```
In [2]:  import kaggle
         !kaggle datasets download aakashsaroop/movie-genreplotposter
```

```
Downloading movie-genreplotposter.zip to /mount/arbeitsdaten31/studenten1/knowledge-disc
overy/project/movie_plot_classification/movie-classification-single-label
 68%|██████████████████████████████         | 10.0M/14.7M [00:00<00:00, 101MB/s]
100%|████████████████████████████████████████| 14.7M/14.7M [00:00<00:00, 52.6MB/s]
```

```
In [3]:  !unzip movie-genreplotposter.zip
```

```
Archive:  movie-genreplotposter.zip
  inflating: Movie_combined.csv
```

Data pre-processing

```
In [4]:  import pandas as pd
```

```
In [5]:  movie_df = pd.read_csv("Movie_combined.csv", index_col=0)
```

```
In [6]:  movie_df
```

Loading [MathJax]/extensions/Safe.js

Out[6]:

	Title	Genre	Plot	Poster
0	the great train robbery	action	The film opens with two bandits breaking into ...	https://m.media-amazon.com/images/M/MV5BOWE4M2...
1	the suburbanite	comedy	The film is about a family who move to the sub...	https://m.media-amazon.com/images/M/MV5BMTUwOT...
2	how brown saw the baseball game	comedy	Before heading out to a baseball game at a nea...	0
5	the adventures of dollie	drama	On a beautiful summer day a father and mother ...	https://m.media-amazon.com/images/M/MV5BMDlmYz...
6	the black viper	drama	A thug accosts a girl as she leaves her workpl...	https://m.media-amazon.com/images/M/MV5BZDVjYT...
...	...	...	...	...
21223	selam	drama	The film opens with a Senegalese boy named Kha...	https://m.media-amazon.com/images/M/MV5BMzEyMz...
21224	particle (film)	drama	Zeynep lost her job at weaving factory, and he...	NaN
21225	çalgı çengi ikimiz	comedy	Two musicians, Salih and Gürkan, described the...	https://m.media-amazon.com/images/M/MV5BNWM0N2...
21226	olanlar oldu	comedy	Zafer, a sailor living with his mother Döndü i...	https://m.media-amazon.com/images/M/MV5BYWViZj...
21227	istanbul kırmızısı	romance	The writer Orhan Şahin returns to İstanbul aft...	https://m.media-amazon.com/images/M/MV5BYzA0Yj...

17303 rows × 4 columns

In [7]:
```python
movie_df = movie_df[["Title","Genre","Plot"]]
movie_df
```

Out[7]:

	Title	Genre	Plot
0	the great train robbery	action	The film opens with two bandits breaking into ...
1	the suburbanite	comedy	The film is about a family who move to the sub...
2	how brown saw the baseball game	comedy	Before heading out to a baseball game at a nea...
5	the adventures of dollie	drama	On a beautiful summer day a father and mother ...
6	the black viper	drama	A thug accosts a girl as she leaves her workpl...
...	...	...	...
21223	selam	drama	The film opens with a Senegalese boy named Kha...
21224	particle (film)	drama	Zeynep lost her job at weaving factory, and he...
21225	çalgı çengi ikimiz	comedy	Two musicians, Salih and Gürkan, described the...
21226	olanlar oldu	comedy	Zafer, a sailor living with his mother Döndü i...
21227	istanbul kırmızısı	romance	The writer Orhan Şahin returns to İstanbul aft...

