

Team 4 - Project 3

Christine Jauregui

Michael Leonard

Jacqueline Meyer

Eric Zhou

ETL to Create







Movie

Recommendation

Engine



Data Source

https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata/data?select=tmdb_5000_...      New Chrome available 

 Search

Sign In

Register

TMDB 5000 Movie Dataset

Metadata on ~5,000 movies from TMDb

Data Card

Code (2038)

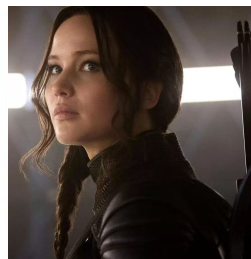
Discussion (64)

Suggestions (0)

About Dataset

Background

What can we say about the success of a movie before it is released? Are there certain companies (Pixar?) that have found a consistent formula? Given that major films costing over \$100 million to produce can still flop, this question is more important than ever to the industry. Film aficionados

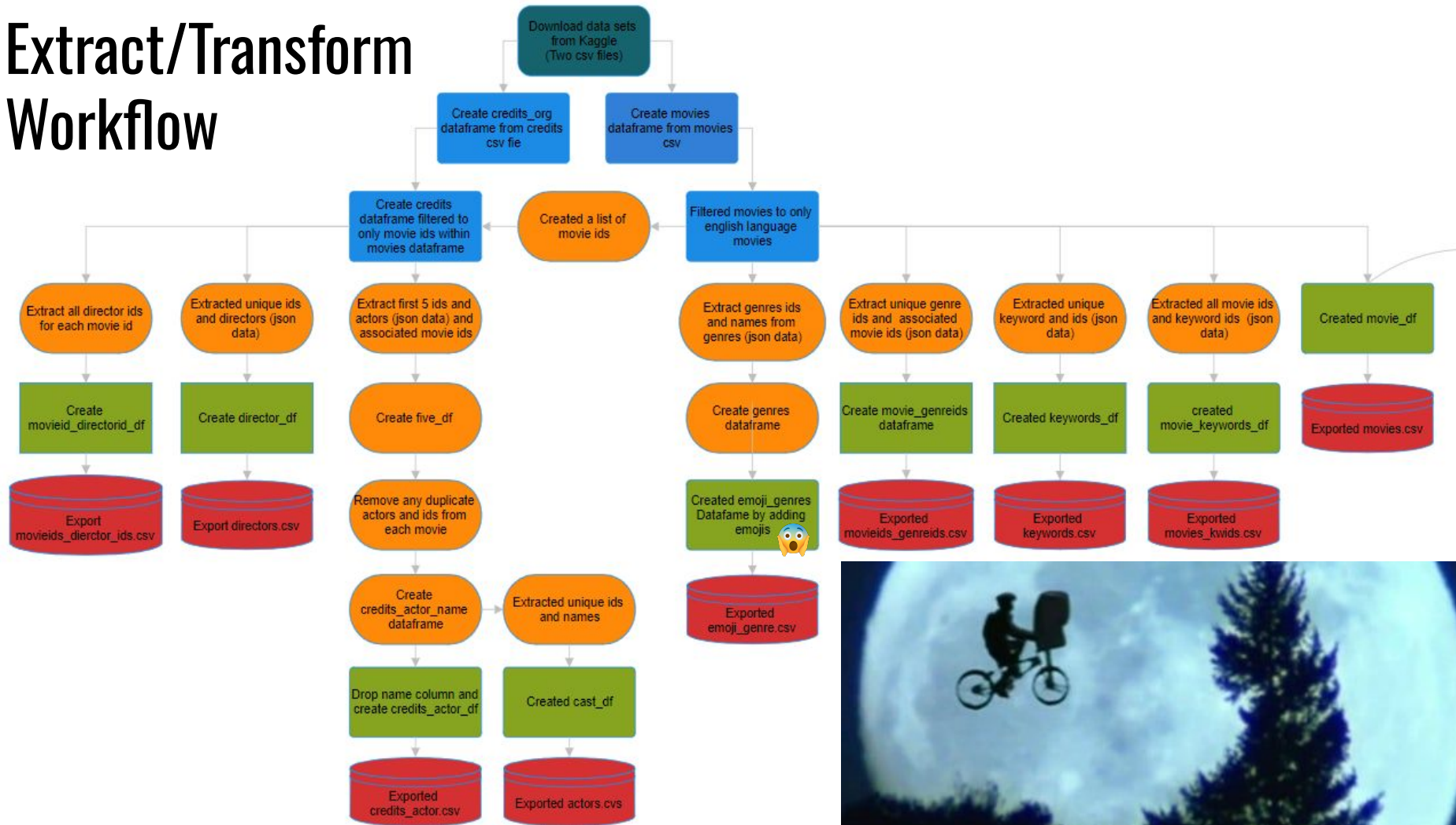


License

Other (specified in description)

