

## RAPPORT DE PROJET

### NoSQL Databases - Projet Exploration et Interrogation de Bases de Données NoSQL avec Python

Semestre 8 - 2025

Moustapha FAYE  
Pierre GARNIER  
Vanessa PHAM  
TD45  
30/03/2025

## Sommaire

<b>Code Source : lien github.....</b>	<b>3</b>
<b>1. Objectifs.....</b>	<b>3</b>
<b>2. Démarche.....</b>	<b>3</b>
a. Fonctionnalités.....	3
b. Installation.....	4
c. Structure du projet.....	4
d. Lancement.....	4
e. Configuration de l'application.....	5
<b>3. Requêtes implémentées.....</b>	<b>6</b>
MongoDB : requêtes dans "mongodb_queries.py" .....	6
Neo4j en Cypher : requêtes dans "neo4j_queries.py" .....	14
Questions Transversales : "neo4j_queries.py" et "mongodb_queries.py" .....	21
Visualisations.....	24
<b>4. Difficultés Rencontrées.....</b>	<b>26</b>
<b>5. Solutions Adoptées.....</b>	<b>27</b>
<b>6. Conclusion.....</b>	<b>27</b>

## Code Source : lien github

<https://github.com/Pieerr/NoSQL-project>

Dans ce répertoire github, vous retrouverez les codes sources du projet. Les fichiers avec le code des requêtes se trouvent dans le dossier "queries". [lien](#)

Vous retrouverez également une vidéo démo de l'application. [lien](#)

## 1. Objectifs

Ce projet a pour but de se familiariser avec deux bases de données NoSQL : MongoDB (base orientée document) et Neo4j (base orientée graphe), à travers le développement d'une application Python (via Streamlit) permettant :

- La connexion sécurisée à ces bases en cloud,
- L'exécution de requêtes complexes,
- L'analyse des résultats,
- Et leur visualisation.

L'application offre une interface utilisateur développée avec Streamlit pour interagir avec les données, exécuter des requêtes et visualiser les résultats.

## 2. Démarche

### a. Fonctionnalités

- Connexion aux bases de données MongoDB et Neo4j (locales ou cloud)
- Import de données depuis un fichier JSON
- Conversion de données de MongoDB vers Neo4j
- Exécution de requêtes sur MongoDB :
  - Analyses statistiques des films (années, votes, revenus, etc.)
  - Visualisations des distributions et corrélations
  - Filtrage et regroupement des données
- Exécution de requêtes sur Neo4j :
  - Analyse des relations entre acteurs et films
  - Recherche de chemins et de connexions
  - Détection de communautés d'acteurs
  - Recommandations de films

## b. Installation

Cloner le dépôt :

```
git clone https://github.com/Pieerr/NoSQL-project.git
cd NoSQL-project
```

Installer les dépendances :

```
pip install -r requirements-file.md
```

Configurer les bases de données :

- MongoDB (local ou Atlas)
- Neo4j (local ou cloud)

## c. Structure du projet

```
projet_nosql/
├── app.py
├── mongodb_connect.py
├── neo4j_connect.py
├── queries/
│   ├── __init__.py (fichier vide)
│   ├── mongodb_queries.py
│   └── neo4j_queries.py
├── data/
└── movies.json
```

## d. Lancement

Lancer l'application Streamlit :

```
streamlit run app.py
```

Dans l'interface web :

- Configurer les connexions aux bases de données
- Importer les données
- Exécuter des requêtes
- Visualiser les résultats

