

Q1: (10 Points) Perform exploratory data analysis (EDA) over the shared dataset. Analysis may include 1) finding frequently occurring values in categorical features, 2) Counting of Nan values per column, 3) Plotting correlations between features, etc.. Report at least three insights about the dataset.

Checking Nan values in all the 4 data sets -

For ratings dataset -

```
0 0
1 0
2 0
3 0
dtype: int64
```

For tags dataset -

```
0 0
1 0
2 0
3 0
dtype: int64
```

For links dataset -

```
0 0
1 0
2 8
dtype: int64
```

For movies dataset -

```
0 0
1 0
2 0
dtype: int64
```

Frequently occurring values in categorical dataset :

1) Frequency of categorical values of genre in movie dataset :

```
(no genres listed) 34
```

Action	1828
Adventure	1263
Animation	611
Children	664
Comedy	3756
Crime	1199
Documentary	440
Drama	4361
Fantasy	779
Film-Noir	87
Horror	978
IMAX	158
Musical	334
Mystery	573
Romance	1596
Sci-Fi	980
Thriller	1894
War	382
Western	167

Insights about each datasets:

- **Movie dataset**

	0	1	2
count	9742	9742	9742
unique	9742	9737	951
top	130083	War of the Worlds (2005)	Drama
freq	1	2	1053

- **Frequency dataset:**

	0	1	2	3
count	100837	100837	100837	100837
unique	611	9725	11	85044
top	414	356	4.0	1459787998
freq	2698	329	26818	128

- Ratings dataset:

	0	1	2	3
count	100837	100837	100837	100837
unique	611	9725	11	85044
top	414	356	4.0	1459787998
freq	2698	329	26818	128

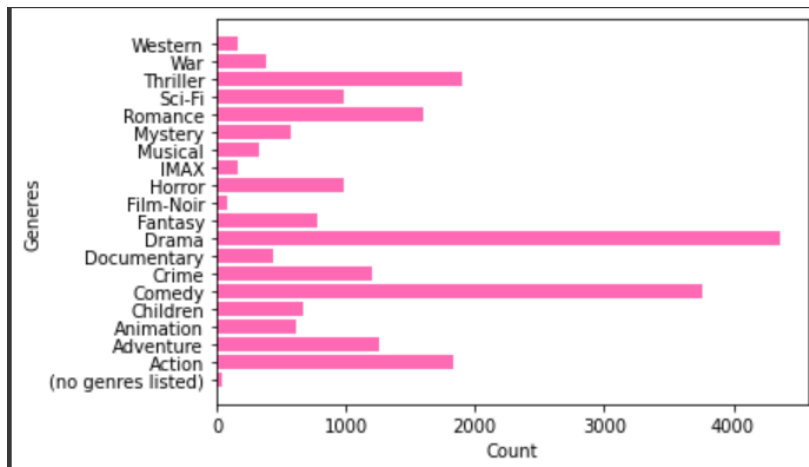
- Tags dataset:

	0	1	2	3
count	3684	3684	3684	3684
unique	59	1573	1590	3412
top	474	296	In Netflix queue	1138307125
freq	1507	181	131	5

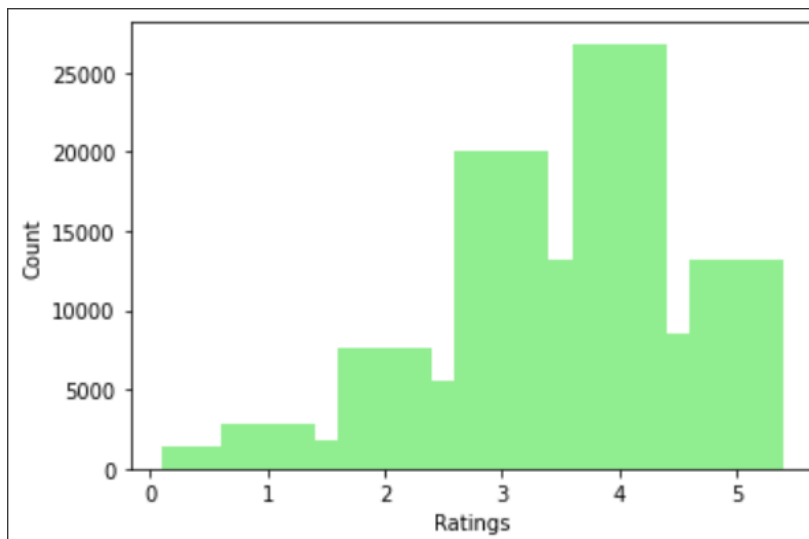
Note - There were no missing values present in the dataset so no need to handle it

