

Projet : NChoice



Quel est le film idéal à voir ce soir ?

RENDU PROJET NCHOICE : «SDD4»

(NIVEAU L3)

présenté par :

Kassandra MARTIN, Léa MARIE

Noa THÉBAUT, Joan SCHMIT VALLES

réalisé sous la direction de :

Namrata PATEL, Sandra BRINGAY

SOMMAIRE :

SOMMAIRE :	2
1. Descriptif de la table	3
1.1 Filtrage	3
2 Modèles MCD et MOD	5
2.1 Modèle MCD	5
2.2 Modèle MOD	5
3 Import des données et requêtes	6
3.1 Import des données	6
3.2 Nos requêtes	6

1. Descriptif de la table

<https://www.kaggle.com/datasets/shivamb/netflix-shows>

À propos de cet ensemble de données : Netflix est l'une des plateformes de streaming multimédia et vidéo les plus populaires. Ils proposent plus de 8000 contenus cinématographiques. A la mi-2021, ils ont plus de 200 millions d'abonnés dans le monde. Cet ensemble de données tabulaires se compose de listes de tous les contenus disponibles sur Netflix, ainsi que de détails tels que la distribution, les réalisateurs, les notes, l'année de sortie, la durée etc.

Notre base de données est donc une liste de films et séries disponibles sur Netflix.

1.1 Filtrage

Le filtrage de notre base de données s'est déroulé via un raisonnement par déduction logique de ce qui allait nous servir dans le but de répondre à notre problématique. En effet, nous avons commencé avec Excel en supprimant les colonnes que l'on considérait inutile pour notre application puis en modifiant tous les caractères spéciaux (ex: "é", "à", "û"...).

Ainsi, les colonnes conservées sont les suivantes :

show_id : CHAR(5)

type : CHAR(7)

title : VARCHAR

director : VARCHAR

cast : VARCHAR

country : VARCHAR

release_year (renommé year) : DATE (selon un format de type "AAAA")

duration : CHAR(10)

listed_in (renommé genre) : CHAR(28)

description : VARCHAR

Afin de continuer, nous avons créé une feuille pour chaque table et chaque relation déséquilibrée entre celles-ci. Pour cela, nous avons dû créer des identifiants pour chacune d'elles. Par la suite, nous avons séparé les genres, director et cast car ils étaient sous forme de listes scindées par des virgules. Pour effectuer cette manipulation, toujours sur Excel, nous avons mis chacune de ces colonnes sur une nouvelle feuille puis avons utilisé la fonction "séparer par des virgules" afin de copier et coller les valeurs dans leur feuille respective.

Ultérieurement, nous avons supprimé les doublons et nous nous sommes rendu compte que certains ne se supprimaient pas (dans les feuilles genre, director et cast). En effet, un espace se trouvait devant certaines instances ce qui empêchait le processus. Afin de résoudre ce problème, nous avons utilisé le langage Python sur un document Google Colab, avec lequel nous avons supprimé ces espaces du début à l'aide de la fonction "`str.lstrip()`".

```

import pandas as pd

data_show = pd.read_csv('ensemble-show.csv', sep=";")
data_cast = pd.read_csv('cast.csv', sep=";")
data_jouer = pd.read_csv('jouer.csv', sep=";")
data_director = pd.read_csv('director.csv', sep=";")
data_genre = pd.read_csv('genre.csv', sep=";")

data_cast["cast"] = data_cast["cast"].str.lstrip()

data_cast.to_csv("data_new.csv")

data_director["director"] = data_director["director"].str.lstrip()

data_director.to_csv("director_new.csv")

data_genre["genre"] = data_genre["genre"].str.lstrip()

data_genre.to_csv("genre_new.csv")

```

Pour finir, sur le même colab Python, nous avons codé une boucle afin de relier chaque identifiants des acteurs, directeurs et genres associés aux films et séries.