## e. Configuration de l'application

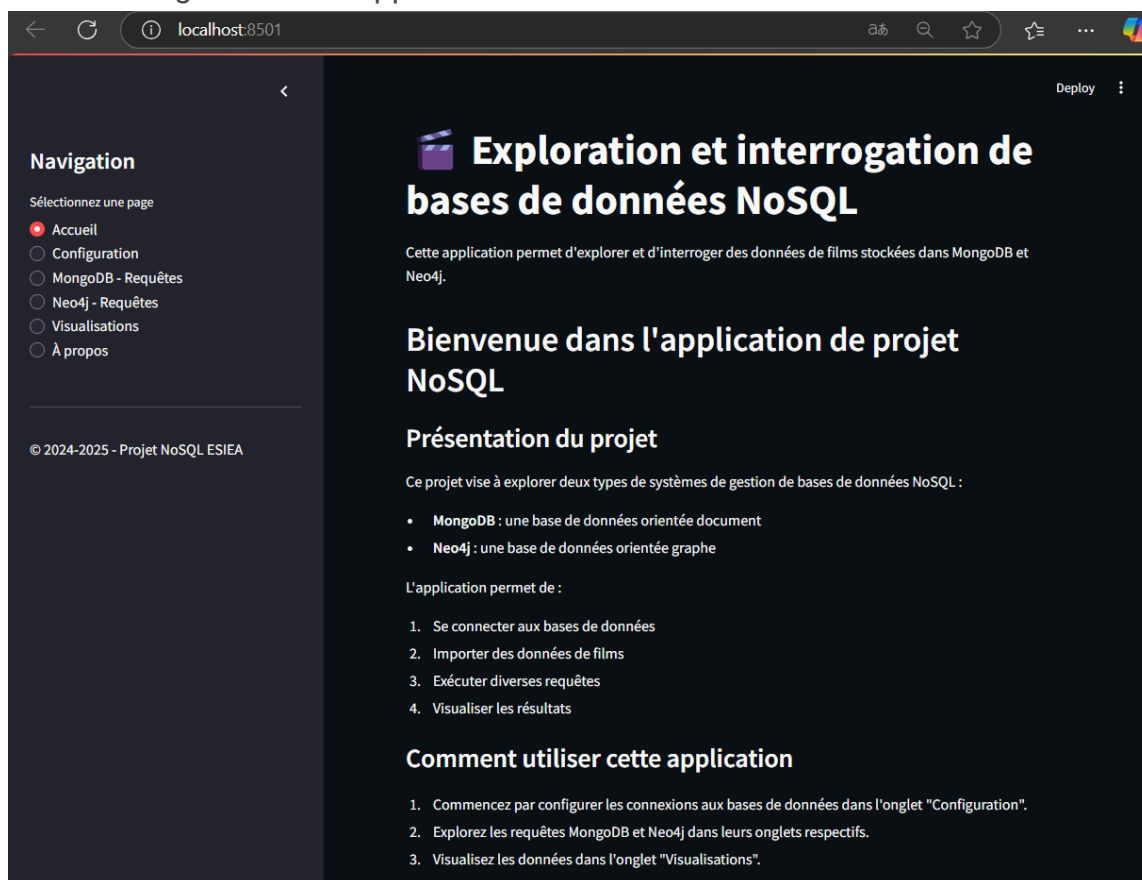


Figure 1 : Accueil

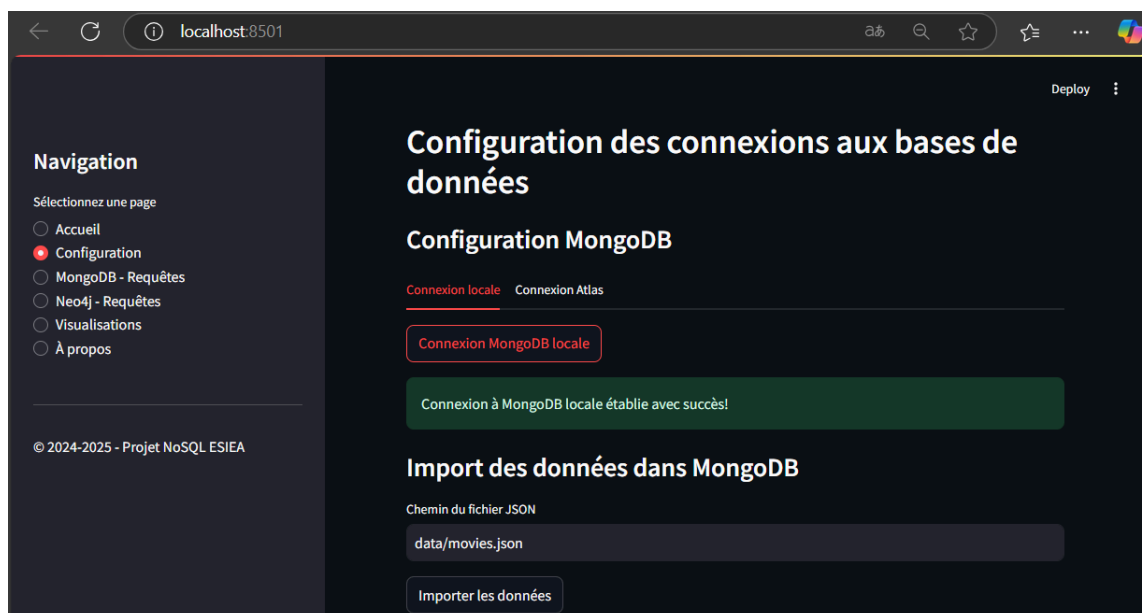


Figure 2 : Configuration : connexion MongoDB

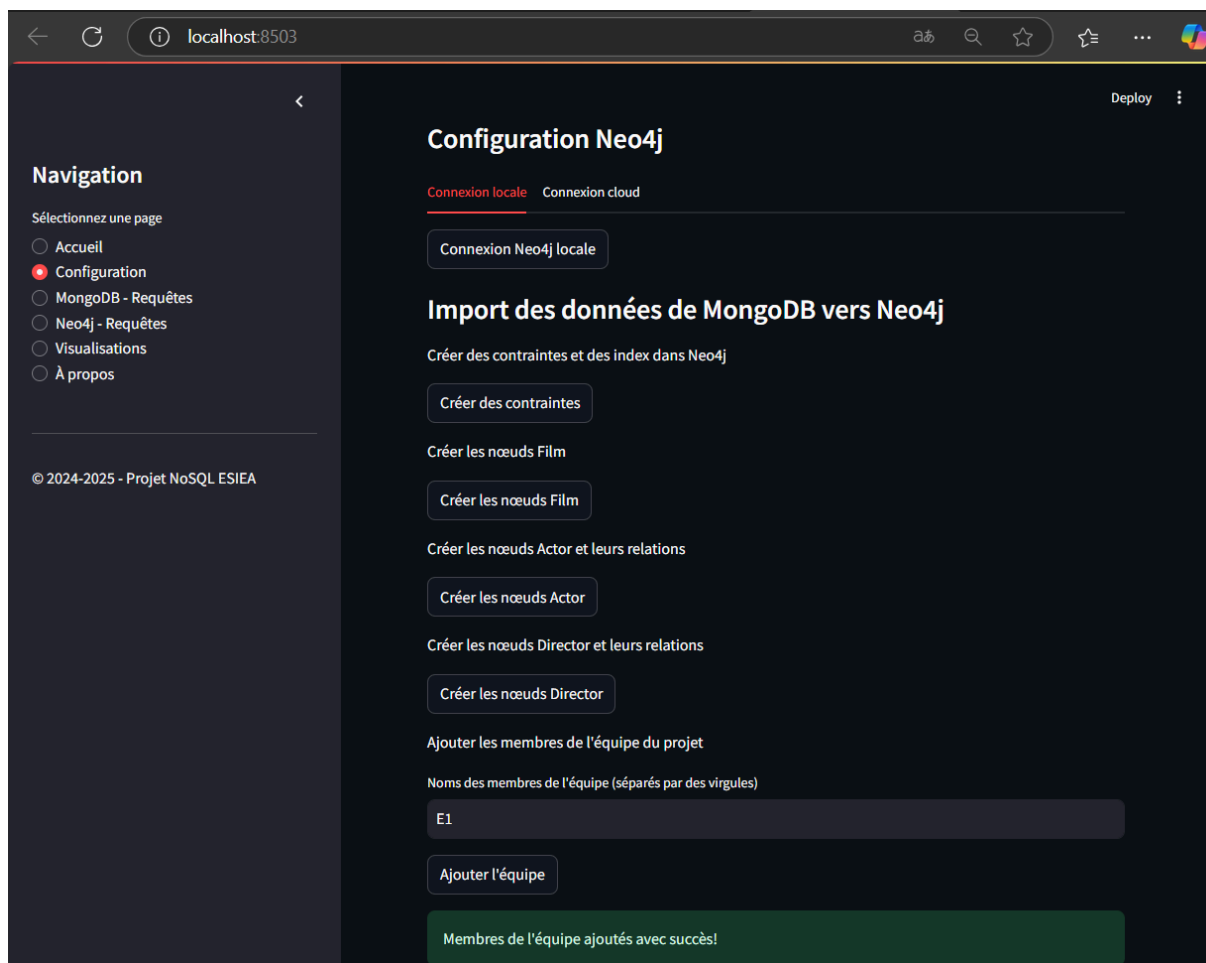


Figure 3 : Configuration : connexion MongoDB vers Neo4j

### 3. Requêtes implémentées

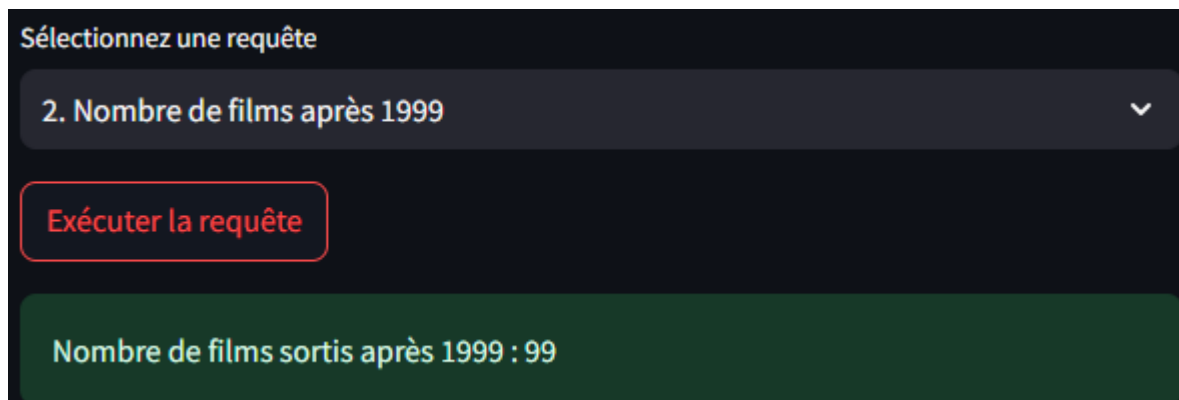
L'application implémente les 30 requêtes demandées dans l'énoncé du projet, dont :