17303 rows × 3 columns

In [8]:
```python
from collections import Counter
```

In [9]:
```python
Counter(movie_df["Title"]).most_common()
```

Loading [MathJax]/extensions/Safe.js

```
Out[9]:  [('prey', 9),
          ('i love you', 9),
          ('paradox', 7),
          ('love', 6),
          ('ride', 6),
          ('zero', 6),
          ('sanctuary', 5),
          ('the end', 5),
          ('the village', 5),
          ('the box', 5),
          ('lucky', 5),
          ('retribution', 5),
          ('euphoria', 5),
          ('forever', 4),
          ('the valiant', 4),
          ('summertime', 4),
          ('sleeping beauty', 4),
          ('limbo', 4),
          ('bloodline', 4),
          ('lovesick', 4),
          ('angel', 4),
          ('anna', 4),
          ('teenage mutant ninja turtles', 4),
          ('rush', 4),
          ('drive', 4),
          ('hush', 4),
          ('awake', 4),
          ('doubt', 4),
          ('sherlock holmes', 4),
          ('beneath', 4),
          ('erased', 4),
          ('the guest', 4),
          ('the forest', 4),
          ('time after time', 4),
          ('rogue', 4),
          ('undercover', 4),
          ('the importance of being earnest', 4),
          ('the trap', 4),
          ('bad blood', 4),
          ('desire', 4),
          ('drifters', 4),
          ('blue', 4),
          ('chase', 4),
          ('orange', 4),
          ('shadow', 4),
          ('the escape', 3),
          ('romance', 3),
          ('i do', 3),
          ('the river', 3),
          ('grand hotel', 3),
          ("dante's inferno", 3),
          ('boy meets girl', 3),
          ('hostages', 3),
          ('dream girl', 3),
          ('the heiress', 3),
          ('five', 3),
          ('monsoon', 3),
          ('the trial', 3),
          ('the trip', 3),
          ('riot', 3),
          ('the candidate', 3),
          ('sisters', 3),
          ('the deep', 3),
          ('coma', 3),
```

```
('lost and found', 3),
('exposed', 3),
('finders keepers', 3),
('reckless', 3),
('the bride', 3),
('labyrinth', 3),
('power', 3),
('alice', 3),
('hook', 3),
('shattered', 3),
('switch', 3),
('whore', 3),
('threesome', 3),
('the cure', 3),
('contact', 3),
('fallen', 3),
('godzilla', 3),
('life', 3),
('the cell', 3),
('the crew', 3),
('the gift', 3),
('screwed', 3),
('darkness', 3),
('identity', 3),
('twisted', 3),
('dreamland', 3),
('the woods', 3),
('the list', 3),
('spiral', 3),
('the deal', 3),
('deception', 3),
('monsters vs. aliens', 3),
('the unborn', 3),
('watchmen', 3),
('the sitter', 3),
('after', 3),
('beyond', 3),
('eden', 3),
('the signal', 3),
('deadpool', 3),
('the darkness', 3),
('lights out', 3),
('the girl on the train', 3),
('dead of night', 3),
('eyewitness', 3),
('darling', 3),
('invasion', 3),
('the fixer', 3),
('naked', 3),
('inferno', 3),
('trauma', 3),
('three', 3),
('adulthood', 3),
('triangle', 3),
('phobia', 3),
('careful', 3),
('last night', 3),
('into the forest', 3),
('together', 3),
('bliss', 3),
('the foreigner', 3),
('the protector', 3),
('heroes', 3),
('blood money', 3),
```

```
                ('bodyguard', 3),
                ('classmates', 3),
                ('tiger', 3),
                ('another', 3),
                ('asura', 3),
                ('dust', 3),
                ('blind', 3),
                ('honey', 3),
                ('great expectations', 2),
                ('the cook', 2),
                ('sex', 2),
                ('the boat', 2),
                ('resurrection', 2),
                ('doomsday', 2),
                ('goldie', 2),
                ('tabu', 2),
                ('forgotten', 2),
                ('welcome home', 2),
                ('reunion', 2),
                ('first lady', 2),
                ('the chaser', 2),
                ('go west', 2),
                ('pinocchio', 2),
                ('texas', 2),
                ('bluebeard', 2),
                ('oh, what a night', 2),
                ('the uninvited', 2),
                ('divorce', 2),
                ('the picture of dorian gray', 2),
                ('death valley', 2),
                ('the guilty', 2),
                ('the window', 2),
                ('the capture', 2),
                ('crisis', 2),
                ('alice in wonderland', 2),
                ('m', 2),
                ('i, the jury', 2),
                ('salome', 2),
                ('second chance', 2),
                ('bait', 2),
                ('betrayed', 2),
                ('diane', 2),
                ('the mountain', 2),
                ('another time, another place', 2),
                ('the hunters', 2),
                ('the journey', 2),
                ('the miracle', 2),
                ('tarzan, the ape man', 2),
                ('mysterious island', 2),
                ('rampage', 2),
                ('dead ringer', 2),
                ('lilith', 2),
                ('the outrage', 2),
                ('mirage', 2),
                ('never too late', 2),
                ('penelope', 2),
                ('caprice', 2),
                ('the incident', 2),
                ('flesh', 2),
                ('jigsaw', 2),
                ('the party', 2),
                ('the arrangement', 2),
                ('the dunwich horror', 2),
                ('the last warrior', 2),
```

```
            ('see no evil', 2),
            ('payday', 2),
            ('scarecrow', 2),
            ('obsession', 2),
            ('first love', 2),
            ('julia', 2),
            ('the fury', 2),
            ('the black hole', 2),
            ('head over heels', 2),
            ('the visitor', 2),
            ('kill or be killed', 2),
            ("mother's day", 2),
            ('simon', 2),
            ('the border', 2),
            ('harry tracy', 2),
            ('the thing', 2),
            ('breathless', 2),
            ('class', 2),
            ('firstborn', 2),
            ('just the way you are', 2),
            ('the karate kid', 2),
            ('splash', 2),
            ('flesh and blood', 2),
            ('into the night', 2),
            ('perfect', 2),
            ('revolution', 2),
            ('target', 2),
            ('armed response', 2),
            ('crawlspace', 2),
            ('crossroads', 2),
            ('the fly', 2),
            ('american gothic', 2),
            ('the night before', 2),
            ('some girls', 2),
            ('after midnight', 2),
            ('arena', 2),
            ('breaking in', 2),
            ('intruder', 2),
            ('little monsters', 2),
            ('the package', 2),
            ('renegades', 2),
            ('wild at heart', 2),
            ('the super', 2),
            ('unforgiven', 2),
            ('fearless', 2),
            ('the firm', 2),
            ('for love or money', 2),
            ('much ado about nothing', 2),
            ('junior', 2),
            ('wolf', 2),
            ('haunted', 2),
            ('the hunted', 2),
            ('restoration', 2),
            ('rob roy', 2),
            ('seven', 2),
            ('top dog', 2),
            ('wild bill', 2),
            ('bullet', 2),
            ('emma', 2),
            ('eye for an eye', 2),
            ('house arrest', 2),
            ('last man standing', 2),
            ('mother', 2),
            ('swingers', 2),
```

```
('george of the jungle', 2),
('metro', 2),
('dance with me', 2),
('vampires', 2),
('instinct', 2),
('storm', 2),
('thicker than water', 2),
("charlie's angels", 2),
('frequency', 2),
('the girl', 2),
('supernova', 2),
('driven', 2),
('chance', 2),
('spider-man', 2),
('the room', 2),
('criminal', 2),
('the forgotten', 2),
('the woodsman', 2),
('abominable', 2),
('the contract', 2),
('the covenant', 2),
('curious george', 2),
('expiration date', 2),
('the illusionist', 2),
('little man', 2),
('open season', 2),
('over the hedge', 2),
("pirates of the caribbean: dead man's chest", 2),
('something new', 2),
('special', 2),
('ghost rider', 2),
('the invisible', 2),
('meet the robinsons', 2),
('the messengers', 2),
('the mist', 2),
('p2', 2),
('primeval', 2),
('the reaping', 2),
('shrek the third', 2),
('spider-man 3', 2),
('war', 2),
('bolt', 2),
('ghost town', 2),
('the incredible hulk', 2),
('jumper', 2),
('madagascar: escape 2 africa', 2),
('meet the browns', 2),
('mirrors', 2),
('the mummy: tomb of the dragon emperor', 2),
('quarantine', 2),
('the ruins', 2),
('space chimps', 2),
('speed racer', 2),
('star wars: the clone wars', 2),
('taken', 2),
('9', 2),
('brothers', 2),
('fragments', 2),
('the messenger', 2),
('spread', 2),
('terminator salvation', 2),
('transformers: revenge of the fallen', 2),
('up', 2),
('buried', 2),
```

```
('cyrus', 2),
('game of death', 2),
('hereafter', 2),
('inception', 2),
('iron man 2', 2),
('let me in', 2),
('bad teacher', 2),
('the double', 2),
('dream house', 2),
('the good doctor', 2),
('insidious', 2),
('like crazy', 2),
('limitless', 2),
('red state', 2),
('the reunion', 2),
('transformers: dark of the moon', 2),
('the way', 2),
('bad ass', 2),
('excision', 2),
('gambit', 2),
('the master', 2),
('wanderlust', 2),
('hansel & gretel', 2),
('last call', 2),
('life of crime', 2),
('we are what we are', 2),
('welcome to the jungle', 2),
('non-stop', 2),
('jinn', 2),
('the amazing spider-man 2', 2),
('deliver us from evil', 2),
('guardians of the galaxy', 2),
('mercenaries', 2),
('the prince', 2),
('no good deed', 2),
('vice', 2),
('the boy next door', 2),
('wild card', 2),
('the voices', 2),
('it follows', 2),
('true story', 2),
('inside out', 2),
('ant-man', 2),
('the choice', 2),
('ratchet & clank', 2),
('nerve', 2),
('friend request', 2),
('gifted', 2),
('patrick', 2),
('restraint', 2),
('black water', 2),
('the black balloon', 2),
('the square', 2),
('crush', 2),
('the clinic', 2),
('the hunter', 2),
('mystery road', 2),
('sweetheart', 2),
('possession', 2),
('london', 2),
('blackmail', 2),
('no limit', 2),
('queen of hearts', 2),
('jericho', 2),
```

```
                ('old mother riley', 2),
                ('daybreak', 2),
                ('torment', 2),
                ('mandy', 2),
                ('the innocents', 2),
                ('the damned', 2),
                ('hysteria', 2),
                ('accident', 2),
                ('sebastian', 2),
                ('the reckoning', 2),
                ('pressure', 2),
                ('the disappearance', 2),
                ('absolution', 2),
                ('enigma', 2),
                ('bent', 2),
                ('wonderland', 2),
                ('paranoid', 2),
                ('another life', 2),
                ('asylum', 2),
                ('death at a funeral', 2),
                ('outpost', 2),
                ('popcorn', 2),
                ('pimp', 2),
                ('shame', 2),
                ('star wars: the force unleashed', 2),
                ('clean', 2),
                ('lucid', 2),
                ('sidekick', 2),
                ('adoration', 2),
                ('cairo time', 2),
                ('altitude', 2),
                ('the disappeared', 2),
                ('enemy', 2),
                ('bite', 2),
                ('natasha', 2),
                ('remember', 2),
                ('dirt', 2),
                ('still life', 2),
                ('somewhere only we know', 2),
                ('the road', 2),
                ('circle', 2),
                ('love o2o', 2),
                ('the victim', 2),
                ('the mirror', 2),
                ('fighter', 2),
                ('call girl', 2),
                ('genesis', 2),
                ('mafia', 2),
                ('dev', 2),
                ('anwar', 2),
