

IMDB DATA ANALYSIS USING SQL

Submitted by

Nelson Joseph (1002050500)

Jeeva Manavalan (1002067528)

in partial fulfilment for the award of the degree of

M.S. IN DATA SCIENCE



THE UNIVERSITY OF TEXAS AT ARLINGTON

DASC 5300/CSE 5300 FOUNDATIONS OF COMPUTING

INSTRUCTOR: Prof. Sharma Chakravarthy

06 DECEMBER 2022

Honor Code:

Nelson **Joseph** and Jeeva **Manavalan** did not give or receive any assistance on this project, and the report submitted is wholly on our own.

Table of Contents

Introduction	3
File Description	3
Division of Labour	3
Pre-Processing	4
Analysis 1	5
Analysis 1 A	5
Analysis 1 B	6
Analysis 1 C	8
Total number of movies each Actor/Actress acted during their career	8
Average rating of each Actor/Actress during their career	8
Sylvester Stallone	9
Jim Carrey	9
Kate Winslet	9
Emma stone	9
Analysis 2	10
Analysis based on each year within year range 1977-1986	10
Maximum number of movies acted by Actor/Actress in each year.	10
Minimum number of movies acted by Actor/Actress in each year.	10
Analysis based on the entire year range 1977-1986	11
Total number of movies acted by Actor/Actress	11
Minimum number of movies acted by Actor/Actress	11
Problems Encountered	12
Conclusion	13

Introduction

The project's objective was to do analysis on the IMDB database using SQL queries to discover details about the actors' and actresses' professional development. Using the SQL developer, we ran SQL queries on the Oracle database. Pre-processing was first carried out to fully comprehend the database schema, the tables, and their attributes. According to the parameter file provided, which contains the names of one actor and one actress as well as the years 1977 to 1986, we did two studies. Three subcategories were considered in the initial analysis. The actor and actress listed in the parameter file were Sylvester Stallone and Kate Winslet. To identify the actors and actresses who had the best year of their careers and who advanced the most professionally, we also employed two other actors and actresses in our initial research. The total number of films that each actor appeared in annually as well as their average career ratings were examined. Discovered which movies the chosen actor or actress is best known for. Finally learned which actor and actress appeared in the most and least amount of films from 1977 to 1986.

File Description

File Type	File Name
Sql Files	preprocessing_1.sql, analysis_1_A.sql, analysis_1_B.sql, analysis_2_.sql
Text files	analysis_1_A.txt, analysis_1_B.txt, analysis_2_a.txt, analysis_2_b.txt, analysis_2_c.txt, analysis_2_d.txt, analysis_2_e.txt, analysis_2_entire.txt, Knownfortitles spool text.txt, preprocessing.txt, All_imdb_tables.txt
CSV file	Analysis_1_C.csv
Jupyter Notebook	Project_3_Analysis_1c_Visualizations.ipynb

Table 1.1 displays information about the various files used in the overall analysis

Division of Labour

TASK	MEMBER
Pre-processing	Pre-processing is completed collectively.
Analysis 1. A	Nelson
Analysis 1. B	Jeeva
Analysis 1. C	Nelson
Analysis 2	Jeeva
Report	The task is completed collaboratively

Table 1.2 Division of Labour for the Analysis Project

Pre-Processing

- First, we used SQL queries to determine the table with the owner as IMDB00 IMDB Database so that we could select the information needed for the analysis.
- The attributes of each IMDB00 table were identified using sample queries for properly understanding the data.
- Using the data above, we were able to determine the best movies that each actor or actress is best known for.
- Extracted the top 4 movies of Sylvester Stallone, Kate Winslet, Jim Carrey, and Emma Stone along with the movies start year and Genre.
- Pre-processing results are shown below.

