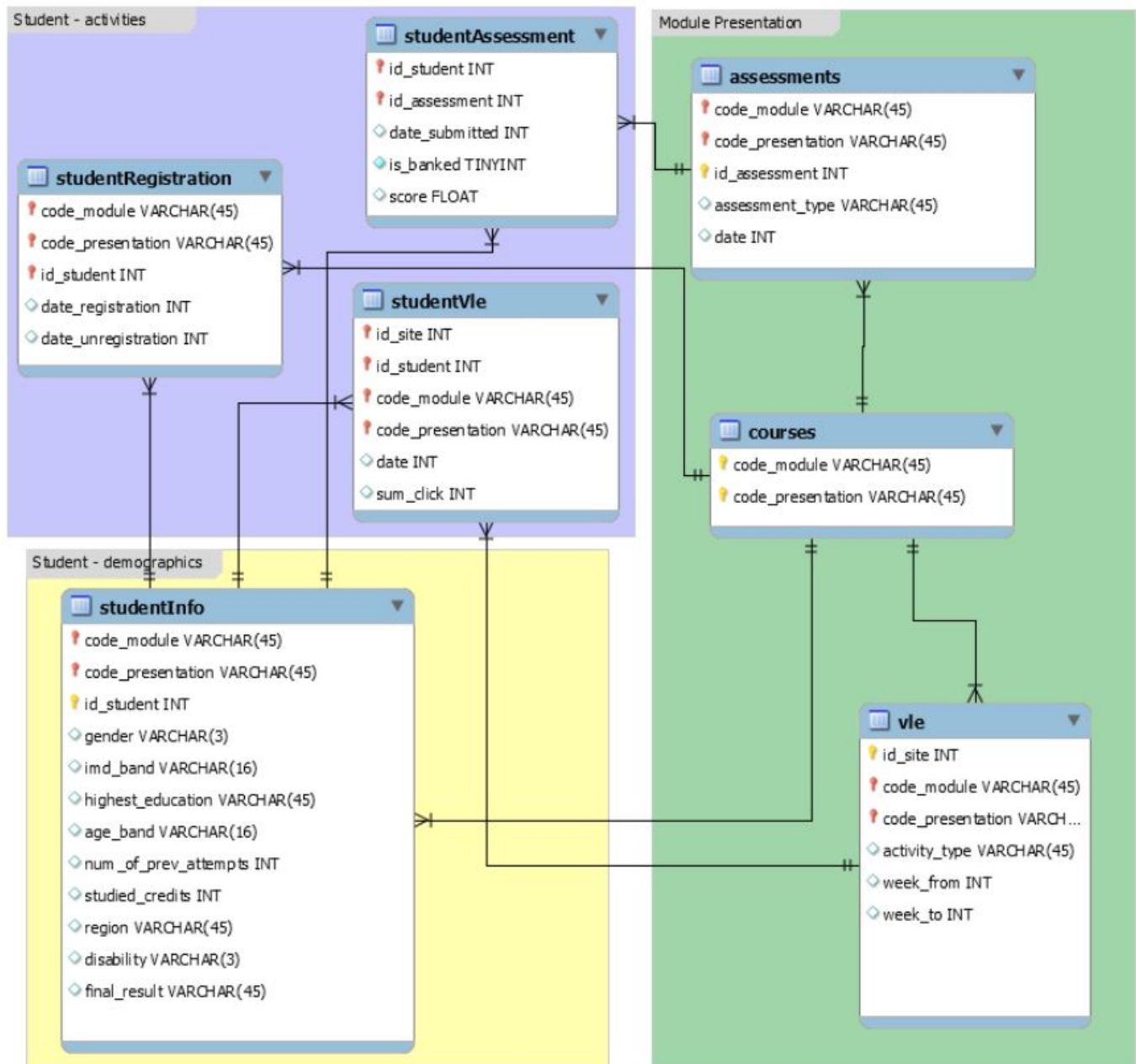


# Open University Learning Analytics dataset

## Présentation des données

### Schema de la base de données



### courses.csv

Le fichier contient la liste de tous les modules disponibles et leurs présentations. Les colonnes sont les suivantes :

- **code\_module** - nom de code du module, qui sert d'identifiant.
- **code\_presentation** - nom de code de la présentation. Il se compose de l'année et de "B" pour la présentation commençant en février et de "J" pour la présentation commençant en octobre.

- **length** - longueur du module-présentation en jours.

La structure des présentations B et J peut être différente, c'est pourquoi il est bon d'analyser les présentations B et J séparément. Néanmoins, pour certaines présentations, la présentation B/J précédente correspondante n'existe pas et donc la présentation J doit être utilisée pour informer la présentation B ou vice versa. Dans l'ensemble de données, c'est le cas des modules CCC, EEE et GGG.

