In [ ]:
```
#!pip install mysql-connector-python pandas
```

In [2]:
```
import mysql.connector
import pandas as pd
```

1. How many movies are in data set ?

In [23]:
```
connection = mysql.connector.connect(
    host='mysql-db',   # container name of the MySQL container
    port=3306,
    user='root',
    password='root',
    database='movielens'
)

query = '''SELECT COUNT(*) AS number_of_movies FROM movies;'''

df = pd.read_sql(query, connection)

connection.close()
print("There are", df.loc[0,"number_of_movies"],"movies in the database")
```

```
There are 528 movies in the database
```

```
/tmp/ipykernel_311/4245464557.py:11: UserWarning: pandas only supports SQLAlchemy
connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connectio
n. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, connection)
```

1. What is the most common genre of movie?

In [24]:
```
connection = mysql.connector.connect(
    host='mysql-db',
    port=3306,
    user='root',
    password='root',
    database='movielens'
)

query = '''SELECT genres, COUNT(*) AS frequency FROM movies
            GROUP BY genres
            ORDER BY frequency DESC
            LIMIT 1;'''

df = pd.read_sql(query, connection)

connection.close()

df.head()
```

```
/tmp/ipykernel_311/1721752559.py:14: UserWarning: pandas only supports SQLAlchemy
connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connectio
n. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, connection)
```

Out[24]:

	genres	frequency
0	Drama	70

1. What are top 10 movies with highest rate ?

In [25]:
```
connection = mysql.connector.connect(
    host='mysql-db',
    port=3306,
    user='root',
    password='root',
    database='movielens'
)

query = '''SELECT ratings.movieId, movies.title, SUM(rating)/COUNT(rating) AS avg_r
            FROM ratings LEFT JOIN movies ON ratings.movieId=movies.movieId
            GROUP BY movieId, title
            ORDER BY avg_rating DESC
            LIMIT 10;'''

df = pd.read_sql(query, connection)

connection.close()

df.head(10)
```

```
/tmp/ipykernel_311/608167798.py:15: UserWarning: pandas only supports SQLAlchemy c
onnectable (engine/connection) or database string URI or sqlite3 DBAPI2 connectio
n. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, connection)
```

Out[25]:

	movieId	title	avg_rating
0	69860	None	5.0
1	6201	None	5.0
2	112512	None	5.0
3	90943	None	5.0
4	26928	None	5.0
5	86668	None	5.0
6	95149	None	5.0
7	114265	None	5.0
8	124851	None	5.0
9	26401	None	5.0

1. What are 5 most often rating users ?

In [26]:
```
connection = mysql.connector.connect(
    host='mysql-db',
    port=3306,
    user='root',
    password='root',
    database='movielens'
)

query = '''SELECT userId, COUNT(*) AS frequency FROM ratings
            GROUP BY userId
            ORDER BY frequency DESC
            LIMIT 5;'''
```

```
df = pd.read_sql(query, connection)

connection.close()

df.head(5)
```

```
/tmp/ipykernel_311/2039589586.py:14: UserWarning: pandas only supports SQLAlchemy
connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connectio
n. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, connection)
```

Out[26]:

	userId	frequency
0	414	2698
1	599	2478
2	474	2108
3	448	1864
4	274	1346

1. When was done first and last rate included in data set and what was the rated movie tittle?

In [34]:

```
connection = mysql.connector.connect(
    host='mysql-db',
    port=3306,
    user='root',
    password='root',
    database='movielens'
)

query = '''SELECT ratings.movieId, movies.title,  FROM_UNIXTIME(timestamp) AS date
            FROM ratings LEFT JOIN movies ON ratings.movieId=movies.movieId
            WHERE ratings.timestamp = (SELECT MIN(timestamp) FROM ratings)
            OR ratings.timestamp = (SELECT MAX(timestamp) FROM ratings)
            ORDER BY date ASC'''

df = pd.read_sql(query, connection)

connection.close()

df.head(25)
```

```
/tmp/ipykernel_311/1370260207.py:15: UserWarning: pandas only supports SQLAlchemy
connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connectio
n. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, connection)
```

Out[34]:

	movieId	title	date
0	150	Apollo 13 (1995)	1996-03-29 18:36:55
1	595	Beauty and the Beast (1991)	1996-03-29 18:36:55
2	592	Batman (1989)	1996-03-29 18:36:55
3	590	Dances with Wolves (1990)	1996-03-29 18:36:55
4	588	Aladdin (1992)	1996-03-29 18:36:55
5	468	Englishman Who Went Up a Hill But Came Down a ...	1996-03-29 18:36:55
6	434	Cliffhanger (1993)	1996-03-29 18:36:55
7	432	City Slickers II: The Legend of Curly's Gold (...	1996-03-29 18:36:55
8	421	Black Beauty (1994)	1996-03-29 18:36:55
9	420	Beverly Hills Cop III (1994)	1996-03-29 18:36:55
10	351	Corrina, Corrina (1994)	1996-03-29 18:36:55
11	349	Clear and Present Danger (1994)	1996-03-29 18:36:55
12	343	Baby-Sitters Club, The (1995)	1996-03-29 18:36:55
13	227	Drop Zone (1994)	1996-03-29 18:36:55
14	225	Disclosure (1994)	1996-03-29 18:36:55
15	222	Circle of Friends (1995)	1996-03-29 18:36:55
16	218	Boys on the Side (1995)	1996-03-29 18:36:55
17	165	Die Hard: With a Vengeance (1995)	1996-03-29 18:36:55
18	161	Crimson Tide (1995)	1996-03-29 18:36:55
19	22	Copycat (1995)	1996-03-29 18:36:55
20	162	Crumb (1994)	2018-09-24 14:27:30

1. Find all movies released in 1990

In [28]:
```
connection = mysql.connector.connect(
    host='mysql-db',
    port=3306,
    user='root',
    password='root',
    database='movielens'
)

query = '''SELECT * FROM movies
            WHERE title LIKE "%(1990)";'''

df = pd.read_sql(query, connection)

connection.close()

df.head()
```

```
/tmp/ipykernel_311/2477095695.py:12: UserWarning: pandas only supports SQLAlchemy
connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connectio
n. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, connection)
```

Out[28]:

	movieId	title	genres
0	586	Home Alone (1990)	Children|Comedy
1	587	Ghost (1990)	Comedy|Drama|Fantasy|Romance|Thriller
2	590	Dances with Wolves (1990)	Adventure|Drama|Western
3	597	Pretty Woman (1990)	Comedy|Romance