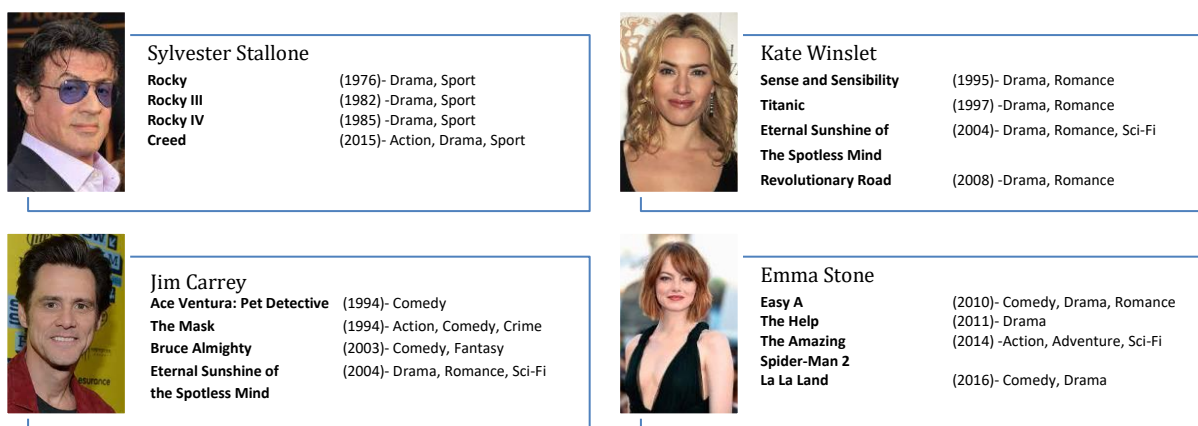


Figure 1.1 Summarization of Pre-processing results

- Based on our preliminary study, we discovered that Jim Carrey and Kate Winslet are both well-known for their roles in the 2004 film "**Eternal Sunshine of the Spotless Mind**".
- In 2007, Kate Winslet and Sylvester Stallone co-starred in the film "**Vanity Fair: Killers Kill, Dead Men Die**".
- Kate Winslet and Emma Stone appeared in a brief web series called "**Movie 43**".

Analysis 1

Based on the parameter file, Sylvester Stallone and Kate Winslet were chosen for this analysis, and our favourite actor and actress were taken as Jim Carrey and Emma Stone respectively.

Analysis 1 A

Each of the four actors and actresses' acting career spans in years, as well as the total number of movies produced in each year, was taken from the IMDb00 database for this study and visualized below.

