

8STT108 – Analyse statistique des données de masse

TP #1

Objectif : Extraire des informations pertinentes et décrire une base de données par l'exploration multidimensionnelle, la régression et la classification.

Objectifs du TP

Mise en situation : vous commencez dans une nouvelle entreprise en tant que scientifique des données. Lors de votre période d'apprentissage, on vous donne comme mission de choisir une base de données en vous demandant d'y extraire le maximum d'information pertinente pour présenter aux gestionnaires du département d'intelligence d'affaire l'essence de la base des données.

L'objectif du travail pratique est d'évaluer l'autonomie des étudiants avec une base de données inconnue pour en faire sortir des informations pertinentes et décrire au mieux celle-ci.

Planification

Chaque groupe d'étudiant devrait suivre les étapes suivantes :

- 21 janvier 2025 : formation des groupes et choix de la base de données. Un courriel devra être envoyé par équipe avant le 28 janvier afin de présenter l'équipe et la base de données choisie.
- 11 février 2025 : rencontre avec chaque groupe pour faire le point.
- 25 février 2025 : présentation orale
- 28 février 2025 : remise du rapport écrit.

Description

Chaque groupe d'étudiant est composé d'un maximum de 5 étudiants. Dans un groupe, chaque étudiant devra travailler et parler lors de la présentation orale.

Le 11 février 2025, chaque groupe pourra rencontrer l'enseignant afin de faire un état d'avancement de leur TP#1.

Le 28 février 2025 chaque groupe d'étudiant devra présenter ces travaux.

Données

Voici une liste non exhaustive de référence ou trouver votre sujet/données :

- Kaggle (<https://www.kaggle.com/datasets>)
- Google Dataset Search (<https://datasetsearch.research.google.com/>)
- Canadien govt open data (<https://open.canada.ca/en/open-data>)
- Github Public Dataset (<https://github.com/awesomedata/awesome-public-datasets>)
- Québec données (<https://www.donneesquebec.ca/>)

Il est à noter que votre choix de base de données et sujet devra être approuvé par l'enseignant. Vous pouvez aussi choisir une base de données déjà utilisé dans un autre cours.

Évaluation

- Rapport écrit (70%) :
 - ✓ Choix des méthodologies statistiques appropriées
 - ✓ Visualisation
 - ✓ Description des conclusions
- Présentation orale (30%)
 - ✓ Originalité de la présentation
 - ✓ Message clair et pouvant être compris par tous

Aucun retard ne sera accepté. Les travaux en retard se verront attribuer la note 0. La note de passage minimale est fixée à 60%. Les travaux remis devront être conformes aux exigences de la politique institutionnelle en matière de maîtrise du français. Une pénalité allant jusqu'à 5% de la note du travail pourra être infligée en raison de la qualité du français.