CRAZY THURSDAY

ZHENG SHUWAN ZHANG SHUYI CHENG YULU LIN YIMENG

2023

November 19th

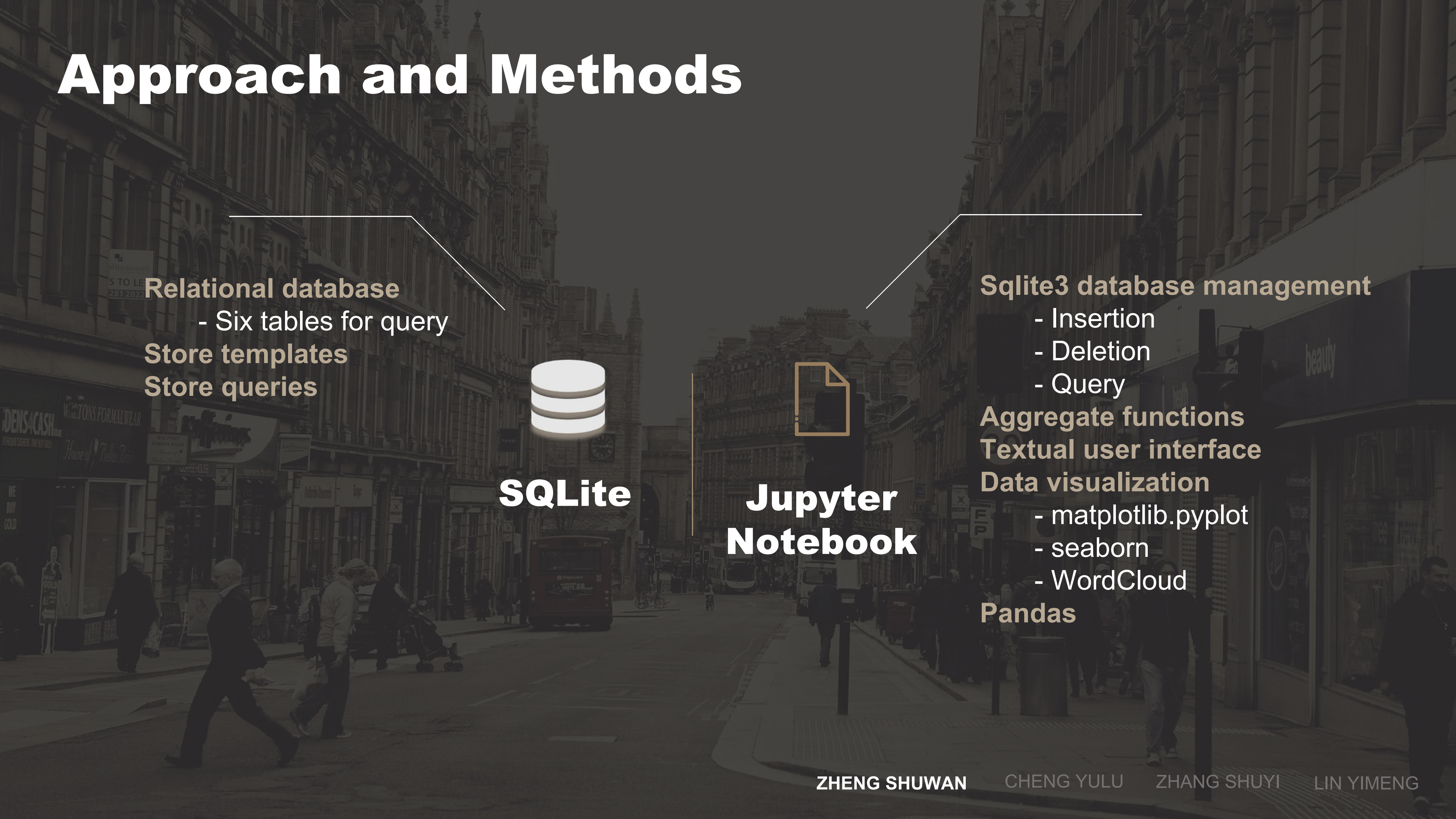
Movie Recommendation & News Integration System

Covered issue

- ***Are global film industry is declining year by year?***
- ***How is the performance of film markets in different countries?***
- ***Which movie genres is currently popular?***
- ***How to recommend movies according to individual preferences?***

Our project, Movie Recommendations and News Integration, is designed to solve the problems above. This project aims to movie recommendations based on users' input preferences. Additionally, this platform integrates historical data to offer insights for movie enthusiasts, researchers, movie creators, and industry investors, providing them with industries relevant information and recommendations.