                ('hattrick', 2),
                ('jumbo', 2),
                ('fox', 2),
                ('striker', 2),
                ('help', 2),
                ('thank you', 2),
                ('singham', 2),
                ('joker', 2),
                ('heartless', 2),
                ('2 states', 2),
                ('ugly', 2),
                ('boss', 2),
                ('bachchan', 2),
                ('winter', 2),
```

```
('traffic', 2),
("father's day", 2),
('premam', 2),
('lavender', 2),
('the silence', 2),
('baby', 2),
('king', 2),
('maska', 2),
('kick', 2),
('ishq', 2),
('rebel', 2),
('tommy', 2),
('gentleman', 2),
('good morning', 2),
('boy', 2),
("don't look back", 2),
('nobody knows', 2),
('the girl who leapt through time', 2),
('confessions', 2),
('ouran high school host club', 2),
('the complex', 2),
('killers', 2),
('blue spring ride', 2),
('cuckoo', 2),
('the inhabited island', 2),
('black lightning', 2),
('you and i', 2),
('quick', 2),
('the great train robbery', 1),
('the suburbanite', 1),
('how brown saw the baseball game', 1),
('the adventures of dollie', 1),
('the black viper', 1),
('a calamitous elopement', 1),
('the fight for freedom', 1),
('at the altar', 1),
("a drunkard's reformation", 1),
('the golden louis', 1),
('an arcadian maid', 1),
('hemlock hoax, the detective', 1),
('the house with closed shutters', 1),
('a lad from old ireland', 1),
('the black arrow: a tale of the two roses', 1),
('brown of harvard', 1),
('sweet memories', 1),
('the musketeers of pig alley', 1),
('petticoat camp', 1),
('put yourself in his place', 1),
("barney oldfield's race for a life", 1),
("bob's baby", 1),
('cohen saves the flag', 1),
('the quakeress', 1),
('the restless spirit', 1),
('the telephone girl and the lady', 1),
('a busy day', 1),
('the call of the north', 1),
('caught in a cabaret', 1),
('caught in the rain', 1),
('cruel, cruel love', 1),
('dough and dynamite', 1),
('gentlemen of nerve', 1),
('getting acquainted', 1),
("his father's rifle", 1),
('the knockout', 1),
```

```
('mabel at the wheel', 1),
("mabel's blunder", 1),
("mabel's married life", 1),
("mabel's strange predicament", 1),
('making a living', 1),
('the masquerader', 1),
('the new janitor', 1),
('the property man', 1),
('rose of the rancho', 1),
('the rounders', 1),
('salomy jane', 1),
('should a woman divorce?', 1),
('those love pangs', 1),
("tillie's punctured romance", 1),
('the wrath of the gods', 1),
('after five', 1),
('carmen', 1),
('enoch arden', 1),
('a fool there was', 1),
('four feathers', 1),
('fresh from the farm', 1),
('giving them fits', 1),
('love, loot and crash', 1),
('the soul of broadway', 1),
('the toast of death', 1),
('the warrens of virginia', 1),
('work', 1),
('the abandonment', 1),
('the bondman', 1),
('civilization', 1),
('a daughter of the gods', 1),
('divorce and the daughter', 1),
('hulda from holland', 1),
('lights of new york', 1),
('manhattan madness', 1),
("the soul's cycle", 1),
('the stepping stone', 1),
('under two flags', 1),
("youth's endearing charm", 1),
('all aboard', 1),
('bucking broadway', 1),
('camille', 1),
("cheyenne's pal", 1),
('the devil-stone', 1),
('easy street', 1),
('golden rule kate', 1),
('her right to live', 1),
('his wedding night', 1),
('the immigrant', 1),
('in again, out again', 1),
('the jury of fate', 1),
('the man without a country', 1),
('a modern musketeer', 1),
('oh doctor!', 1),
("peggy, the will o' the wisp", 1),
('the poor little rich girl', 1),
('reaching for the moon', 1),
('rebecca of sunnybrook farm', 1),
('a reckless romeo', 1),
('the rough house', 1),
('souls in pawn', 1),
('straight shooting', 1),
('a strange transgressor', 1),
('the woman god forgot', 1),
```

```
('all woman', 1),
('amarilly of clothes-line alley', 1),
('are crooks dishonest?', 1),
('believe me, xantippe', 1),
("the bride's awakening", 1),
('cecilia of the pink roses', 1),
('the craving', 1),
('the deciding kiss', 1),
('face value', 1),
('the forbidden city', 1),
('fuss and feathers', 1),
('the goddess of lost lake', 1),
('he comes up smiling', 1),
('hell bent', 1),
('the legion of death', 1),
('moonshine', 1),
('the non-stop kid', 1),
('old wives for new', 1),
('out west', 1),
('the romance of tarzan', 1),
('salomé', 1),
('the scarlet drop', 1),
('set free', 1),
('shoulder arms', 1),
('stella maris', 1),
('tarzan of the apes', 1),
('till i come back to you', 1),
('under the yoke', 1),
("we can't have everything", 1),
('when do we eat?', 1),
('which woman?', 1),
('the whispering chorus', 1),
('ask father', 1),
('better times', 1),
('bolshevism on trial', 1),
('bumping into broadway', 1),
('daddy-long-legs', 1),
('dangerous hours', 1),
("don't change your husband", 1),
('the echo of youth', 1),
('the exquisite thief', 1),
('the false faces', 1),
('for better, for worse', 1),
('from hand to mouth', 1),
('the girl who stayed at home', 1),
('the grim game', 1),
("a gun fightin' gentleman", 1),
('happy though married', 1),
('the haunted bedroom', 1),
('hawthorne of the u.s.a.', 1),
("heart o' the hills", 1),
('his majesty, the american', 1),
('the law of men', 1),
('male and female', 1),
("a man's fight", 1),
('the miracle of love', 1),
('pistols for breakfast', 1),
('the probation wife', 1),
('ravished armenia', 1),
('the roaring road', 1),
('a romance of happy valley', 1),
('scarlet days', 1),
('a society exile', 1),
('soldiers of fortune', 1),
```

```
('sunnyside', 1),
('the test of honor', 1),
('true heart susie', 1),
('the turn in the road', 1),
('the unpainted woman', 1),
('the virtuous thief', 1),
('what every woman learns', 1),
('when the clouds roll by', 1),
('the white heather', 1),
('the world and its woman', 1),
('yankee doodle in berlin', 1),
("you're fired", 1),
('april folly', 1),
('blind youth', 1),
('a child for sale', 1),
('convict 13', 1),
("the devil's pass key", 1),
('the false road', 1),
('the family honor', 1),
('the girl in number 29', 1),
('hairpins', 1),
('haunted spooks', 1),
("her husband's friend", 1),
('high and dizzy', 1),
('huckleberry finn', 1),
('the idol dancer', 1),
('in the heart of a fool', 1),
('the jack-knife man', 1),
("judy of rogue's harbor", 1),
("lady rose's daughter", 1),
('the love flower', 1),
('number, please?', 1),
('pollyanna', 1),
('remodeling her husband', 1),
('the round-up', 1),
('the saphead', 1),
('shore acres', 1),
('silk hosiery', 1),
('silk husbands and calico wives', 1),
('the slim princess', 1),
('something to think about', 1),
('a splendid hazard', 1),
('suds', 1),
('the virgin of stamboul', 1),
('why change your wife?', 1),
('within our gates', 1),
('the woman in the suitcase', 1),
('the affairs of anatol', 1),
('among those present', 1),
('the big punch', 1),
('the blot', 1),
('experience', 1),
('forbidden fruit', 1),
('the four horsemen of the apocalypse', 1),
('the goat', 1),
('the haunted house', 1),
('heedless moths', 1),
('the lotus eater', 1),
('the love light', 1),
('love never dies', 1),
('the lucky dog', 1),
("mother o' mine", 1),
('never weaken', 1),
('no woman knows', 1),
```

```
('the nut', 1),
('orphans of the storm', 1),
('the playhouse', 1),
('roads of destiny', 1),
('a sailor-made man', 1),
('the sea lion', 1),
('seven years bad luck', 1),
('sheltered daughters', 1),
('the silver lining', 1),
('the sky pilot', 1),
('sure fire', 1),
("tol'able david", 1),
('uncharted seas', 1),
("woman's place", 1),
('beyond the rocks', 1),
('a blind bargain', 1),
('daydreams', 1),
('dr. jack', 1),
('dusk to dawn', 1),
('the electric house', 1),
('fair lady', 1),
('foolish wives', 1),
('the loaded door', 1),
('moran of the lady letty', 1),
('pay day', 1),
('peacock alley', 1),
("peg o' my heart", 1),
("ridin' wild", 1),
('trifling women', 1),
('wildness of youth', 1),
('the woman he married', 1),
('the young rajah', 1),
("ambrose applejohn's adventure", 1),
('ashes of vengeance', 1),
('the balloonatic', 1),
('black oxen', 1),
('the call of the canyon', 1),
('the daring years', 1),
('the extra girl', 1),
('gentle julia', 1),
('our hospitality', 1),
('the pilgrim', 1),
('safety last!', 1),
('the shock', 1),
('the shriek of araby', 1),
('the silent command', 1),
('slippy mcgee', 1),
('souls for sale', 1),
('three ages', 1),
('why worry?', 1),
('the arab', 1),
('circe, the enchantress', 1),
('the enchanted cottage', 1),
('feet of clay', 1),
('greed', 1),
('he who gets slapped', 1),
("helen's babies", 1),
('her night of romance', 1),
('his hour', 1),
('hot water', 1),
('the iron horse', 1),
("isn't life wonderful", 1),
('manhandled', 1),
('monsieur beaucaire', 1),
```

```
('the navigator', 1),
('one night in rome', 1),
('the red lily', 1),
('romola', 1),
('sherlock, jr.', 1),
('the shooting of dan mcgrew', 1),
('wild oranges', 1),
('the yankee consul', 1),
('the big parade', 1),
('cheaper to marry', 1),
('confessions of a queen', 1),
('the dark angel', 1),
('fifty-fifty', 1),
("hogan's alley", 1),
('kentucky pride', 1),
('the lady', 1),
("lady windermere's fan", 1),
('little annie rooney', 1),
('the lucky horseshoe', 1),
('the mad whirl', 1),
('the mystic', 1),
('old clothes', 1),
('phantom of the opera', 1),
('the rag man', 1),
('sally, irene and mary', 1),
('sally of the sawdust', 1),
('the salvation hunters', 1),
('seven chances', 1),
('smouldering fires', 1),
('stage struck', 1),
('time, the comedian', 1),
('too many kisses', 1),
('the tower of lies', 1),
('the trail rider', 1),
('wolf blood', 1),
('battling butler', 1),
("the better 'ole", 1),
('beverly of graustark', 1),
('the blackbird', 1),
('la bohème', 1),
('the boy friend', 1),
('cruise of the jasper b', 1),
('dance madness', 1),
('don juan', 1),
('ella cinders', 1),
('the exquisite sinner', 1),
('fine manners', 1),
('flesh and the devil', 1),
("for heaven's sake", 1),
('the gay deceiver', 1),
('the general', 1),
('the great k & a train robbery', 1),
('hands up!', 1),
('high steppers', 1),
('the joy girl', 1),
('kid boots', 1),
('looking for trouble', 1),
('the magician', 1),
('mare nostrum', 1),
("miss brewster's millions", 1),
('old ironsides', 1),
('private izzy murphy', 1),
('the sea beast', 1),
('so this is paris', 1),
```
Loading [MathJax]/extensions/Safe.js

```
("so's your old man", 1),
('the strong man', 1),
('the temptress', 1),
('torrent', 1),
('tramp, tramp, tramp', 1),
('the winning of barbara worth', 1),
('adam and evil', 1),
('children of divorce', 1),
('the enemy', 1),
('the fair co-ed', 1),
('for the love of mike', 1),
('foreign devils', 1),
('the garden of allah', 1),
('his first flame', 1),
('the kid brother', 1),
('knockout reilly', 1),
('a little journey', 1),
('london after midnight', 1),
('long pants', 1),
('the love of sunya', 1),
('the red mill', 1),
('the road to romance', 1),
('rookies', 1),
('the show', 1),
('sunrise: a song of two humans', 1),
('the way of all flesh', 1),
('why girls love sailors', 1),
('wings', 1),
('the yankee clipper', 1),
('4 devils', 1),
('the air circus', 1),
('the barker', 1),
('beware of bachelors', 1),
('the cardboard lover', 1),
('the cavalier', 1),
('the crimson city', 1),
('the crowd', 1),
('the divine woman', 1),
('the docks of new york', 1),
("don't marry", 1),
('drums of love', 1),
('dry martini', 1),
('forbidden hours', 1),
('four sons', 1),
('four walls', 1),
('the garden of eden', 1),
('glorious betsy', 1),
('the godless girl', 1),
('half a bride', 1),
('heart trouble', 1),
('a lady of chance', 1),
('the last command', 1),
('laugh, clown, laugh', 1),
('the law of the range', 1),
('the legion of the condemned', 1),
('lilac time', 1),
('the lion and the mouse', 1),
('love me and the world is mine', 1),
('the magnificent flirt', 1),
('the man who laughs', 1),
('the matinee idol', 1),
('the mating call', 1),
('my man', 1),
('naughty baby', 1),
```

