



MovieHut.com

the Essential Movie Search Engine

M.EIC PRI 2023/2024

Group G72

Main Motivation



Nowadays, it's **IMPOSSIBLE** to know every movie that's currently available.

As we are living in the age of social media, it's not uncommon, while we scroll through one of the many apps, to come across:



**Quotes from
movies**




**Still images from
a movie scene**



**Full movies
divided in parts**

that we have never heard of/know very little about.



Data Selection

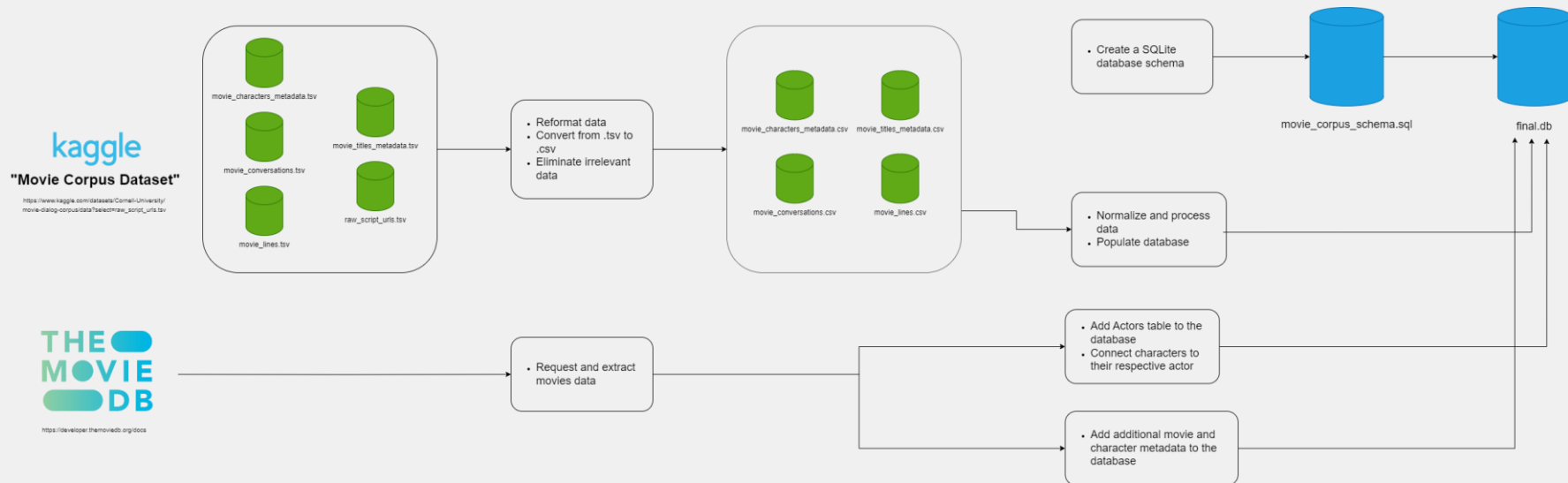


kaggle

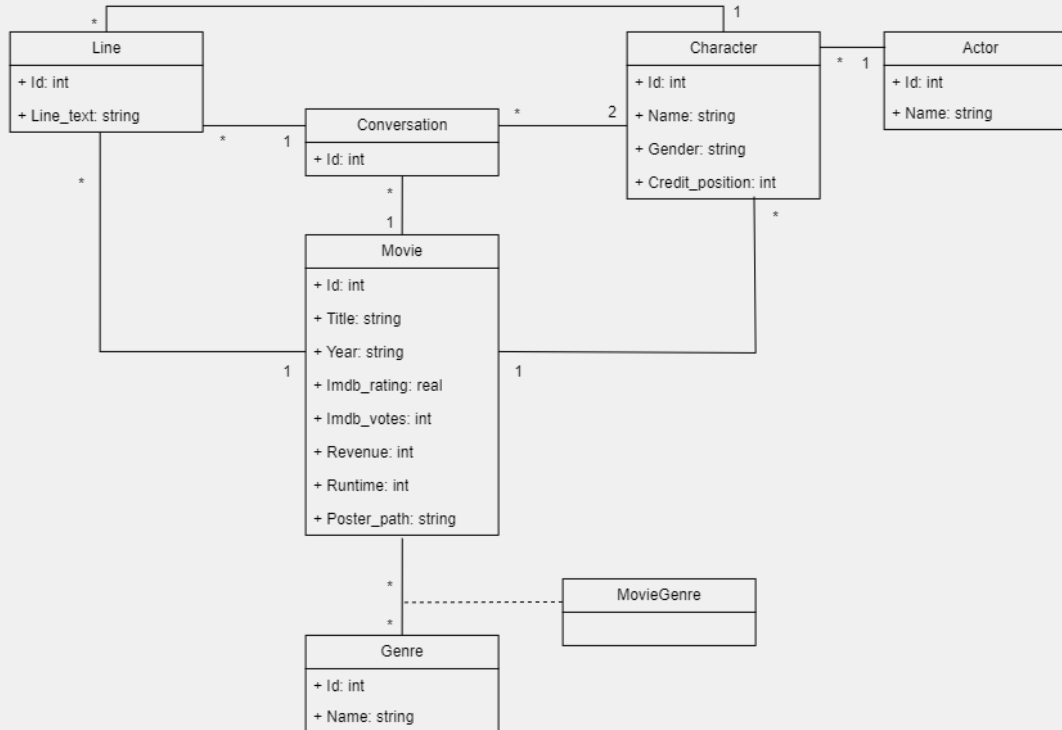
THE
MOVIE
DB



Data Processing and Pipeline

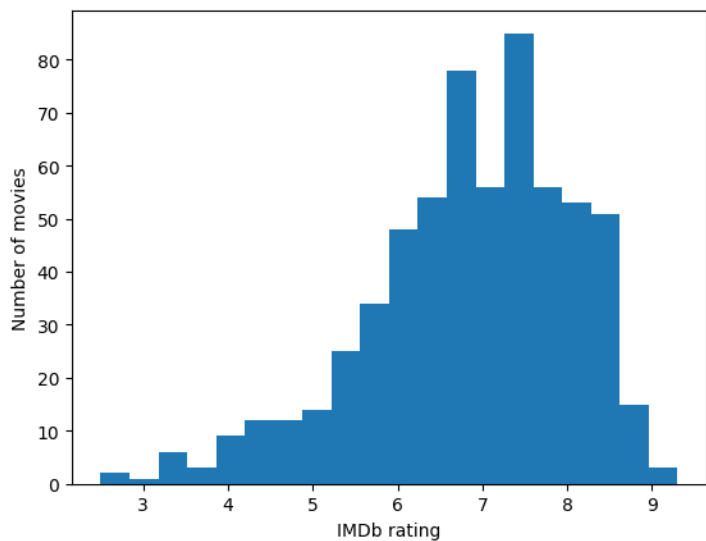


