# Movie Recommended Engine:

The Data set link: <https://www.kaggle.com/zeeshanmulla/recommendation-system-movie>

# About this Dataset:

The dataset contains 2 files, one file name is Recommendation movie and the file extension is .csv. The second file is Movie title and this file extension is .text. The recommendation movie file has a 100003 record and 4 columns. There are four columns in recommendation file and each column names are user\_id, item\_id, rating and timestamp. In the first column, each user has a unique id. The second column is item\_id each movie has a unique item\_id. The third column is rating in this column each user gives a movie rating between (1.0 – 5.0). The Fourth column name is the timestamp in this column the numbers tell how many seconds are in each movie. The second file has a 1682 record and 2 columns. The first column name is item\_id, in this column each movie has a unique item\_id. The second column name is a movie title, in this column, each movie has a unique title.

# Problem Statement:

I'm watching one movie, after watching this movie which movies I should watch based on past behaviour? After combining these two files I want to make a movie recommendation system using collaborative filtering.

# What is Collaborative filtering method:

Collaborative filtering methods can be either user-based, or item-based, and work with user behaviour to analyse other users' preferences, then make recommendations to the new user.

# Summary:

Work with user behaviours to analyse other user's preferences, then make a robust movie recommendation to the new user.

# Technology:

Programming Language: Python 3.7.

Software: Jupyter notebook.

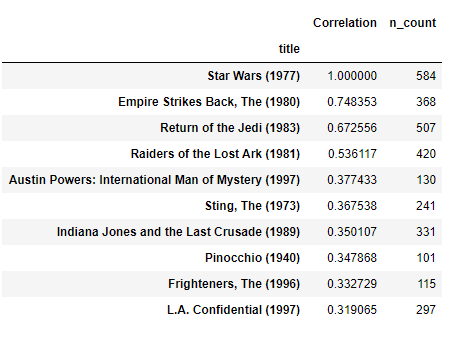
Dataset Name: Movie Recommended Engine (from Kaggle).

Library use in python: Pandas, NumPy, Matplotlib.

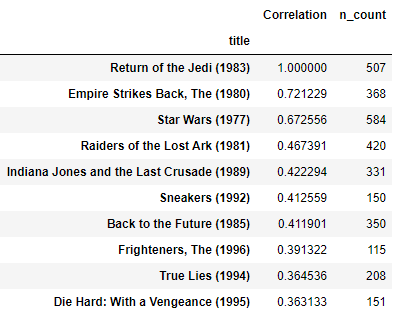
# Explanation:

I used a python programming language to perform a data analysis task in this data set. I have use pandas and NumPy libraries to perform some data analysis tasks such as descriptive statistics of data, data-type of the features, cleaning of the data, merging two files based on common column names in these datasets. After completing the data analysis tasks next task was visualization of the data. Plot some features to understand the insight into the data. After completing the visualization task next task is to make a pivot table, in the pivot table we would give some hyperparameters like index, column, and values to understand the relation between User Id and movie title. Also, users have given a rating of some movies. After making a pivot table, the next step is to find a recommend movies based on users rating behaviours. Example after watching Star Wars movie which movies I should watch based on a user's rating behaviour so, I would make a recommendation system based on the Collaborative filtering method. So, I am considered one movie whose name is Star Wars. First step I should make a correlation table with other movies (how other movies correlate with Star Wars movie). After making a correlation table, I get all movies with correlation values which are correlated with the Star Wars movie. Also, I have made a data frame which stores all movie title and its correlation value. Next step is to remove all the null value in the data frame because some movies does not make a correlation with Star War movie. After removing null values, the next task is to add the user's total ratings for each movie in the data frame. In the last step, after watching the Star Wars movie I should watch only those movies which make the highest correlation with Star Wars movie and the number of ratings is greater than 100.

# Output:



**(Top 9 movie after watching Star Wars (1977) movie)**



**(Top 9 movie after watching Returns of Jedi (1983) movie)**

# Output Explanation:

In the first picture, after watching the Star Wars movie, I would recommend the top 9 movies after Star Wars movie.

In the second picture, after watching the Returns of Jedi (1983), I would recommend the top 9 movies after Returns of Jedi movie.