Plotting correlations between features

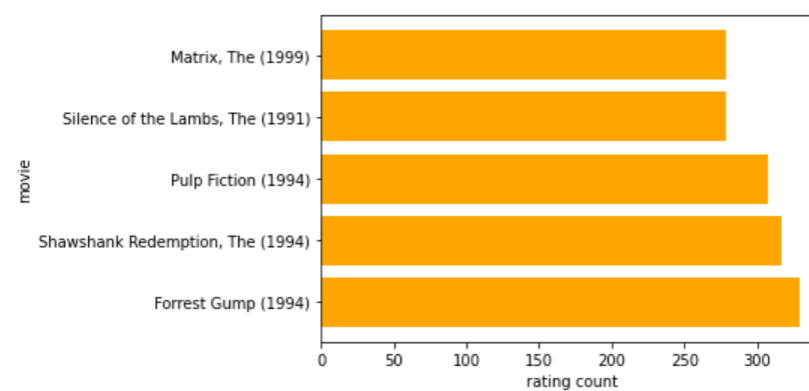
- Common genres



- Ratings given by people



- Top 5 most Rated movies



Three insights about dataset -

- The Mostly watched genre present in dataset is drama
- Mostly Rating given by audience to the movies is 4 so the dataset contains movies which are generally liked by audience
- The Most rated movie is Forrest Gump (1994) which is liked by the audience

Q2: (60 points) Using the association rule mining techniques, build a recommendation system for the dataset provided to you. Your system should be able to recommend movies given a customer profile. I.e. What set of movies a customer has watched in the past? The set can contain any non-zero number of movies. Let us say a customer has seen "avengers, iron man", the system can recommend the following movies "captain america, hulk, thor, doctor strange". The final scores will be computed using the metric "Precision@k". For a given customer profile, you will recommend four best movies, i.e. k=4.

Assumption :

1. All the dataset are stored in drive
 2. Model is too big so its uploaded on drive its link - https://drive.google.com/file/d/1-D5GAPp69W2_GRWlvf4c4QG8zY8MDDJt/view?usp=sharing
 3. If we dont get any recommendation we provide the top rated movies in that genre as recommendation
 4. output is saved in drive
- Firstly changed the movies dataset i.e added columns of each genre and added 0 or 1 according to the movies - if the movies is of a particular genre or not.
 - Then merged the movies and ratings dataset on the basis of movie_id
 - Dropped the duplicate columns and duplicate rows
 - Then reshaped the data using index as User_Id (rows) and columns as movie titles
 - Converted the values to 0 or 1 depending on whether the user has watched the movie or not
 - Used the apriori algorithm of mlexend library and trained the model using association rule of mlexend library.
 - Then sorted the order of recommendation on the basis of lift.
 - Then returned the suggested movies according to the input

Recommendation for Godfather, The (1972) :

1. Godfather: Part II, The (1974)
2. Monty Python and the Holy Grail (1975)

3. Star Wars: Episode IV - A New Hope (1977)
4. Schindler's List (1993)

Q3: (30 points) Visualize the maximal frequent pattern set.