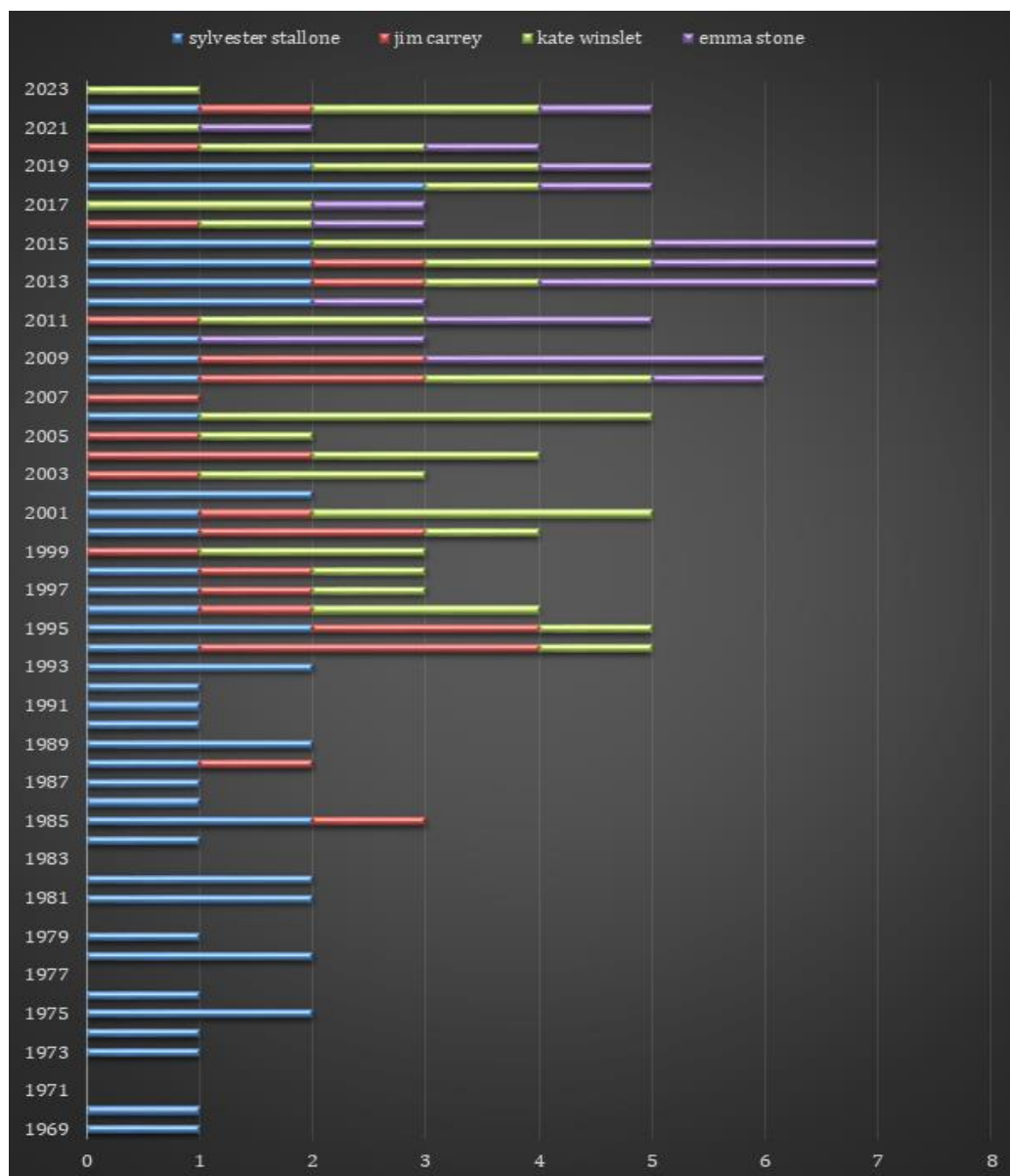


Figure A.1 Total number of movies acted by Actor/Actress in each year

Above diagram represent movie count of Actor/Actress in each year over their acting career. Movie counts represented are visualized by each colour, not by line's length.

Analysis 1 B

In order to track and illustrate the actor's or actress's career growth for Analysis 1 B, we divided the span into seven discrete time periods.

NAME	RANGE	MOVIE COUNT	RATING	NAME	RANGE	MOVIE COUNT	RATING
Emma Stone	2001-2010	7	5.72	Kate Winslet	2001-2010	16	6.89
Emma Stone	2011- 2020	15	6.90	Kate Winslet	2011- 2020	18	6.43
Emma Stone	2021-2030	1	7.30	Kate Winslet	2021-2030	2	7.90
Jim Carrey	1981-1990	3	5.45	Sylvester Stallone	1961-1970	2	4.05
Jim Carrey	1991-2000	11	6.84	Sylvester Stallone	1971-1980	8	6.16
Jim Carrey	2001-2010	10	6.73	Sylvester Stallone	1981-1990	13	5.92
Jim Carrey	2011- 2020	5	5.86	Sylvester Stallone	1991-2000	11	5.93
Jim Carrey	2021-2030	1	6.60	Sylvester Stallone	2001-2010	7	5.68
Kate Winslet	1991-2000	9	7.02	Sylvester Stallone	2011- 2020	13	5.80

Table B.1 Total number of movies acted by each actor/actress during each decade

The first of the four superstars to begin an acting career was Sylvester Stallone in 1969. As a result, we divided the time period into decades beginning in 1961. **1961–1970**, **1971–1980**, **1981–1990**, **1991–2000**, **2001–2010**, **2011–2020**, and **2021–2030** are the seven separate time eras.