ZHENG SHUWAN CHENG YULU ZHANG SHUYI LIN YIMENG



Approach and Methods

Relational database
- Six tables for query

Store templates

Store queries

SQLite

Jupyter Notebook

Sqlite3 database management
- Insertion
- Deletion
- Query

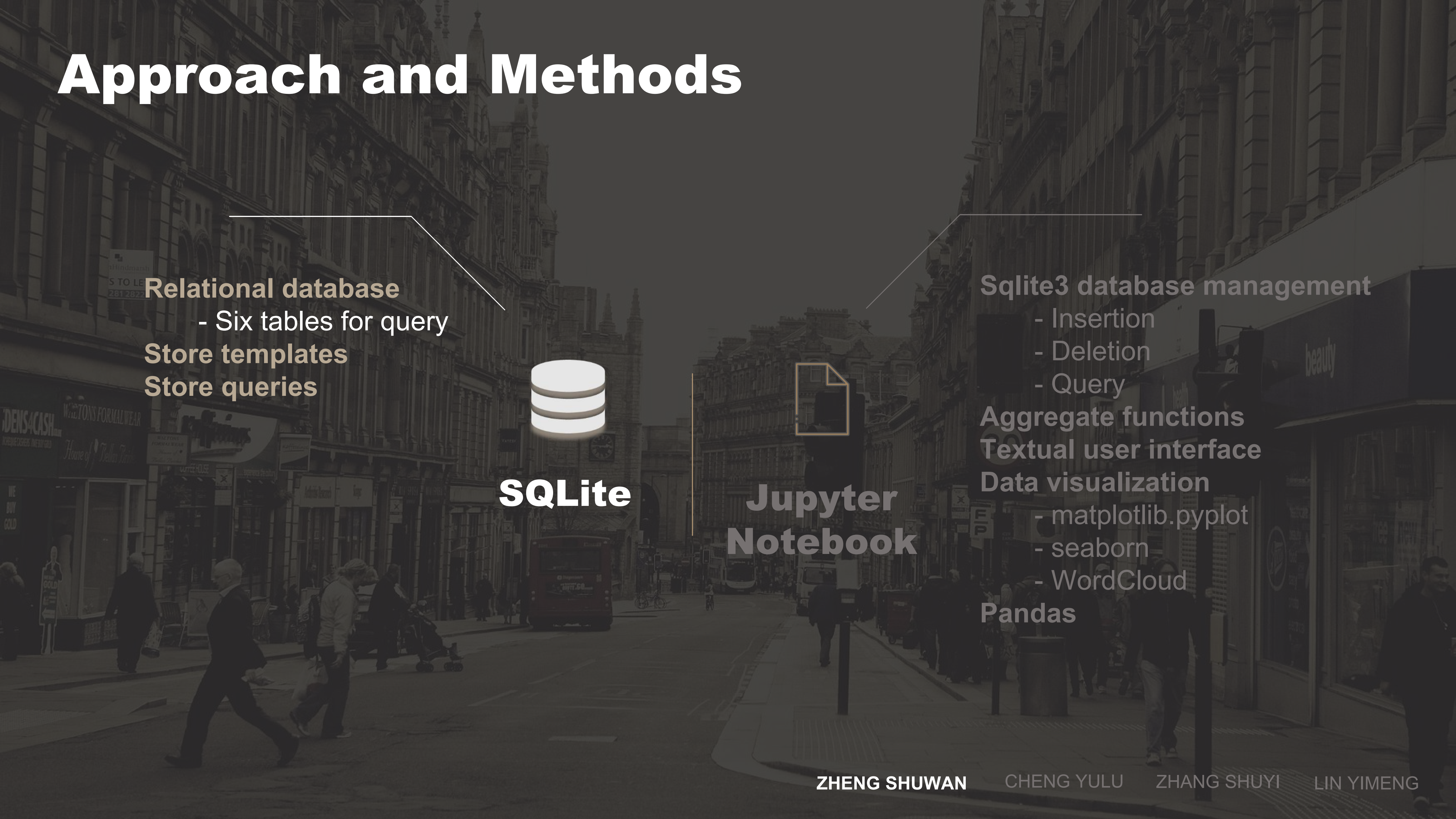
Aggregate functions

Textual user interface

Data visualization
- matplotlib.pyplot
- seaborn
- WordCloud

Pandas

Approach and Methods



Relational database

- Six tables for query

Store templates

Store queries

SQLite

Jupyter Notebook

Sqlite3 database management

- Insertion
- Deletion
- Query

Aggregate functions

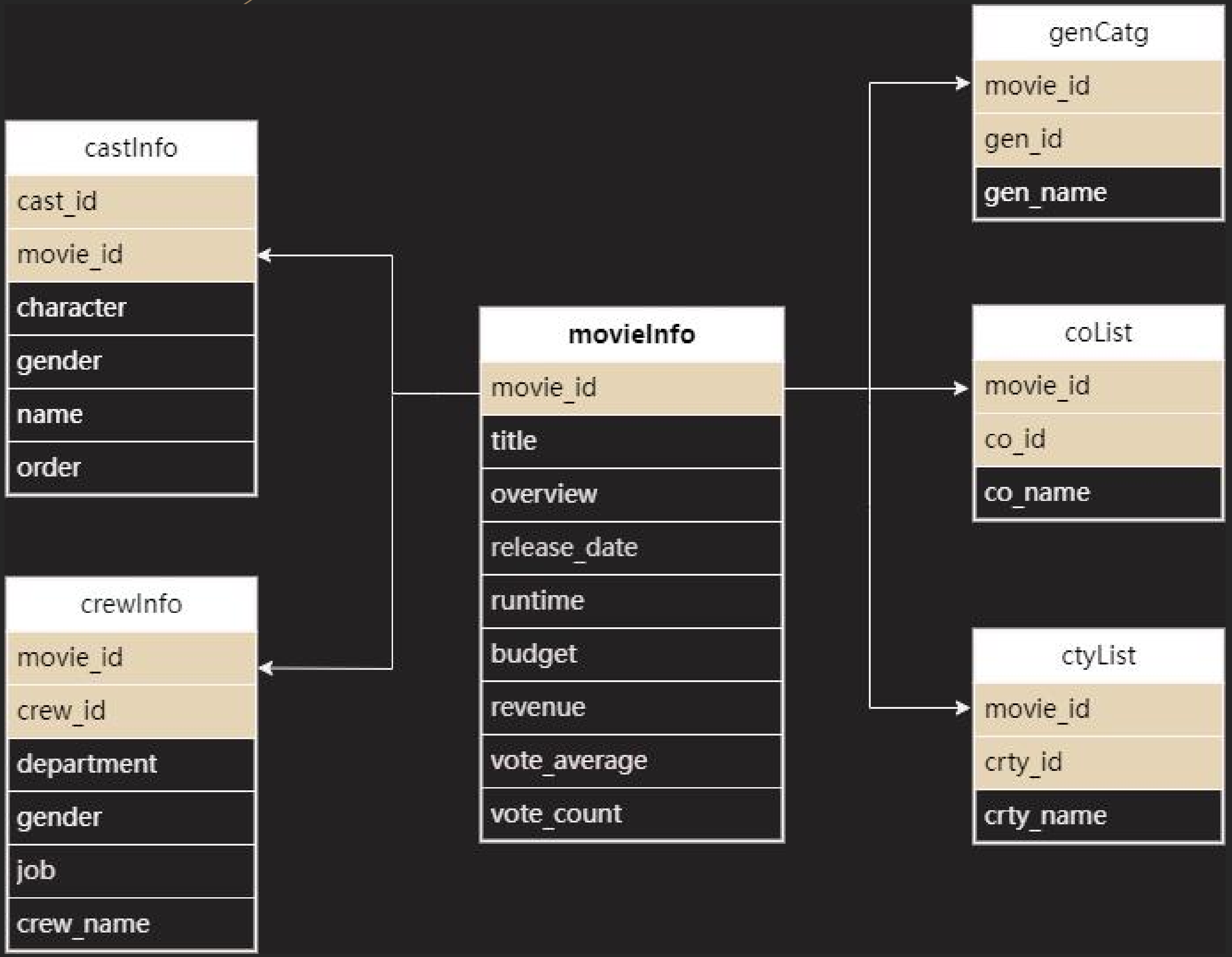
Textual user interface

Data visualization

- matplotlib.pyplot
- seaborn
- WordCloud

Pandas

Schema Design



castInfo

- cast_id
- movie_id
- character
- gender
- name
- order

crewInfo

- movie_id
- crew_id
- department
- gender
- job
- crew_name

movieInfo

- movie_id
- title
- overview
- release_date
- runtime
- budget
- revenue
- vote_average
- vote_count

genCatg

- movie_id
- gen_id
- gen_name

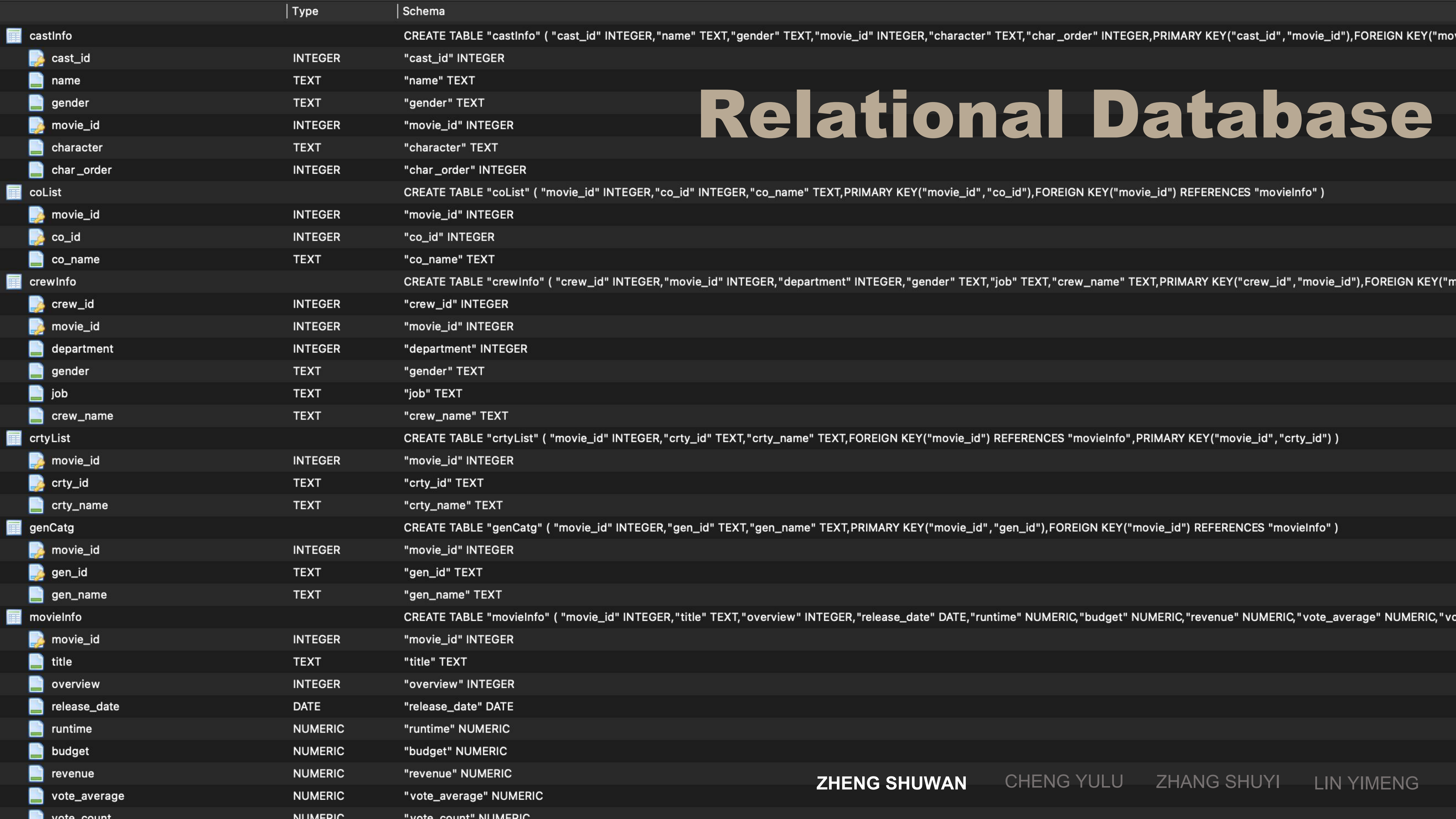
coList

- movie_id
- co_id
- co_name

ctyList

- movie_id
- crty_id
- crty_name

Relational Database



	Type	Schema
castInfo		CREATE TABLE "castInfo" ("cast_id" INTEGER,"name" TEXT,"gender" TEXT,"movie_id" INTEGER,"character" TEXT,"char_order" INTEGER,PRIMARY KEY("cast_id","movie_id"),FOREIGN KEY("mo
cast_id	INTEGER	"cast_id" INTEGER
name	TEXT	"name" TEXT
gender	TEXT	"gender" TEXT
movie_id	INTEGER	"movie_id" INTEGER
character	TEXT	"character" TEXT
char_order	INTEGER	"char_order" INTEGER
coList		CREATE TABLE "coList" ("movie_id" INTEGER,"co_id" INTEGER,"co_name" TEXT,PRIMARY KEY("movie_id","co_id"),FOREIGN KEY("movie_id") REFERENCES "movieInfo")
movie_id	INTEGER	"movie_id" INTEGER
co_id	INTEGER	"co_id" INTEGER
co_name	TEXT	"co_name" TEXT
crewInfo		CREATE TABLE "crewInfo" ("crew_id" INTEGER,"movie_id" INTEGER,"department" INTEGER,"gender" TEXT,"job" TEXT,"crew_name" TEXT,PRIMARY KEY("crew_id","movie_id"),FOREIGN KEY("m
crew_id	INTEGER	"crew_id" INTEGER
movie_id	INTEGER	"movie_id" INTEGER
department	INTEGER	"department" INTEGER
gender	TEXT	"gender" TEXT
job	TEXT	"job" TEXT
crew_name	TEXT	"crew_name" TEXT
crtyList		CREATE TABLE "crtyList" ("movie_id" INTEGER,"crty_id" TEXT,"crty_name" TEXT,FOREIGN KEY("movie_id") REFERENCES "movieInfo",PRIMARY KEY("movie_id","crty_id"))
movie_id	INTEGER	"movie_id" INTEGER
crty_id	TEXT	"crty_id" TEXT
crty_name	TEXT	"crty_name" TEXT
genCatg		CREATE TABLE "genCatg" ("movie_id" INTEGER,"gen_id" TEXT,"gen_name" TEXT,PRIMARY KEY("movie_id","gen_id"),FOREIGN KEY("movie_id") REFERENCES "movieInfo")
movie_id	INTEGER	"movie_id" INTEGER
gen_id	TEXT	"gen_id" TEXT
gen_name	TEXT	"gen_name" TEXT
movieInfo		CREATE TABLE "movieInfo" ("movie_id" INTEGER,"title" TEXT,"overview" INTEGER,"release_date" DATE,"runtime" NUMERIC,"budget" NUMERIC,"revenue" NUMERIC,"vote_average" NUMERIC,"vo
movie_id	INTEGER	"movie_id" INTEGER
title	TEXT	"title" TEXT
overview	INTEGER	"overview" INTEGER
release_date	DATE	"release_date" DATE
runtime	NUMERIC	"runtime" NUMERIC
budget	NUMERIC	"budget" NUMERIC
revenue	NUMERIC	"revenue" NUMERIC
vote_average	NUMERIC	"vote_average" NUMERIC
vote_count	NUMERIC	"vote_count" NUMERIC

ZHENG SHUWAN CHENG YULU ZHANG SHUYI LIN YIMENG

Stored Queries & Templates

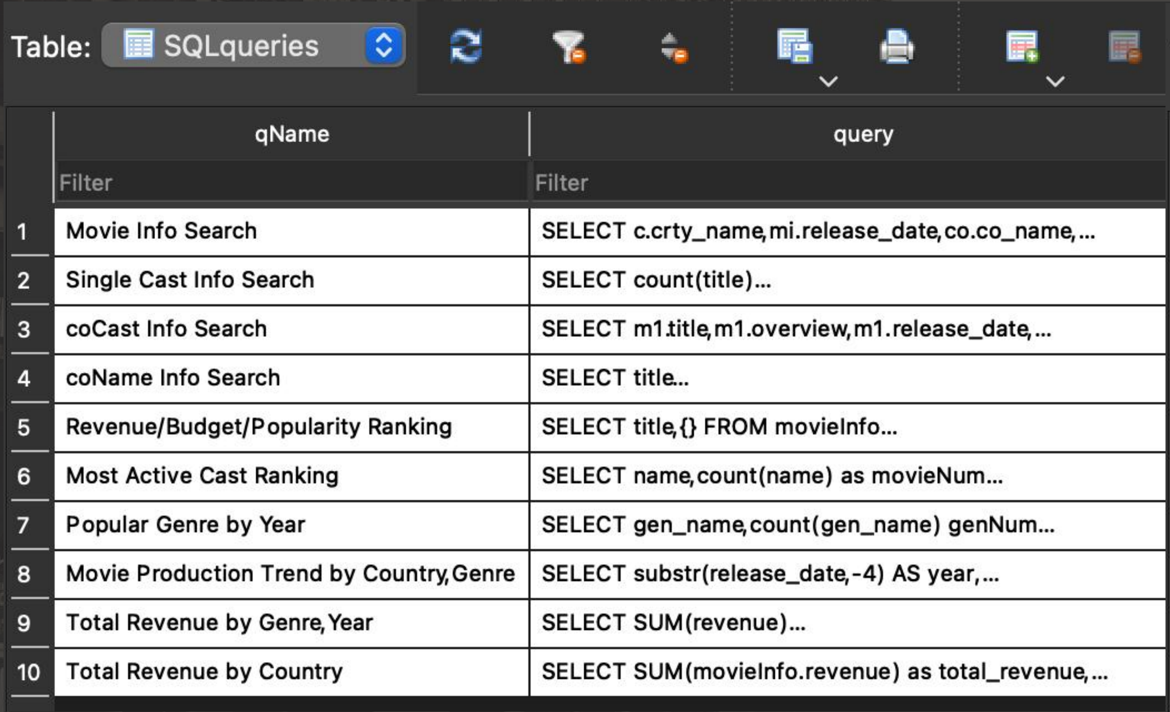


Table: SQLqueries

	qName	query
	Filter	Filter
1	Movie Info Search	SELECT c.crty_name,mi.release_date,co.co_name,...
2	Single Cast Info Search	SELECT count(title)...
3	coCast Info Search	SELECT m1.title,m1.overview,m1.release_date,...
4	coName Info Search	SELECT title...
5	Revenue/Budget/Popularity Ranking	SELECT title,{} FROM movieInfo...
6	Most Active Cast Ranking	SELECT name,count(name) as movieNum...
7	Popular Genre by Year	SELECT gen_name,count(gen_name) genNum...
8	Movie Production Trend by Country,Genre	SELECT substr(release_date,-4) AS year,...
9	Total Revenue by Genre,Year	SELECT SUM(revenue)...
10	Total Revenue by Country	SELECT SUM(movieInfo.revenue) as total_revenue,...