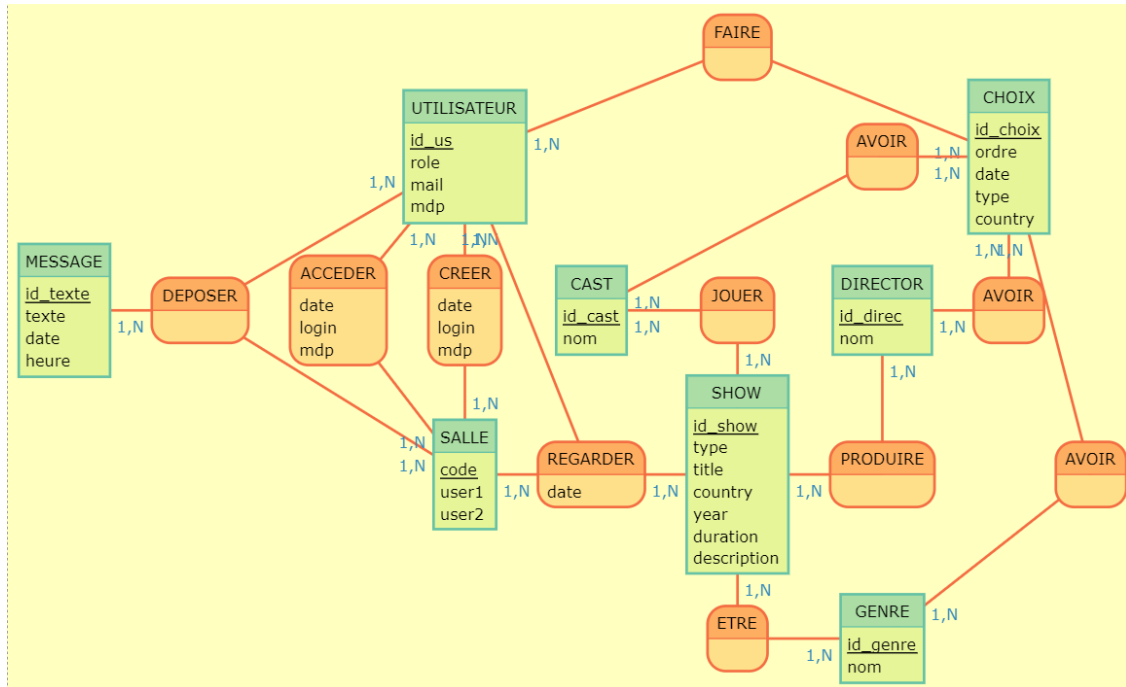
```

etre=pd.DataFrame({"id_genre":[],
                  "id_show":[] })
for i in range (0,len(genre["genre"])):
    for j in range (0,len(all["genre"])):
        if genre["genre"][i] in all["genre"][j]:
            df1 = pd.DataFrame({"id_genre": [genre['id_genre'][i]],
                              "id_show": [all["show_id"][j]]})
            etre=pd.concat([etre,df1])

```

2 Modèles MCD et MOD

2.1 Modèle MCD



2.2 Modèle MOD

GENRE (id_genre, nom)
 CAST (id_cast, nom)
 SHOW (id_show, type, title, country, year, duration, description)
 DIRECTOR (id_direct, nom)
 MESSAGE (id_texte, texte, date, heure)
 SALLE (code, user1, user2)
 UTILISATEUR (id_us, role, mail, mdp)
 CHOIX (id_choix, ordre, date, type, country)
 ETRE (id_genre, id_show)
 PRODUIRE (id_direct, id_show)
 JOUER (id_cast, id_show)
 REGARDER (date, id_show code, id_us)
 ACCEDER (date, login, mdp, code, id_us)
 CREER (date, login, mdp, code, id_us)
 DEPOSER (code, texte, id_us)
 FAIRE (id_us, id_choix)
 AVOIR (id_genre, id_choix)
 AVOIR (id_direct, id_choix)
 AVOIR (id_cast, id_choix)

3 Import des données et requêtes

3.1 Import des données

Une fois toutes ces étapes effectuées, nous avons exporté nos tables sur phpMyAdmin afin de pouvoir y faire des requêtes.