Assumption:

- Dataset used is the merged dataset of frequency and merged

This is the visualization of our maximal frequency of our merged dataset. Merged dataset contains the merge of movie and rating dataset

```
[frozenset({'7'}),  
frozenset({'15'}),  
frozenset({'17'}),  
frozenset({'27'}),  
frozenset({'33'}),  
frozenset({'40'}),  
frozenset({'43'}),  
frozenset({'63'}),  
frozenset({'73'}),  
frozenset({'76'}),  
frozenset({'82'}),  
frozenset({'89'}),  
frozenset({'93'}),  
frozenset({'112'}),  
frozenset({'119'}),  
frozenset({'137'}),  
frozenset({'141'}),  
frozenset({'144'}),  
frozenset({'151'}),  
frozenset({'159'}),  
frozenset({'166'}),  
frozenset({'167'}),  
frozenset({'169'}),  
frozenset({'201'}),  
frozenset({'216'}),  
frozenset({'219'})]
```

```
frozenset({'220'}),
frozenset({'233'}),
frozenset({'234'}),
frozenset({'239'}),
frozenset({'240'}),
frozenset({'247'}),
frozenset({'254'}),
frozenset({'266'}),
frozenset({'279'}),
frozenset({'280'}),
frozenset({'282'}),
frozenset({'314'}),
frozenset({'322'}),
frozenset({'328'}),
frozenset({'330'}),
frozenset({'332'}),
frozenset({'334'}),
frozenset({'337'}),
frozenset({'353'}),
frozenset({'372'}),
frozenset({'373'}),
frozenset({'385'}),
frozenset({'411'}),
frozenset({'412'}),
frozenset({'420'}),
frozenset({'432'}),
frozenset({'436'}),
frozenset({'488'}),
frozenset({'490'}),
frozenset({'492'}),
frozenset({'517'}),
frozenset({'522'}),
frozenset({'524'}),
frozenset({'559'}),
frozenset({'562'}),
frozenset({'567'}),
frozenset({'570'}),
frozenset({'572'}),
frozenset({'601'}),
frozenset({'605'}),
frozenset({'607'}),
frozenset({'58'}),
frozenset({'100'}),
frozenset({'117'}),
frozenset({'308'}),
frozenset({'410'}),
frozenset({'552'}),
```

```
frozenset({'594'}),
frozenset({'11'}),
frozenset({'23'}),
frozenset({'24'}),
frozenset({'109'}),
frozenset({'181'}),
frozenset({'195'}),
frozenset({'286'}),
frozenset({'313'}),
frozenset({'317'}),
frozenset({'362'}),
frozenset({'437'}),
frozenset({'553'}),
frozenset({'577'}),
frozenset({'602'}),
frozenset({'4'}),
frozenset({'16'}),
frozenset({'41'}),
frozenset({'129'}),
frozenset({'136'}),
frozenset({'187'}),
frozenset({'198'}),
frozenset({'204'}),
frozenset({'222'}),
frozenset({'227'}),
frozenset({'230'}),
frozenset({'268'}),
frozenset({'301'}),
frozenset({'326'}),
frozenset({'346'}),
frozenset({'351'}),
frozenset({'352'}),
frozenset({'354'}),
frozenset({'369'}),
frozenset({'370'}),
frozenset({'415'}),
frozenset({'419'}),
frozenset({'424'}),
frozenset({'464'}),
frozenset({'542'}),
frozenset({'551'}),
frozenset({'29'}),
frozenset({'80'}),
frozenset({'215'}),
frozenset({'246'}),
frozenset({'287'}),
frozenset({'309'}),
```



```
frozenset({'331'}),
frozenset({'435'}),
frozenset({'482'}),
frozenset({'495'}),
frozenset({'510'}),
frozenset({'70'}),
frozenset({'361'}),
frozenset({'405'}),
frozenset({'428'}),
frozenset({'571'}),
frozenset({'59'}),
frozenset({'83'}),
frozenset({'115'}),
frozenset({'235'}),
frozenset({'265'}),
frozenset({'376'}),
frozenset({'393'}),
frozenset({'479'}),
frozenset({'256'}),