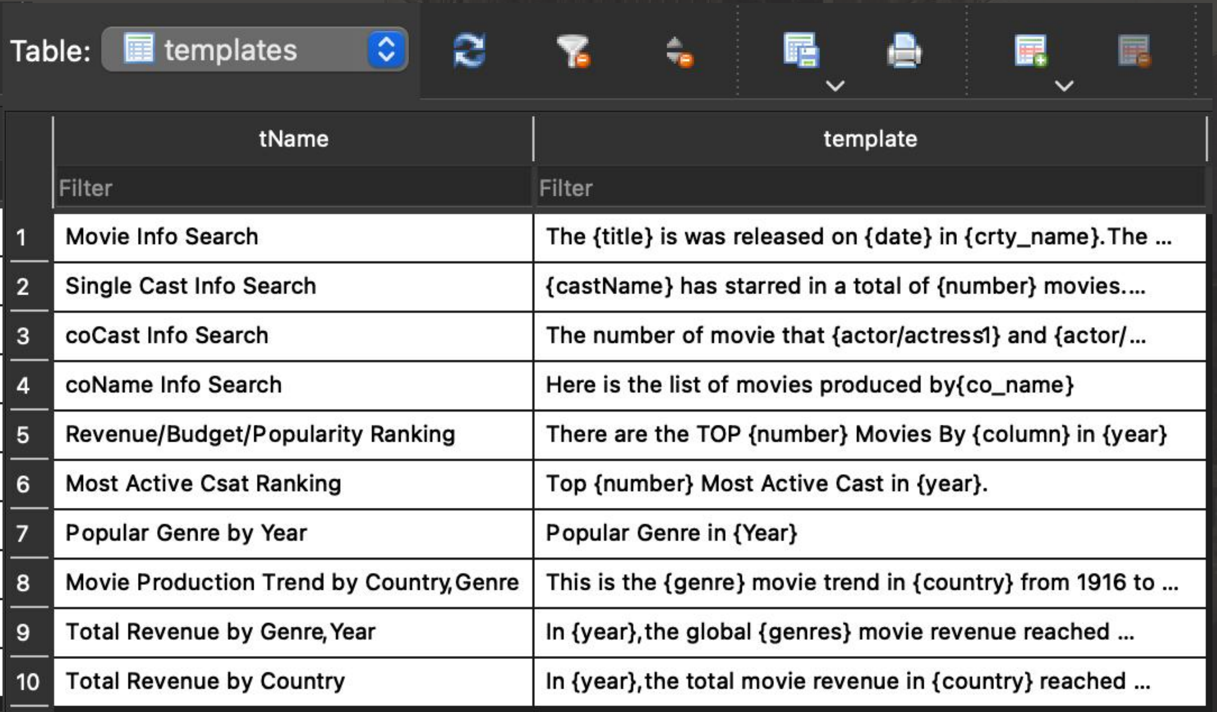
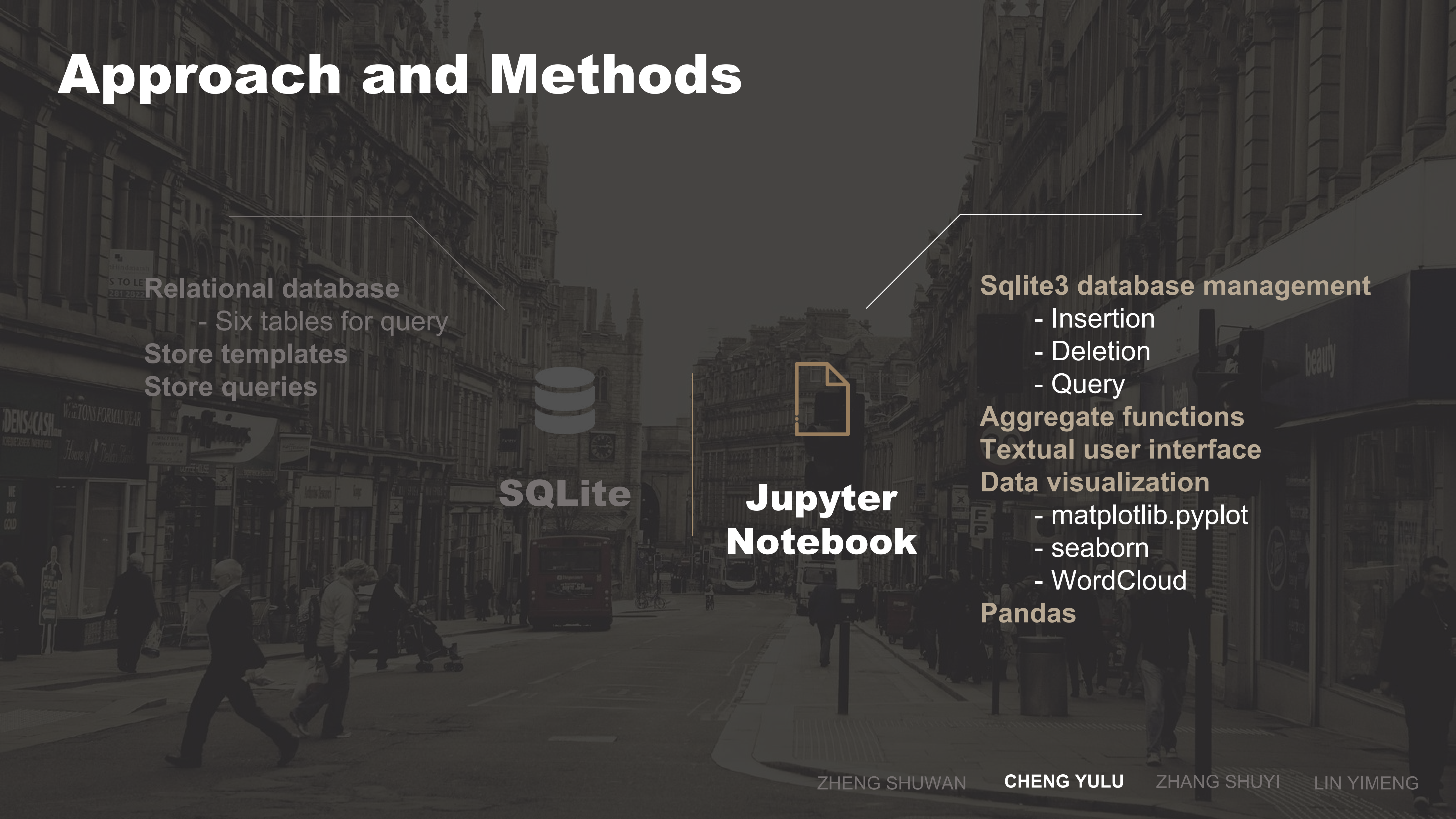


Table: templates

	tName	template
	Filter	Filter
1	Movie Info Search	The {title} is was released on {date} in {crty_name}.The ...
2	Single Cast Info Search	{castName} has starred in a total of {number} movies....
3	coCast Info Search	The number of movie that {actor/actress1} and {actor/...
4	coName Info Search	Here is the list of movies produced by{co_name}
5	Revenue/Budget/Popularity Ranking	There are the TOP {number} Movies By {column} in {year}
6	Most Active Csat Ranking	Top {number} Most Active Cast in {year}.
7	Popular Genre by Year	Popular Genre in {Year}
8	Movie Production Trend by Country,Genre	This is the {genre} movie trend in {country} from 1916 to ...
9	Total Revenue by Genre,Year	In {year},the global {genres} movie revenue reached ...
10	Total Revenue by Country	In {year},the total movie revenue in {country} reached ...

