

Proposition de sujet pour la SAÉ 2.04

Exploitation d'une base de données

1 Contexte professionnel

L'entreprise fictive **SocialConnect**, un réseau social en pleine croissance, souhaite disposer d'un outil performant pour centraliser, gérer et analyser les données liées à l'engagement des utilisateurs. Actuellement, les données sont dispersées dans différents fichiers, ce qui complique leur exploitation. Le projet vise à concevoir une base de données relationnelle regroupant ces informations et à produire des visualisations claires pour faciliter la prise de décision stratégique.

2 Problématique professionnelle

À partir d'un cahier des charges fourni, vous devez :

- Concevoir une base de données relationnelle permettant de centraliser les données des utilisateurs et de leurs interactions.
- Alimenter cette base avec un jeu de données d'essai.
- Réaliser des visualisations claires des données pour en tirer des analyses utiles à la stratégie de l'entreprise.

3 Objectifs pédagogiques

- Créer une base de données relationnelle pour centraliser les informations de l'entreprise.
- Mettre en œuvre un jeu d'essais pour tester la base et valider son fonctionnement.
- Proposer des visualisations graphiques et des analyses pertinentes pour répondre aux problématiques spécifiques de l'entreprise.

Exemples de problématiques à traiter

1. *Quels utilisateurs sont les plus actifs sur la plateforme ?*
2. *Quels types de publications génèrent le plus d'engagement ?*
3. *Quelle est la répartition des interactions par région géographique ?*
4. *Comment évolue l'engagement des utilisateurs au fil du temps ?*

4 Cahier des charges de l'entreprise

SocialConnect collecte des données sur plusieurs aspects de l'engagement des utilisateurs avec les caractéristiques suivantes :

- **Utilisateurs** : Identifiant, nom, âge, sexe, localisation, type d'abonnement.
- **Interactions** : Likes, partages, commentaires, vues de vidéos, clics sur publicités, etc.
- **Groupes et pages** : Créations, membres, publications, réactions, etc.
- **Événements** : Participations, téléchargements, etc.

5 Évaluations intégrées

Voici les évaluations qui seront intégrées dans cette SAÉ :

1. Évaluation collective sur la création de la BDD :

- Déposer sur Moodle le diagramme entité-association de la base de données. (**Correcteur:** M.hébert)

2. Évaluation individuelle sur la maîtrise de Pandas et Matplotlib :

- Manipulation des données avec Pandas et création de visualisations avec Matplotlib. (**Correcteur:** M. Hébert/Mme Nassiet)

3. Évaluation collective sur le rapport d'analyse des données : Ce rapport comprendra plusieurs éléments :

- (a) Un script Python permettant de générer un fichier SQL pour charger la base de données avec les données du fichier Excel (5 points).
- (b) Trois visualisations différentes mettant en avant des éléments significatifs de la base de données (3 visualisations à 2 points chacune).
- (c) Une analyse approfondie pour chacune des visualisations, contextualisant les données dans leur ensemble (3 analyses à 1 point chacune).
- (d) Originalité : des travaux trop similaires ou manquant d'intelligence pratique ou approfondie seront pénalisés (jusqu'à -3 points). Cette évaluation est laissée à la discrétion des enseignants.
- (e) Partie autoévaluative : Vous rédigerez une conclusion de 1000 mots maximum justifiant la note que vous mettriez sur votre travail. Le système de notation suivant sera appliqué :
 - Si la différence entre MaNote et votre note est de 0, alors vous gagnez +3 points.
 - Si la différence est inférieure ou égale à 1, alors vous gagnez +1.5 points.
 - Si la différence est inférieure ou égale à 2, alors vous obtenez 0 point.
 - Si la différence est inférieure ou égale à 3, alors vous perdez -1.5 points.
 - Si la différence est supérieure à 3, vous perdez -3 points.

4. Évaluation individuelle finale :

- Vous devrez utiliser votre diagramme entité-association et écrire, en SQL, les requêtes répondant à des questions de requêtage. Voici quelques exemples :
 - Comment obtenir la moyenne de messages privés des utilisateurs avec un abonnement gratuit ?
 - Comment récupérer l'ensemble des hommes de plus de 41 ans, classés par ordre croissant de leur nombre de commentaires ?

6 Planning indicatif des séances

Semaine	Activité	Durée
S8	Amphi: Présentation	30 min
S10	Autonomie: Création de la base de données Dépôt sur Moodle avant le 08/03/2025 à 22h	6h
S11	Évaluation BD <i>Apporter votre modèle Entité-Relation sous format papier</i>	1h
S11	Autonomie: TP Matplotlib	3h
S13	Évaluation Panda-Matplotlib	1h
S12	Autonomie: TP Panda	3h
S12 & 13	Proposer des visualisations graphiques Analyser les données à partir des visualisations Dégager des tendances en trouvant les éventuelles corrélations	6h
S14-S15	Rédaction du rapport	3h
S15	Dépôt du rapport avant le 11/04 à 12h00	

7 Jeu d'essai fourni

1. Liste d'utilisateurs avec leurs caractéristiques (âge, sexe, localisation, etc.).
2. Données d'interactions (likes, partages, commentaires, etc.).
3. Informations sur les groupes et pages (créations, membres, publications, etc.).
4. Données sur les événements (participations, téléchargements, etc.).

8 Rôles des étudiants

Les étudiants travailleront en groupes de 3 à 4 personnes. Chaque membre est attendu pour contribuer activement à toutes les étapes du projet et sera évalué individuellement pour garantir une juste répartition des efforts.

9 Résultats attendus

Une base de données fonctionnelle, des analyses pertinentes, et des visualisations qui aident SocialConnect à mieux comprendre ses données et à prendre des décisions stratégiques.

10 Exemples de problématiques à traiter

1. Quels utilisateurs sont les plus actifs sur la plateforme ?
2. Quels types de publications génèrent le plus d'engagement ?
3. Quelle est la répartition des interactions par région géographique ?
4. Comment évolue l'engagement des utilisateurs au fil du temps ?