

Présentation orale projet

Lepetit Lenny Kravchenko Daria Palazy Alric

2025-01-14

Introduction / Présentation du Projet

- ▶ Notre projet vise à reprendre les données du site "<https://www.programme-tv.net/>" qui nous donne le programme télévisé des principales chaînes françaises afin de pouvoir effectuer des filtres sur les films pour savoir quel est le film qu'on souhaite regarder qui correspond le plus à nos critères.
- ▶ Les critères des films qu'on a réussi à scraper sont le nom des films, la chaîne sur laquelle ils ont été diffusés, leur durée, leur heure de départ et le type de film.

Etapes pour réussir ce projet

- ▶ Partie 1 : Scrapping des données du site et mise en forme sur un tableau.
- ▶ Partie 2 : Création d'une application Shiny pour pouvoir filtrer les films.
- ▶ Partie 3 : Inclusion d'un onglet Mes_films_Regardés pour pouvoir voir notre liste de films regardés et y laisser un commentaire.

Les packages utilisés pour notre projet

- ▶ tibble
- ▶ rvest
- ▶ shiny
- ▶ dplyr
- ▶ DT

Partie 1 : Le scrapping de données

- ▶ Pour effectuer le scrapping, nous nous sommes aidés du livre “R pour l'extraction de données sur le web” de Armel Soubeiga ainsi que du cours.
- ▶ Notre projet a une dimension seulement personnelle et non commerciale, nous n'utilisons pas d'API pour scraper les données. Nous ne comptons également pas laisser notre code accessible au public après cette présentation pour respecter les conditions d'utilisations.
- ▶ L'absence d'API dans notre scrapping peut poser des problèmes, cependant comme on ne présente pas un projet public en dehors de cette présentation, nous ne risquons pas grand chose si ce n'est l'interdiction d'accès au site, qui nous importe assez peu finalement.

Comment a-t-on fait le scrapping ?

5 Grandes étapes pour faire ce scrapping :

- ▶ Installation des packages (rvest pour le scrapping)
- ▶ Lecture de la page avec `read_html()` puis le contenu est récupéré avec `html_nodes`
- ▶ Extraction des données
- ▶ Nettoyage des données (notamment pour les formats des horaires)
- ▶ Création du tableau de données (tibble)

Affichage des parties de code.

Nous n'avons pas réussi à résoudre le problème associé à tinytex pour la présentation de ce powerpoint.

Quand on met des chunk sur le diapo on est censé installer tinytex ou installer MikTeX mais peu importe la tentative, cela ne fonctionne pas et le site continue de nous dire que Latex n'est pas installé.

Nous allons vous présenter le résultat de notre code à la fin, avec d'une part le résultat qu'on obtient avec le scrapping et de l'autre avec l'ensemble du code.

Exemple de partie du code de scrapping pour la correction du format des horaires

```
-[1] data <- data %>%  
-[2] mutate(Heure_de_départ = gsub("h", ":",  
Heure_de_départ)) %>%  
-[3] mutate(Heure_de_départ = as.POSIXct(Heure_de_départ,  
format = "%H:%M", tz = "UTC")) %>%  
-[4] mutate(Heure_de_départ = format(Heure_de_départ,  
"%H:%M"))
```


Partie 2 : L'application Shiny

Pour créer notre application shiny, on a implémenté un user interface ainsi qu'un serveur.

Sur cette application shiny, on a dans l'onglet principal tout un ensemble de filtres à gauche, les films et leurs informations à droite. Dans un autre onglet Mes_films_regardés, on peut voir la liste des films que l'on a regardé et y laisser un commentaire.

Commentaires sur notre projet

Points positifs : On a appris le web scrapping dans un exemple concret, notre interface shiny est satisfaisante et nos données sont scrappées dans un tableau de manière fluide, claire et correcte. (grâce au formatage des horaires par exemple)

Points négatifs : On a pas fait attention aux règles d'utilisations des données avant de terminer notre projet donc on ne pourra pas publier notre projet. Il sera nécessaire de la faire dans les prochains projets. Aussi, nous avons eu des difficultés avec le Markdown pour la présentation, ce qui nous a empêché de montrer des morceaux de code de façon correcte.

Execution de notre projet

Nous allons vous montrer le rendu final