MongoDB : requêtes dans "mongodb\_queries.py"

1. Afficher l'année où le plus grand nombre de films ont été sortis.



2. Quel est le nombre de films sortis après l'année 1999.



Sélectionnez une requête

2. Nombre de films après 1999

Exécuter la requête

Nombre de films sortis après 1999 : 99

3. Quelle est la moyenne des votes des films sortis en 2007.



Sélectionnez une requête

3. Moyenne des votes en 2007

Exécuter la requête

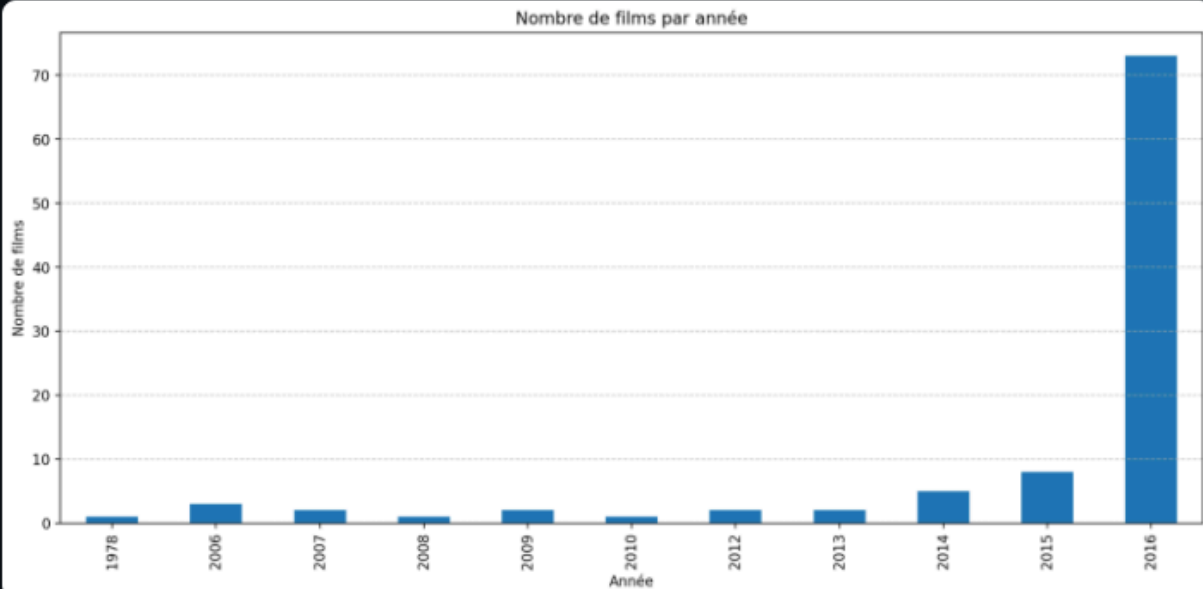
Moyenne des votes des films de 2007 : 192.50

4. Afficher un histogramme qui permet de visualiser le nombre de films par année.

Sélectionnez une requête

4. Histogramme des films par année

Exécuter la requête



5. Quelles sont les genres de films disponibles dans la bases



Sélectionnez une requête

5. Genres disponibles

Exécuter la requête

Genres disponibles dans la base :

```
[
  0 : "Action"
  1 : "Adventure"
  2 : "Animation"
  3 : "Biography"
  4 : "Comedy"
  5 : "Crime"
  6 : "Drama"
  7 : "Family"
  8 : "Fantasy"
  9 : "History"
  10 : "Horror"
  11 : "Music"
  12 : "Mystery"
  13 : "Romance"
  14 : "Sci-Fi"
  15 : "Thriller"
  16 : "War"
  17 : "Western"
]
```

6. Quel est le film qui a généré le plus de revenus.

Sélectionnez une requête

6. Film avec le plus de revenus

Exécuter la requête

Le film avec le plus de revenus est 'Star Wars: Episode VII - The Force Awakens' avec 936.63 millions de dollars.

7. Quels sont les réalisateurs ayant réalisé plus de 5 films dans la base de données ?

Sélectionnez une requête

7. Réalisateur avec plus de 5 films

Exécuter la requête

Aucun réalisateur n'a réalisé plus de 5 films.

8. Quel est le genre de film qui rapporte en moyenne le plus de revenus ?

Sélectionnez une requête

8. Genre avec le plus de revenus en moyenne

Exécuter la requête

Le genre 'Fantasy' rapporte en moyenne le plus de revenus avec 265.83 millions de dollars par film.

9. Quels sont les 3 films les mieux notés (rating) pour chaque décennie (1990-1999, 2000-2009, etc.) ?

Sélectionnez une requête

9. Top 3 films par décennie

Exécuter la requête

Top 3 films par décennie :

## Décennie 1970-1979

	title	year	Metascore
0	Top Dog	1978	67

## Décennie 2000-2009

	title	year	Metascore
0	The Departed	2006	85
1	Avatar	2009	83
2	The Dark Knight	2008	82

## Décennie 2010-2019

	title	year	Metascore
0	Moonlight	2016	99
1	Manchester by the Sea	2016	96
2	La La Land	2016	93

10. Quel est le film le plus long (Runtime) par genre ?

Sélectionnez une requête

10. Film le plus long par genre

Exécuter la requête

Film le plus long par genre :

	title	genre	Runtime (Minutes)
0	Pirates of the Caribbean: At World's End	Action	169
1	Interstellar	Adventure	169
2	Sing	Animation	108
3	The Wolf of Wall Street	Biography	180
4	The Wolf of Wall Street	Comedy	180
5	The Hateful Eight	Crime	187
6	The Hateful Eight	Drama	187
7	Fantastic Beasts and Where to Find Them	Family	133
8	Pirates of the Caribbean: At World's End	Fantasy	169
9	Silence	History	161
10	Split	Horror	117
11	La La Land	Music	128
12	The Hateful Eight	Mystery	187
13	Fifty Shades of Grey	Romance	125
14	Interstellar	Sci-Fi	169
15	The Departed	Thriller	151
16	Inglourious Basterds	War	153
17	The Magnificent Seven	Western	132

11. Créer une vue MongoDB affichant uniquement les films ayant une note supérieure à 80 (Metascore) et générant plus de 50 millions de dollars.