Data Conceptual Model

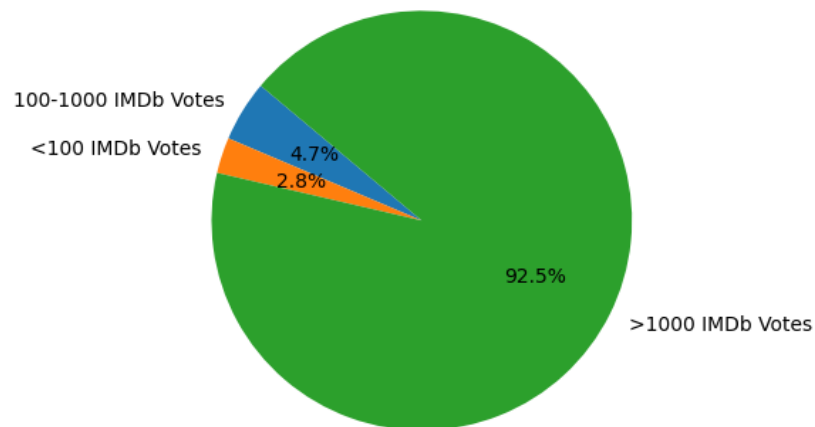


Data Characterization

Distribution of IMDb ratings

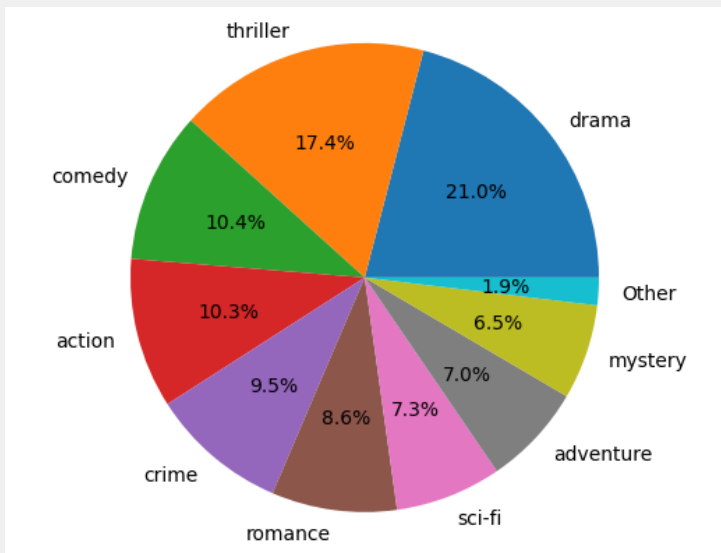


Distribution of the quantity of IMDb votes

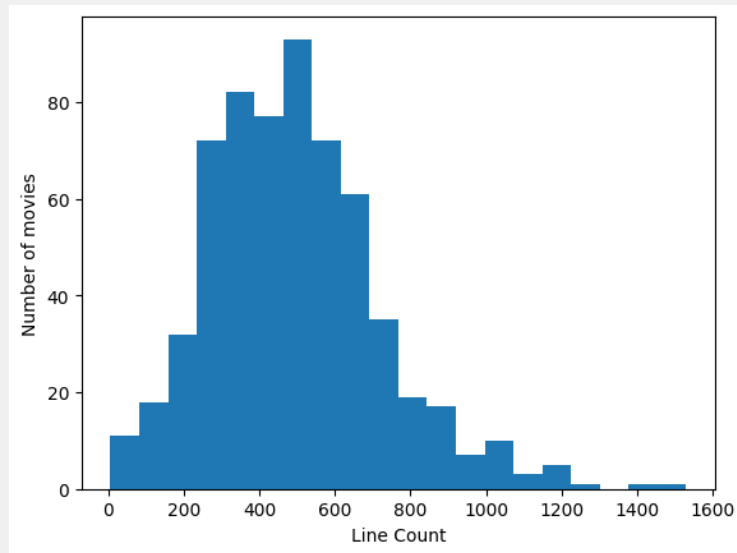


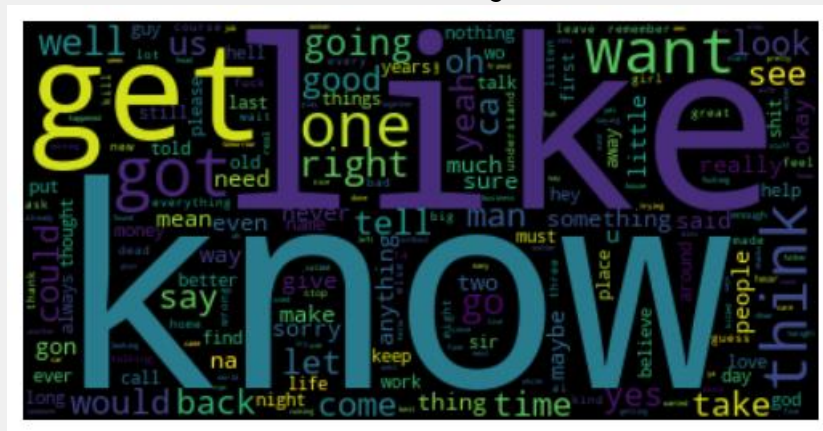
Data Characterization

Distribution of movie genres



Distribution of Lines in the Movies Dataset





Actor Name	Character Count
Nicolas Cage	10
Robert De Niro	8
Johnny Depp	8
Harrison Ford	8
William Shatner	7
Sigourney Weaver	7
Patricia Arquette	7
William H. Macy	6
Philip Baker Hall	6
Dustin Hoffman	6

Actor Name	Character Count
Nicolas Cage	10
Robert De Niro	8
Johnny Depp	8
Harrison Ford	8
William Shatner	7
Sigourney Weaver	7
Patricia Arquette	7
William H. Macy	6
Philip Baker Hall	6
Dustin Hoffman	6

Search Scenarios



“In what movie did the actor X say Y?”



“Who said X?”



“What did X say to Y in movie Z?”



“What character is played by X in movie Y?”



“What movies did actor X play in?”

Document Description



Movie

- Title
- Characters
- Cast
- Genres
- Poster
- Runtime
- Revenue
- IMDb Rating
- IMDb Votes

Actor

- Name
- Movies
- Characters
- Other Related Actors

Conversation

- Lines
 - Movie
 - Characters
 - Cast
 - Previous Dialogue
 - Next Dialogue
-