# DASC 5300 Foundation of Computing Project 3: Disk-based Data Analysis

**Overall Status:**

Before writing the queries, we understood the requirement given in the project description and started learning the basics of SQL like writing simple queries and using group by functions. When we were confident about the project understanding and writing queries, we started the project using SQL Developer and then started to analysis the result.

# File Descriptions:

# We used text file- imdb00-schema-and-content for understanding of tables name and attributes in the table. For Analysis 1a we used a text file- actor-year-freq and visualizes accordingly. Exported the files from SQL developer and excel sheet of all the data we got with clear representation in it.

# Division of Labor:

Both of us equally contributed to the project equally. We started learning how to write the queries together and started to write the queries in the project. We connected every single and discussed about the work we did and what work needs to be done further. After that we started with report. Complete project took us one week for understanding, implementation and report work.

# Problems Encountered and Handling:

1. Connection to omega server and establishing the connection to VPN we faced few issues. Later on reading the guidelines in the website and problem got fixed.
2. Writing join queries referring to the tables to fetch the analysis was quite difficult.
3. While doing the query for Analysis 2 felt little bit difficult. Selecting the required table and checking the suitable information that we need for analysis.

**Analysis 1A:**

Robin Williams(Actor)

* Actor given to our team is Robin Williams who is American actor and Comedian.
* Robin Williams acted 46 movies in total from 1980 to 2014.
* He acted total of 4 movies in 2006 which is highest in his total career.
* On an average he acted one movie per year.
* His primary profession is Actor, soundtrack, writer and nconst is nm0000245
* He is born in 1951 and started his acting career in 1980. That means at 29 years of age he started to act.

Additional analysis:

Robin is one of topmost actor in period 2006 acted 4 films in a year.

Robin acted 46 movies and he is also standup comedian. In that list top rated film of Robin is

|  |  |  |
| --- | --- | --- |
| **Primary title** | **Top Rating** | **Start year** |
| The Night of 1,000 Wogans | 9.6 | 1991 |

Reese Witherspoon(Actress)

* Actress given to our team is Reese Witherspoon, who is American actress
* Reese Witherspoon acted 35 movies in total from 1991 to till date. She is presently acting two movies where one movie is released in 2023 and one in 2025, same is represented in graph.
* She acted total of 3 movies in 1999, which is highest in her total career.
* On an average, she acted one movie per year.
* Her primary profession is Actor, soundtrack, writer and nconst is nm0000702
* She is born in 1976 and started her acting career in 1991. That means at 15 years of age she started to act.

Additional analysis:

Reese Witherspoon is one of topmost actor in period 1999 acted 3 films in a year.

Reese Witherspoon acted 35 movies. In that list top rated film of Reese Witherspoon is

|  |  |  |
| --- | --- | --- |
| **Primary title** | **Rating** | **Start year** |
| The Interview | 9.6 | 2019 |

Favorite Actor: Venkatesh

* Our favorite actor is Venkatesh Daggubati who is Indian actor and Comedian.
* Venkatesh Daggubati acted 85 movies in total from 1986 to 2022.
* He acted total of 7 movies in 1996 which is highest in his total career.
* On an average he acted 3 movie per year.
* His primary profession is Actor, producer and nconst is nm0893142
* He is born in 1960 and started his acting career in 1986. That means at 26 years of age he started to act.

Additional analysis:

Venkatesh Daggubati is one of the top Telugu actors acted 7 movies only in one year.

Venkatesh Daggubati acted 85 movies. In that list top rated film of Venkatesh Daggubati is

|  |  |  |
| --- | --- | --- |
| **Primary title** | **Rating** | **Start year** |
| Nuvvu Naaku Nachchav | 8.8 | 2001 |

Favorite Actress: Soundarya

* Our favorite actress is Soundarya, who is Indian actress, Producer
* Soundarya acted 87 movies in total from 1993 to 2004.
* She acted total of 11 movies in year 1995 and 1996, which is highest and peek years of her total career.
* On an average, she acted 7 movie per year.
* Her primary profession is Actor, producer and nconst is nm0815897
* She is born in 1976 and started her acting career in 1993. That means at 18 years of age she started to act.
* Soundarya died in 2004 in a plane crash.

Additional analysis:

Soundarya was one of the top Telugu actresses acted 11 movies only in one year. Her acting span is very less but acted many movies per year

Soundarya acted 87 movies. In that list top rated film of Soundarya are

|  |  |  |
| --- | --- | --- |
| **Primary title** | **Rating** | **Start year** |
| Padaiyappa | 8.2 | 1999 |
| Aapthamitra | 8.2 | 2004 |

**Analysis 1B:**

* Robin started his career in 1980 and he got good opportunities and could act one movie per year initially.
* His career graph gradually increased yearly and started acting two movies per year and started becoming as successful actor.
* At period of 1998 to 2002 is his peek career period where he acted about three movies in a year and received awards for his movies and at the same time continued as comedian.
* Later, from 2004, career was not good. Gradual decrease of opportunities and in 2014 he died.
* Reese Witherspoon started her career as actress in 1991 and acted one movie per year.
* At very initial stage of her career only she got good opportunities and acted like 3 movies in a year.
* Then gradually count of number of movies acted became less, but then she is still acting.

**Analysis 1C:**

* This graph is the comparison of actor and actress’s career period from 1980 to till date.
* Robin is elder than Reese and started to act in 1980 but Reese in 1991.
* Robin acted more number of movies when compared to Reese till date and in 2006 robin acted 4 movies but Reese highest number of movies is only 3 in a year.
* Reese is still in acting field whereas Robin died in 2014 only.
* Overall, Robin acted 46 movies in span of 34 year and continued as standup comedian, which is great number and good career progress.
* Reese acted 35 movies in the span of 31 years and relatively a good number.
* From 2019 to 2021, no movies of her are released may be because of Covid.

**Analysis 2:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | year | Category | Movie count | Rank value |
| Tom Byron | 1990 | actor | 63 | 1 |
| Sharon Kane | 1990 | actress | 35 | 1 |
| Eleazar García | 1984 | actor | 4 | 2260 |
| Eric Ryan | 1984 | actor | 4 | 2260 |
| Eric Tsang | 1984 | actor | 4 | 2260 |
| Farooq Shaikh | 1984 | actor | 4 | 2260 |
| Feng Ku | 1984 | actor | 4 | 2260 |
| Fernando Poe Jr. | 1984 | actor | 4 | 2260 |
| Fred Williamson | 1984 | actor | 4 | 2260 |
| George Rudy | 1984 | actor | 4 | 2260 |

* Highest number of movies acted in the given range of years is Tom Byron with the count of 63 movies and rank value is 1, in the category of Actors
* Highest number of movies acted in the given range of years is Sharon Kane with the count of 35 movies and rank value is 1, in the category of Actress
* We considered minimum movie count as 4 and there is huge list, so attached the excel in file and here are few samples of the result
* In the given span we could see the rank value of about 2260 and acted only 4 movies in complete period for all the actors.
* Tom Byron and Sharon Kane are very famous actor and actress and received many awards for their acting and production.
* We got total count of 303 actors who acted minimum number of movies in the span of 1984 to 1993.
* Even the rating is very high in IMDB database. Most of them are unrecognized actors and actress in these results.

**Conclusion:**

This project helped us have clear understanding of Relational database and how to write the queries and subquery .We could analyze the actors and actress information like period of acting, ratings, number of movies, and comparison between actors etc. through this database.