

# Informatique S4 : Bases de Données

## Projet pratique N°2

### 1 Explications :

1. Les étudiants par groupe de TD doivent constituer des groupes projet et choisir un ensemble de données sur une thématique et sur lequel les étapes du projet 1 vont s'appliquer. Les étudiants ne peuvent pas constituer un groupe projet avec des étudiants d'un autre groupe TD. Les groupes sont constitués de **1 à 2 personnes maximum. Même groupe pour les parties 1 et 2. Merci aux redoublants de changer de sources de données par rapport à l'an dernier.**
2. Chaque groupe fera un **fichier Word** composé de copies d'écrans sur le code et les résultats d'exécution, avec des commentaires. Le fichier Word portera le nom des membres du groupe. Les fichiers sources commentés seront déposés avec le fichier Word sur l'EPI (ou lien filex) ;
3. Chaque groupe va choisir un lot de données différent.
4. Il y a plusieurs thématiques correspondants à différentes sources de données sur kaggle.com. Chaque groupe choisit un data set parmi les différentes sources de donnée (.csv) comme par exemple : diabetes, heart, covid\_19\_data, HAM10000\_metadata (skin cancer), netflix\_titles, googleplaystore, vgsales (video game), data (european soccer), StudentsPerformance, winequality-red, winemag-data-130k-v2 (wine somelier values), beers (beer craft), olist\_customers\_dataset (brazilian e-commerce)...
5. Le projet est divisé en 2 étapes :
  - a. **ETAPE 1 : Structuration des données, lien avec Python**
    - i. Téléchargement d'un data set Kaggle
    - ii. Rétro conception
      1. Faites un retro engineering pour déterminer le modèle EA (utilisez Looping)
      2. Ecrire les relations et schémas relationnels
      3. Veillez à ce qu'ils soient normalisés en 3FN
      4. Faire un modèle physique
    - iii. Data Analyse : Faites un lien avec EXCEL
      1. Utilisez les fonctions mathématiques simples d'EXCEL, et éventuellement les tableaux croisés dynamiques
      2. Créez des figures (camemberts, histogrammes, ...)
    - iv. Faites un lien avec Python :
      1. Créez une 4eme table et alimentez-la à partir d'un programme Python
      2. Supprimez un enregistrement à partir du programme Python
      3. Prévoyez des contrôles de saisie pour garantir les contraintes de la base.
    - v. Déposez sur l'EPI
  - b. **ETAPE 2 : SQL**
    1. Ajoutez les données Kaggle dans la base
    2. Faites les copies d'écrans, lisibles, avec les commentaires nécessaires à la compréhension.

3. Faites les copies d'écrans, lisibles, avec les commentaires nécessaires à la compréhension.
4. Introduction des requêtes SQL (vous ferez également la modélisation en langage algébrique)
  - 1) Créez 3 requêtes avec un select from where et 3 conditions de types différents (like, between, opérateurs, jointures ...)
  - 2) Créez 2 requêtes avec 2 select from where imbriqués utilisant des conditions de types différents
  - 3) Utilisez un group by
  - 4) Utilisez un group by having count
  - 5) Créez des requêtes réalisant une : différence, union, intersection, division, ALL et ANY
  - 6) Testez vos requêtes et validez (ou pas) les résultats
  - 7) Faites les copies d'écrans, lisibles, avec les commentaires nécessaires à la compréhension.
- ii. Introduction de requêtes SQL dans PYTHON (vous ferez également la modélisation en langage algébrique)
  1. Introduire dans le code Python de l'étape 1 une requête SQL du type select from where différente du ii)
  2. Testez et validez votre requête
  3. Faites les copies d'écrans, lisibles, avec les commentaires nécessaires à la compréhension.
- iii. Implémentation des propriétés ACID de la base : transaction  
(<https://pynative.com/python-mysql-transaction-management-using-commit-rollback/>)
  1. Implémentez dans votre programme Python un commit
  2. Implémentez dans votre programme Python un roll back
  3. Faites les copies d'écrans, lisibles, avec les commentaires nécessaires à la compréhension.
- iv. QUESTION BONUS : Usage de l'outil Power BI de Microsoft ou tableau
  1. Chargez la version gratuite de l'outil sur votre machine
  2. Utilisez votre base de données pour faire des graphiques BI, des calculs de tendances, ...
  3. Choisissez des KPI et lien avec votre source de données et modélisez vos résultats