## assessments.csv

Ce fichier contient des informations sur les évaluations dans les présentations de modules. En général, chaque présentation comporte un certain nombre d'évaluations suivies de l'examen final. Le CSV contient des colonnes :

- **code\_module** - code d'identification du module, auquel l'évaluation appartient.
- **code\_presentation** - code d'identification de la présentation, à laquelle l'évaluation appartient.
- **id\_assessment** - numéro d'identification de l'évaluation.
- **assessment\_type** - type d'évaluation. Il existe trois types d'évaluation : Tutor Marked Assessment (TMA), Computer Marked Assessment (CMA) et Final Exam (Exam).
- **date** - information sur la date de soumission finale de l'évaluation calculée comme le nombre de jours depuis le début de la présentation du module. La date de début de la présentation porte le numéro 0 (zéro).
- **poids** - poids de l'évaluation en %. En général, les examens sont traités séparément et ont le poids de 100% ; la somme de toutes les autres évaluations est de 100%.

Si l'information sur la date de l'examen final est manquante, elle se situe à la fin de la dernière semaine de présentation.

## vle.csv

Le fichier csv contient des informations sur les matériels disponibles dans le VLE. Il s'agit généralement de pages html, de fichiers pdf, etc. Les étudiants ont accès à ces matériels en ligne et leurs interactions avec ces matériels sont enregistrées. Le fichier vle.csv contient les colonnes suivantes :

- **id\_site** - un numéro d'identification du matériel.
- **code\_module** - un code d'identification du module.
- **code\_presentation** - le code d'identification de la présentation.
- **activity\_type** - le rôle associé au matériel du module.
- **week\_from** - la semaine à partir de laquelle il est prévu d'utiliser le matériel.
- **week\_to** - semaine jusqu'à laquelle il est prévu d'utiliser le matériel.

## studentInfo.csv

Ce fichier contient des informations démographiques sur les étudiants ainsi que leurs résultats. Le fichier contient les colonnes suivantes :

- **code\_module** - code d'identification d'un module sur lequel l'étudiant est inscrit.

- **code\_présentation** - le code d'identification de la présentation pendant laquelle l'étudiant est inscrit au module.
- **id\_student** - un numéro d'identification unique pour l'étudiant.
- **gender** - le sexe de l'étudiant.
- **region** - identifie la région géographique où l'étudiant vivait pendant la présentation du module.  
highest\_education - niveau d'éducation le plus élevé de l'étudiant à l'entrée de la présentation du module.
- **imd\_band** - spécifie la bande de l'indice de déprivation multiple de l'endroit où l'étudiant vivait pendant la présentation du module.
- **age\_band** - bande de l'âge de l'étudiant. num\_of\_prev\_attempts - le nombre de fois où l'étudiant a tenté ce module.
- **studied\_credits** - le nombre total de crédits pour les modules que l'étudiant étudie actuellement.
- **disability** - indique si l'étudiant a déclaré un handicap.
- **final\_result** - résultat final de l'étudiant dans la présentation du module.

## studentRegistration.csv

Ce fichier contient des informations sur le moment où l'étudiant s'est inscrit à la présentation du module. Pour les étudiants qui se sont désinscrits, la date de désinscription est également enregistrée. Le fichier contient cinq colonnes :

- **code\_module** - le code d'identification d'un module.
- **code\_presentation** - le code d'identification de la présentation.
- **id\_student** - un numéro d'identification unique pour l'étudiant.
- **date\_registration** - la date d'inscription de l'étudiant à la présentation du module, c'est le nombre de jours mesuré par rapport au début du module-présentation (par exemple, la valeur négative -30 signifie que l'étudiant s'est inscrit à la présentation du module 30 jours avant son début).
- **date\_unregistration** - date de désinscription de l'étudiant de la présentation du module, c'est le nombre de jours mesurés par rapport au début de la présentation du module. Les étudiants qui ont terminé le cours ont ce champ vide. Les étudiants qui se sont désinscrits ont la valeur de la colonne final\_result dans le fichier studentInfo.csv.

## studentAssessment.csv

Ce fichier contient les résultats des évaluations des élèves. Si l'étudiant ne soumet pas l'évaluation, aucun résultat n'est enregistré. Si le résultat des évaluations n'est pas enregistré dans le système, le résultat de l'examen final est absent. Ce fichier contient les colonnes suivantes :

- **id\_évaluation** - le numéro d'identification de l'évaluation.
- **id\_student** - le numéro d'identification unique de l'étudiant. date\_submitted - la date de soumission de l'étudiant, mesurée en nombre de jours depuis le début de la présentation du module.
- **is\_banked** - un drapeau d'état indiquant que le résultat de l'évaluation a été transféré depuis une présentation précédente.
- **score** - le score de l'étudiant dans cette évaluation. La fourchette est comprise entre 0 et 100. Un score inférieur à 40 est interprété comme un échec. Les notes se situent dans une fourchette de 0 à 100.

## studentVle.csv

The studentVle.csv file contains information about each student's interactions with the materials in the VLE. This file contains the following columns:

- **code\_module** – an identification code for a module.
- **code\_presentation** - the identification code of the module presentation.
- **id\_student** – a unique identification number for the student.
- **id\_site** - an identification number for the VLE material.
- **date** – the date of student's interaction with the material measured as the number of days since the start of the module-presentation.
- **sum\_click** – the number of times a student interacts with the material in that day.