Name: Taher Paratha
ID: 1001979730

DASC 5300 Project 3 Report

1. Overview

This project was perfect to get clear understanding of sql queries and how to use it to get the required data. I was able to connect to real online database server and query data and perform analysis on it, just like in real world cases.

2. File Descriptions

i. **Project3.ipynb**: This is the python notebook file created containing code for the queries written and analysis done.
ii. **Proj_3.sql:** sql file containing just the queries for the data acquisition.
iii. **Analysis1.csv**: analysis 1 queried data saved to csv.
iv. **Analysis1.png**: analysis 1 visualisation.
v. **Analysis2.csv**: analysis 2 queried data saved to csv.
vi. **Analysis3a.csv**: analysis 3 queried data saved to csv for actor.
vii. **Analysis3a.jpg**: analysis 3 visualisation for actor.
viii. **Analysis3b.csv**: analysis 3 queried data saved to csv for actress.
ix. **Analysis3b.csv**: analysis 3 visualisation for actress.

3. Division of Labour

Since I was the only one in the team, I completed the project alone. I divided the work among myself, that is, by creating milestones and assigning time to each section. I was able to complete the project in timely manner, by dividing the work into milestones.
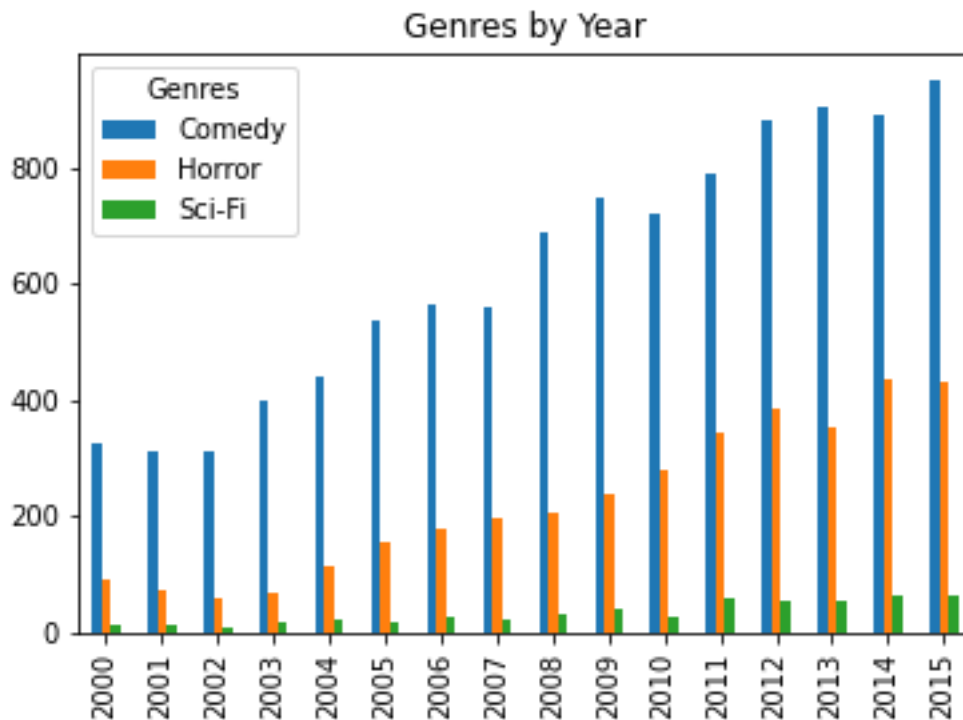
4. Problems Encountered

a. Connecting to Oracle and omega was difficult, but with some research and IT support I was able to do it.
b. Understanding the table was difficult since I couldn't open a csv file to look at the data. But I understood that this is how it is in the real world. The table schemas provided in the project description really helped.
c. Handling such huge amount of data was tricky. Writing a wrong query and not limiting the result would lead to continuous flow of results leading to crash. Dr. Sharma's tip really help "Fetch top 10 rows only".

5. Analysis

i. Analysis 1:

The query was written to get data on movies titletype from year 2000 to 2015 on genres Comedy, Horror, and Sci-fi. As we can see from the plot below, all genres have an upward trend in number of movies produced. Comedy out of all three have the highest increase.