Expected update frequency

Extract/Transform Workflow



Example of Json Fields

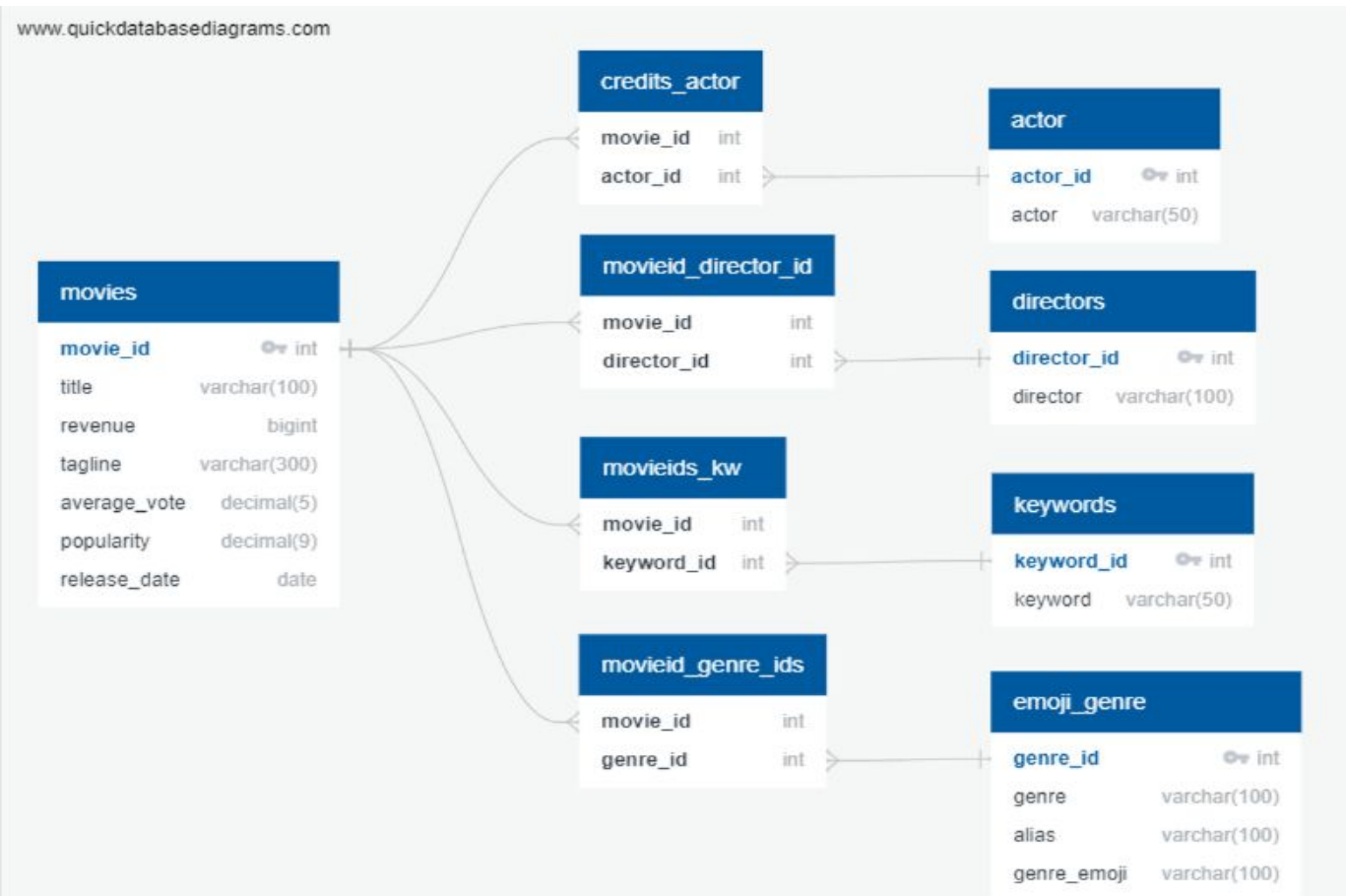


movie_id	title	keywords	genre
19995	Avatar	[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "society"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"id": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "alien"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "na...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]
285	Pirates of the Caribbean: At World's End	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "drug abuse"}, {"id": 911, "name": "exotic island"}, {"id": 1319, "name": "east india trading company"}, {"id": 2038, "name": "love of one's life"}, {"id": 2052, "name": "traitor"}, {"id": 2580, "name": "shipwreck"}, {"id": 2660, "name": "strong woman"}, {"id": 3799, "name": "ship"}, {"id": 5740, "name": "alliance"}, {"id": 5941, "name": "caly...	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 28, "name": "Action"}]

movie_id	title	cast	crew
0	19995	Avatar	
		[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington", "order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana", "order": 1}, {"cast_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52fe48009251416c750aca39", "gender": ...	[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": "Editor", "name": "Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "job": "Sound Designer", "na...
1	285	Pirates of the Caribbean: At World's End	