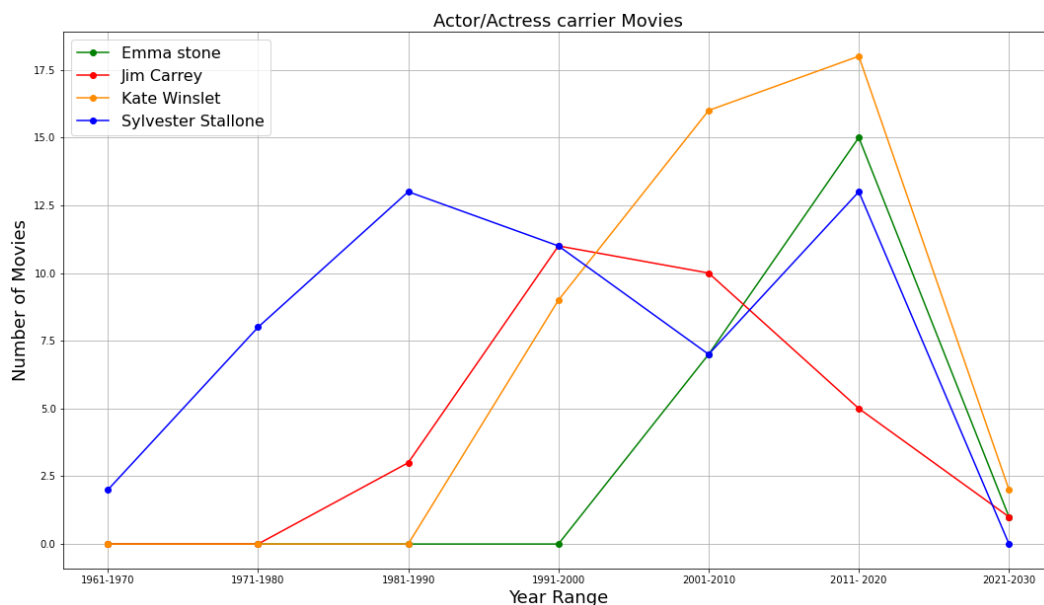


Figure B.1 Visualization of total number of movies each actor/actress acted in each decade.

- Since Sylvester Stallone was the first to begin his career among the others, he appeared in the most films of the four. Kate Winslet, Jim Carrey, and Emma Stone are listed in that sequence after him.
- Even though Kate Winslet's career just began in 1991-2000, she appeared in more films than Jim Carrey, whose career began in 1981 and ended in 1990.
- When compared to other actors and actresses, Kate Winslet appeared in the most movies from 2011 to 2020, followed by Emma Stone.

- Emma Stone, who is in her early 30s, was the last member of the group to begin her acting career.
- The only actresses to get an Academy Award were Kate Winslet and Emma Stone; other actors received only nominations. Emma Stone was nominated for an Academy Award three times and won once.
- In our analysis, Jim Carrey is the only actor who hasn't even received an Academy Award nomination.
- With a total of seven nominations and one win, Kate Winslet has received the most academy award nominations.
- In his career, Sylvester Stallone received three Academy Award nominations, although he was never successful.
- According to our analysis, Emma Stone earns more than any other actor or actress.

Sylvester Stallone, having begun his career in 1968 with the film "**square root**" he rose to international popularity with the release of "**Rocky**" in 1976, for which he was nominated for an Academy Award for best actor and an Academy Award for best original screenplay in 1977.

Jim Carrey started out in stand-up comedy in 1977, and from 1990 to 1994, he was a cast member of the television show "**In Living Colour**". He became well-known in 1994 thanks to the movies "**Ace Ventura: Pet Detective**", "**The Mask**", and "**Dumb & Dumber**".