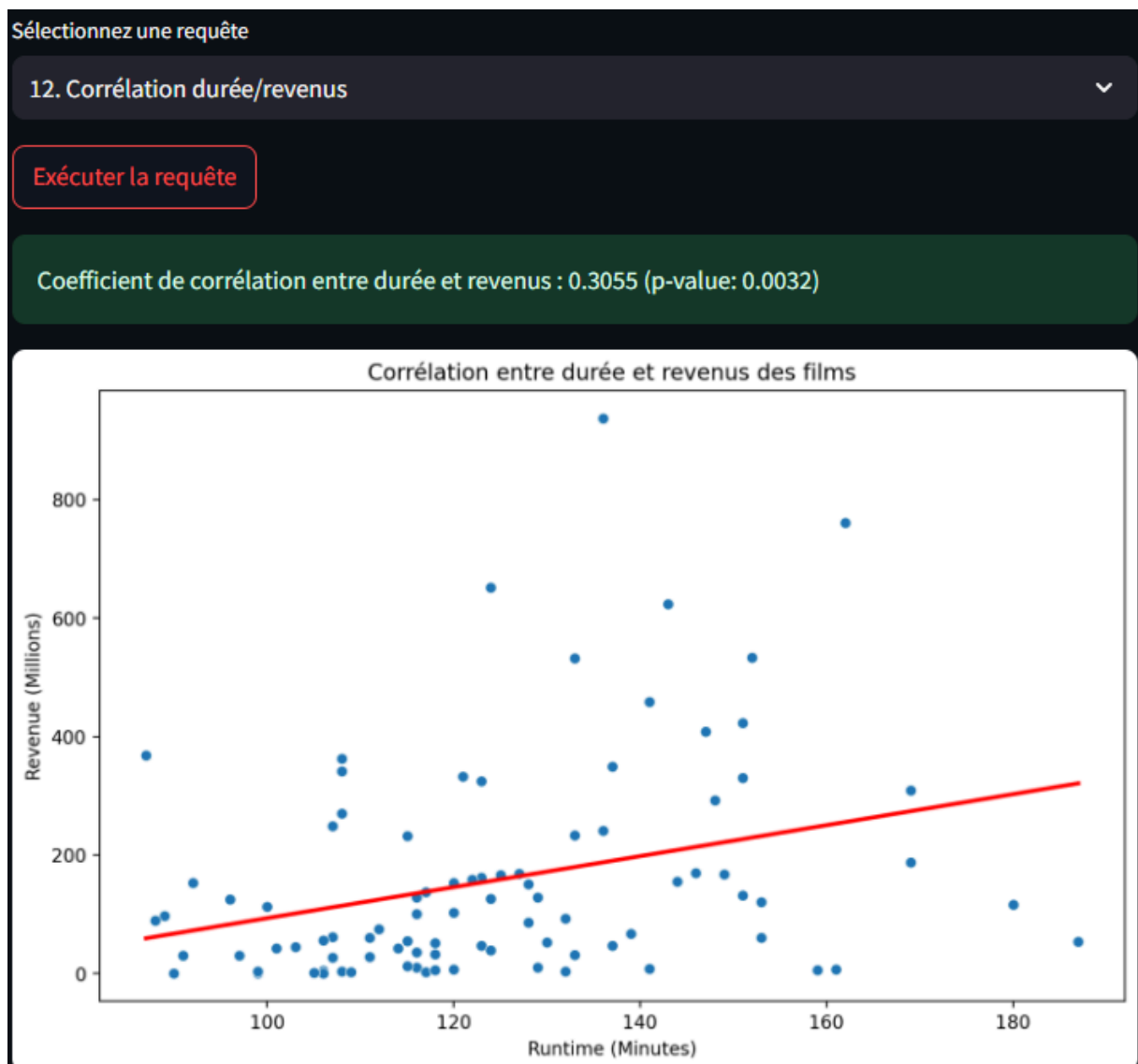
Sélectionnez une requête

11. Vue des films avec notes > 80 et revenus > 50M

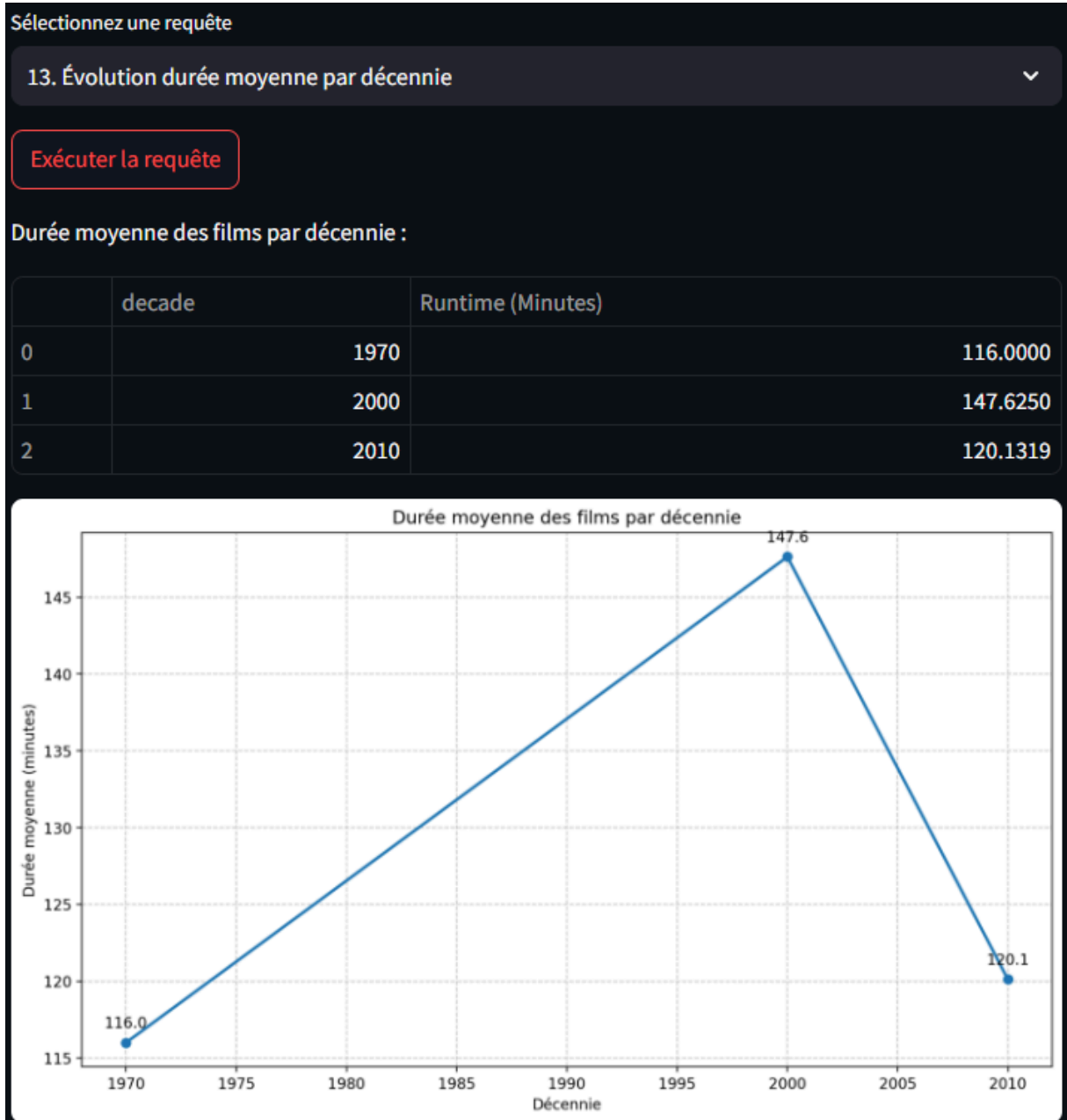
Exécuter la requête

Vue créée avec 8 films ayant une note > 80 et des revenus > 50M.

12. Calculer la corrélation entre la durée des films (Runtime) et leur revenu (Revenue).  
(réaliser une analyse statistique.)



13. Y a-t-il une évolution de la durée moyenne des films par décennie ?



Neo4j en Cypher : requêtes dans “neo4j\_queries.py”

14. Quel est l'acteur ayant joué dans le plus grand nombre de films ?

Sélectionnez une requête

14. Acteur avec le plus de films

Exécuter la requête

L'acteur ayant joué dans le plus grand nombre de films est 'Matthew McConaughey' avec 4 films.

15. Quels sont les acteurs ayant joué dans des films où l'actrice Anne Hathaway a également joué ?

Sélectionnez une requête

15. Acteurs ayant joué avec Anne Hathaway

Exécuter la requête

Acteurs ayant joué avec Anne Hathaway :

	actor_name	films
0	Austin Stowell	Colossal
1	Jason Sudeikis	Colossal
2	Jessica Chastain	Interstellar
3	Mackenzie Foy	Interstellar
4	Matthew McConaughey	Interstellar
5	Tim Blake Nelson	Colossal

16. Quel est l'acteur ayant joué dans des films totalisant le plus de revenus ?

Sélectionnez une requête

16. Acteur avec le plus de revenus

Exécuter la requête

L'acteur ayant joué dans des films totalisant le plus de revenus est 'Chris Evans' avec 1490.35 millions de dollars.

17. Quelle est la moyenne des votes ?

Sélectionnez une requête

17. Moyenne des votes

Exécuter la requête

La moyenne des votes est de 259963.74.

18. Quel est le genre le plus représenté dans la base de données ?

Sélectionnez une requête

18. Genre le plus représenté

Exécuter la requête

Aucun résultat trouvé ou pas de nœuds Genre.

19. Quels sont les films dans lesquels les acteurs ayant joué avec vous ont également joué ?



Sélectionnez une requête

19. Films des acteurs ayant joué avec vous

Nom du membre de l'équipe

E1

Exécuter la requête

Films des acteurs ayant joué avec E1 :

	film_title	year
0	A Monster Calls	2016
1	Hacksaw R_idge	2016
2	Sausage Party	2016
3	Star Trek Beyond	2016
4	Guardians of the Galaxy	2014
5	The Wolf of Wall Street	2013

20. Quel réalisateur Director a travaillé avec le plus grand nombre d'acteurs distincts ?

Sélectionnez une requête

20. Réalisateur avec le plus d'acteurs

Exécuter la requête

Le réalisateur ayant travaillé avec le plus d'acteurs est 'Christopher Nolan' avec 14 acteurs.