		[{"cast_id": 4, "character": "Captain Jack Sparrow", "credit_id": "52fe4232c3a36847f800b50d", "gender": 2, "id": 85, "name": "Johnny Depp", "order": 0}, {"cast_id": 5, "character": "Will Turner", "credit_id": "52fe4232c3a36847f800b511", "gender": 2, "id": 114, "name": "Orlando Bloom", "order": 1}, {"cast_id": 6, "character": "Elizabeth Swann", "credit_id": "52fe4232c3a36847f800b515", "gender": ...	[{"credit_id": "52fe4232c3a36847f800b579", "department": "Camera", "gender": 2, "id": 120, "job": "Director of Photography", "name": "Dariusz Wolski"}, {"credit_id": "52fe4232c3a36847f800b4fd", "department": "Directing", "gender": 2, "id": 1704, "job": "Director", "name": "Gore Verbinski"}, {"credit_id": "52fe4232c3a36847f800b54f", "department": "Production", "gender": 2, "id": 770, "job": "Pr...

“When structuring a Database, follow the Data” ~Randy~

“And use the Force” ~Michael~




```

"movies"
--
"movie_id" PK int
"title" varchar(100)
"revenue" bigint
>tagline" varchar(300)
"average_vote" decimal(5)
"popularity" decimal(9)
"release_date" date

"actor"
--
"actor_id" PK int
"actor" varchar(50)

"credits_actor"
--
"movie_id" int FK >- "movies"."movie_id"
"actor_id" int FK >- "actor"."actor_id"

"directors"
--
"director_id" PK int FK
"director" varchar(100)

"movieid_director_id"
--
"movie_id" int FK >- "movies"."movie_id"
"director_id" int FK >- directors.director_id

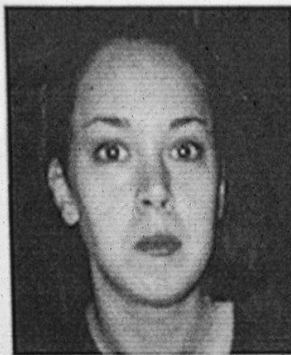
"emoji_genre"
--
"genre_id" PK int
"genre" varchar(100)
"alias" varchar(100)
"genre_emoji" varchar(100)

"movieid_genre_ids"
--
"movie_id" int FK >- "movies"."movie_id"
"genre_id" int FK >- emoji_genre.genre_id

"keywords"