Approach and Methods



Relational database
- Six tables for query

Store templates

Store queries

SQLite

Jupyter Notebook

Sqlite3 database management
- Insertion
- Deletion
- Query

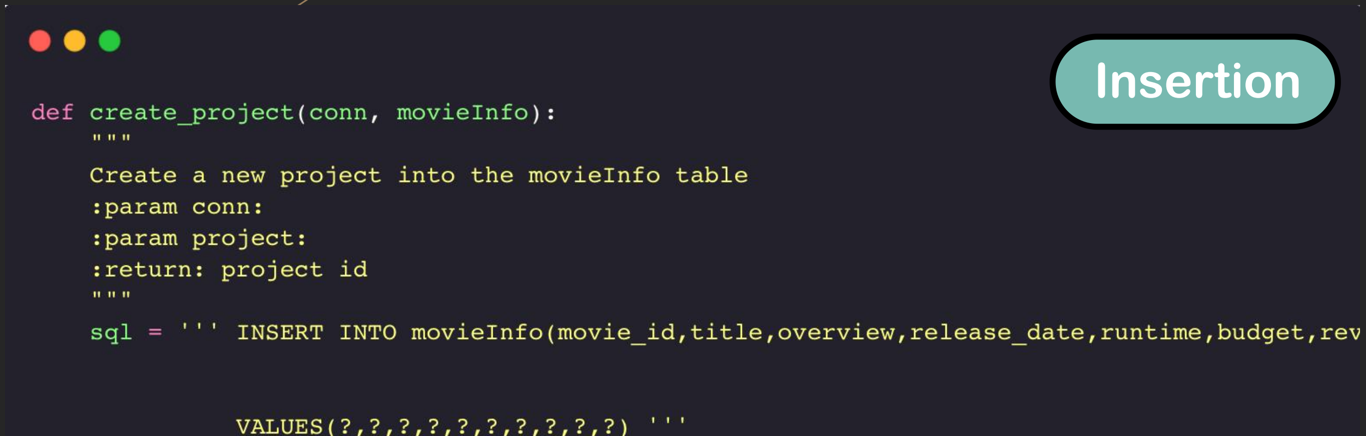
Aggregate functions

Textual user interface

Data visualization
- matplotlib.pyplot
- seaborn
- WordCloud

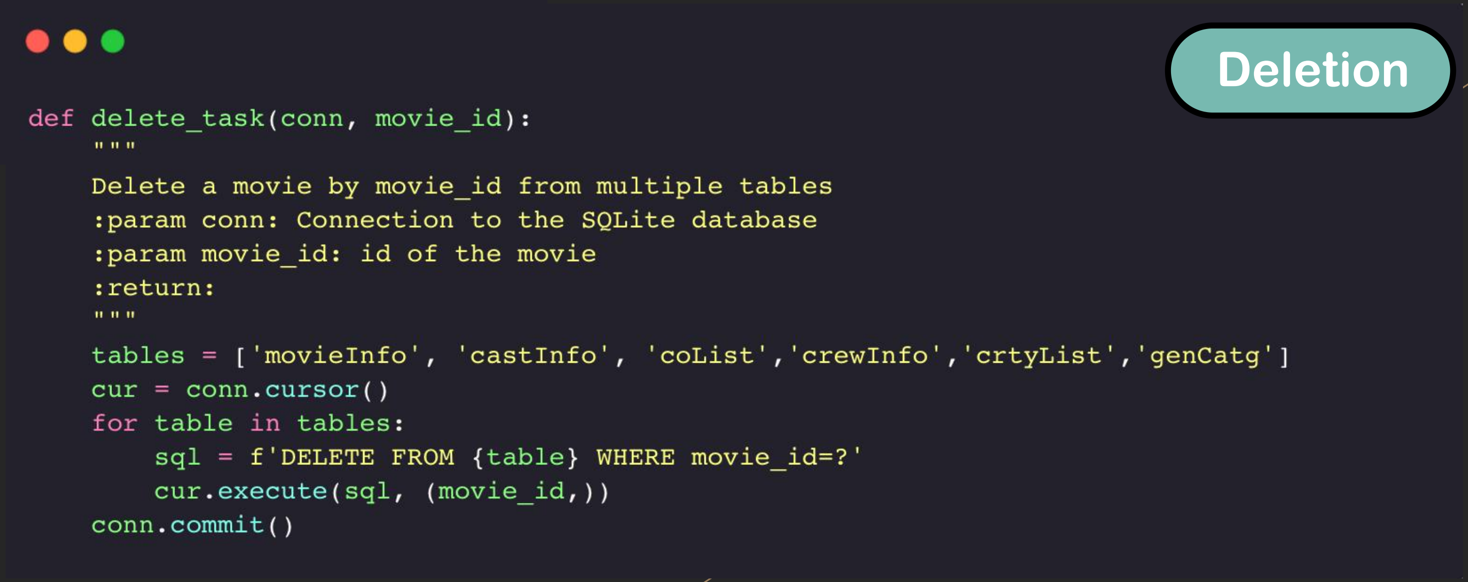
Pandas

Method 1: Database management operations



Insertion

```
def create_project(conn, movieInfo):
    """
    Create a new project into the movieInfo table
    :param conn:
    :param project:
    :return: project id
    """
    sql = ''' INSERT INTO movieInfo(movie_id,title,overview,release_date,runtime,budget,rev


              VALUES(?,?,?,?,?,?,?,?,?,?) '''
    cur = conn.cursor()
    cur.execute(sql, movieInfo)
    conn.commit()
    return cur.lastrowid
```

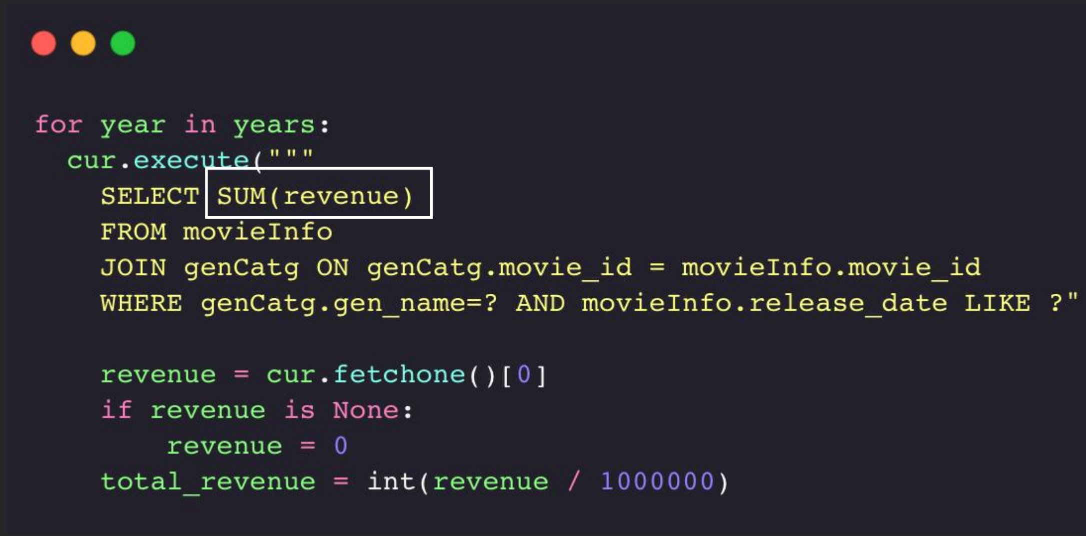


Deletion

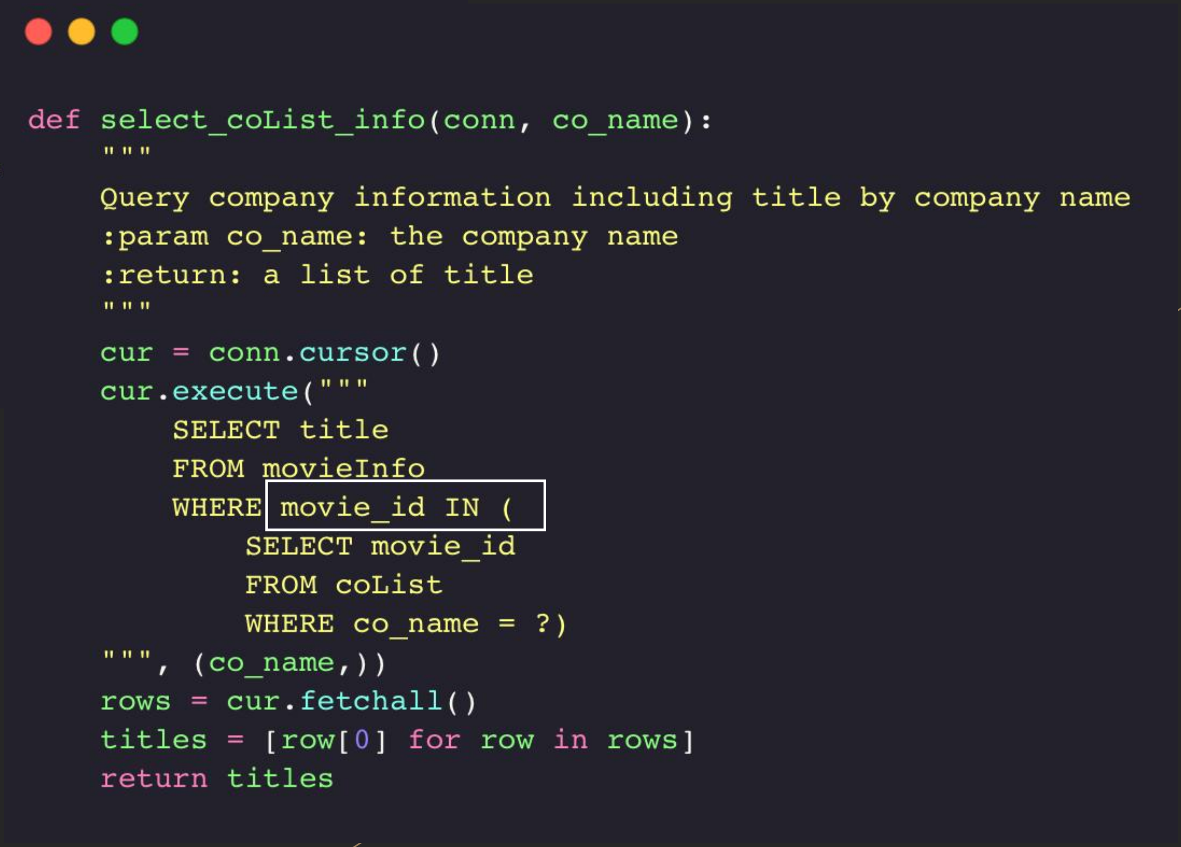
```
def delete_task(conn, movie_id):
    """
    Delete a movie by movie_id from multiple tables
    :param conn: Connection to the SQLite database
    :param movie_id: id of the movie
    :return:
    """
    tables = ['movieInfo', 'castInfo', 'coList','crewInfo','crtyList','genCatg']
    cur = conn.cursor()
    for table in tables:
        sql = f'DELETE FROM {table} WHERE movie_id=?'
        cur.execute(sql, (movie_id,))
    conn.commit()
```

- **Add new movies**
- **Remove outlines**

Method 2: Subquery & Aggregate functions



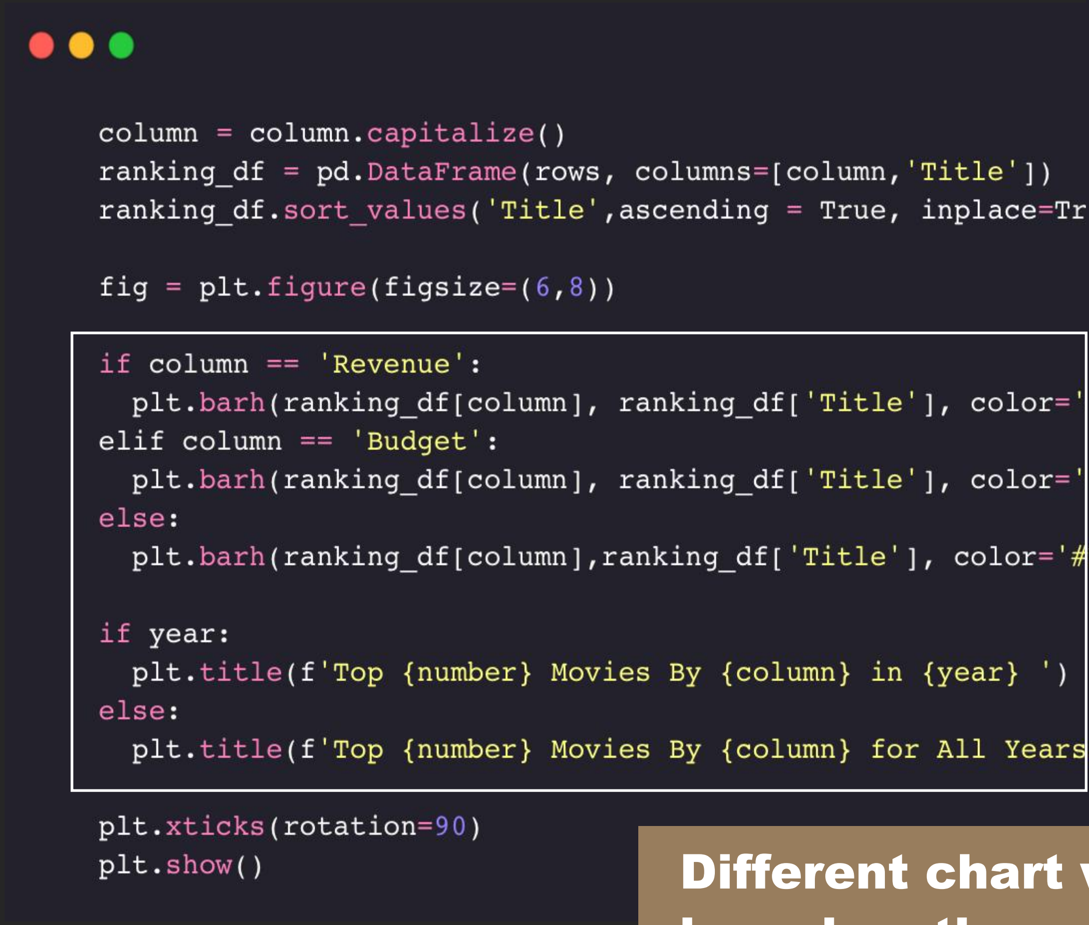
```
for year in years:
  cur.execute("""
    SELECT SUM(revenue)
    FROM movieInfo
    JOIN genCatg ON genCatg.movie_id = movieInfo.movie_id
    WHERE genCatg.gen_name=? AND movieInfo.release_date LIKE ?"

    revenue = cur.fetchone()[0]
    if revenue is None:
        revenue = 0
    total_revenue = int(revenue / 1000000)
```



