

Cycle pluridisciplinaire d'études supérieures Sciences des données, arts et cultures

Syllabus 2023/24 (L1) - Titre du cours :

L'analyse de données appliquée à l'étude de données historiques, littéraires et artistiques

Objectifs du cours

Ce cours a pour objectif de doter les étudiants des outils et les méthodes centraux de l'analyse de données, ainsi que de leur donner l'occasion d'appliquer ces compétences à des données historiques, littéraires et artistiques dans le cadre de projets concrets. Les étudiants développeront leur capacité à construire un projet de recherche, à collecter, nettoyer et analyser des données, et à interpréter les résultats. Les étudiants seront amenés à développer ces compétences afin d'utiliser les données comme un moyen de répondre à des questions interdisciplinaires complexes.

Organisation

- Cours magistraux : 20%
- Travaux pratiques en classe : 30%
- Projets interdisciplinaires individuels ou en groupe : 30%
- Séances de tutorat en petits groupes : 20%

Exemples de projets

- **Question** : Quels sont les facteurs qui influencent l'appréciation des différentes catégories de musique ?

Données : Liste et description des musiques disponibles sur Spotify

Méthodes :

- Catégorisation des musiques selon leurs propriétés acoustiques (valence, énergie, rythme...) et leurs genres (rap, électro, classique...) afin d'identifier les grandes tendances.
- Identification des caractéristiques musicales qui peuvent prédire les préférences pour chaque catégorie de musique à l'aide d'une modélisation statistique.

- **Question** : Comment la représentation des femmes dans la littérature chinoise a-t-elle évolué au cours de l'histoire ?

Données : Résumés de fictions chinoises anciennes et modernes

Méthodes :

- Comptage des occurrences de mots pour mesurer la présence et la proportion de personnages féminins et identifier les thèmes associés à la présence ou l'absence de femmes.
- Comparaison des résultats avec la littérature en anthropologie de la famille et des relations de genre pour analyser les évolutions observées.

- **Question** : Comment les protagonistes de films sont-ils représentés sur les affiches ?

Données : Collection des affiches de films diffusés aux États-Unis et en France

Méthodes :

- Extraction automatique des éléments de chaque affiche (présence des personnages, émotions des personnages, couleurs et contrastes).
- Construction de modèles statistiques pour relier ces éléments aux genres de films, et pour étudier l'évolution des affiches au cours du XXe et du début du XXIe siècle en fonction du pays.

Contenu du cours

Introduction aux méthodes d'analyse de données

- Définitions et concepts de base de l'analyse de données
- Types de données et méthodes d'analyse correspondantes
- Outils de programmation pour l'analyse de données

Collecte de données

- Sources de données historiques, littéraires et artistiques
- Méthodes de collecte de données
- Éthique de la collecte de données

Préparation des données

- Nettoyage des données
- Transformation des données
- Normalisation des données

Analyse exploratoire des données

- Visualisation graphique des données
- Mesures de tendance centrale et de dispersion
- Analyse de la corrélation

Analyse de données avancée

- Régression linéaire simple et multiple
- Classification et regroupement
- Analyse factorielle

Applications interdisciplinaires

- Spécificités de l'application des méthodes d'analyse de données à l'étude de données historiques, littéraires et artistiques
- Élaboration de questions de recherche

Réalisation du projet interdisciplinaire en groupe

- Choix d'un sujet de recherche interdisciplinaire
- Collecte et préparation des données
- Analyse exploratoire et avancée des données
- Interprétation des résultats
- Rédaction d'un rapport de projet
- Présentation orale du projet

Évaluation

- Rendus intermédiaires : 30%
 - Projet interdisciplinaire individuel ou collectif : 30%
 - Présentation orale final : 40%
-