```

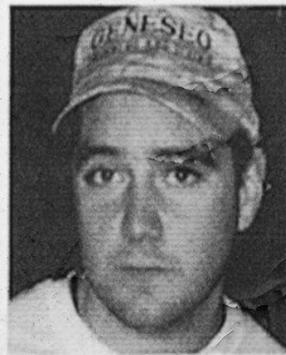
MISSING



Heather Donahue
Age: 22 Height: 5'6" Weight: 127lb
Eyes: hazel Hair: brown



Joshua Leonard
Age: 23 Height: 5'10" Weight: 162lb
Eyes: blue Hair: blonde



Michael Williams
Age: 24 Height: 5'8" Weight: 162lb
Eyes: brown Hair: brown

Last seen camping in the Black Hills Forrest area, near Black Hills, South Dakota.

**PLEASE CALL FREDERICK COUNTY SHERIFF'S OFFICE WITH
ANY INFORMATION YOU MAY HAVE!**

(301) 555-1870

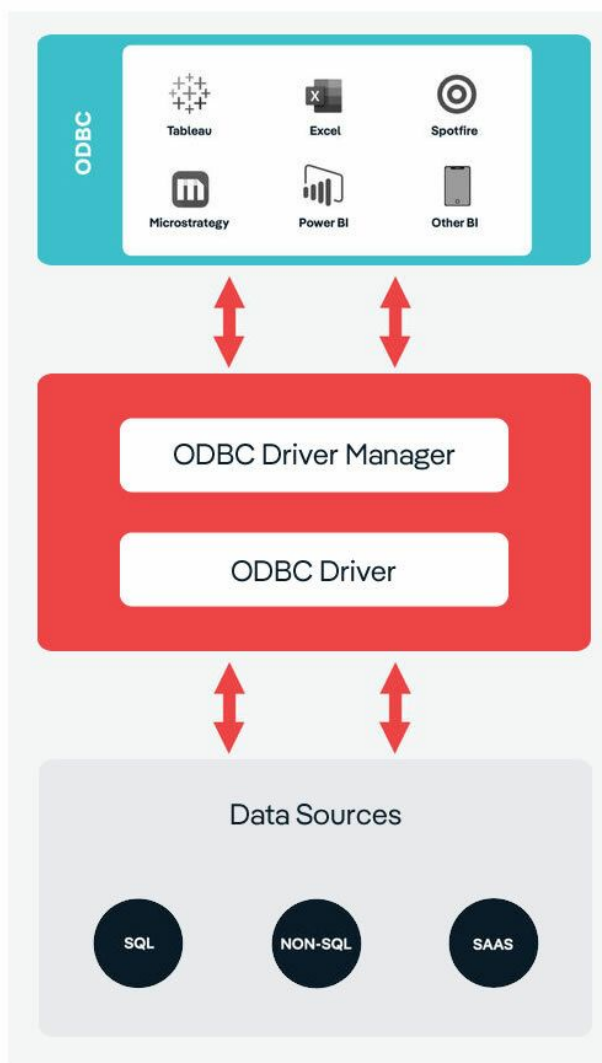
Michael's
Little
Cousin Josh

Open Database Connectivity (ODBC)

ODBC is a standard API for connecting to a database

Query PostgreSQL in Python with:

- PostgreSQL ODBC Driver
- PyODBC library

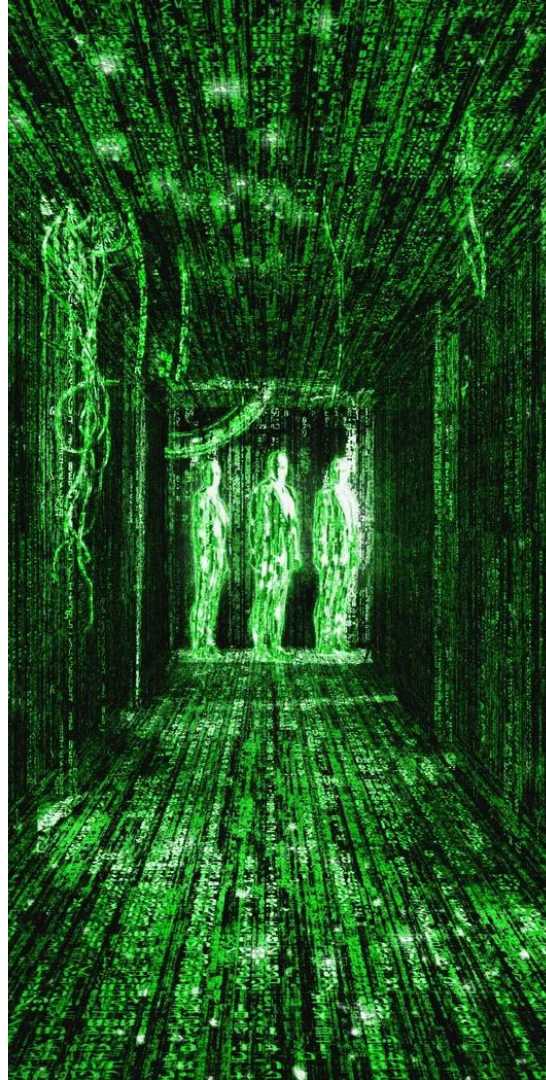


PyODBC: Connect to SQL Server

```
# create dictionary containing connection information to local postgresql db
config = dict(
    server = 'localhost',
    port = 5432,
    database = 'movie_db',
    username = 'postgres',
    password = 'postgres')

# create connection string to be fed into connect function
conn_str = (
    'SERVER={server},{port};' +
    'DATABASE={database};' +
    'UID={username};' +
    'PWD={password}')

# connect to our movie database using configuration string
conn = pyodbc.connect(
    r'DRIVER={PostgreSQL Unicode(x64)};' +
    conn_str.format(**config))
```



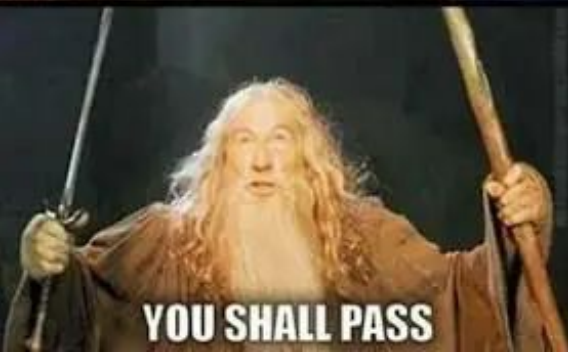
PyODBC: Run a SQL Query in Python

Example: Search for top rated Christmas movies

```
# create a cursor and run query through the cursor
christmas_movies = conn.cursor()
christmas_movies.execute("SELECT DISTINCT m.title, m.average_vote, m.popularity, m.tagline FROM \
movies m \
JOIN movieids_kw mkw \
ON m.movie_id = mkw.movie_id \
JOIN keywords kw \
ON mkw.keyword_id = kw.keyword_id \
WHERE kw.keyword LIKE '%christmas%' \
ORDER BY m.average_vote DESC, m.popularity DESC")
```

	Title	Average Vote	Popularity	Tagline
0	The Apartment	8.1	22.889294	Movie-wise, there has never been anything like...
1	It's a Wonderful Life	8.0	42.868495	It's a wonderful laugh! It's a wonderful love!
2	Spotlight	7.8	41.503588	Break the story. Break the silence.
3	Harry Potter and the Philosopher's Stone	7.5	109.984351	Let the Magic Begin.
4	Die Hard	7.5	66.169628	40 Stories. Twelve Terrorists. One Cop.





How Did We Do?