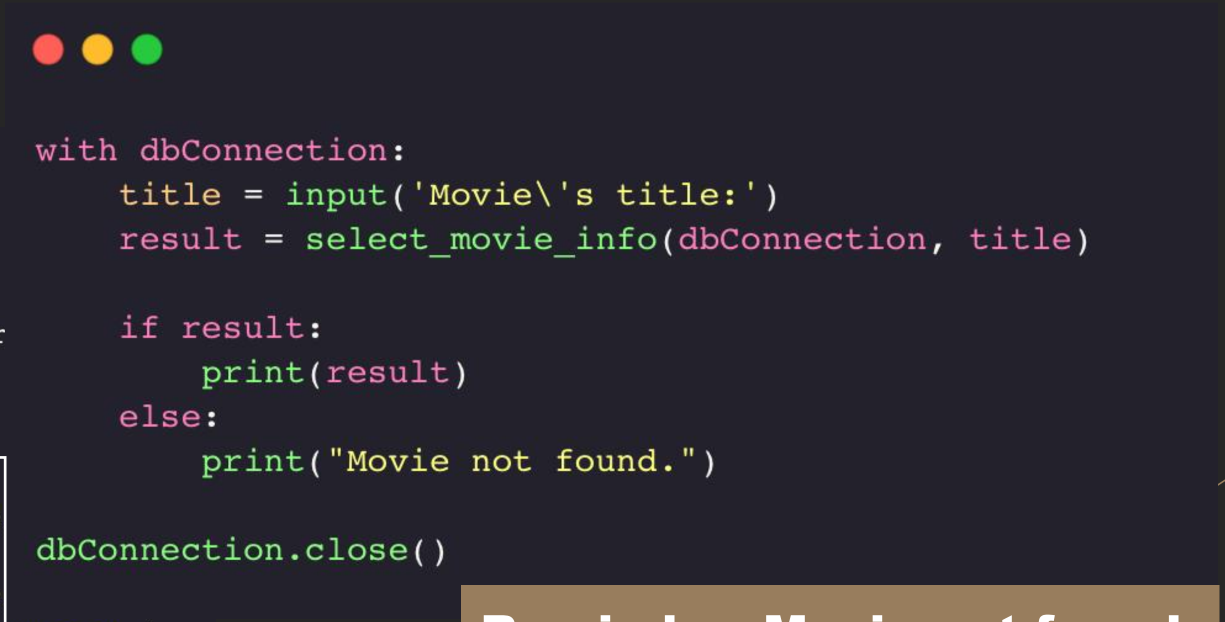
```
def select_coList_info(conn, co_name):
    """
    Query company information including title by company name
    :param co_name: the company name
    :return: a list of title
    """
    cur = conn.cursor()
    cur.execute("""
        SELECT title
        FROM movieInfo
        WHERE movie_id IN (
            SELECT movie_id
            FROM coList
            WHERE co_name = ?)
    """, (co_name,))
    rows = cur.fetchall()
    titles = [row[0] for row in rows]
    return titles
```

- **Processed mutiple queries**

Method 3: if…elif…else statement



```
column = column.capitalize()
ranking_df = pd.DataFrame(rows, columns=[column,'Title'])
ranking_df.sort_values('Title',ascending = True, inplace=Tr

fig = plt.figure(figsize=(6,8))

if column == 'Revenue':
  plt.barh(ranking_df[column], ranking_df['Title'], color='
elif column == 'Budget':
  plt.barh(ranking_df[column], ranking_df['Title'], color='
else:
  plt.barh(ranking_df[column],ranking_df['Title'], color='#F5C518')

if year:
  plt.title(f'Top {number} Movies By {column} in {year} ')
else:
  plt.title(f'Top {number} Movies By {column} for All Years ')

plt.xticks(rotation=90)
plt.show()
```

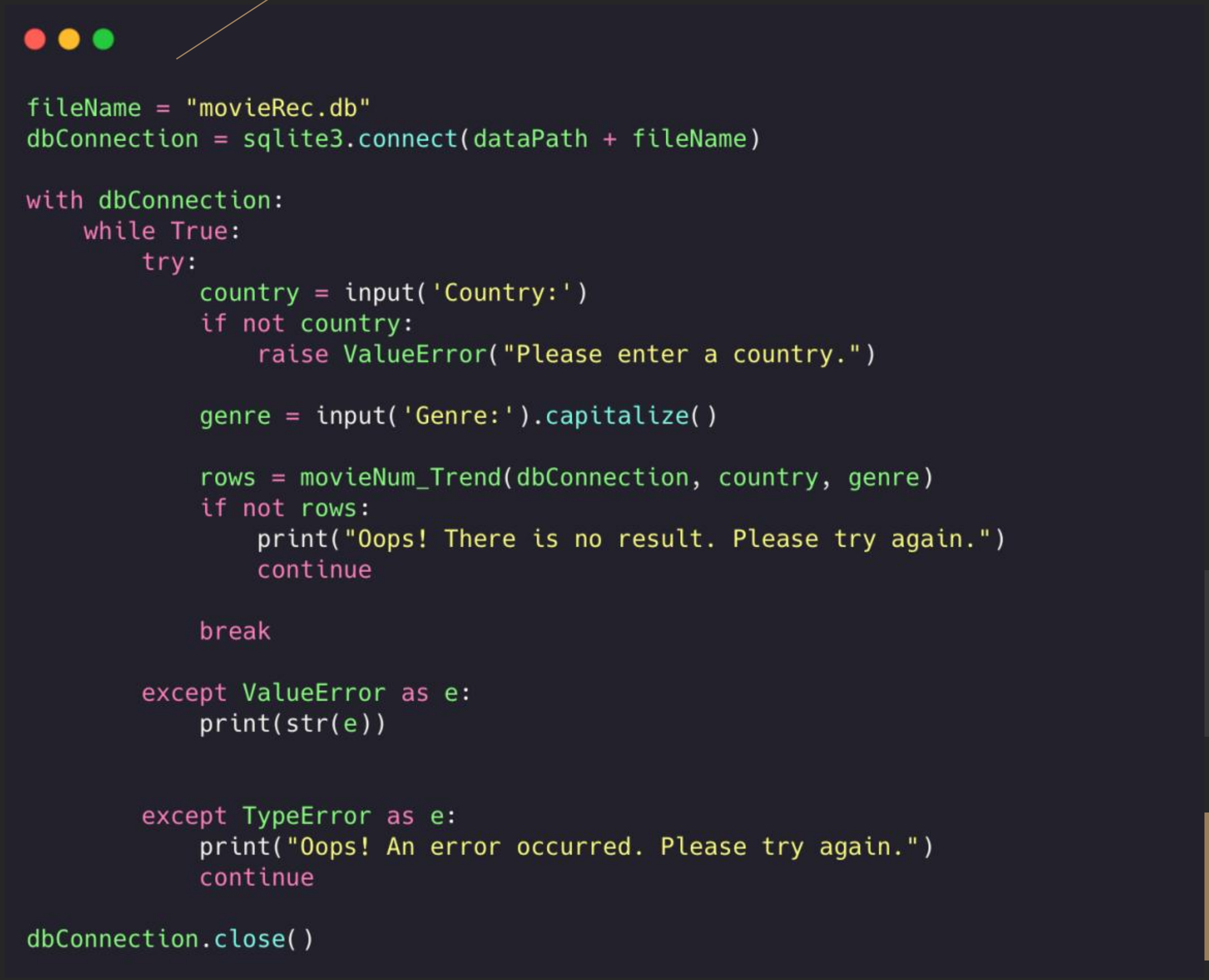


```
with dbConnection:
    title = input('Movie\'s title:')
    result = select_movie_info(dbConnection, title)

    if result:
        print(result)
    else:
        print("Movie not found.")

dbConnection.close()
```

