

DASC-5300-001-Foundation of Computing

Project-3: IMDb Database and Analysis

Overall Status:

We have a general knowledge of the project after carefully reading the problem statement and the project-3 description file. We have determined the initial goals for each team member's participation to the project and assigned timestamps for each assignment in order to stay on schedule. We were able to develop SQL queries that could retrieve the requisite rows of data for data analysis, which we then plotted and analyzed to identify certain trends and patterns.

File Descriptions:

- Project3_Analysis1.sql: This file contains the SQL used to get the results for Analysis1.
- Project3_Analysis2.sql: This file contains the SQL used to get the results for Analysis2.
- Project3_Analysis1_spool.txt: This file contains the output data for Analysis 1 together with the SQL queries.
- Project3_Analysis2_spool.txt: This file contains the output data for Analysis 2 together with the SQL queries.

Division of Labor:

We made sure that the assignments were divided equitably based on their difficulty and required amount of time. The project's coding and analysis components were each broken down into doable assignments with deadlines.

- **Rohith Kumar Nayakanti (1002024866)** – Performed the initial analysis together, designed queries for both the Analysis 1 and Analysis 2.
- **Sasank Kinnera (1001874178)** – Worked on analyzing the outputs of both the analysis after performing the initial analysis together. Derived patterns, and a final report was prepared.

It took about 15 hours to complete the initial understanding and coding phase. Over 15 hours were spent on the report's approach and analysis.

Problems Encountered:

- For Analysis 1(b), there was a problem incorporating disjoint periods in the same query, but we quickly conducted study and found a solution.

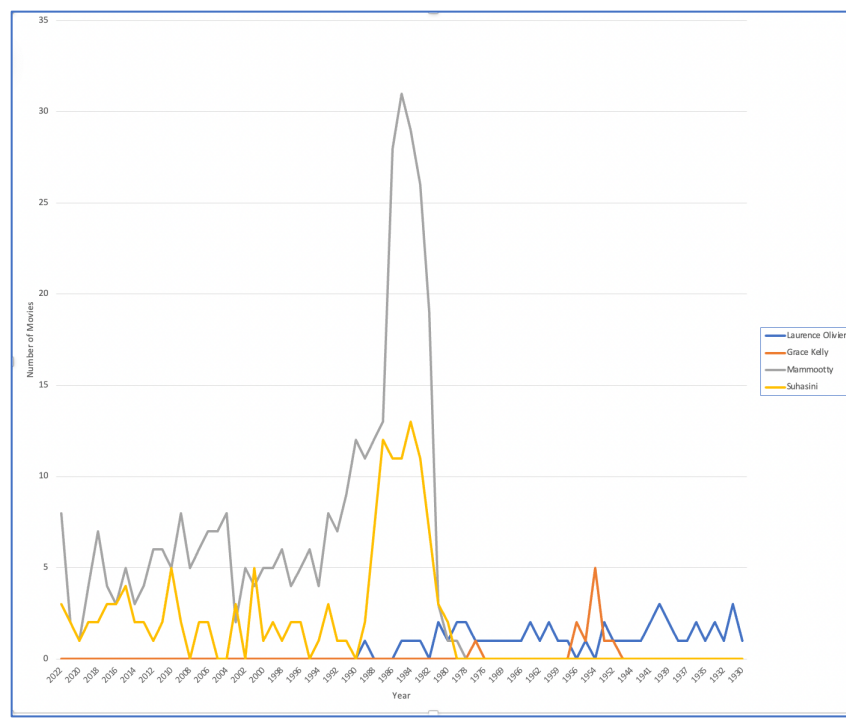
- For Analysis 2, there was a little problem when using the order by command for the year with a union function. The problem was eventually resolved by adding a temporary relation.

Analysis 1:

We were provided two names for this study (one actor and one actress, in the param file), and we chose our two personal favorites, one actor and one actress.

I. Analysis 1(a):

We have obtained the total number of movies made in each year by each actor/actress, who has been participating using the IMDb00 database.



Observations:

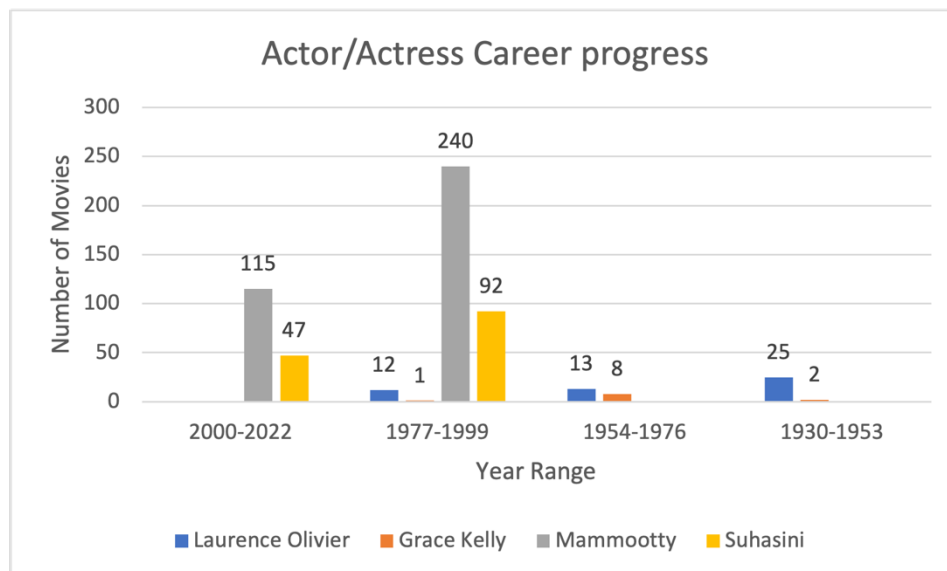
- From the above graph, it can be observed that the actor 'Mammooty' has acted highest number of movies in the year 1985.