frozenset({'22'}),
frozenset({'365'}),
frozenset({'225'}),
frozenset({'409'}),
frozenset({'25'}),
frozenset({'543'}),
frozenset({'520'}),
frozenset({'52'}),
frozenset({'95'}),
frozenset({'113'}),
frozenset({'139'}),
frozenset({'210'}),
frozenset({'212'}),
frozenset({'408'}),
frozenset({'466'}),
frozenset({'475'}),
frozenset({'74'}),
frozenset({'168'}),
frozenset({'465'}),
frozenset({'300'}),
frozenset({'563'}),
frozenset({'564'}),
frozenset({'20'}),
frozenset({'442'}),
frozenset({'36'}),
frozenset({'440'}),
frozenset({'184'}),
frozenset({'306'}),
```

```
frozenset({'1', '4.0'}),
frozenset({'1', '5.0'}),
frozenset({'1', 'Toy Story (1995)'}),
frozenset({'18', '3.5'}),
frozenset({'18', '4.0'}),
frozenset({'19', '3.0'}),
frozenset({'19', '2.0'}),
frozenset({'21', '3.5'}),
frozenset({'21', '4.0'}),
frozenset({'4.0', '45'}),
frozenset({'45', '5.0'}),
frozenset({'3.0', '45'}),
frozenset({'3.0', '50'}),
frozenset({'3.0', '57'}),
frozenset({'4.0', '57'}),
frozenset({'4.0', '64'}),
frozenset({'3.5', '64'}),
frozenset({'4.0', '66'}),
frozenset({'2.5', '68'}),
frozenset({'2.0', '68'}),
frozenset({'4.0', '68'}),
frozenset({'3.0', '68'}),
frozenset({'4.5', '68'}),
frozenset({'3.5', '68'}),
frozenset({'4.0', '91'}),
frozenset({'3.0', '91'}),
frozenset({'103', '4.0'}),
frozenset({'132', '3.0'}),
frozenset({'135', '4.0'}),
frozenset({'140', '3.0'}),
frozenset({'140', '4.0'}),
frozenset({'156', '4.0'}),
frozenset({'156', '3.0'}),
frozenset({'1.0', '160'}),
frozenset({'177', '4.0'}),
frozenset({'177', '2.5'}),
frozenset({'177', '3.0'}),
frozenset({'177', '3.5'}),
frozenset({'182', '4.0'}),
frozenset({'182', '4.5'}),
frozenset({'182', '3.5'}),
frozenset({'182', '3.0'}),
frozenset({'186', '4.0'}),
frozenset({'200', '4.0'}),
frozenset({'202', '4.0'}),
frozenset({'202', '3.0'}),
frozenset({'2.0', '217'})
```

```
frozenset({'217', '3.0'}),
frozenset({'226', '4.0'}),
frozenset({'232', '3.5'}),
frozenset({'232', '3.0'}),
frozenset({'232', '4.0'}),
frozenset({'2.5', '232'}),
frozenset({'249', '4.0'}),
frozenset({'249', '3.0'}),
frozenset({'249', '4.5'}),
frozenset({'249', '3.5'}),
frozenset({'263', '4.0'}),
frozenset({'274', '4.0'}),
frozenset({'274', '3.5'}),
frozenset({'274', '3.0'}),
frozenset({'2.5', '274'}),
frozenset({'2.0', '274'}),
frozenset({'275', '5.0'}),
frozenset({'275', '4.0'}),
frozenset({'288', '4.0'}),
frozenset({'288', '3.5'}),
frozenset({'288', '3.0'}),
frozenset({'2.0', '288'}),
frozenset({'290', '4.0'}),
frozenset({'292', '4.0'}),
frozenset({'292', '3.0'}),
frozenset({'2.0', '298'}),
frozenset({'298', '3.0'}),
frozenset({'2.5', '298'}),
frozenset({'0.5', '298'}),
frozenset({'304', '4.0'}),
frozenset({'307', '4.0'}),
frozenset({'3.5', '307'}),
frozenset({'2.0', '307'}),
frozenset({'2.5', '307'}),
frozenset({'3.0', '307'}),
frozenset({'339', '4.0'}),
frozenset({'357', '4.0'}),
frozenset({'367', '4.0'}),
frozenset({'380', '5.0'}),
frozenset({'380', '4.0'}),
frozenset({'3.0', '380'}),
frozenset({'2.0', '380'}),
frozenset({'3.5', '381'}),
frozenset({'381', '4.0'}),
frozenset({'3.5', '382'}),
frozenset({'391', '4.0'}),