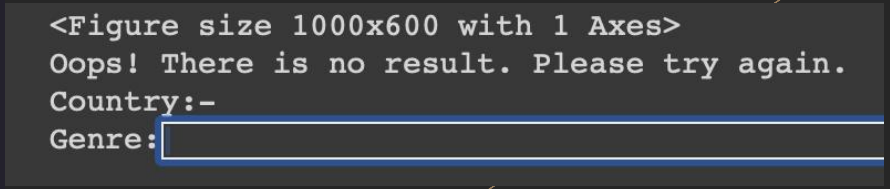
Reminder: Movie not found

Different chart will be displayed based on the selection from users

Method 4: try-except module



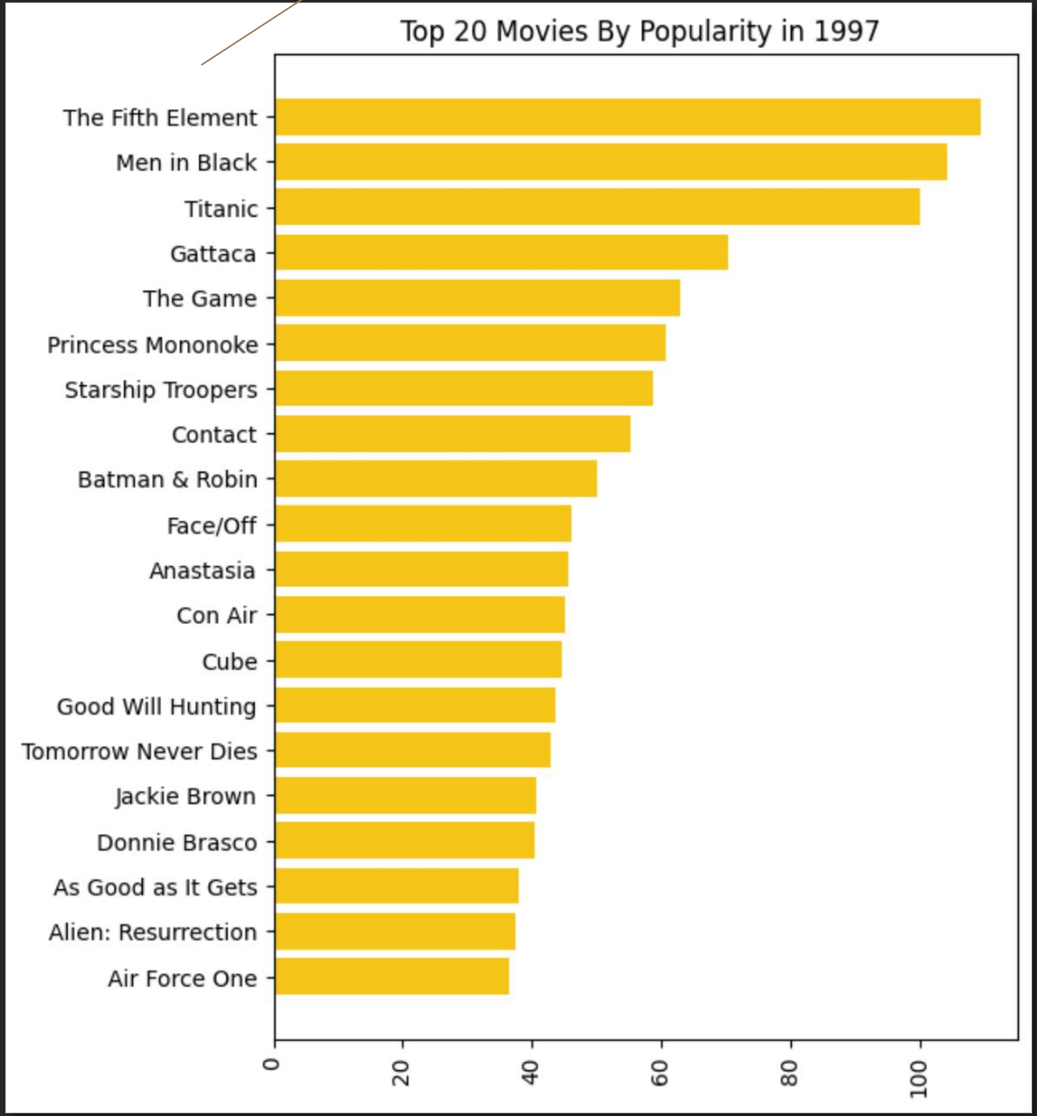
```
fileName = "movieRec.db"
dbConnection = sqlite3.connect(dataPath + fileName)

with dbConnection:
    while True:
        try:
            country = input('Country:')
            if not country:
                raise ValueError("Please enter a country.")

            genre = input('Genre:').capitalize()

            rows = movieNum_Trend(dbConnection, country, genre)
            if not rows:
                print("Oops! There is no result. Please try again.")
                continue

            break

        except ValueError as e:
            print(str(e))


        except TypeError as e:
            print("Oops! An error occurred. Please try again.")
            continue

dbConnection.close()
```

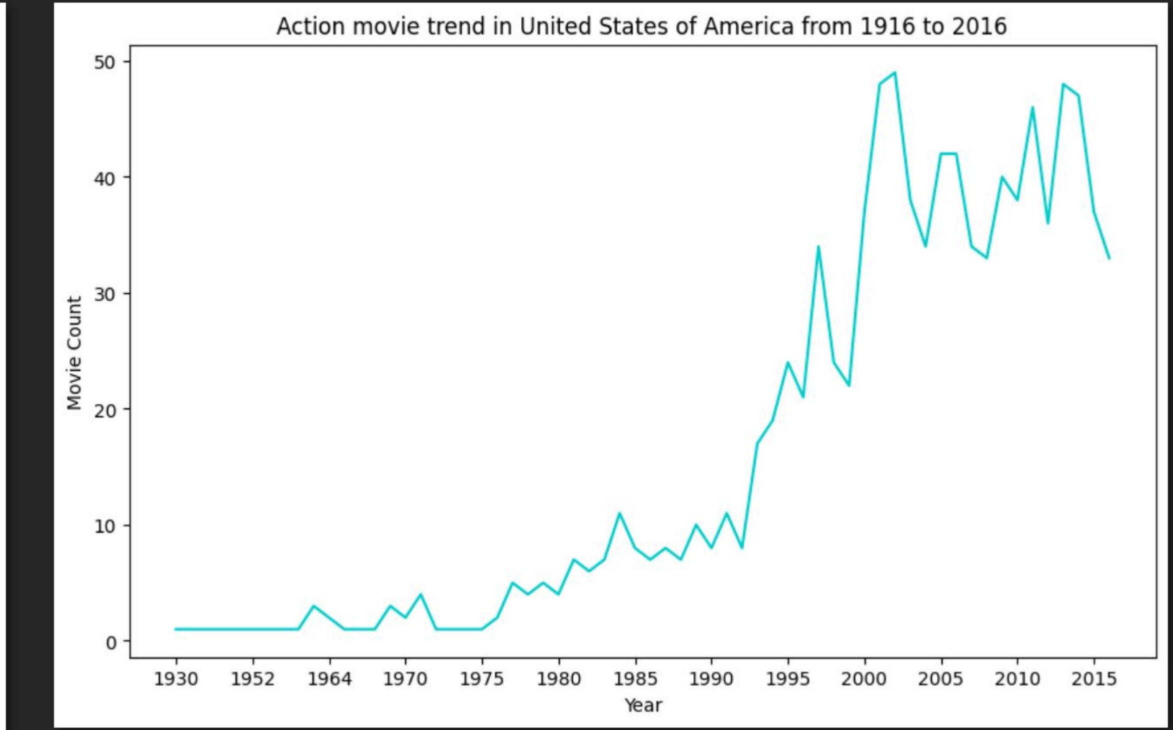


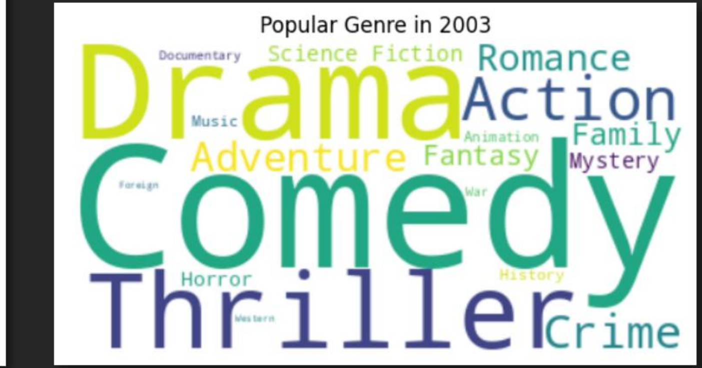
```
<Figure size 1000x600 with 1 Axes>
Oops! There is no result. Please try again.
Country:-
Genre:
```

- **Handle the ValueError and TypeError exceptions**

Method 5: Visualization







- Concise
- More readable

Templates Design

Basic Info Search

1.1 Movie Info
1.2 Actor Info
1.3 Actor collobration
1.4 Company Info

Annual Ranking

2.1 Revenue/Budget/Popularity Ranking
2.2 Most Active Actor Ranking
2.3 Popular Genre by Year

Data News

3.1 Production Trend by Country & Genre
2.2 Total Revenue by Genre & Year
2.3 Total Revenue by Country & Year

ZHENG SHUWAN CHENG YULU **ZHANG SHUYI** LIN YIMENG

MOVIE SEARCH

EXPLORE THE MOVIES YOU LIKE

Template 1.1: Movie Info Search

User Input:
movie title

Output Content
The *{title}* was released on *{date}* in *{crty_name}*. The film was produced by *{company}* and directed by *{director}*.It has already grossed *{revenue}* revenue in total.

1. What is the release date, director, production company of the movie?

2. How much revenue was generated this year?

Movie Information Search

movie title

Hi! Here what you want:

The **Avatar** was released on **10/12/2009** in **United States of America**.The film was produced by **Ingenious Film Partners** and directed by **James Cameron**. It has already grossed **2787.97** million in total.

Template 1.2: Single Actor Info Search

User Input:
Cast name

Output Content
{castName} has starred in a total of *{number}* movies. Among them, *{movieName}* achieved the highest revenue of *{amount}* USD, in which the character he/she played is *{characterName1}*. *{movieName}*also has a popularity rate of *{popularity}*, in which the character he/she played is *{characterName2}*. Additionally, *{castName}* has frequently collaborated with *{coStarName}*, and they have acted together in*{coStarNumber}* movie(s).

1. How many movies has he/she starred in?

2. Which movie achieved the highest revenue, how much revenue was generated this year, and what is the character name of this movie?

3. Which movie got the highest popularity? what is its popularity rate and character name?

4. Who is the frequent collaborator? How many movies have they made together?

Single Actor Search

cast name

Leonardo DiCaprio has starred in a total of 22 movies. Among them, Titanic achieved the highest revenue of 184.5 million USD, in which the character he/she played is Jack Dawson. And Inception has the highest popularity rate of 167.58, in which the character he/she played is Dom Cobb. Additionally, Leonardo DiCaprio has frequently collaborated with Nellie Sciutto, and they have acted together in 3 movies.

Template 1.3: Actor Collobration

User Input:
Two cast name

Output Content
The number of movie ***{actor/actress1}*** and ***{actor/actress2}*** have acted together is ***{number}***.
The first movie they collaborated on was ***"{movie_title}"***, which was released on ***{release_date}*** and directed by ***{direcctor}***.
The movie's overview is: ***{overview}***.

1. How many movies have they acted together?

2. What was the first movie they collaborated? the director, release date, and overview of this movie?

Actor Collobration

actor name 1

actor name 2

Search

The number of movie that George Lucas and Harrison Ford have acted together is 1.
The first movie they collaborated on was "Indiana Jones and the Temple of Doom", which was released on 23/5/1984 and directed by Steven Spielberg.
The movie's overview is: After arriving in India, Indiana Jones is asked by a desperate village to find a mystical stone. He agrees – and stumbles upon a secret cult plotting a terrible plan in the catacombs of an ancient palace.

ZHENG SHUWAN CHENG YULU **ZHANG SHUYI** LIN YIMENG

Template 1.4: Company Info Search

User Input:
Company name

Output Content
Here is the list of movies produced by ***{co_name}***.
{title}

1. Which movies are produced by this company?

Company Info Search

company name

Here is the list of movies produced by Walt Disney Pictures:

- **Pirates of the Caribbean: Dead Man's Chest**
- **Pirates of the Caribbean: At World's End**
- **Inspector Gadget**
- **The Straight Story**
- **Monsters, Inc.**
- **Fantasia**
- **Aladdin**

……

ZHENG SHUWAN CHENG YULU **ZHANG SHUYI** LIN YIMENG

Template 2.1: Revenue/Budget/Popularity Ranking

User Input:
Revenue/Budget/Popularity
Year
Number

Output Content
Bar chart

There are the TOP ***{movies}***
By ***{column}*** in ***{year}***.