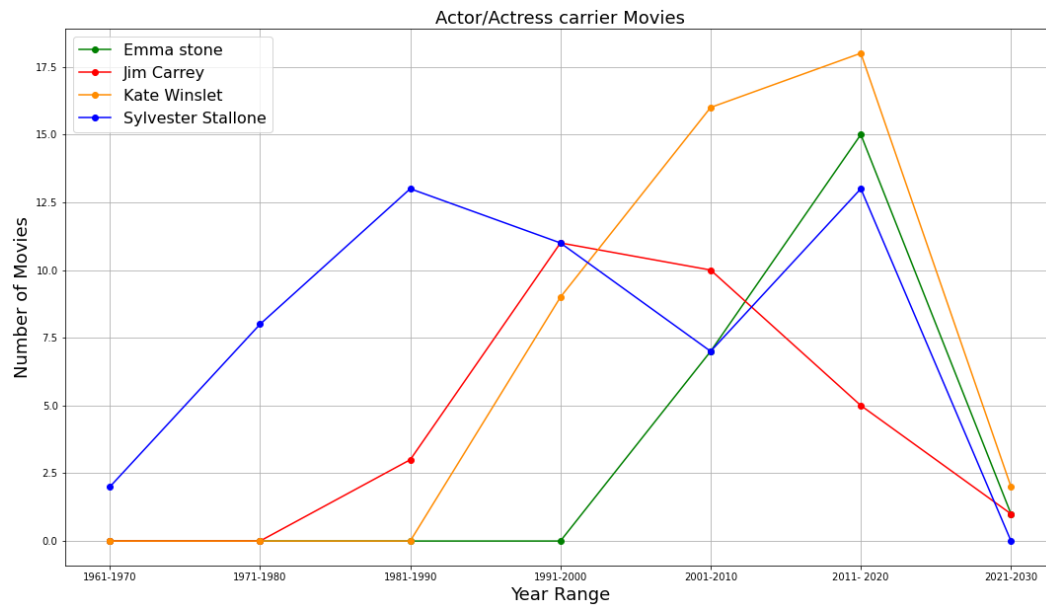
Kate Winslet made her acting debut in "**Heavenly Creature**" in 1994, and in 1995, she won the BAFTA Award for the movie *Sense and Sensibility*. She had a breakthrough with the 1997 movie "**Titanic**", for which she was nominated for an Academy Award and won widespread recognition.

Emma Stone made a breakthrough with three movies: "**Ghosts of girlfriend's past**", "**Paper Man**", and "**Zombieland**". Emma Stone started her acting career in 2007 with the movie "**Superbad**". She starred in the movie *Easy A* in 2010, for which she was nominated for the Golden Globe Award for best actress, the BAFTA Rising Star Award for best actress, and the MTV Movie Award for best humorous performance.

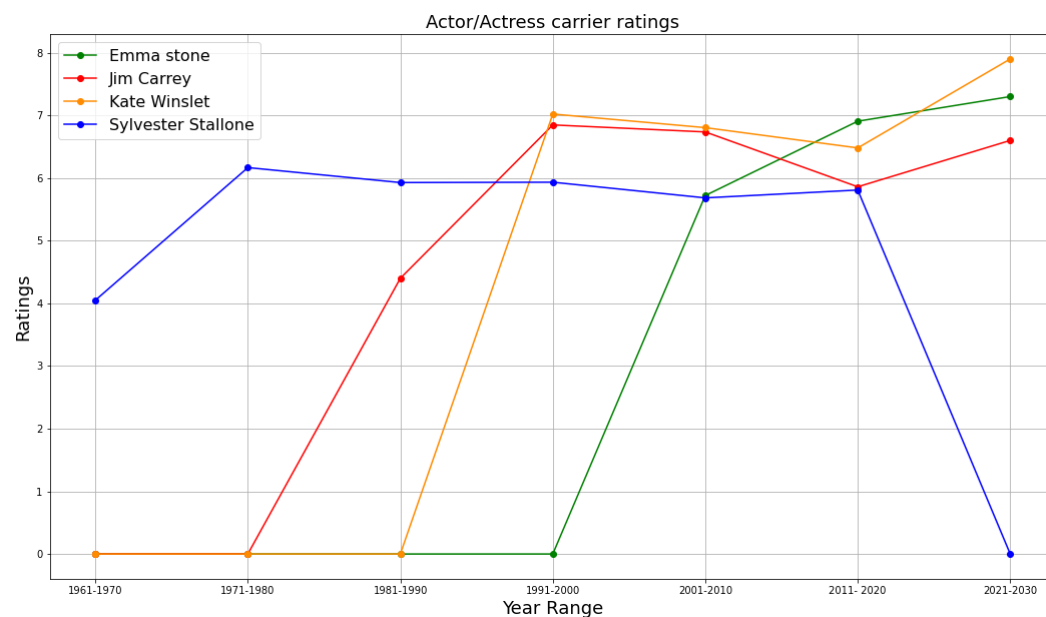
Analysis 1 C

Compared the acting career of Sylvester Stallone, Kate Winslet, Jim Carrey and Emma Stone for the 7 disjoint period using visualization.