frozenset({'391', '5.0'}),
```

```
frozenset({'4.0', '414'}),
frozenset({'3.0', '414'}),
frozenset({'414', '5.0'}),
frozenset({'2.0', '414'}),
frozenset({'2.5', '414'}),
frozenset({'3.5', '414'}),
frozenset({'3.0', '438'}),
frozenset({'3.5', '438'}),
frozenset({'4.0', '438'}),
frozenset({'3.0', '448'}),
frozenset({'4.0', '448'}),
frozenset({'3.5', '448'}),
frozenset({'2.0', '448'}),
frozenset({'2.5', '448'}),
frozenset({'4.0', '453'}),
frozenset({'1.5', '599'}),
frozenset({'1.5', '448'}),
frozenset({'4.0', '462'}),
frozenset({'4.0', '469'}),
frozenset({'3.0', '469'}),
frozenset({'469', '5.0'}),
frozenset({'4.0', '474'}),
frozenset({'3.0', '474'}),
frozenset({'3.5', '474'}),
frozenset({'4.5', '474'}),
frozenset({'2.5', '474'}),
frozenset({'2.0', '474'}),
frozenset({'4.0', '477'}),
frozenset({'3.5', '477'}),
frozenset({'4.5', '477'}),
frozenset({'3.0', '480'}),
frozenset({'2.5', '480'}),
frozenset({'4.0', '480'}),
frozenset({'3.5', '480'}),
frozenset({'4.0', '483'}),
frozenset({'3.5', '483'}),
frozenset({'4.5', '483'}),
frozenset({'4.0', '484'}),
frozenset({'3.5', '509'}),
frozenset({'3.0', '509'}),
frozenset({'4.0', '514'}),
frozenset({'4.0', '525'}),
frozenset({'3.5', '525'}),
frozenset({'4.0', '534'}),
frozenset({'4.5', '534'}),
frozenset({'3.5', '534'}),
frozenset({'4.0', '555'})
```

```
frozenset({'5.0', '555'}),
frozenset({'3.0', '555'}),
frozenset({'4.0', '560'}),
frozenset({'3.5', '560'}),
frozenset({'4.0', '561'}),
frozenset({'3.0', '561'}),
frozenset({'4.5', '573'}),
frozenset({'4.0', '580'}),
frozenset({'4.0', '587'}),
frozenset({'4.0', '590'}),
frozenset({'3.0', '590'}),
frozenset({'3.5', '590'}),
frozenset({'4.0', '596'}),
frozenset({'3.5', '596'}),
frozenset({'4.0', '597'}),
frozenset({'3.0', '597'}),
frozenset({'5.0', '597'}),
frozenset({'3.0', '599'}),
frozenset({'4.0', '599'}),
frozenset({'3.5', '599'}),
frozenset({'2.5', '599'}),
frozenset({'2.0', '599'}),
frozenset({'2.5', '600'}),
frozenset({'4.0', '600'}),
frozenset({'3.0', '600'}),
frozenset({'3.5', '600'}),
frozenset({'4.0', '603'}),
frozenset({'5.0', '603'}),
frozenset({'3.0', '603'}),
frozenset({'3.0', '606'}),
frozenset({'4.5', '606'}),
frozenset({'4.0', '606'}),
frozenset({'3.5', '606'}),
frozenset({'4.5', '608'}),
frozenset({'3.0', '608'}),
frozenset({'4.0', '608'}),
frozenset({'3.5', '608'}),
frozenset({'5.0', '610'}),
frozenset({'4.0', '610'}),
frozenset({'4.5', '610'}),
frozenset({'3.0', '610'}),
frozenset({'3.5', '610'}),
frozenset({'4.0', '6'}),
frozenset({'3.0', '6'}),
frozenset({'4.0', '42'}),
frozenset({'5.0', '51'}),
frozenset({'294', '3.0'})
```

```
frozenset({'2.0', '294'}),
frozenset({'3.0', '368'}),
frozenset({'2.0', '368'}),
frozenset({'6', 'Heat (1995)'}),
frozenset({'28', '3.5'}),
frozenset({'28', '3.0'}),
frozenset({'2.5', '28'}),
frozenset({'4.5', '62'}),
frozenset({'4.0', '62'}),
frozenset({'4.0', '84'}),
frozenset({'3.0', '84'}),
frozenset({'105', '4.0'}),
frozenset({'105', '5.0'}),
frozenset({'105', '3.5'}),
frozenset({'105', '4.5'}),
frozenset({'199', '4.0'}),
frozenset({'305', '4.0'}),
