Az információk a programomhoz használt filmekről a „movies.db” adatbázisban kerültek letárolásra. Az adatbázishoz SQLite formátumot használtam annak egyszerűsége, hordozhatósága miatt, illetve egy nagy előnye, hogy megbízhatóan tud működni gyakorlatilag szinte minden programozási környezetben.

A felhasznált adatok egy kaggle.com-ról származó adatbázisból származnak, melynek forrása a következő: <https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata>

A felhasznált adatbázis a Tmdb-ről származó adatokkal lett feltöltve, így az általam készített adatbázis adatainak az eredeti forrása a Tmdb-n szereplő adatok.

Az adatbázis 1 darab táblát tartalmaz, a „Movies”-t, mely a következő mezőkből áll:

* ID:
  + Típus: INTEGER
  + Leírás: A film egyedi azonosítója, ez a primary key (elsődleges kulcs).
* Title:
  + Típus: TEXT
  + Leírás: A film címe angolul, amely az esetek nagy részében megegyezik a filmek eredeti címével.
* Genre:
  + Típus: TEXT
  + Leírás: A film műfaja(i) felsorolás szerűen, vesszővel elválasztva. Egyes filmeknél csak 1 darab van belőle, más, több műfajba is beleillő filmeknél több is lehet.
* Released:
  + Típus: INTEGER
  + Leírás: A film megjelenésének éve.
* Runtime:
  + Típus: INTEGER
  + Leírás: A film játékideje percben megadva.
* Gender\_of\_the\_protagonist:
  + Típus: INTEGER
  + Leírás: A főszereplő nevét jelzi. Az értéke 1, ha a főszereplő nő, ha férfi akkor pedig 2 az értéke.
* Main\_actor:
  + Típus: TEXT
  + Leírás: A főszereplő színész neve.
* Keywords:
  + Típus: TEXT
  + Leírás: Az adott filmhez kapcsolódó, a Tmdb-ről származó kulcsszavak vesszővel ellátott felsorolása, melyek használatával könnyebben találhatunk hasonlóságokat a filmek között.
* Director:
  + Típus: TEXT
  + Leírás: Az adott film rendezőjének/rendezőinek neve.
* Language:
  + Típus: TEXT
  + Leírás: A film eredeti nyelve, illetve a benne elhangzó nyelvek felsorolása
* Pruduction\_countries:
  + Típus: TEXT
  + Leírás: A készítő országok neveinek felsorolása.
* Tmdb\_score:
  + Típus: REAL
  + Leírás: Az adott film Tmdb-n szereplő értékelése, amely a felhasználói értékelések átlaga. Az értékelés egy 0-tól 10-ig terjedő skálán történik, így a pontszám is a két érték közé eshet csak.
* Number\_of\_ratings:
  + Típus: INTEGER
  + Leírás: Megmutatja, hogy a Tmdb oldalán eddig hányan értékelték az adott filmet, tehát a „Tmdb\_score” nevű mező hány értékelésnek az átlaga.
* Popularity:
  + Típus: REAL
  + Leírás: Az adott film Tmdb szerinti népszerűségi indexe, melyet a felhasználók tevékenységeiből számolnak.
* Budget:
  + Típus: INTEGER
  + Leírás: Az adott film költségvetésének összege amerikai dollárban megadva.
* Revenue:
  + Típus: INTEGER
  + Leírás: Az adott film bevételének összege amerikai dollárban megadva.