CREATE VIEW partie3.vue\_voyelles\_m2105 AS

SELECT

i.prenom,

ra.code\_nip,

ra.annee\_univ,

ra.m2105,

ra.m1101,

ra.m1102,

ra.m1201,

ra.m2101

FROM partie3.\_individu i

JOIN partie3.\_etudiant e ON i.ine = e.ine

JOIN partie3.resultats\_annee ra ON e.code\_nip = ra.code\_nip

WHERE ra.m2105 IS NOT NULL;

Nous cherchons à expliquer la moyenne générale (‘moy’) des étudiants au semestre 3 à partir de leurs résultats dans différents modules et de leur type de bac.

CREATE VIEW vue\_patients\_consultations AS

SELECT patients.id\_patient, patients.nom, patients.prenom, patients.date\_naissance, services.nom\_service, consultations.date\_consultation, consultations.compte\_rendu

FROM patients

LEFT JOIN consultations ON patients.id\_patient = consultations.id\_patient

LEFT JOIN services ON patients.id\_service = services.id\_service;

annee\_univ;no\_semestre;code\_nip;rg;civilite;nom;prenom;bac;**groupe\_td\_2a;**groupe\_tp\_2a;moy;ue31;m3101;m3102;m3103;m3104;m3105;m3106c;**ue32;**m3201;m3202c;m3203;m3204;m3205;m3206;ue33;m3301;m3302;**m3303**

WbExport -file=mon\_fichier\_csv\_de\_sortie.csv

-outputDir=.

-type=text

-sourceTable=ma\_table\_a\_exporter

-schema=mon\_schema

-delimiter=';'

-header=true

-keyColumns=No\_Dossier,formation

-dateFormat='d/M/y'

-timestampFormat='d/M/y H:m:s'

;

Parfait ! À partir des consignes de *Livrable3.pdf* et du *Modèle3A.pdf*, je peux adapter la **structure de présentation attendue** à **tes propres données** (issues du fichier .csv avec les colonnes comme annee\_univ, moy, m3101, m3102, etc.).

### **✍️ Voici la syntaxe finale adaptée à ton jeu de données :**

## **0 Les données - Problématique**

### **(a) Présentation des données**

Le fichier utilisé contient plusieurs séries statistiques concernant des étudiant·es d’un cursus universitaire. Chaque étudiant·e est représenté·e de manière unique par son code NIP, accompagné de ses informations personnelles et de ses résultats académiques.

* **La population** est l’ensemble des étudiant·es figurant dans la base de données, chacun identifié par un code\_nip, avec les champs nom, prénom, et civilité.
* **La 1ʳᵉ variable statistique** sur cette population est la **moyenne générale (moy)** de chaque étudiant·e pour le semestre.
* **La 2ᵉ** est la **note obtenue à l’UE 31**, composée des matières m3101 à m3106c.
* **La 3ᵉ** est la **note obtenue à l’UE 32**, comprenant m3201 à m3206.
* **La 4ᵉ** est la **note obtenue à l’UE 33**, incluant m3301 à m3303.
* **La 5ᵉ** est le **type de baccalauréat obtenu** par l’étudiant·e (bac), variable qualitative qui peut être codée si besoin.
* **La 6ᵉ** est le **groupe de TD** dans lequel est inscrit l’étudiant·e (groupe\_td\_2a).
* **La 7ᵉ** est le **groupe de TP** de l’étudiant·e (groupe\_tp\_2a).

*Voici une capture d’écran des premières lignes du fichier* :

*(capture à insérer ici)*

### **(b) Problématique**

À partir de ces données, on cherche à répondre à la problématique suivante :

**Certaines variables (UE, groupes, type de bac…) permettent-elles d’expliquer les différences de moyennes générales entre les étudiant·es ?**

Plus précisément, on souhaite comprendre **quelles caractéristiques favorisent une bonne moyenne semestrielle** (moy) chez les étudiant·es.

La variable à expliquer (variable endogène) est donc : **la moyenne générale moy**.

Souhaites-tu que je t’aide également à écrire la suite (import, mise en forme, corrélation, régression, etc.) en suivant ce modèle ?