Comedy has the highest increase because comedy is a more digestible and lighthearted genre which anyone can enjoy in any mood. Therefore, more and more people prefer a comedy movie, and hence more and more movies are made in this genre since it has a higher probability of being successful. Horror is also a very liked genre and people who like horror movies usually watch it for the adrenaline rush and therefore doesn't matter if the movie is actually good or not. Therefore, lots of movies are made in this genre too. Sci-fi is a niche genre and the people who like it prefer high quality movie with high graphics and visuals which is costly to produce (Eg. Dune, Interstellar, etc) and therefore not many movies are made in this genre.
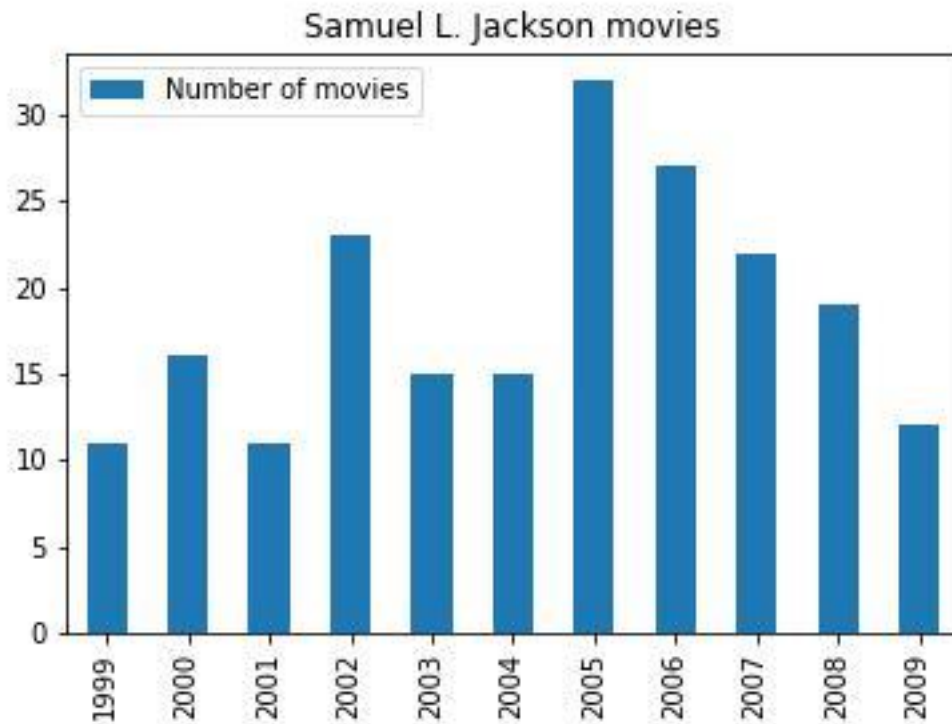
ii. Analysis 2:

For the second analysis I found out all the movies where the director was also an actor in the movie. This query required the condition that the Nconst in the name_title_mapping and the name code in the director column of Title_crew_dir must be same. Hence it gave us only movies where actor and director were the same person.

It gives us movies like spaceballs, Airplane!, Life of Brian etc. This table is saved to csv file "Analysis2.csv"

iii. Analysis 3:

For this part I chose Samuel L. Jackson as the actor and Jennifer Aniston as the actor and actress. I queried their Nconst and found out all the movies they were involved in from 1999 to 2009. I found the following result.

We can see that for Samuel L Jackson, the movies keep increasing as the year increase and peaks in 2005. This is because he was in his prime and getting better and better recognition for his acting. After 2005, the movies reduce since his age has also started to increase and after working in such high amount of movies, he also must have started to limit movies that he accepts due to age.

For Jennifer Aniston, she became famous after she started acting in the hit TV series Friends. The show started in 1994 and ended in 2004. This can be seen in the plot below where she is involved in huge amount of production till 2004 and after that the number of movies reduce and hitting a huge low in 2007. As her age also increase, this could be another reason for the decrease as the years go by.

Jennifer Aniston movies