1. Which movies are included in the top number by revenue/budget/popularity in this year?

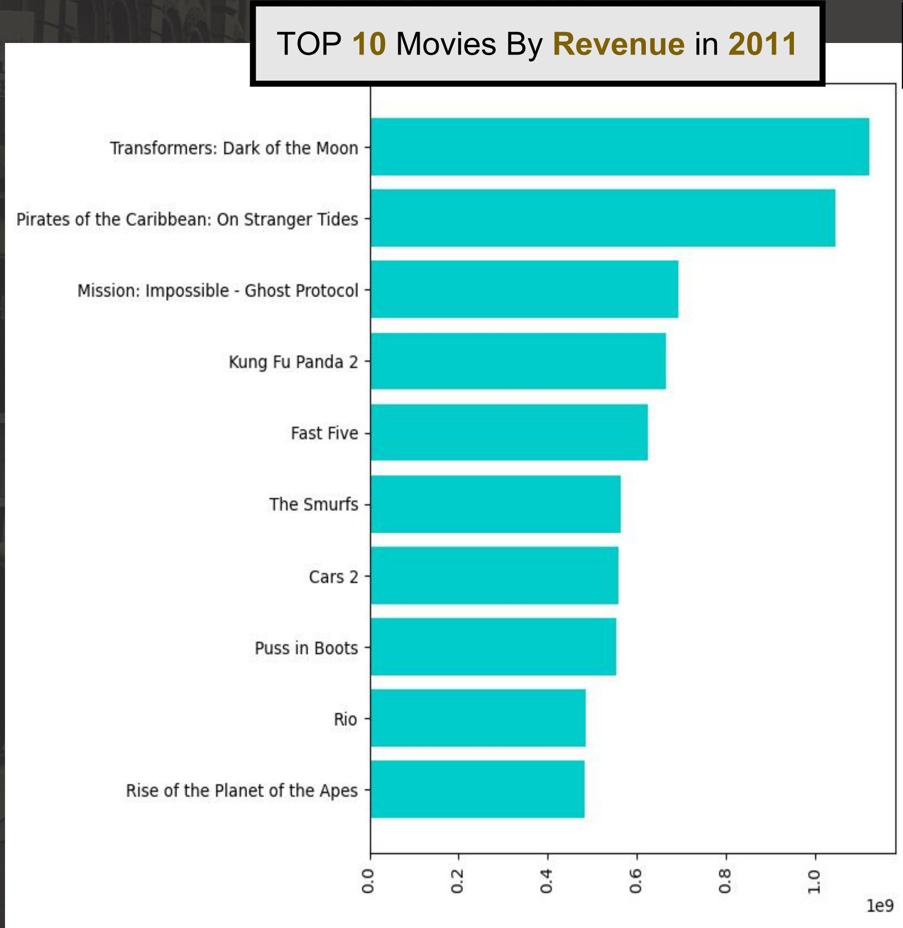
Revenue/Budget/Popularity Ranking

Revenue/Budget/Popularity

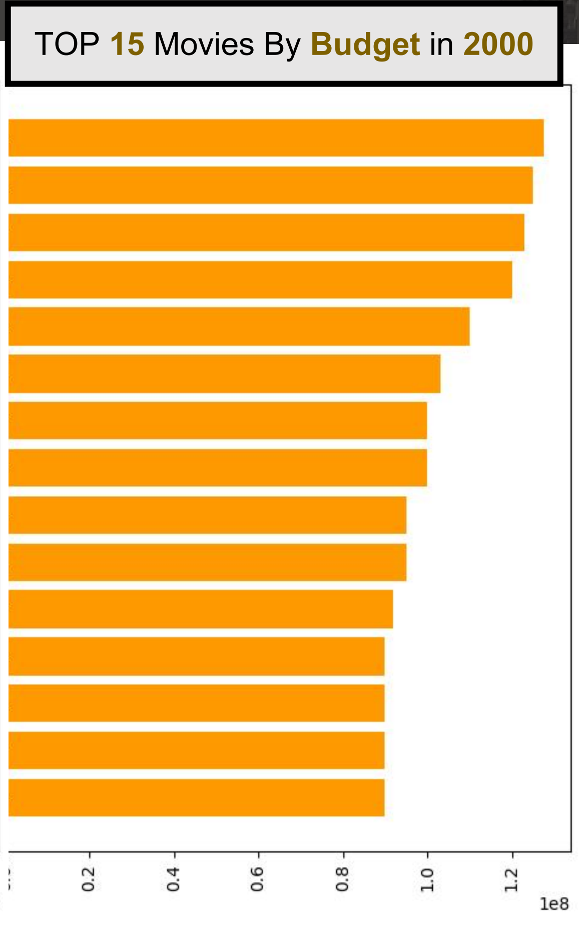
Year

Top 10/15/30/50

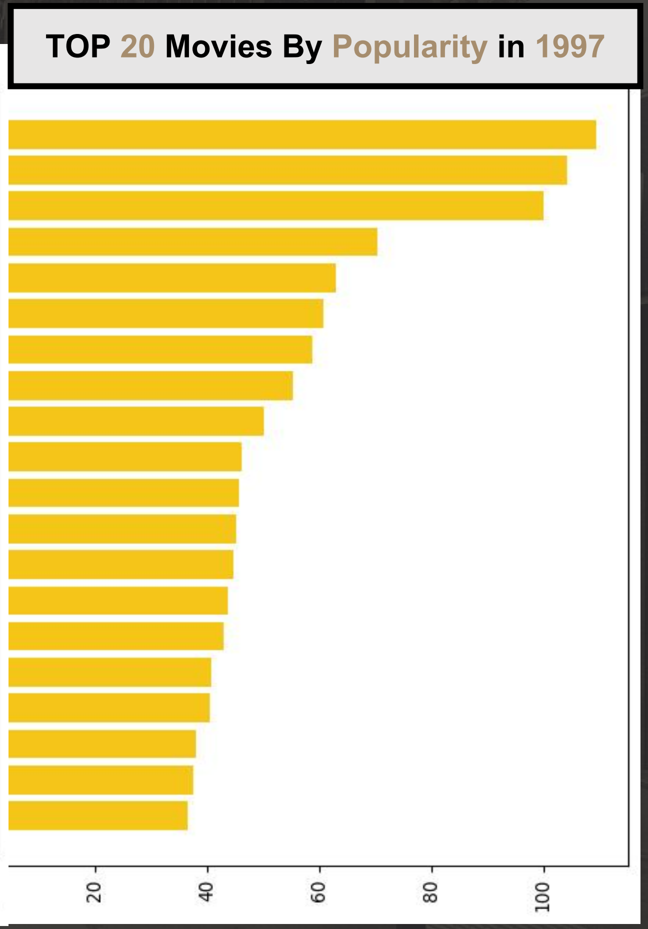
Search



TOP 10 Movies By Revenue in 2011



TOP 15 Movies By Budget in 2000



TOP 20 Movies By Popularity in 1997

Template 2.2: Most Active Actor Ranking

User Input:
Year

Output Content:
Table

Top ***{number}*** Most Active Cast in ***{year}***.

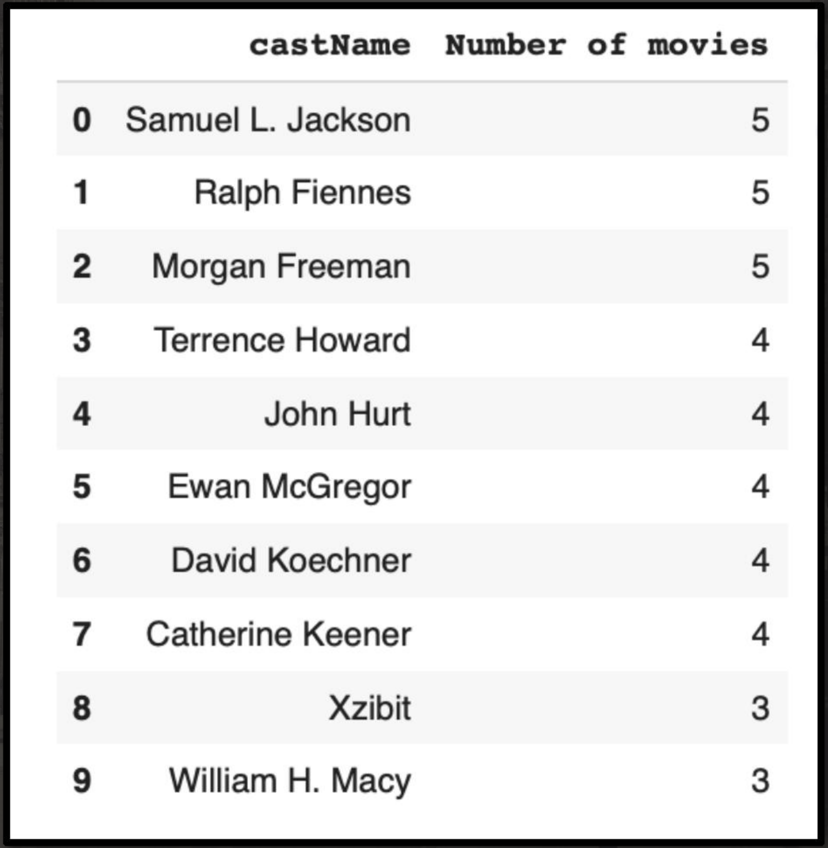
1. Who is the most active actor in this year?
2. How many movies does he/her played?

Most Active Actor Ranking

Year

Top 10/15/30/50

Search



	castName	Number of movies
0	Samuel L. Jackson	5
1	Ralph Fiennes	5
2	Morgan Freeman	5
3	Terrence Howard	4
4	John Hurt	4
5	Ewan McGregor	4
6	David Koechner	4
7	Catherine Keener	4
8	Xzibit	3
9	William H. Macy	3