21. Quels sont les films les plus "connectés", c'est-à-dire ceux qui ont le plus d'acteurs en commun avec d'autres films ?

Sélectionnez une requête

21. Films les plus connectés

Exécuter la requête

Films les plus connectés (avec le plus d'acteurs en commun) :

	film1	film2	common_actor_count	common_actors
0	Pirates of the Caribbean: Dead Man's Chest	Pirates of the Caribbean: The Curse of the Black Pearl	3	Johnny Depp, Orlando Bloom, Keira Knightley
1	Captain America: Civil War	The Avengers	3	Chris Evans, Robert Downey Jr., Chris Hemsworth
2	The Dark Knight	The Prestige	2	Christian Bale, Michael Caine
3	Captain America: Civil War	Avengers: Age of Ultron	2	Chris Evans, Chris Hemsworth
4	Pirates of the Caribbean: On Stranger Tides	Pirates of the Caribbean: The Curse of the Black Pearl	2	Johnny Depp, Orlando Bloom
5	The Avengers	Avengers: Age of Ultron	2	Chris Evans, Chris Hemsworth
6	The Dark Knight	Sully	1	Aaron Eckhart
7	Gold	Interstellar	1	Matthew McConaughey
8	Top Dog	Arrival	1	Amy Adams
9	Gold	Sing	1	Matthew McConaughey

22. Trouver les 5 acteurs ayant joué avec le plus de réalisateurs différents.

Sélectionnez une requête

22. 5 acteurs avec le plus de réalisateurs

Exécuter la requête

Acteurs ayant joué avec le plus de réalisateurs différents :

	actor_name	director_count	directors
0	E1	5	Anthony Russo Shawn Burkett James Cameron Todd Ph
1	Scarlett Johansson	4	Anthony Russo Christopher Nolan Christophe Lourdelet
2	Chris Pratt	4	Morten Tyldum James Gunn Antoine Fuqua Colin Trevoi
3	Matthew McConaughey	4	Martin Scorsese Stephen Gaghan Christopher Nolan Chi
4	Ben Affleck	4	Dav_id Fincher Ben Affleck Gavin O'Connor Zack Snyder

23. Recommander un film à un acteur en fonction des genres des films ou il a déjà joué.

Sélectionnez une requête

23. Recommandation de film pour un acteur

Nom de l'acteur

Anne Hathaway

Exécuter la requête

Films recommandés pour Anne Hathaway :

	film_title	year	rating	votes
0	Pirates of the Caribbean: Dead Man's Chest	2006	unrated	552027
1	Pirates of the Caribbean: At World's End	2016	unrated	498821
2	Prometheus	2012	unrated	485820
3	Batman v Superman: Dawn of Justice	2016	unrated	472307
4	Jurassic World	2015	unrated	455169

24. Créer une relation INFLUENCE PAR entre les réalisateurs en se basant sur des similarités dans les genres de films qu'ils ont réalisés.

Sélectionnez une requête

24. Relation d'influence entre réalisateurs

Exécuter la requête

0 relations d'influence créées entre réalisateurs.

