



TP Noté - Atelier R Rstudio

Gaël Roustan (Argonautes)

2024

Abstract

Ce document contient les instructions du devoir noté.

TP Noté atelier R Studio

Prise en main de R Studio (Durée : 1 heure)

- Installation et configuration de R Studio
- Manipulation de l'interface de R Studio
- Utilisation des raccourcis clavier
- Création d'un projet R

Importation et manipulation de données

- Importation de données à partir de fichiers CSV, Excel, etc.
- Exploration et nettoyage des données
- Manipulation de données avec les fonctions de base de R
- Filtrage, tri et regroupement des données

Visualisation des données

- Création de graphiques statiques (histogrammes, diagrammes en barres, etc.)
- Utilisation des packages de visualisation (ggplot2, plotly, etc.)
- Personnalisation des graphiques (couleurs, légendes, titres)
- Création de graphiques interactifs

Analyse de données avec R

- Application de techniques statistiques (moyenne, médiane, écart-type, etc.)
- Utilisation de fonctions pour l'analyse des données
- Réalisation de tests statistiques (test t, ANOVA, corrélation)
- Modélisation statistique avec des techniques de régression

Projets pratiques

- Réalisation de projets pratiques basés sur des données réelles
- Manipulation avancée des données
- Visualisation avancée et analyse exploratoire
- Application de techniques statistiques

Projet Fil Rouge

Analyse exploratoire de données sur les utilisateurs d'une plateforme de streaming vidéo

Description

Dans ce projet, vous allez utiliser R Studio pour réaliser une analyse exploratoire de données sur un jeu de données d'utilisateurs d'une plateforme de streaming vidéo axée sur les nouvelles technologies. L'objectif est de comprendre les préférences des utilisateurs et d'extraire des informations pertinentes à partir des données.

Étapes du projet

1. Importation des données : Importez le jeu de données contenant des informations sur les utilisateurs, telles que l'âge, le sexe, l'emplacement géographique, les préférences de contenu, etc. Vous pouvez utiliser la source de données ouverte [MovieLens](#) qui fournit des données sur les évaluations et les tags de films pour créer votre jeu de données. A des fins pédagogiques, vous prendrez l'[échantillon de données](#).
2. Exploration des données : Effectuez une exploration initiale des données en calculant des statistiques descriptives (moyenne, médiane, écart-type) et en identifiant les valeurs manquantes ou aberrantes.
3. Visualisation des données : Créez des graphiques pour représenter les relations entre les variables clés, tels que des diagrammes en secteurs pour les préférences de contenu, des cartes pour la répartition géographique des utilisateurs, etc.
4. Analyse des tendances : Identifiez les tendances d'utilisation de la plateforme en analysant les variations saisonnières, les habitudes de visionnage, les heures de pointe, etc.
5. Modélisation : Utilisez des techniques de clustering pour regrouper les utilisateurs en fonction de leurs préférences et de leur comportement. Analysez les clusters obtenus pour mieux comprendre les segments d'utilisateurs.
6. Présentation des résultats : Créez un rapport ou une présentation synthétisant vos principales découvertes et conclusions à partir de l'analyse des données. Mettez l'accent sur les insights pertinents pour la plateforme de streaming vidéo axée sur les nouvelles technologies.

Pour ce projet, vous pouvez utiliser le jeu de données "MovieLens" qui est une source de données ouverte largement utilisée dans le domaine de l'analyse de données liée aux films et aux recommandations. Cela

vous permettra de vous immerger dans l'analyse de données liées aux nouvelles technologies et d'appliquer les compétences acquises lors de l'atelier R Studio.

Une fois la découverte du jeu de données, répondez aux questions du [formulaire](#). Etant donné que le formulaire demande d'envoyer des graphiques, il vous faut un compte Google pour envoyer vos réponses.