Top 10 Most Active Actors in 2005

Template 2.3: Popular Genre by Year

User Input:
Year

Output Content:
World Cloud

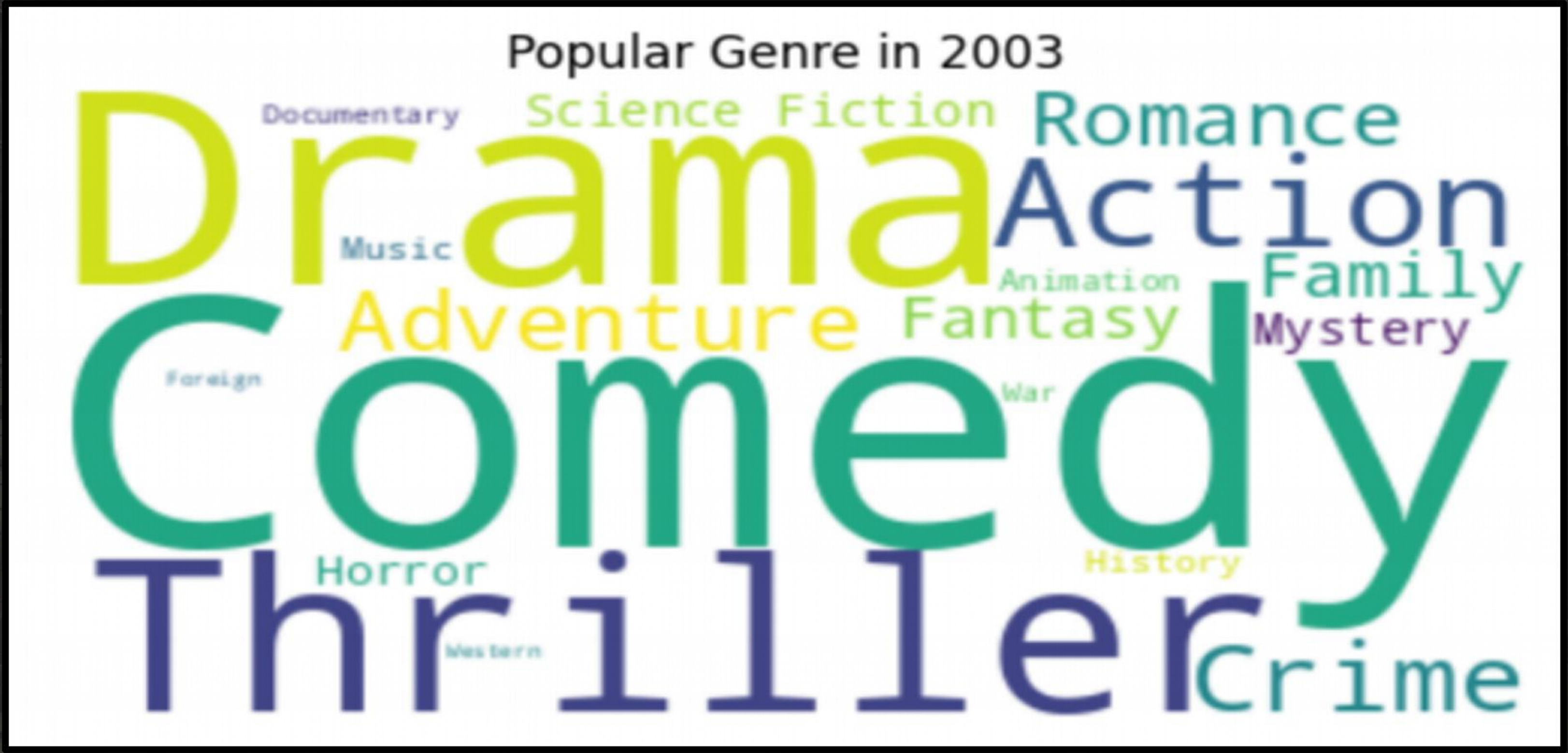
Popular Genre in *{Year}*

What is the popular genre in this year?

Popular Genre by Year

Year

Search



Popular Genre in 2003

Template 3.1: Production Trend by Country & Genre

User Input:
Country
Genre

Output Content
Line Chart

This is the ***{genre}*** movie trend in ***{country}*** from 1916 to 2016

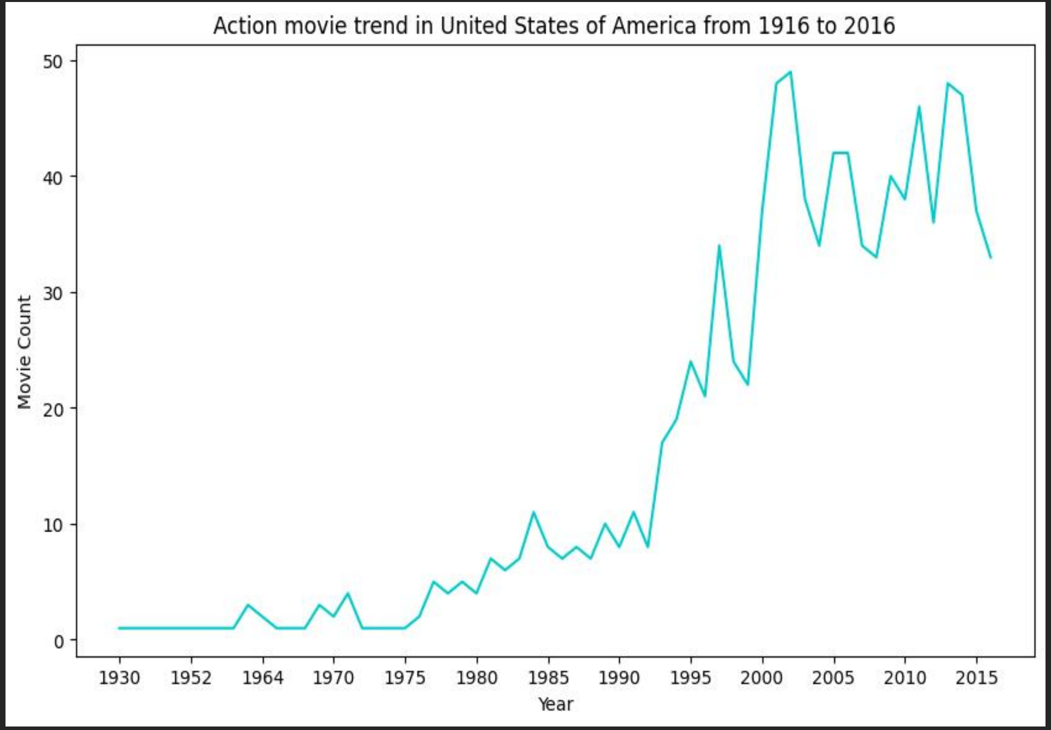
Show the trend for this genre in this country.

Production Trend by Country & Genre

Country

Genre

Search



ZHENG SHUWAN CHENG YULU ZHANG SHUYI **LIN YIMENG**

Template 3.2: Total Revenue by Genre & Year

User Input:
Year
Genre

Output Content:
In *{year}*, the global *{genres}* movie revenue reached *{revenue data}* dollars, which represents a YoY *{increase/decrease}* of *{percentage}* compared to *{last_year}* with a total revenue of *{revenue}* dollars.
Among the films released, the top 5 highest-revenue films were *{title}*.

1. How much total revenue of this genre of all the movies was generated this year?
2. What the trend compared to last year?
3. Which movies got top5 highest-revenue?

Total Revenue by Genre & Year

Year

Genre

Search

In 2004, the global Action movie revenue reached 5711 million dollars, which represents a YoY decrease of 22.74% compared to 2003 with a total revenue of 7392 million dollars.

The top 5 movies in 2004 with the Action genre are: Spider-Man 2, The Incredibles,The Day After Tomorrow, Shark Tale, National Treasure

Template 3.3: Total Revenue by Country

User Input:
Year
Country

Output Content:
In *{year}*, the total movie revenue in *{country}* reached *{revenue}* dollars, which represents a YoY *{increase/decrease}* of *{percentage}* compared to *{last_year}* with a total revenue of *{revenue}* dollars. Among the films released, the top 5 highest-revenue films were *{title}*.

1. How much total revenue of movies in this country was generated this year?
2. What the trend compared to last year?
3. Which movies got top5 highest-revenue?

Total Revenue by Country

Year   Country   Search

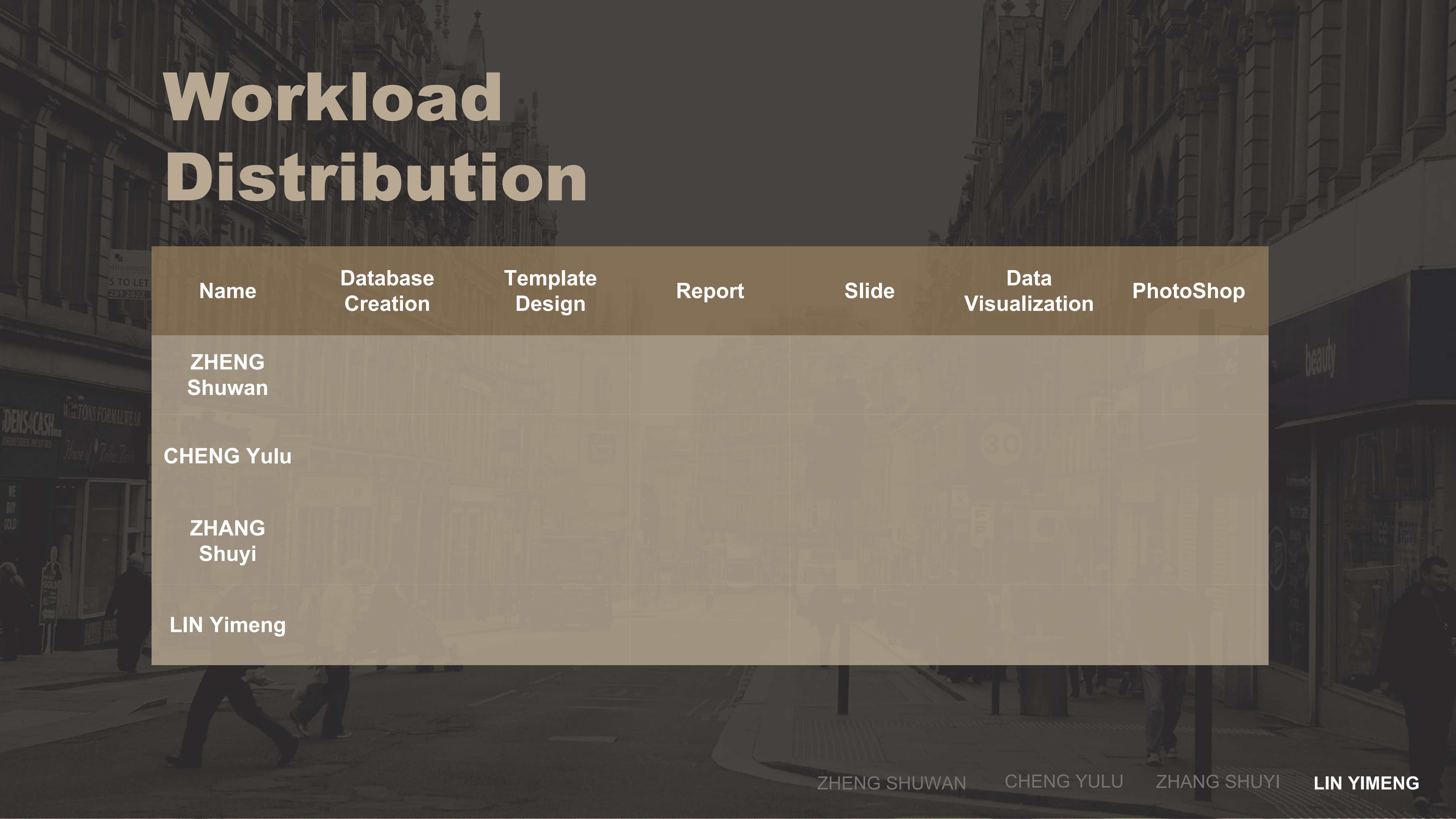
In 2015, the total movie revenue in US reached 22074 million dollars, which represents a YoY decrease of 5.63% compared to 2014 with a total revenue of 23390 million dollars.

Among the films released, the top 5 highest-revenue films were: **Jurassic World, Furious 7, Avengers: Age of Ultron, Minions, Spectre**

ZHENG SHUWAN   CHENG YULU   ZHANG SHUYI   **LIN YIMENG**

Workload Distribution

Name	Database Creation	Template Design	Report	Slide	Data Visualization	PhotoShop
ZHENG Shuwan	🎬🎬	🎬	🎬	🎬🎬		
CHENG Yulu	🎬	🎬🎬	🎬	🎬🎬	🎬	🎬
ZHANG Shuyi		🎬	🎬🎬	🎬		🎬🎬
LIN Yimeng		🎬	🎬🎬	🎬		



Workload Distribution

Name	Database Creation	Template Design	Report	Slide	Data Visualization	PhotoShop
ZHENG Shuwan						
CHENG Yulu						
ZHANG Shuyi						
LIN Yimeng						

Hindmarsh
S TO LET
281 2822

30
beauty
health

THANK YOU

30
beauty
health