Loading [MathJax]/extensions/Safe.js

```
("noah's ark", 1),
('on trial', 1),
('our dancing daughters', 1),
('the red dance', 1),
('red hair', 1),
('sadie thompson', 1),
('say it with sables', 1),
('sharp shooters', 1),
('show girl', 1),
('show people', 1),
('the singing fool', 1),
('the smart set', 1),
('so this is love?', 1),
('speedy', 1),
('steamboat bill, jr.', 1),
('stop that man!', 1),
('street angel', 1),
('that certain thing', 1),
('we faw down', 1),
('the wedding march', 1),
('west of zanzibar', 1),
('west point', 1),
('wife savers', 1),
('the wind', 1),
('a woman of affairs', 1),
('acquitted', 1),
('the aviator', 1),
('berth marks', 1),
('captain lash', 1),
('careers', 1),
('the case of lena smith', 1),
('children of the ritz', 1),
('coquette', 1),
('desert nights', 1),
('disraeli', 1),
('drag', 1),
('dynamite', 1),
('fancy baggage', 1),
('the forward pass', 1),
('the gamblers', 1),
('the great divide', 1),
('half marriage', 1),
('hallelujah!', 1),
('hardboiled rose', 1),
('his glorious night', 1),
('jazz heaven', 1),
('lady of the pavements', 1),
('the letter', 1),
('the locked door', 1),
('masked emotions', 1),
("new year's eve", 1),
('night parade', 1),
('one hysterical night', 1),
('our modern maidens', 1),
('the pagan', 1),
('queen kelly', 1),
('redskin', 1),
('seven faces', 1),
('seven footprints to satan', 1),
('show boat', 1),
('the sky hawk', 1),
("smilin' guns", 1),
('smiling irish eyes', 1),
('spite marriage', 1),
```

```
           ('the squall', 1),
           ('stark mad', 1),
           ('street girl', 1),
           ('synthetic sin', 1),
           ('this is heaven', 1),
           ('the trespasser', 1),
           ('under the southern cross', 1),
           ('the very idea', 1),
           ('wedding rings', 1),
           ('welcome danger', 1),
           ('where east is east', 1),
           ('why be good?', 1),
           ('wild orchids', 1),
           ('the wolf song', 1),
           ('woman trap', 1),
           ('the younger generation', 1),
           ('abraham lincoln', 1),
           ('animal crackers', 1),
           ('anna christie', 1),
           ('beau bandit', 1),
           ('behind the make-up', 1),
           ('the benson murder case', 1),
           ('the big house', 1),
           ('the big trail', 1),
           ('borrowed wives', 1),
           ('the boudoir diplomat', 1),
           ('the case of sergeant grischa', 1),
           ('check and double check', 1),
           ('courage', 1),
           ('the devil to pay!', 1),
           ("the devil's holiday", 1),
           ('the divorcee', 1),
           ('doughboys', 1),
           ('dumbbells in ermine', 1),
           ('extravagance', 1),
           ('the fall guy', 1),
           ('feet first', 1),
           ('the florodora girl', 1),
           ('for the defense', 1),
           ('general crack', 1),
           ...]
```

