**Final Project Report - INSY 5336 Python Programming Spring 2021**

**Name: Nitish Prabhu Kota | Student ID: 1001738851**

In this project, I created a Python program that pulls movie information from Metacritic's "Movies of All Time" page for the top 500 most popular films.

A python script has been written to collect movie information for the top 500 films from each website and store it in a comma separated file. The data is also saved in a SQLite database called MovieInfoDatabase, which is located in the directory where the Jupyter Notebook code is run. MovieInfoTable is a table in the MovieInfoDatabase. The movie information is collected by looping through all the 5 pages of the movies and store the names of the movies in the list, similarly we do this for director list, cast list and principal cast list.

The program can provide the movie details like director or cast names on entering the movie name. Analyzed how many times has each actor/actress appeared in these top 500 movies, analyze how many times has each director appeared in these top 500 movies.

We can calculate the success factor of the directors by calculating the number of movies directed by them, A director is more successful if he/she has directed more movies and similarly a director is least successful if he/she has directed least movies. Similarly, we can calculate the success factor of the actors by calculating the number of movies acted by them, An actor is more successful if he/she has acted in more movies and vice versa.

**From our analysis, we have the following data.**

**Total number of director(s) who are in the Top 500 movie list: 411.**

**Total number of cast(s) who are in the Top 500 movie list: 6503.**

**Total Count of director(s) who have directed 6 or more than 6 movies: 4. 407 directors have directed less than 6 movies.**

**Total Count of cast(s) who have acted in 5 or more than 5 movies: 20. 6483 cast(s) have acted in less than 5 movies.**

Using the cosine similarity, (metric used to measure how similar the documents), we figured out the similarity between different directors, different casts. High similarity will result in a value close to 1 and a low similarity would result in a similarity value of 0.

Also using a list of favorite actors, we found a cosine similarity amongst the different combinations between them, which signifies the similarity between the casts based on the supporting actors that they have worked with.

Top 5 most popular supporting actor/actress among them all are: Rudy bond, Kim hunter, Lee j. cobb, V leigh and Richard s. castellano. (for the set of actors considered)

Based on the analysis, we can see that there are very few directors out of 411, who have directed 6 or more than 6 movies i.e., 4 directors. More the movies directed by a director, more the count of actors that he/she has worked with. And we can also understand that the director has directed more movies than others.

Similarly with the cast, out of 6503 actors, there are only 20 actors who has acted in 5 or more than 5 movies. More the movies acted in by a actor, more the count of sub actors that he/she has worked with. And we can also understand that the actor has acted in more movies than others.

Most of the directors and actors ( around 60 to 80%) have directed/acted in 1 or 2 movies and are still in the beginning phase of their career.