frozenset({'305', '5.0'}),
frozenset({'3.0', '305'}),
frozenset({'312', '4.0'}),
frozenset({'318', '4.0'}),
frozenset({'318', '4.5'}),
frozenset({'3.5', '318'}),
frozenset({'3.0', '318'}),
frozenset({'325', '4.0'}),
frozenset({'3.5', '425'}),
frozenset({'3.0', '592'}),
frozenset({'47', 'Seven (a.k.a. Se7en) (1995)'}),
frozenset({'104', '4.0'}),
frozenset({'122', '5.0'}),
frozenset({'260', '4.0'}),
frozenset({'3.5', '387'}),
frozenset({'387', '4.0'}),
frozenset({'2.5', '387'}),
frozenset({'3.0', '387'}),
frozenset({'452', '5.0'}),
frozenset({'3.0', '489'}),
frozenset({'3.5', '489'}),
frozenset({'4.0', '489'}),
frozenset({'50', 'Usual Suspects, The (1995)'}),
frozenset({'221', '4.5'}),
frozenset({'356', '4.0'}),
frozenset({'5.0', '593'}),
frozenset({'4.0', '593'}),
frozenset({'110', 'Braveheart (1995)'}),
frozenset({'125', '4.0'}),
frozenset({'223', 'Clerks (1994)'}),
```

```
frozenset({'231', 'Dumb & Dumber (Dumb and Dumber) (1994)'}),
frozenset({'260', '5.0', 'Star Wars: Episode IV - A New Hope (1977)'}),
frozenset({'5.0', '527'}),
frozenset({'4.0', '527'}),
frozenset({'296', '5.0', 'Pulp Fiction (1994)'}),
frozenset({'316', 'Stargate (1994)'}),
frozenset({'349', 'Clear and Present Danger (1994)'}),
frozenset({'356', '5.0', 'Forrest Gump (1994)'}),
frozenset({'111', '4.0'}),
frozenset({'111', '3.5'}),
frozenset({'111', '3.0'}),
frozenset({'367', 'Mask, The (1994)'}),
frozenset({'457', 'Fugitive, The (1993)'}),
frozenset({'480', 'Jurassic Park (1993)'}),
frozenset({'500', 'Mrs. Doubtfire (1993)'}),
frozenset({'527', 'Schindler's List (1993)'}),
frozenset({'590', 'Dances with Wolves (1990)'}),
frozenset({'592', 'Batman (1989)'}),
frozenset({'593', 'Silence of the Lambs, The (1991)'}),
frozenset({'608', ' Fargo (1996)'}),
frozenset({'648', 'Mission: Impossible (1996)'}),
frozenset({'733', 'Rock, The (1996)'}),
frozenset({'736', 'Twister (1996)'}),
frozenset({'780', 'Independence Day (a.k.a. ID4) (1996)'}),
frozenset({'1073', 'Willy Wonka & the Chocolate Factory (1971)'}),
frozenset({'1089', 'Reservoir Dogs (1992)'}),
frozenset({'1097', 'E.T. the Extra-Terrestrial (1982)'}),
frozenset({'1136', 'Monty Python and the Holy Grail (1975)'}),
frozenset({'1196', 'Star Wars: Episode V - The Empire Strikes Back (1980)'}),
frozenset({'1197', 'Princess Bride, The (1987)'}),
frozenset({'1198',
'Raiders of the Lost Ark (Indiana Jones and the Raiders of the
Lost Ark) (1981)'}),
frozenset({'1206', 'Clockwork Orange, A (1971)'}),
frozenset({'1208', 'Apocalypse Now (1979)'}),
frozenset({'1210', 'Star Wars: Episode VI - Return of the Jedi (1983)'}),
frozenset({'1213', 'Goodfellas (1990)'}),
frozenset({'1214', 'Alien (1979)'}),
frozenset({'1222', 'Full Metal Jacket (1987)'}),
frozenset({'1240', 'Terminator, The (1984)'}),
frozenset({'1258', 'Shining, The (1980)'}),
frozenset({'1265', 'Groundhog Day (1993)'}),
frozenset({'1270', 'Back to the Future (1985)'}),
frozenset({'1291', 'Indiana Jones and the Last Crusade (1989)'}),
frozenset({'1580', 'Men in Black (a.k.a. MIB) (1997)'}),
frozenset({'1732', 'Big Lebowski, The (1998)'}),
```