25. Quel est le "chemin" le plus court entre deux acteurs donnés (ex : Tom Hanks et Scarlett Johansson) ?

Sélectionnez une requête

25. Chemin le plus court entre deux acteurs

Premier acteur

Tom Hanks

Deuxième acteur

Scarlett Johansson

Exécuter la requête

Aucun chemin trouvé entre Tom Hanks et Scarlett Johansson.

26. Analyser les communautés d'acteurs : Quels sont les groupes d'acteurs qui ont tendance à travailler ensemble ? (Utilisation d'algorithmes de détection de communautés comme Louvain.)

Sélectionnez une requête

26. Analyse des communautés d'acteurs

Exécuter la requête

Communautés d'acteurs détectées :

	communityId	↓ actors	film_count	community_size
1	1	Stephen Lang Jane Levy Dylan Minnette Daniel Z	1	5
2	2	Chris Evans Robert Downey Jr. Scarlett Johansson	1	5
4	4	Charlie Hunnam Robert Pattinson Sienna Miller T	1	5
0	0	Brittany Blanton Ayse Howard Roman Jossart Na	1	5
3	3	Andrew Garfield Adam Driver Liam Neeson Tadar	1	5

Une analyse plus détaillée des communautés d'acteurs a été effectuée. Voir le rapport pour les visualisations complètes.

Questions Transversales : “neo4j\_queries.py” et “mongodb\_queries.py”

27. Quels sont les films qui ont des genres en commun mais qui ont des réalisateurs différents ?

Sélectionnez une requête

27. Films avec genres communs et réalisateurs différents

Exécuter la requête

Films avec des genres communs mais des réalisateurs différents :

	Film 1	Film 2	Genres communs
0	The Departed	Gold	["Thriller","Drama"]
1	The Departed	Passengers	["Drama"]
2	The Departed	Top Dog	["Thriller","Drama"]
3	The Departed	Manchester by the Sea	["Drama"]
4	The Departed	Hidden Figures	["Drama"]
5	The Departed	Hacksaw Ridge	["Drama"]
6	The Departed	Arrival	["Drama"]
7	The Departed	The Dark Knight	["Crime","Drama"]
8	The Departed	Moonlight	["Drama"]
9	The Departed	The Girl on the Train	["Crime","Drama"]
10	The Departed	Interstellar	["Drama"]
11	The Departed	Sully	["Drama"]
12	The Departed	The Autopsy of Jane Doe	["Thriller"]
13	The Departed	Jason Bourne	["Thriller"]
14	The Departed	Fifty Shades of Grey	["Thriller","Drama"]
15	The Departed	Split	["Thriller"]
16	The Departed	Their Finest	["Drama"]
17	The Departed	Dead Awake	["Thriller"]
18	The Departed	Mother's Day	["Drama"]
19	The Departed	Deepwater Horizon	["Thriller","Drama"]

Total trouvé : 100 paires de films (limité à 20 pour l'affichage)

28. Recommander des films aux utilisateurs en fonction des préférences d'un acteur donné.

Sélectionnez une requête

28. Recommandation basée sur les préférences d'un acteur

Nom de l'acteur

Tom Hanks

Exécuter la requête

Films recommandés basés sur les préférences de Tom Hanks :

	title	year	common_actor_count
0	The Dark Knight	2008	1

29. Créer une relation de "concurrence" entre réalisateurs ayant réalisé des films similaires la même année.

Sélectionnez une requête

29. Relation de concurrence entre réalisateurs

Exécuter la requête

0 relations de concurrence créées entre réalisateurs.

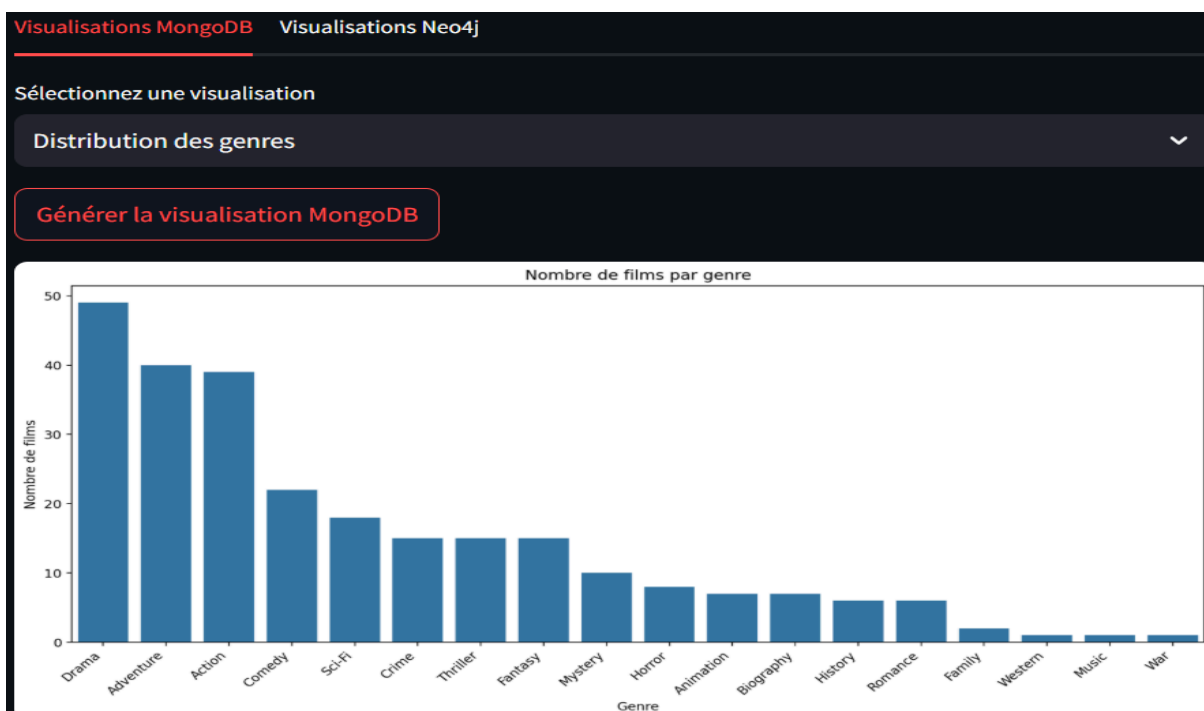
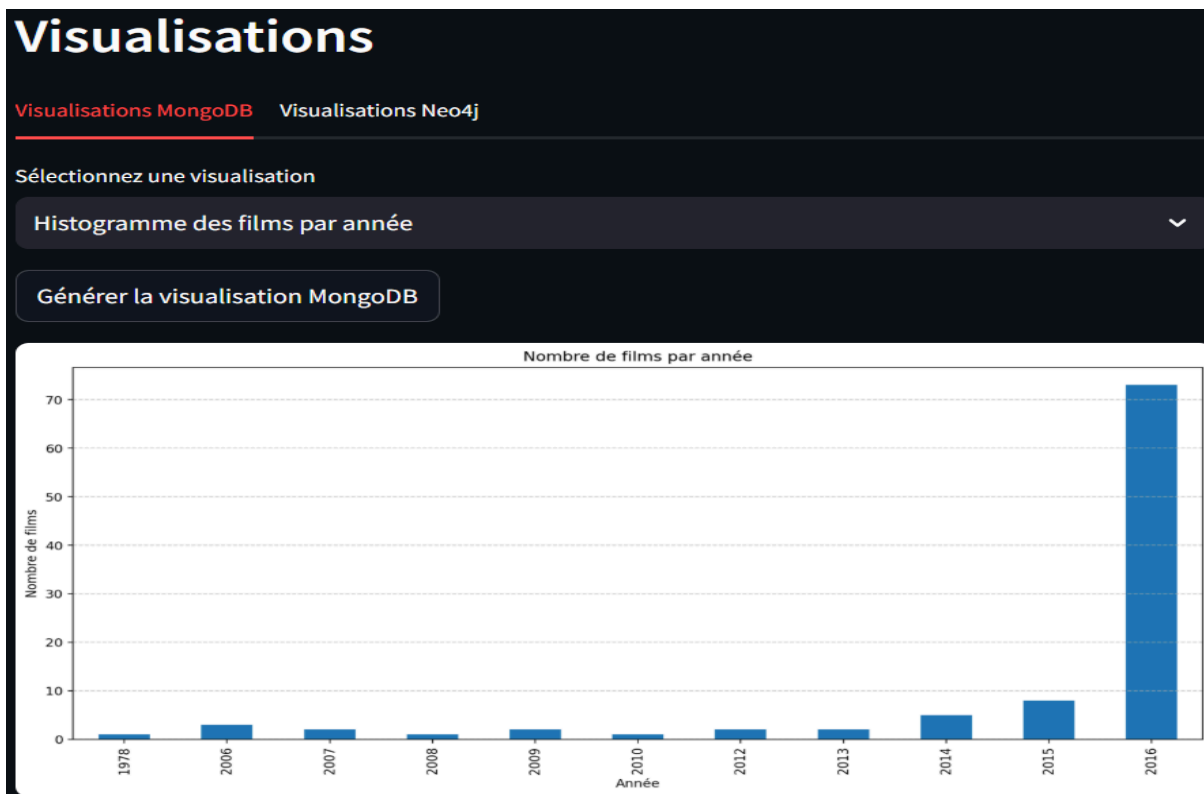
30. Identifier les collaborations les plus fréquentes entre réalisateurs et acteurs, puis analyser si ces collaborations sont associées à un succès commercial ou critique