Total number of movies each Actor/Actress acted during their career



Average rating of each Actor/Actress during their career



Sylvester Stallone

He had average ratings thorough out his career films; in 2006, he made a comeback and acted in "**Expendables**" in 2010, and since then, he had an average career growth.

Jim Carrey

He had a successful career as the lead in the comedy-drama film The Truman Show, for which he received a Golden Globe nomination. He transitioned from non-comedy films to different genres in 2018. He returned to his comedic style after 2018, making a comeback in the 2018 film Sonic the Hedgehog. His career average ratings in each decade was increasing till 1991-2000. Then it started decreasing.

Kate Winslet

She made tremendous growth in her career from 2008 to 2011 because to roles in movies like "**The Reader**", for which she received the Academy Award for Best Actress in a Leading Role. Her low career averages from 2012 to 2019 are a result of her personal issues and her difficulty finding excellent roles. Currently, she is attempting a comeback while having a mediocre career.

Emma stone

From 2012 to 2015, she had a successful career, including roles in the films Amazing Spiderman 1,2 as well as the critically acclaimed film '**Birdman**' for which she received an Academy Award nomination. She was at her peak in 2016 with her performance in "**La La Land**", for which she won an Academy Award, and in 2017 she was the highest-paid actress in the world, and Time magazine named her one of the world's 100 most influential people. As of 2022, she has a successful career with numerous prominent roles, including the box office hit Cruella, for which she received a Golden Globe nomination.

Analysis 2

The actor/actress who has done the maximum and minimum number of movies between 1977 and 1986 is found for this analysis. We chose a minimum of 5 movies to avoid listing a large number of actors/actresses. The outputs were sorted in ascending order by year. The actor/actress who appeared in the maximum and minimum number of films within each year from 1977 to 1986 has also been identified.

Analysis based on each year within year range 1977-1986

Maximum number of movies acted by Actor/Actress in each year.

Name	Year	Category	Movie's count
Richard Allan	1978	Actor	43
Zerrin Egeliler	1979	Actress	42

Richard Allan ([hyperlink](#)) was born in Marseille, France. He is a producer and an actor.

Zerrin Egeliler ([hyperlink](#)) is an actress born in Istanbul, Turkey.

Minimum number of movies acted by Actor/Actress in each year.

Name	Year	Category	Movie's count
A.K. Hangal	1981,	Actor	5
Adoor Bhasi	1986,	Actor	5
Aachi Manorama	1981,	Actress	5
Aiko Matsuoka	1986,	Actress	5

A.K. Hangal ([Hyperlink](#)) was a very late 30 years old, he made his Hindi film debut in 1946's "**Teesri Kasam**", directed by Basu Bhattacharya. He has since made numerous Hindi cinema appearances, often portraying the stereotypically humble and oppressed old man. He has played many notable characters in movies, most of them were good. He has so far made an appearance in over 125 films.

Adoor Bhasi ([Hyperlink](#)) was a director as well as an actor, with credits for "**Athirathram**" (1984), "**Raagam**" (1975), and "**Kodiyettam**" (1977). The 29th of March 1990 saw his passing.

Aachi Manorama ([Hyperlink](#)), a veteran actress from Tamil Nadu, has performed in more than 1500 films, 1000 plays, and a few television programs. She started her career as a drama artist, which was more typical at the time. Like most actors at the period, she made a gradual move from dramas to movies. When she first entered the film industry, she tended to concentrate more on comedies.

Aiko Matsuoka ([Hyperlink](#)) was not well-documented in the IMDb database; however, she has appeared in a total of 8 films. She appeared in 5 of the 8 films in 1986.