In [10]: `Counter(movie_df["Plot"].apply(lambda x: x[:50])).most_common()`

```
Out[10]:   [('At night, a carnivorous, shape-shifting alien name', 9),
           ('When Hong Kong police negotiator Lee Chung-chi (Lo', 7),
           ('New York City film school graduate Leta Evans (De ', 6),
           ('A writer fakes his own death and establishes a new', 6),
           ('In 1928, in the county of Yoknapatawpha, Mississip', 5),
           ('Wendell "Sonny" Lawson (Reynolds), an unscrupulous', 5),
           ('Residents of the small, isolated Pennsylvania vill', 5),
           ('In December 1976, financially-strapped couple Arth', 5),
           ('The storyline is light-hearted and revolves around', 5),
           ('Yoshioka, an experienced detective, investigates t', 5),
           ('The story unfolds in the Eurasian Steppes. Vera li', 5),
           ('Peter Ibbetson (Reid) is an orphan raised by his u', 4),
           ('The story focuses on Jane Hudson (Katharine Hepbur', 4),
           ('After many childless years, King Stefan and Queen ', 4),
           ('Three women in Florida have husbands serving in Vi', 4),
           ('Sam Roffe, President of Roffe & Sons Pharmaceutica', 4),
           ('Psychoanalyst Saul Benjamin takes on a patient tem', 4),
           ('As noted above, Jackie Chan re-edited The Protecto', 4),
           ('The story of a Czech actress, looking for work in ', 4),
           ('As a crime wave rises in New York City, reporter A', 4),
           ('In Texas, seasoned undercover narcotics police off', 4),
           ('A machine/human-hybrid prototype formerly working ', 4),
           ('Helen and Jackson live together in New York City. ', 4),
           ('Young billionaire Clay Beresford, Jr. (Hayden Chri', 4),
           ('In 1964 at a Catholic church in The Bronx, New Yor', 4),
           ('In 1890, private detective Sherlock Holmes and his', 4),
           ('Six high school seniors head out to a secluded lak', 4),
           ('Ben Logan is an American single parent who has rec', 4),
           ('Spencer and Laura Peterson with their children Luk', 4),
           ('The majority of the story is set in and around the', 4),
           ('A constantly bickering Anglo-Irish family live in ', 4),
           ('While carrying out his research as a travel journa', 4),
           ('The film is based on the Yugoslav resistance movem', 4),
           ('Fur trapper Jean La Bête (Oliver Reed) paddles his', 4),
           ('In October 1941 Stan Graham, a Westland smallholde', 4),
           ('The plot revolves around the players on a hockey t', 4),
           ('The film stars Zachary Bennett as Francis Waterson', 4),
           ('Hong Yunsheng is a jobless wanderer in Fujian. Att', 4),
           ('An opening scene shows a famous ship, Lady in Blue', 4),
           ('The main player in Chase to get to the "truth" is ', 4),
           ('Yakobi (Kalabhavan Mani) is an orphan who works in', 4),
           ('The story is a melodrama concerning a young woman,', 4),
           ('The film begins with a short prologue explaining t', 3),
           ('A young woman, Natalie Storm (played by Glaum), wo', 3),
           ('The Boy meets and marries The Girl. A year later, ', 3),
           ('Allen John Spender (Charles Farrell) is a virile o', 3),
           ('Doctor Otternschlag, a disfigured veteran of World', 3),
           ('Jim Carter, a former stoker, takes over a fairgrou', 3),
           ('Two screenwriters, Law and Benson (James Cagney an', 3),
           ('A group of 26 Czechoslovakian citizens are jailed ', 3),
           ("A daydreamer, Georgie Allerton's imagination goes ", 3),
           ('Catherine Sloper (Olivia de Havilland) is a plain,', 3),
           ('Roseanne Rogers (Susan Douglas Rubeš) trudges from', 3),
           ('A well established gentleman falls in love with hi', 3),
           ('Josef K. (Anthony Perkins) is sleeping in his bedr', 3),
           ('Paul Groves (Peter Fonda), a television commercial', 3),
           ('While the warden (real-life warden Frank A. Eyman)', 3),
           ('Marvin Lucas (Peter Boyle), a political election s', 3),
           ('Advertising salesman Philip Woode (Lisle Wilson) i', 3),
           ('While scuba-diving near shipwrecks off Bermuda, va', 3),
           ('Dr. Susan Wheeler (Geneviève Bujold) is a surgical', 3),
           ('While visiting Switzerland, an American college pr', 3),
           ('By the end of the 19th century, Butch Cassidy, the', 3),
           ("The subject of her professor's romantic designs, E", 3),
```
Loading [MathJax]/extensions/Safe.js

```
("Con man Michael Rangeloff (Michael O'Keefe) is a f", 3),
('Teenage outcast Johnny Rourke (Aidan Quinn) falls ', 3),
('The film begins with Baron Charles Frankenstein (S', 3),
('Fifteen-year-old Sarah Williams rehearses a play i', 3),
('Pete St. John (Richard Gere), a ruthless and highl', 3),
('Alice Tate is an upper-class New York housewife, w', 3),
('Peter Banning is a successful corporate lawyer liv', 3),
('While driving at night along the northern Californ', 3),
('A promiscuous and rather misogynistic man, Steve (', 3),
('Liz is a Los Angeles street prostitute. The audien', 3),
('Due to an administrative error two male college st', 3),
('Set in a small town of Stillwater, Minnesota, Erik', 3),
('Dr. Ellie Arroway works for the SETI program at th', 3),
('An opening voiceover announces that this is the st', 3),
('Elderly inmate Willie Long (Obba Babatundé) attend', 3),
('Child psychologist Catherine Deane (Jennifer Lopez', 3),
('Four retired mobsters Bobby (Richard Dreyfuss) - t', 3),
('In the town of Brixton, Georgia, widow Annie Wilso', 3),
('Willard (Norm Macdonald) is an overworked, underpa', 3),
('Forty years after an unfinished occult ritual resu', 3),
('A convict named Malcolm Rivers awaits execution fo', 3),
('Having solved a high-profile case involving serial', 3),
('Eighteen-year-old Audrey (Agnes Bruckner) lives wi', 3),
('In 1965, after burning down a tree in her yard, re', 3),
('After the American Civil War, a group of Southern ', 3),
('The story follows lonely introvert Mason, a telesa', 3),
('Struggling Hollywood film producer Charlie Berns i', 3),
('Timid accountant Jonathan McQuarry (Ewan McGregor)', 3),
('In the far reaches of space, a planet explodes sen', 3),
('Casey Beldon has nightmarish hallucinations of str', 3),
('Utilizing an alternative timeline and applying mul', 3),
('The film opens with Noah Griffith (Jonah Hill) giv', 3),
('In 2012, Ana, a failed writer who works as a nurse', 3),
('A detective searching for a missing child is appro', 3),
('Hyun Jae, or Jae for short, is an 18-year-old Kore', 3),
('Three MIT students – Jonah, Nic, and Haley – are o', 3),
('Wade Wilson is a dishonorably discharged special f', 3),
('Peter and Bronny Taylor go on a vacation to the Gr', 3),
('In a textile factory, an employee named Esther enc', 3),
('Rachel Watson is an alcoholic who rides a train ai', 3),
('Architect Walter Craig (Mervyn Johns) wakes up aft', 3),
('After an argument with her husband, Lucy Church st', 3),
('An alien spacecraft crash-lands on Earth, near a s', 3),
('It is the story of a man named Yakov Bok, a Jew li', 3),
('After a rough sexual encounter with a married woma', 3),
('Set inside a "Quake" like video game, one of the g', 3),
('Ben (Colin Firth) awakens from a coma to discover ', 3),
('Jennifer Peters (Justine Waddell) attempts to save', 3),
('Six years after being jailed for killing Trife, Sa', 3),
('While preparing to take her autistic son Tommy on ', 3),
('Psychiatrist Dr. Peter Ross (Glaser) M.D., is usin', 3),
('Careful is set in Tolzbad, a fictional mountain to', 3),
('In Toronto, a group of friends and family prepare ', 3),
('In the near future, two teenage sisters, Nell and ', 3),
('Du Xiaoju (Xu Jinglei) and Wang Yi (Tong Dawei) ar', 3),
('Liu Cheng (Liu Peiqi) is a widowed cook making his', 3),
('Bliss tells the story of a modern and non-traditio', 3),
('Ngoc Minh Quan, a Chinese Nuang chef, runs a Chine', 3),
('Rahul (Dev) is a rich, city boy, born to billionai', 3),
('Jailed as a juvenile for killing his abusive fathe', 3),
('A poor boy falls in love with a rich girl but the ', 3),
('Two film academy students, Ali (Vatsal Sheth) and ', 3),
('The story is about Kunal (Kunal Khemu) who believe', 3),
('Jaihind is the journey of five people, linked by o', 3),
```

```
('Bodyguard revolves around young Jayakrishnan (Dile', 3),
('Ravi (Sumanth), Raaji (Sadha), Satish (Ravi Varma)', 3),
('The story begins with Vishnu (Rahul Ravindran) bei', 3),
('In 1972, Misaki, a popular student of Yomiyama Nor', 3),
('Mid 15th-century Japan. Flood, drought and famine ', 3),
('The protagonist, Aleksei, is a passive young outsi', 3),
('Year 2157. Maxim Kammerer (Stepanov), a space expl', 3),
('A missing person case involving a female universit', 3),
('In the remote and undeveloped eastern Black Sea re', 3),
('A young boy, Pip, runs into an escaped convict at ', 2),
('Fatty is the head chef at the "Bull Pup" restauran', 2),
('The film is a morality story on the evils of marit', 2),
('Buster is married with two children (both of whom ', 2),
('Katyusha, a country girl, is seduced and abandoned', 2),
('In Sussex, England after the Great War, former ari', 2),
('The credits (accompanied by organ music endemic to', 2),
('In every port, sailor Bill, Spencer Tracy, meets g', 2),
('Aged emissary Hitu arrives by Western sailing ship', 2),
('Retired Papa Strauss, a widower, who has been a su', 2),
('Newspapers around the world proclaim the birth of ', 2),
('The granddaughter of a President of the United Sta', 2),
('An ambulance chaser, unethical and disliked by man', 2),
('Confidence man S. Quentin Quale (Groucho) heads we', 2),
('Jiminy Cricket explains that he is going to tell a', 2),
('Hoppity the Grasshopper, after a period spent away', 2),
('Two Confederate veterans, broke and homeless, are ', 2),
('Eastern school teacher Catherine Allen becomes not', 2),
('All Paris is frightened by the murders attributed ', 2),
('In 1937, London music critic and composer Roderick', 2),
('A woman who has been married and divorced five tim', 2),
('It is London, 1886. Dorian Gray (Hurd Hatfield) is', 2),
('A dance hall girl is murdered and her killer flees', 2),
('The film tells the story of two roommates, Mike Ca', 2),
('Set and filmed on location in the tenement section', 2),
('Lin Vanner (Lew Ayres), is manager of an oil compa', 2),
('Dr. Eugene Ferguson (Cary Grant), a renowned Ameri', 2),
('On a riverbank, Alice spots a White Rabbit in a wa', 2),
('Captain Phillip Donlin (Lloyd Bridges) and his sma', 2),
('Martin W. Harrow (David Wayne) is a compulsive chi', 2),
('Although based on the New Testament story, the fil', 2),
('A mob bookkeeper (played by Milburn Stone) is conf', 2),
('Middle-aged Marko (Haas) is searching for a lost g', 2),
('Betrayed is an espionage thriller set in the Nazi-', 2),
('The action is set in 16th-century France.\r\nDiane d', 2),
('When a passenger plane crashes near the top of Mon', 2),
('Jack Flynn is driving down a highway at night, loo', 2),
('Major Cleve "Iceman" Saville (Robert Mitchum), a v', 2),
('In 1956 a group of passengers stranded for days at', 2),
('Teresa (Carroll Baker), a postulant at the convent', 2),
('Paul Winchell plays a father to Jerry Mahoney who ', 2),
('Herbert H. Heebert (Jerry Lewis) is a young man wh', 2),
('In 1865 during the American Civil War, a massive s', 2),
('The story centers on a group of big game hunters w', 2),
('At the funeral of her husband Frank, wealthy widow', 2),
('Set in a private mental institution, Chestnut Lodg', 2),
('Three disparate travelers, a disillusioned preache', 2),
('When the electricity cuts out in the New York City', 2),
('Harry Lambert is a New England lumber company exec', 2),
('Penelope Elcott (Natalie Wood) is the wife of weal', 2),
('In the Swiss Alps, a skier shoots and kills anothe', 2),
('On a late Sunday evening in the Bronx, two deadbea', 2),
('As the film begins, Geraldine ejects Joe from thei', 2),
('After someone places sugar cubes laced with LSD in', 2),
('A film crew is making a Gunga Din-style costume ep', 2),
```

