Project

Choose one of the following topics:

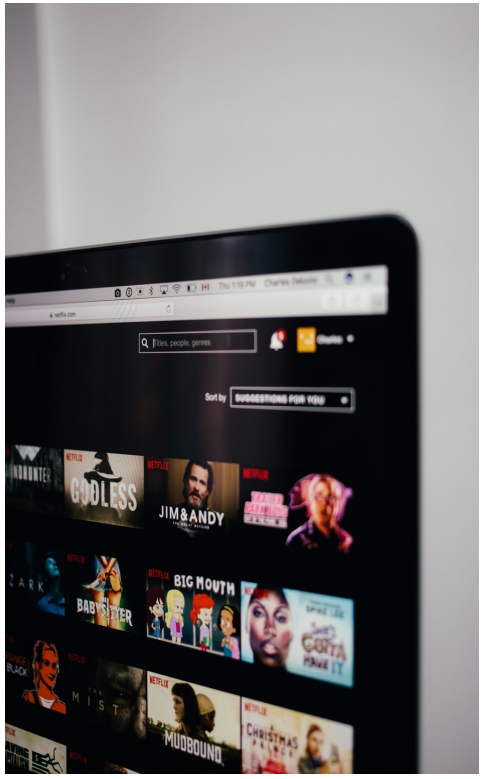
1) Build a movie recommendation system with Neo4j
2) Data analysis with a NoSQL system

Groups of 2 students

Deadline: 4 November

Topic 1

- **Objective**: Build a **movie recommendation** system based on the *movielens* dataset and *Neo4j*
  - Starting from a given user and a given movie, your system calculates the predicted rating that the user would give to this movie
  - Evaluate your system by comparing actual and predicted ratings for a sample of users and movies
- **Deliverable**: Zip file with short report together with code, data and documentation



Topic 1

Cypher Code Example

```
MATCH (u:User) WITH DISTINCT u LIMIT 150
MATCH (u)-[r:RATED]->(m:Movie)
WITH u, collect(r) as rcol
WITH u, head(rcol) as r1
MATCH (u)-[r1]->(m)
WITH u, m, r1.rating AS actual_rating

// Calculate predicted rating of u for m by applying a content-based
//or collaborative filtering technique (do not use actual rating)

WITH u,m, … AS predicted_rating, actual_rating
WITH u,m, (predicted_rating-actual_rating)*(predicted_rating-actual_rating)
as squared_error
WITH count(*) as count, sum(squared_error) as sum
RETURN count, sqrt(tofloat(sum)/count) as RMSE
```

Topic 2

- **Objective**: Perform data analysis using **NoSQL**
  - Identify a research question and relevant datasets (e.g., Kaggle, Open Data, Our World In Data, social networks, Wikipedia)
  - Import the data into a NoSQL database and perform analysis using features of the database
  - Discuss findings in the context of the research question.

Topic 2

- Write a report that describes the research question, methodology, data sources, analysis method, and results
- **Deliverable**: Zip file with the report together with code, data and documentation