Analysis based on the entire year range 1977-1986

Total number of movies acted by Actor/Actress

Name	Category	Movie's count
George Estregan	actor	170
Kayoko Sugi	actress	194

George Estregan ([Hyperlink](#)) mostly acted in Philippines action movies. He won two further FAMAS Awards for Best Supporting actor for **"Kid Kaliwete"** (1978) and **"Lumakad Kang Hubad sa Mundong Ibabaw"** after being named FAMAS Best Actor for **"Sukdulan"** (1980). He was chosen for **"Magkayakap sa Magdamag's"** best supporting actor category (1986). Additionally, he was nominated by the Gawad Urian for Best Actor in the film Hostage: **"Hanapin si Batuigas"** (1977). He died in Santa Mesa, Manila, the Philippines, on August 8, 1988.

Kayoko Sugi ([Hyperlink](#)), an actress best known for her work in the films **"Maru hadaka seishun jigoku"** (1980), **"Yokoku bôkô: Yaru! Sasu"** (1980), and **"Uwaki no shikaku kankei"** (1983).

Minimum number of movies acted by Actor/Actress

Name	Category	Movie's count
Stephen Boyd	actor	3
Gisella Burinato	actress	3

Stephen Boyd ([Hyperlink](#)) is an Irish actor won the Golden Globe for Best Supporting Actor in a Motion Picture. His final three acting roles came from the three films he appeared in in 1977, the year he went away.

Gisella Burinato ([Hyperlink](#)), her performances in the films **"The Crypt Monster"** (2021), **"A Leap in the Dark"** (1980), **"Nel più alto dei cieliand"** (1977), and **"In the Highest of Skies"** (1977) are what people most commonly associate her with as an actress. Three of his four films were released between 1977 and 1986.

Problems Encountered

1. During pre-processing, we attempted to extract the top 4 films of the chosen actor or actress, but the "**knownfortitles**" attribute of the "**name_basics**" table contained the "**title ids**" of the top 4 films collectively. As a result, we had to manually enter the "**title ids**" of the top 4 films into another query to obtain the results we were looking for.
2. The movies that an actor or actress acted, does not just include those that were released in theatres; there were also some films that were released on "Tv and OTT". We overlooked it at first but later used **LIKE '%movie%'** along with the condition to solve the problem.
3. The goal of analysis 2 was to identify the actor or actress who appeared in the most films during the specified time period, which was between **1977** and **1984**. However, the sample output displayed there identified the actor or actress who appeared in the most films during each of those years, so we conducted analyses in both directions to achieve the desired results.

Conclusion

The main goal of project was to use SQL queries to do analysis on the IMDB database to learn more about the actors' and actresses' career progress. Sylvester Stallone, Jim Carrey, Kate Winslet, and Emma Stone were the actors and actresses analyzed. To help viewers better understand these actors' and actresses' performances, their acting careers have been separated into decades. Overall, Sylvester Stallone began his acting career first among the four celebrities selected, and he received an average rating throughout his career. Among these four stars, Emma Stone is currently the most paid actress. Her notoriety skyrocketed when the "**Amazing Spiderman-2**" movie was released in 2014. Her prime acting years were from 2011 through 2020. Despite the fact that they began their careers in different decades, Jim Carrey and Kate Winslet are both recognized for their roles in the film "**Eternal sunshine of spotless mind**". Visualizations of each actor appeared in each year, together with their career ratings, were used to explain the progression of each actor's career. **Richard Allan** appeared in 43 films in the year 1978, according to the study of the parameter file year range, which included the years 1977 to 1986, and **Zerrin Egeliler** appeared in 42 films in the following year, 1979. The actors **A.K. Hangal** and **Adoor Bhasi**, as well as the actresses **Aachi Manorama** and **Aiko Matsuoka**, have appeared only in 5 movies in different years within the year range 1977 to 1986. **George Estregan**, who appeared in 170 films between the years of 1977 and 1986, is the actor who appeared in the most films overall. The actress **Kayoko Sugi**, who has been in 194 films, is similar. The actor **Stephen Boyd** and actress **Gisella Burinato** was found acted in only 3 movies in the year range 1977 to 1986 from our analysis.