frozenset({'2028', 'Saving Private Ryan (1998)'}),
frozenset({'2115', 'Indiana Jones and the Temple of Doom (1984)'}),
frozenset({'2329', 'American History X (1998)'}),
frozenset({'2571', '5.0', 'Matrix, The (1999)'}),
frozenset({'2628', 'Star Wars: Episode I - The Phantom Menace (1999)'}),
frozenset({'2716', 'Ghostbusters (a.k.a. Ghost Busters) (1984)'}),
frozenset({'2858', 'American Beauty (1999)'}),
frozenset({'2959', 'Fight Club (1999)'}),
frozenset({'3147', 'Green Mile, The (1999)'}),
frozenset({'3578', 'Gladiator (2000)'}),
frozenset({'3793', 'X-Men (2000)'}),
frozenset({'318', '5.0', 'Shawshank Redemption, The (1994)'}),
frozenset({'1704', 'Good Will Hunting (1997)'}),
frozenset({'6874', 'Kill Bill: Vol. 1 (2003)'}),
frozenset({'48516', 'Departed, The (2006)'}),
frozenset({'58559', 'Dark Knight, The (2008)'}),
frozenset({'79132', 'Inception (2010)'}),
frozenset({'32', 'Twelve Monkeys (a.k.a. 12 Monkeys) (1995)'}),
frozenset({'357', 'Four Weddings and a Funeral (1994)'}),
frozenset({'539', 'Sleepless in Seattle (1993)'}),
frozenset({'588', 'Aladdin (1992)'}),
frozenset({'595', 'Beauty and the Beast (1991)'}),
frozenset({'1923', 'There's Something About Mary (1998)'}),
frozenset({'1968', 'Breakfast Club, The (1985)'}),
frozenset({'2683', 'Austin Powers: The Spy Who Shagged Me (1999)'}),
frozenset({'2762', 'Sixth Sense, The (1999)'}),
frozenset({'3996',
 'Crouching Tiger, Hidden Dragon (Wo hu cang long) (2000)'}),
frozenset({'4226', 'Memento (2000)'}),
frozenset({'4896',
 'Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter
and the Philosopher's Stone) (2001)'}),
frozenset({'34', 'Babe (1995)'}),
frozenset({'39', 'Clueless (1995)'}),
frozenset({'150', 'Apollo 13 (1995)'}),
frozenset({'153', 'Batman Forever (1995)'}),
frozenset({'253',
 'Interview with the Vampire: The Vampire Chronicles (1994)'}),
frozenset({'344', 'Ace Ventura: Pet Detective (1994)'}),
frozenset({'364', 'Lion King, The (1994)'}),
frozenset({'380', 'True Lies (1994)'}),
frozenset({'589', 'Terminator 2: Judgment Day (1991)'}),
frozenset({'597', 'Pretty Woman (1990)'}),
frozenset({'2', 'Jumanji (1995)'}),
frozenset({'10', 'GoldenEye (1995)'}),
frozenset({'161', 'Crimson Tide (1995)'}),
frozenset({'165', 'Die Hard: With a Vengeance (1995)'}),


```
frozenset({'185', 'Net, The (1995)'}),
frozenset({'208', 'Waterworld (1995)'}),
frozenset({'292', 'Outbreak (1995)'}),
frozenset({'293',
'Léon: The Professional (a.k.a. The Professional) (Léon)
(1994)'}),
frozenset({'329', 'Star Trek: Generations (1994)'}),
frozenset({'377', 'Speed (1994)'}),
frozenset({'434', 'Cliffhanger (1993)'}),
frozenset({'454', 'Firm, The (1993)'}),
frozenset({'587', 'Ghost (1990)'}),
frozenset({'2001: A Space Odyssey (1968)', '924'}),
