Analysis of the IMDB Database

Brief overview of how the project was approached:

The project was approached with no knowledge of SQL and since this project was all about SQL, I had to teach myself SQL through the lecture slides and also videos. We were given a lot of time for this project, so the first few weeks consisted of getting familiar with SQL, and the next few were allocated for testing the actual queries on the terminal.
The overall project did not take much time, and I was able to finish it quickly once I understood the assignment.

Division of Labor

This project was done by myself only, and it took me about two weeks to get this project done. The actual SQL part took about two weeks, and the report took a day.

Analysis 1:

Analysis 1.a

How was this achieved?

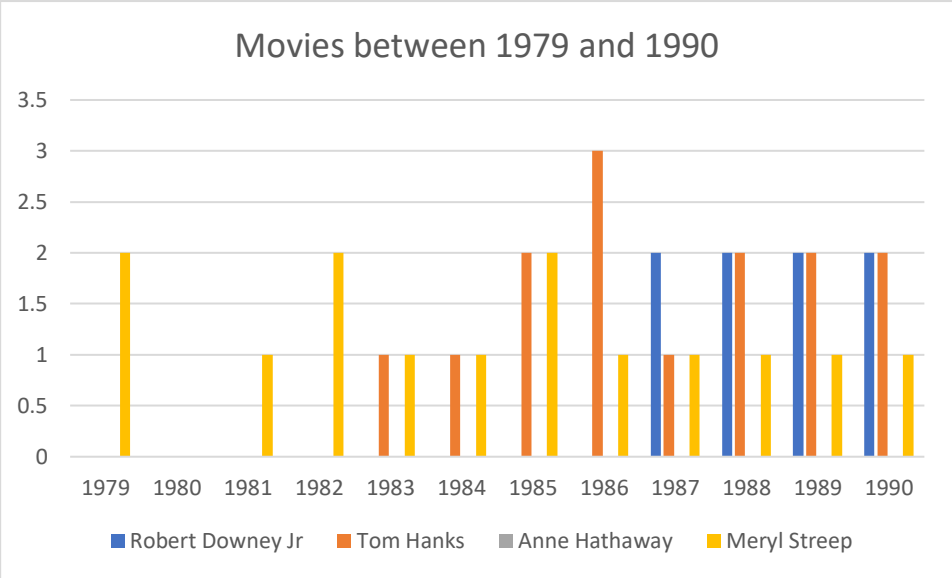
For the first analysis, we were given the code, but I did change it to include the actors and actresses that I was given in order to get my results. The actors I got were Robert Downey Jr and Tom Hanks and for my actress' I got Anne Hathaway and Meryl Streep.

Analysis 1.b

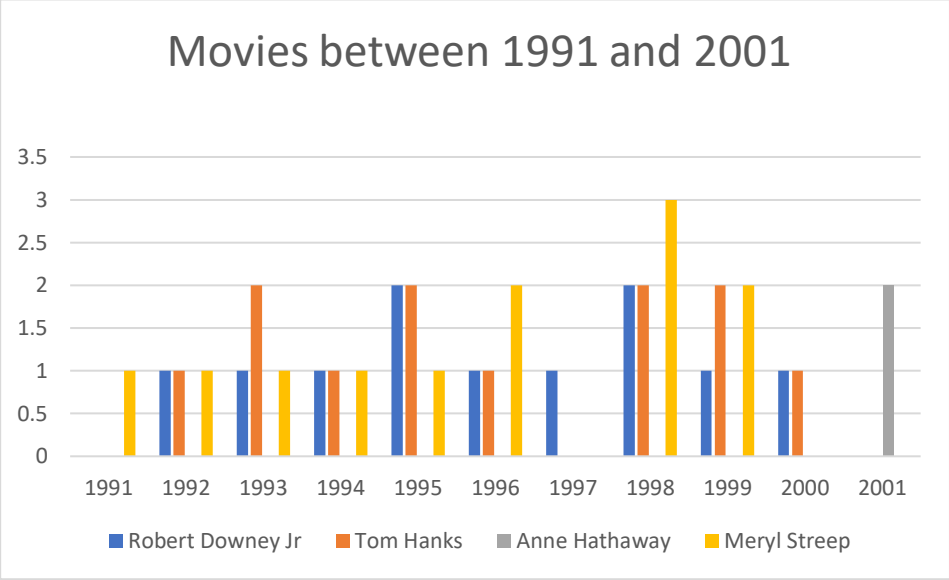
For this analysis, the same code was used but I had to divide the range of years into 4 disjoint periods and got the count of movies each actor and actress did based on those years.