3.2 Nos requêtes

- Lister les noms des contenus qui appartiennent au genre “Kids TV”

```
SELECT show_new.title
FROM show_new, etre, genre_new
WHERE show_new.id_show = etre.id_show
AND genre_new.id_genre = etre.id_genre
AND genre_new.genre = "Kids TV"
```

title
Tayo and Little Wizards
Angry Birds
Chhota Bheem
He-Man and the Masters of the Universe
Numberblocks
Saved by the Bell
Pokemon Master Journeys: The Series
Tilpo Tilpo
Mighty Raju
Kid Cosmic
Octonauts: Above & Beyond
Tobot Galaxy Detectives
Tayo the Little Bus
Bunk d
Dive Club
Sharkdog
Heroes of Goo Jit Zu
Luv Kushh
Pororo - The Little Penguin

- Lister les contenus les plus anciens

```
SELECT *
FROM show_new
WHERE show_new.year = (
    SELECT MIN(show_new.year)
    FROM show_new)
```

id_show	type	title	country	year	duration	description
s4251	TV Show	Pioneers: First Women Filmmakers*		1925	1 Season	This collection restores films from women who tack...

- **Afficher les contenus fait par Quentin Tarantino dans lesquels joue Brad Pitt**

```
SELECT show_new.title
FROM show_new, director_new, produire
WHERE director_new.id_directeur = produire.id_directeur
AND produire.id_show = show_new.id_show
AND director_new.directeur = "Quentin Tarantino"
AND EXISTS(SELECT show_new.title
FROM show_new, jouer, cast_new
WHERE cast_new.id_cast=jouer.id_cast
AND jouer.id_show = show_new.id_show
AND cast_new.cast = "Brad Pitt");
```

title
Django Unchained
The Hateful Eight: Extended Version
The Hateful Eight
Inglourious Basterds
Jackie Brown
Kill Bill: Vol. 1
Kill Bill: Vol. 2
Pulp Fiction

- **Compter le nombre de contenus faits par réalisateur par ordre décroissant**

```
SELECT COUNT(show_new.title) as nb_show, director_new.directeur
FROM show_new, director_new, produire
WHERE produire.id_show = show_new.id_show
AND produire.id_directeur=director_new.id_directeur
GROUP BY director_new.directeur
ORDER BY nb_show DESC
```

nb_show ▼ 1	director
22	Rajiv Chilaka
16	Suhas Kadav
16	Marcus Raboy
15	Jay Karas
13	Cathy Garcia-Molina
12	Martin Scorsese
12	Youssef Chahine
11	Steven Spielberg
10	Don Michael Paul
9	Shannon Hartman
9	Yilmaz Erdogan
9	Anurag Kashyap
9	David Dhawan

- **Afficher tous les contenus de type “Movie”**

```
SELECT title from show_new
WHERE type="Movie"
```

title
Dick Johnson Is Dead
The Starling
Stowaway
Wild Dog
Oloibiri
Tell Me When
Keymon and Nani in Space Adventure
Motu Patlu Dino Invasion
Motu Patlu in Octopus World

Les requêtes suivantes nous serviront durant la sélection des choix des utilisateurs (type, genre, director, country, ...).

- **Afficher tous les genres :**

```
SELECT DISTINCT(genre_new.genre)
FROM genre_new
```

genre
Documentaries
TV Comedies
Thrillers
Horror Movies
Kids TV
Action & Adventure
Reality TV
Anime Series
International Movies
Sci-Fi & Fantasy
Classic Movies
International TV Shows
TV Shows
Stand-Up Comedy
TV Action & Adventure
Movies
Stand-Up Comedy & Talk Shows

- **Afficher tous les types de show_new:**

```
SELECT DISTINCT(show_new.type)
FROM show_new
```

type
r Movie
r TV Show

- **Afficher tous les pays de la table show_new :**

```
SELECT DISTINCT(show_new.country)
FROM show_new
```

country
United States
France, United States
Germany, United States
Canada, Nigeria, United States
Mexico
Italy
Australia
France, United States, United Kingdom, Canada
France, United Kingdom
India
India, United Kingdom
Canada, United States, Mexico
Philippines
Germany
South Korea
United Kingdom, Germany, United States

- **Afficher tous les réalisateurs :**

```
SELECT director_new.director
FROM director_new
```

director
Kirsten Johnson
Bruno Garotti
David Zucker
Zeek Earl
Michael Kampa
Chanda Bell
Jennifer Westcott
Richard Greenberg
Peter Berg
Nicolas Pesce
Christopher Cain
Harald Zwart
Stacia Crawford
Kinka Usher
Remi Weekes
Tope Oshin
Terrie Samundra
Olivier Marchal