Total Number of Movies			
Laurence Olivier	Grace Kelly	Mammooty	Suhasini
50	11	355	139

- Moreover, the actor 'Mammootty' has been on top over the other actors in the total number of movies each of them participated.
- It can be observed that, Mammootty and Suhasini had started their acting career in 1978 and have been continuing at present.
- On the other hand, the actor 'Laurence Olivier' had started his acting career in 1930's and been through till 1990.

II. Analysis 1(b):

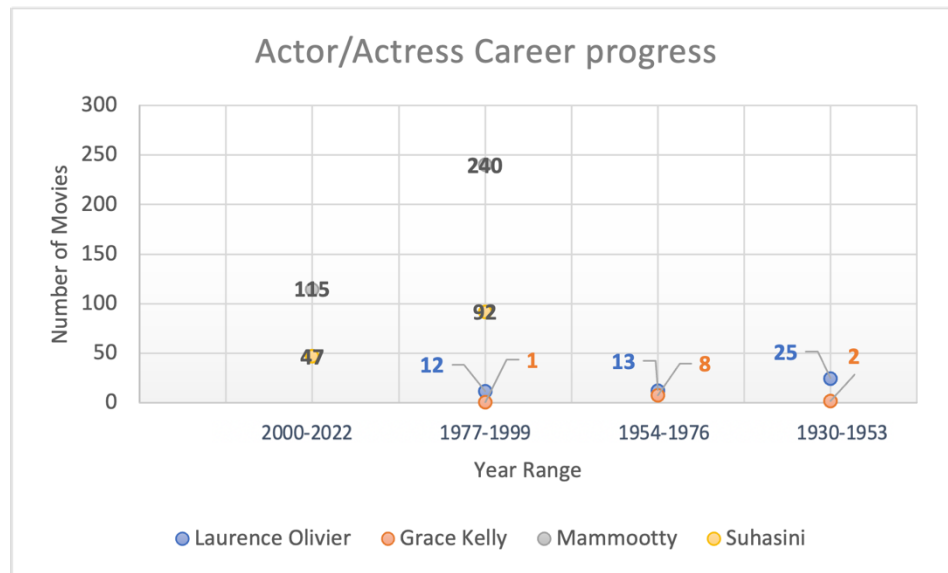
In this analysis, we separated each actor's or actress's performing career into four separate periods, each lasting 23 years, computed the total number of films in which they appeared throughout each time, and monitored and visualized their professional development.



Observations:

- The actor "Laurence Olivier" is seen to have been in the most films during the first period and to have been on a downhill trend.
- Grace Kelly, an actress began her acting career a bit slowly, increased the number of films she appeared in during the second period, and terminated it during the third with a single film.
- The other two actors began their careers in the years 1977-1999 and appeared in more films in the years that followed.

- III. **Analysis 1(c):** In this analysis, we used visualization to compare the actors and actresses selected for each phase of their acting careers.



Observations:

- From the above graph it can be observed that, some of the movies are overlapped for two or more participants and below is the list of common movies.

Actor	Actress	Common Movie	Year
Mammootty	Suhasini	Koodevide?	1983
Mammootty	Suhasini	Ente Upasana	1984
Mammootty	Suhasini	Pranamam	1986
Mammootty	Suhasini	Manivathoorile Ayiram Shivarathrikal	1987
Mammootty	Suhasini	Yatra	2018

- In the period 1977 – 1999, there were 4 movies overlapped by the actor Mammootty and actress Suhasini and 1 common movie had been released for the same participants in the period 2000 – 2022.

Analysis II:

For this analysis, we have been given a range of years in parameter file i.e., 1979-1988 and in that range, we need to find any actor/actress who have done the maximum number of movies and minimum number of movies.

Name-id	Name	Start Year	Movies	Comments	
nm0261738	George Estregan	1979	200	Max for Actor	
nm0993695	Kayoko Sugi	1979	149	Max for Actress	
nm0587944	Barry Miller	1979	4	First Actor with 4 Movies	For a total of 5553 records
nm3158332	Chakram	1988	4	Last Actor with 4 Movies	
nm8065719	Shahji Kale	1979	5	First Actor with 5 Movies	
nm0586347	Kan Mikami	1988	5	Last Actor with 5 Movies	
nm0856187	Laurent Terzieff	1979	4	First Actress with 4 Movies	For a total of 3581 records
nm0194564	Karol Cálík	1988	4	Last Actress with 4 Movies	
nm0509600	Li Wang	1979	5	First Actress with 5 Movies	
nm0699819	Uldis Pucitis	1988	5	Last Actress with 5 Movies	

Observations:

- We have received a total of 9134 records for the actor/actress who have done a total of at least 4 movies in the provided range 1979-1988.
- The maximum number of movies were performed by the actor 'George Estregan' in the given range
- The maximum number of movies were performed by the actress 'Kayoko Sugi' in the given range.

Conclusion:

We are able to understand the relational database and create queries for each analysis after conducting preliminary research on the provided data. We were able to undertake analysis and identify trends after successfully obtaining the needed outputs.