Analysis 1.c

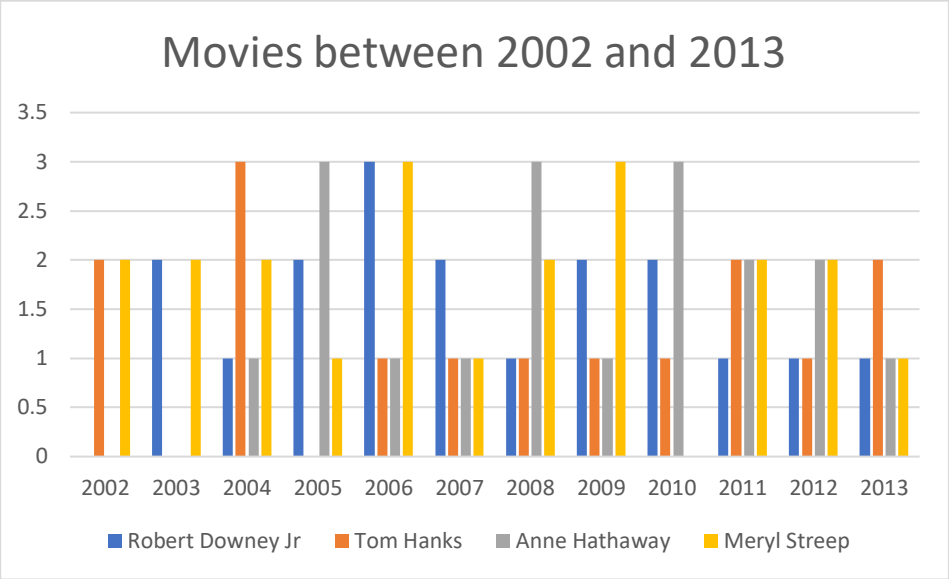
Based on the previous analysis, I did visualizations using Excel to represent each actor and actress' career progress.



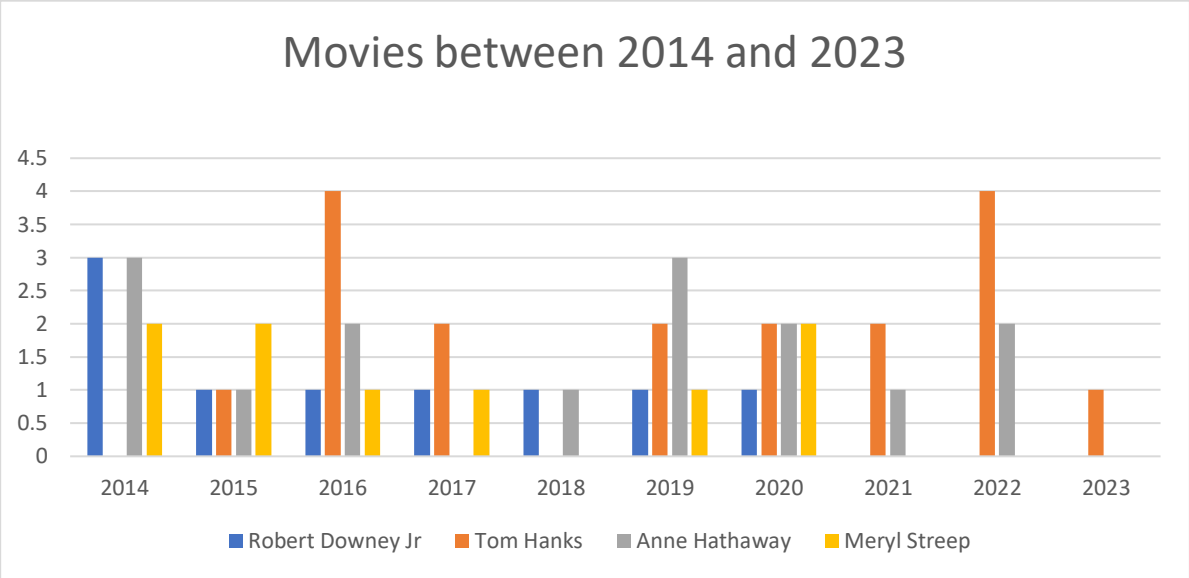
Based on the graph, for my first disjoint period of the range between 1979 and 1990, I can clearly see Meryl Streep did the most movies starting of 1979 till 1982, and that is because she was a popular actress of that time as compared to other actors who didn't start their careers until later on. Tom Hanks started to pick up pace in the year 1986 where he did 3 movies. Robert Downey was a newer actor who caught up with the other two after 1986 and did around 1 to 2 movies from 1987 till 1990. Anne Hathaway's first movie was in 2001 which explains why she isn't on this graph