Sélectionnez une requête

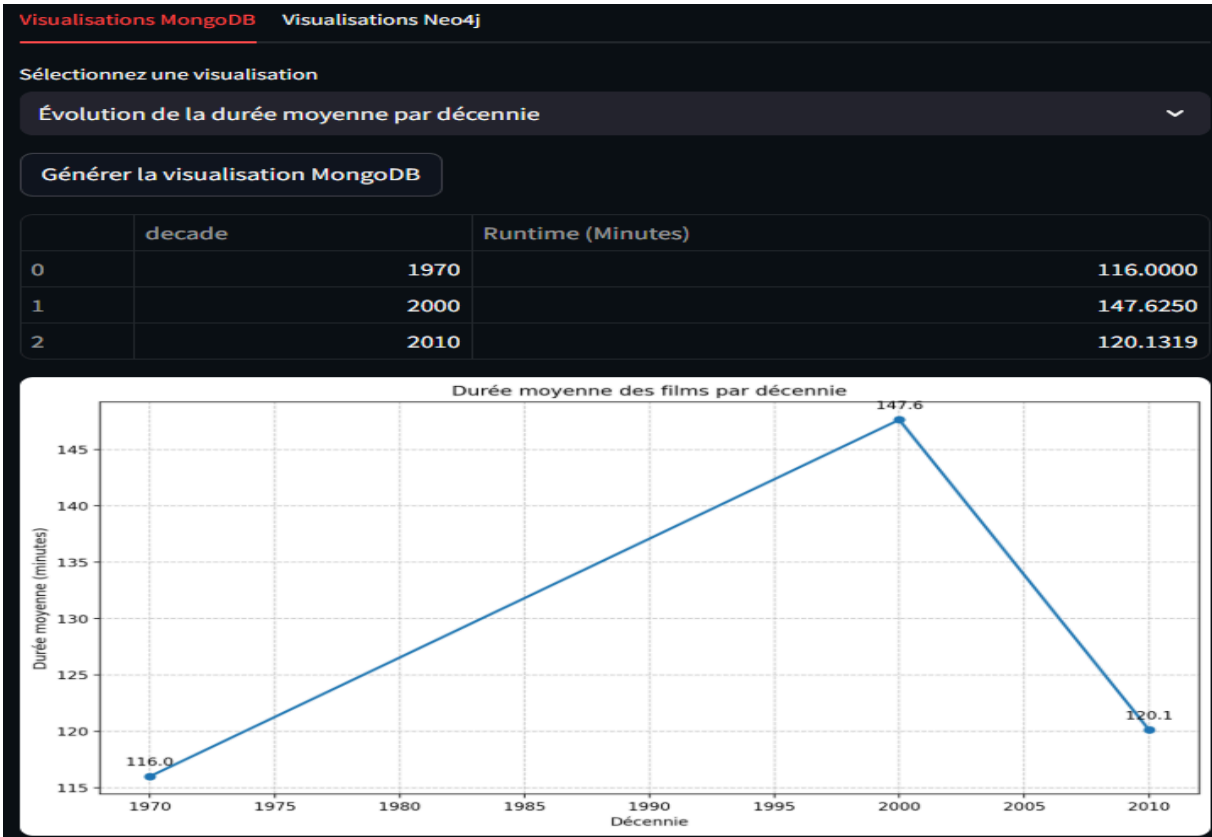
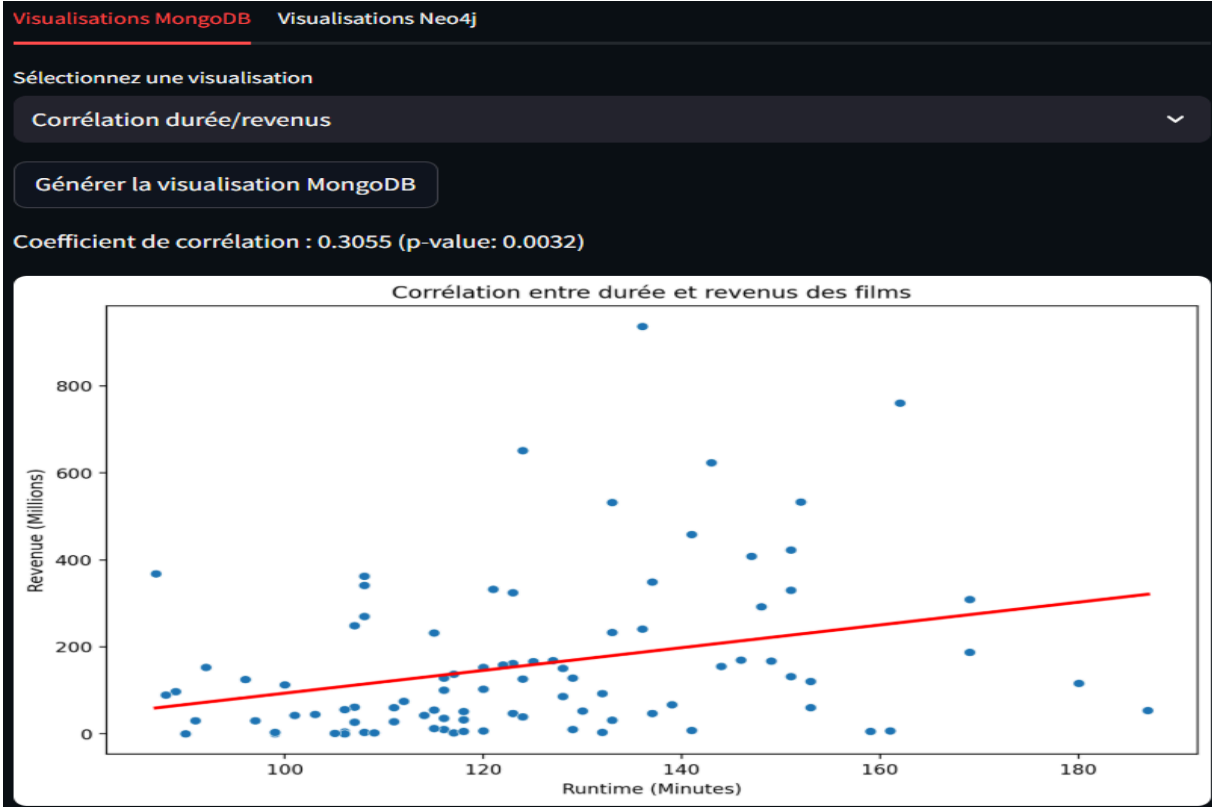
30. Analyse des collaborations réalisateur-acteur