('Wealthy ad man Eddie Anderson makes a failed suici', 2),
('A 38-year-old writer of pornographic novels named ', 2),
('A woman groans and writhes with the pain of childb', 2),
('Flapping Eagle lives on an Indian reservation in t', 2),
('Maury Dann (Rip Torn) is a successful country-west', 2),
(' Two vagabonds, Max Millan, a short-tempered ex-co', 2),
('In the fictional country of Lugash, a mysterious t', 2),
('In 1959, Michael Courtland (Robertson), a New Orle', 2),
('Elgin (William Katt) is a college student who spen', 2),
('The young Lillian and her friend Julia, daughter o', 2),
('Peter Sandza (Kirk Douglas), a former CIA agent, s', 2),
('Nearing the end of a long mission exploring deep s', 2),
('Charles Richardson is a civil servant in his early', 2),
('In an enigmatic, alien landscape, Jerzy Colsowicz ', 2),
('A famous martial artist, Steve Chase (James Ryan),', 2),
('In a Growth Opportunity graduation, a couple named', 2),
('The Institute for Advanced Concepts, a group of sc', 2),
('Following the grisly suicide of the U.S. Ambassado', 2),
('Immigration enforcement agent Charlie Smith lives ', 2),
('In Antarctica, in 1982, a Norwegian helicopter pur', 2),
('Jesse Lujack (Richard Gere) is a cocky, nihilistic', 2),
('When Jonathan Ogner first shows up to prep school,', 2),
('Sixteen-year-old Jake Livingston (Christopher Coll', 2),
('Susan is a professional flautist in Boston who has', 2),
('In 1964, 8-year-old Allen Bauer (David Kreps) is v', 2),
('Bill Markham (Powers Boothe) is an engineer who ha', 2),
('In 1501, a city in Italy has been taken by a coup ', 2),
('Ben White (Reb Brown) attends the funeral of his s', 2),
('Upon discovering that his wife is having an affair', 2),
('Rolling Stone reporter Adam Lawrence (John Travolt', 2),
('Fur trapper Tom Dobb unwillingly participates in t', 2),
('In Dallas, Walter Lloyd (Hackman) runs a lumber bu', 2),
('One of the men of a dangerous Yakuza boss known as', 2),
('A young woman walks through the hallways of an apa', 2),
('Eugene Martone has a fascination for blues music w', 2),
('Seth Brundle, a brilliant but eccentric scientist,', 2),
('Cynthia is traumatized by the death of her baby af', 2),
('We see the origin of Pinhead as a British military', 2),
('While on a routine patrol on United States airspac', 2),
('The storyline is told through a series of flashbac', 2),
('Michael goes to Quebec City, Canada for Christmas ', 2),
('Allison and Cheryl are two college students in a n', 2),
('Steve Armstrong (Paul Satterfield) is working as a', 2),
("Ernie Mullins (Burt Reynolds) is New York's old-pr", 2),
('A supermarket closes, and the workers begin restoc', 2),
("Brian's family has moved to a new town, and he fee", 2),
('Gene Hackman stars as Johnny Gallagher, a Special ', 2),
('Buster McHenry (Kiefer Sutherland) works as an und', 2),
('Larry Wilson (Andrew McCarthy) and Richard Parker ', 2),
('Lovers Lula and Sailor are separated after he is j', 2),
('Joe Pesci stars as Louie Kritski, a heartless slum', 2),
('In 1881 in Big Whiskey, Wyoming, two cowboys, Quic', 2),
('While taking a business trip, architect Max Klein ', 2),
('Mitch McDeere (Tom Cruise) is a young man from an ', 2),
('Doug Ireland is a concierge at the Bradbury, a lux', 2),
('Having just crushed an uprising by his half-brothe', 2),
('Austrian Research geneticist Dr. Alex Hesse (Arnol', 2),
('In the Pride Lands of Africa, a lion rules over th', 2),
('Will Randall is bitten by a wolf while driving hom', 2),
('In 1928 England, David Ash (Aidan Quinn), an Ameri', 2),
('Paul Racine, a computer chip company executive fro', 2),
('A young doctor, Robert Merivel, enters the service', 2),
('In Scotland, 1713, Robert Roy MacGregor is the Chi', 2),
('Soon-to-retire detective William Somerset is partn', 2),