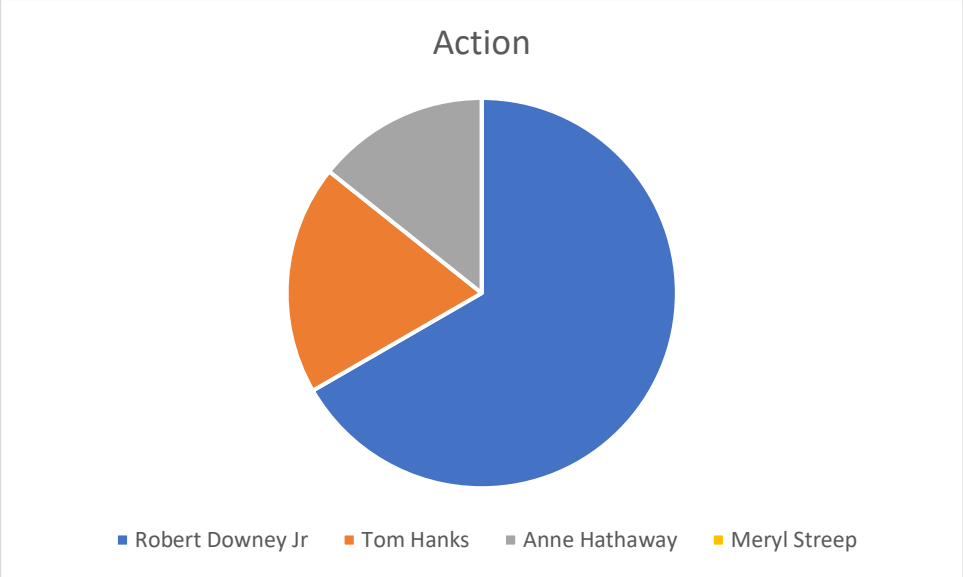
For my second disjoint period of the range between 1991 and 2001, I can clearly see Meryl Streep did the most movies in 1988. Whereas, Tom Hanks did around 1 to 2 movies every year until 2000. Robert Downey did a movie almost every year from 1992 till 2000. However, Anne Hathaway started acting in 2001, and that is shown in the graph as she did 2 movies in that year.



In the disjoint period of years between 2002 and 2013, there were more movies starred by Anne Hathaway. All actors made around 2 or 3 between in almost all the years.



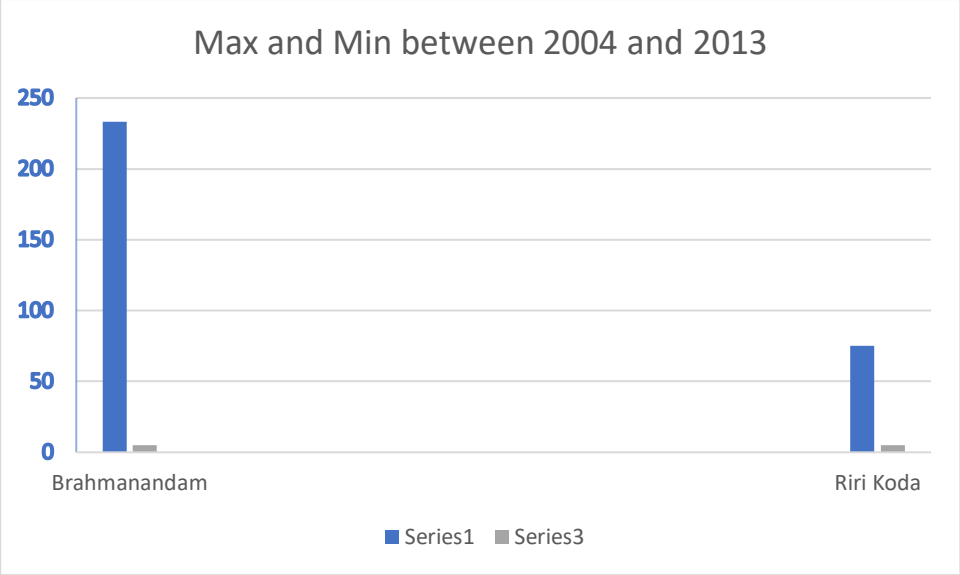
In my last disjoint period between 2014 and 2023, Tom Hanks made around 4 movies in 2016 and 2022. The others made around 2 to 3 almost every year except Meryl Streep and Robert Downey.



Based on all actors and actress's, the pie chart shows that Robert Downey Jr. did the most action movies, followed by Tom Hanks.

Analysis 2

For this analysis, we had to get the maximum and minimum number of movies any actor or actress had done in the range of years I was given which was between 2003 and 2014. The highest number of movies done by an actor was 233 by an actor called Brahmandam. The most movies done by an actress was 75 by someone called Riri Koda.



Overall, for my minimum number of actors and actresses I got a lot of rows, so I changed my query to show actors who had done more than 5 movies and used the fetch the top 10 rows to show my results. The most movies made by for my actors was an Indian actor called Brahmanadam and the highest movies for actress overall was Riri Koda; a Korean actress. For my number of actors in the minimum query, I initially got 1290 rows and for my actress' I got 854 rows.

The three problems I faced

1) The biggest issue was getting familiar with SQL as I had no prior knowledge of it and learning it from scratch within a few weeks was challenging.

2) The second issue was figuring out which software to use and which would help more with transferring of files and writing queries without having to retype them again and again

3) The last issue was getting the number of minimum number of movies for the second analysis as I got a lot of rows and that made it harder to get the final results.

Conclusion:

Overall, the project was confusing at first, but as each task was completed, it made more and more sense. Running the queries was a bit confusing at first, but once I got the hang of it, it made more sense and I was able to complete the project.

References:

https://www.hollywoodinsider.com/anne-hathaway-

https://www.slashfilm.com/1049392/the-14-best-robert-downey-jr-movies-ranked/

https://inshorts.com/en/news/brahmanandam-holds-record-for-maximum-movies-1468469657529