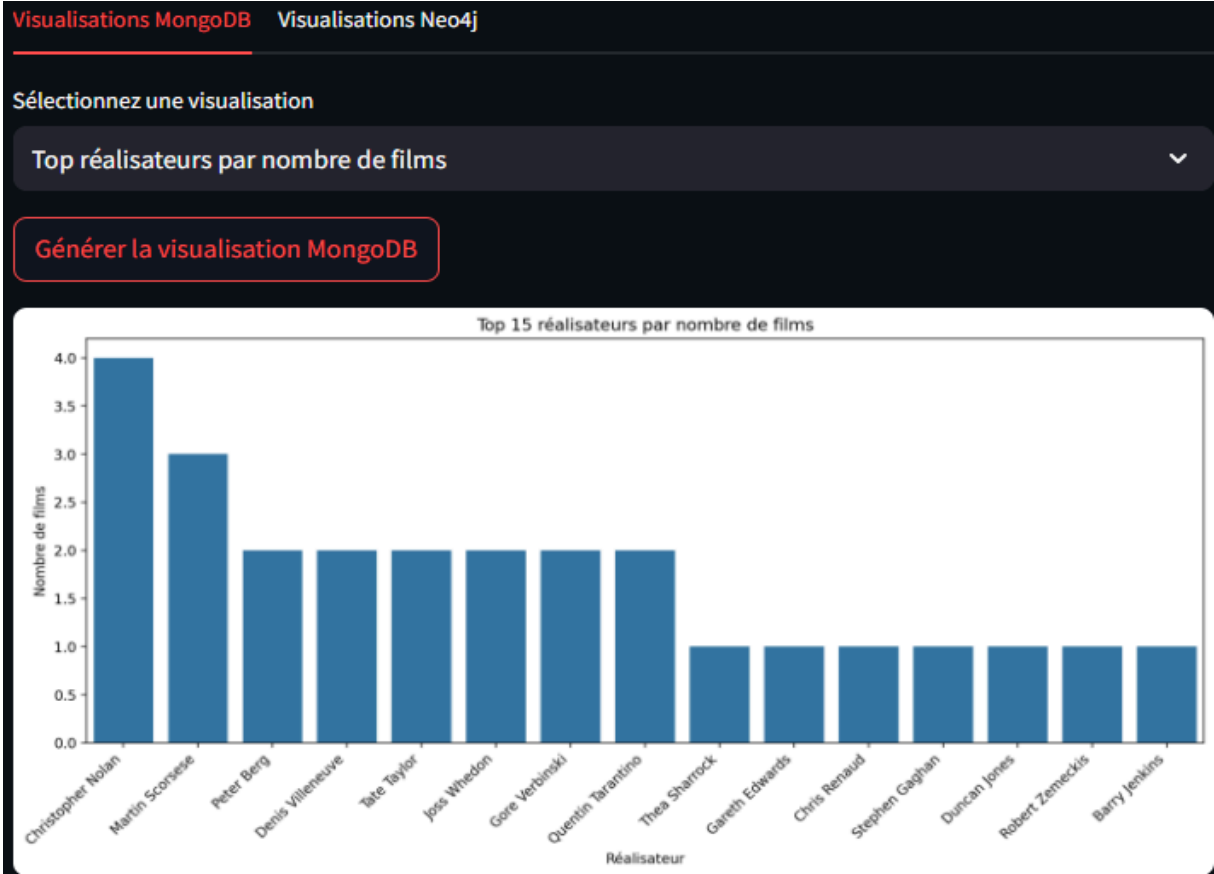
Exécuter la requête

Aucune collaboration fréquente trouvée.









4. Difficultés Rencontrées

Difficulté	Solution
Requêtes agrégées Mongo complexes	Utilisation de la console Compass pour tester avant intégration
Chargement massif vers Neo4j	Préparation préalable des fichiers bien formatés
Visualisation de graphes	Intégration de Neovis.js dans Streamlit avec un wrapper personnalisé

## 5. Solutions Adoptées

- **Modularisation du code** : Séparation en `database.py`, `queries.py`, `visuals.py`.
- **Utilisation de GitHub** : Suivi des versions et travail collaboratif.
- **Docstrings et commentaires** : Chaque fonction est bien documentée.
- **Jeux de données réduits** pour test avant production.
- **Requêtes sauvegardées** dans des fichiers `.cypher` et `.json`.

## 6. Conclusion

Ce projet nous a permis de comprendre les différences entre les bases orientées documents et graphes, d'apprendre à manipuler MongoDB et Neo4j avec Python, et de développer une application d'analyse de données complète avec visualisation.