('The film opens as two white nationalists destroy a', 2),
("A well-known lawman and scout of the 19th Century'", 2),
('Butch "Bullet" Stein (Mickey Rourke) is a 35-year-', 2),
('The film describes a year in the life of Emma Wood', 2),
('Karen McCann is happily married to Mack and has tw', 2),
('The film begins by showing the Beindorfs, Janet (J', 2),
('In 1932 Prohibition-era Texas, aimless wanderer Jo', 2),
('Mike Peters is a struggling comedian who left New ', 2),
('In an animated sequence, a plane flying through th', 2),
('SFPD Inspector Scott Roper (Eddie Murphy) is the b', 2),
('After burying his mother, Rafael Infante (Chayanne', 2),
('The film takes place at an elite New England board', 2),
('A team of Vatican-sponsored vampire hunters led by', 2),
('The film examines the mind of anthropologist Ethan', 2),
('The film begins on 23 August 1992 with the top-sec', 2),
('In Los Angeles, two rival gang leaders are also tr', 2),
('Natalie Cook (Cameron Diaz), Dylan Sanders (Drew B', 2),
('In October 1969, firefighter Frank Sullivan (Quaid', 2),
('The film is narrated by Agathe De La Boulaye as "T', 2),
('Supernova chronicles the search-and-rescue patrol ', 2),
("Greece's Ionian Islands are invaded by the Italian", 2),
('Halfway through the (fictionalized) 2001 Champ Car', 2),
('Chance is a twentysomething slacker living in Los ', 2),
('High-school senior Peter Parker lives with his Aun', 2),
('Johnny is a successful banker who lives in a San F', 2),
('Richard Gaddis is a small-time crook with a pencha', 2),
('Telly Paretta (Julianne Moore) believes that her s', 2),
('Walter (Kevin Bacon), a convicted child molester, ', 2),
('The film begins with a farmer named Billy Hoss (Re', 2),
('Frank Carden (Morgan Freeman) is a professional as', 2),
('In the town of Ipswich, four popular teenage boys ', 2),
('The introduction of the movie is a cartoon short w', 2),
('The story is told by a Native American elderly man', 2),
('In Vienna, Austria-Hungary, 1889, a magician named', 2),
('The film starts off with Calvin "Babyface" Simms (', 2),
('In the peaceful town of Timberline, 900-pound (408', 2),
('While scavenging for food, RJ the raccoon enters t', 2),
('The wedding of Will Turner and Elizabeth Swann is ', 2),
('Kenya McQueen (Sanaa Lathan) is a successful, sing', 2),
('Comic book fan Les Franken (Michael Rapaport) sign', 2),
('Over one hundred years ago, the devil Mephistophel', 2),
('High school senior Nick Powell plans to skip his g', 2),
('Lewis is an aspiring 12-year-old inventor at an or', 2),
('A terrified mother and her young son are packing a', 2),
('Following a violent thunderstorm, David Drayton le', 2),
('Young businesswoman Angela Bridges (Rachel Nichols', 2),
('In Burundi, a British forensic anthropologist is e', 2),
('Katherine Winter (Hilary Swank) and colleague Ben ', 2),
('Prince Charming vows that he will become King of F', 2),
('Three years after the events of the second film, P', 2),
('A White Shepherd puppy named Bolt is adopted by a ', 2),
('The film begins as married New York City businessm', 2),
('At Culver University in Virginia, General Thunderb', 2),
('In Ann Arbor, Michigan, 15-year-old David Rice (Ma', 2),
('In Africa, Zuba the lion tries to teach his son Al', 2),
('In Chicago, Brenda Brown-Davis (Angela Bassett) is', 2),
('A security guard runs through a subway station unt', 2),
('In ancient China, a brutal and tyrannical warlord ', 2),
('On the evening of March 11, 2008, news reporter an', 2),
('Two young American couples — Jeff (Jonathan Tucker', 2),
('Ham III, grandson of the first chimpanzee in space', 2),
('Speed Racer is an 18-year-old whose life and love ', 2),
('During the Clone Wars, Jedi Knights Anakin Skywalk', 2),
('Retired CIA field agent Bryan Mills attempts to bu', 2),