frozenset({'1682', 'Truman Show, The (1998)'}),
frozenset({'4306', 'Shrek (2001)'}),
frozenset({'4886', 'Monsters, Inc. (2001)'}),
frozenset({'4963', 'Ocean's Eleven (2001)'}),
frozenset({'4993',
'Lord of the Rings: The Fellowship of the Ring, The (2001)'}),
frozenset({'4995', 'Beautiful Mind, A (2001)'}),
frozenset({'5349', 'Spider-Man (2002)'}),
frozenset({'5445', 'Minority Report (2002)'}),
frozenset({'5816', 'Harry Potter and the Chamber of Secrets (2002)'}),
frozenset({'5952', 'Lord of the Rings: The Two Towers, The (2002)'}),
frozenset({'5989', 'Catch Me If You Can (2002)'}),
frozenset({'6539',
'Pirates of the Caribbean: The Curse of the Black Pearl
(2003)'}),
frozenset({'7153', 'Lord of the Rings: The Return of the King, The
(2003)'}),
frozenset({'8961', 'Incredibles, The (2004)'}),
frozenset({'33794', 'Batman Begins (2005)'}),
frozenset({'586', 'Home Alone (1990)'}),
frozenset({'6377', 'Finding Nemo (2003)'}),
frozenset({'68954', 'Up (2009)'}),
frozenset({'1721', 'Titanic (1997)'}),
frozenset({'858', 'Godfather, The (1972)'}),
frozenset({'1200', 'Aliens (1986)'}),
frozenset({'1527', 'Fifth Element, The (1997)'}),
frozenset({'7438', 'Kill Bill: Vol. 2 (2004)'}),
frozenset({'60069', 'WALL·E (2008)'}),
frozenset({'111', 'Taxi Driver (1976)'}),
frozenset({'541', 'Blade Runner (1982)'}),
frozenset({'1193', 'One Flew Over the Cuckoo's Nest (1975)'}),
frozenset({'4878', 'Donnie Darko (2001)'}),
frozenset({'4973', 'Amelie (Fabuleux destin d'Amélie Poulain, Le)
(2001)'}),
frozenset({'7361', 'Eternal Sunshine of the Spotless Mind (2004)'}),
```

```
frozenset({'1036', 'Die Hard (1988)'}),  
frozenset({'1221', 'Godfather: Part II, The (1974)'}),  
frozenset({'778', 'Trainspotting (1996)'}),  
frozenset({'2706', 'American Pie (1999)'}),  
frozenset({'2918', 'Ferris Bueller's Day Off (1986)'}),  
frozenset({'5418', 'Bourne Identity, The (2002)'})]
```

Learnings:

- Developed our knowledge on EDA
- Got insights about association rules and how apriori algorithm works
- Developed understanding of using association learning on movie recommendation system
- were able to recommend movies based on previous choices.
- Learnt to visualize the frequency set

References :

- <https://towardsdatascience.com/how-to-find-closed-and-maximal-frequent-itemsets-from-fp-growth-861a1ef13e21>
- <https://towardsdatascience.com/complete-guide-to-association-rules-2-2-c92072b56c84>
- <https://www.kaggle.com/ankits29/movie-recommendation-with-ml-apriori-explained#Train-the-Model>
- <https://www.analyticsvidhya.com/blog/2021/04/mastering-exploratory-data-analysiseda-for-data-science-enthusiasts/>
- <https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>