

Devoir en Groupe – Analyse de Données avec SQL

Objectif du devoir

Dans ce devoir, vous travaillerez en groupe sur un jeu de données de votre choix afin de réaliser une analyse descriptive en utilisant SQL. Vous devrez synthétiser vos résultats dans une présentation PowerPoint claire et structurée.

Consignes Générales

- **Travail en groupe** (3 à 4 étudiants)
- **Durée** : 1 à 2 semaines
- **Livrables** :
 - Un fichier SQL contenant toutes les requêtes exécutées
 - Une présentation PowerPoint (8 à 12 diapositives) comprenant :
 - Le contexte du dataset
 - Des statistiques descriptives
 - Des visualisations simples (graphiques, tableaux)
 - Une interprétation des résultats

Choix du Dataset

Vous pouvez choisir parmi les jeux de données suivants :

1. **Jeu de données e-commerce (Produits, Clients, Commandes)**
 - a. Lien : [E-commerce Dataset sur Kaggle](#)
 - b. Objectif : Analyser les ventes et le comportement des clients.
2. **Jeu de données sur les transactions bancaires**
 - a. Lien : [Synthetic Financial Transactions](#)
 - b. Objectif : Étudier les habitudes de dépenses et détecter des anomalies.
3. **Jeu de données sur les films et les notes IMDb**
 - a. Lien : [IMDb Dataset sur Kaggle](#)
 - b. Objectif : Analyser la popularité des films et leurs notes.

Vous pouvez également proposer votre propre jeu de données, sous réserve de validation.

Analyses et Statistiques Descriptives Attendues

Vous devrez explorer le jeu de données à travers plusieurs axes d'analyse, en utilisant SQL pour répondre aux questions suivantes :

1. Statistiques générales sur le dataset

- a. Nombre total de lignes et de colonnes
- b. Nombre de valeurs manquantes par colonne
- c. Vérification des doublons

2. Analyse des tendances et des distributions

- a. Pour un dataset e-commerce :
 - i. Quels sont les cinq produits les plus vendus ?
 - ii. Quelle est la répartition des ventes par pays ?
 - iii. Quel est le chiffre d'affaires moyen par commande ?
- b. Pour un dataset financier :
 - i. Quelle est la répartition des transactions par type de paiement ?
 - ii. Quel est le montant moyen d'une transaction ?
 - iii. Y a-t-il des pics de transactions à certaines heures de la journée ?
- c. Pour un dataset de films :
 - i. Quel est le genre de film le plus fréquent ?
 - ii. Quelle est la note moyenne des films par décennie ?
 - iii. Quels sont les réalisateurs les plus populaires ?

3. Analyses plus avancées avec SQL

- a. Agrégation des données (moyenne, médiane, écart-type)
- b. Regroupement (GROUP BY) pour trouver les tendances
- c. Filtrage (WHERE, HAVING) pour explorer les valeurs extrêmes
- d. Jointures (JOIN) si plusieurs tables sont disponibles

Présentation PowerPoint attendue (8-12 diapositives)

Structure recommandée :

1. Introduction

- a. Présentation du dataset et des objectifs

2. Statistiques générales

- a. Tableau récapitulatif des premières analyses

3. Exploration des tendances

- a. Graphiques (histogrammes, camemberts)
- b. Explications des principales observations

4. Analyses avancées

- a. Résultats issus des requêtes SQL plus complexes
- b. Éventuelles comparaisons entre sous-groupes

5. Conclusions & Recommandations

- a. Synthèse des résultats
- b. Limites de l'analyse et pistes d'amélioration

Critères d'Évaluation

- **Pertinence des analyses** (choix des indicateurs et justesse des requêtes SQL) – 40%
- **Qualité de la présentation** (clarté, structuration, visuels) – 30%
- **Interprétation des résultats** (capacité à tirer des conclusions) – 20%
- **Travail d'équipe & originalité** – 10%

Deadline et Remise

- **Date limite** : 06/02/2025
- **Livrables à envoyer par mail** : philippebenamar@gmail.com

Ce devoir vous permettra de pratiquer SQL tout en développant des compétences en analyse exploratoire et présentation de données.

Si vous avez des questions, n'hésitez pas à les poser !