('In Germany, 1930, a scientist (Alan Oppenheimer) i', 2),
('Sam Cahill (Tobey Maguire) is a U.S. Marine captai', 2),
('While in a restaurant, Carla Davenport, the restau', 2),
('In 1959, Lucinda Embry hears whispers while starin', 2),
('Will Montgomery (Ben Foster), a lone rebellious U.', 2),
('Narcissistic gigolo Nikki (Ashton Kutcher) lives i', 2),
('In 2003, Dr. Serena Kogan (Carter) of Cyberdyne Sy', 2),
('In 17,000 B.C., the Seven Primes travel across the', 2),
('In 1940, nine-year-old Carl Fredricksen idolizes f', 2),
('On October 23, 2006, Paul Conroy, an American civi', 2),
('Jamie (Catherine Keener) informs her ex-husband Jo', 2),
('The story is told in flashback. CIA agent Marcus J', 2),
('On assignment in Thailand, French television journ', 2),
('Dominick "Dom" Cobb and Arthur are "extractors", w', 2),
("In Russia, the media covers Tony Stark's disclosur", 2),
('In March 1983, in Los Alamos, New Mexico, a disfig', 2),
('Elizabeth Halsey is an immoral English teacher at ', 2),
('Two FBI agents are conducting surveillance at a wa', 2),
('In 2007, Will Atenton (Daniel Craig) leaves his jo', 2),
('British medical student Martin E. Blake transfers ', 2),
('Married couple Josh and Renai Lambert, their child', 2),
('Anna Gardner (Felicity Jones), a British exchange ', 2),
("Eddie Morra is a struggling author with writer's b", 2),
('Travis (Michael Angarano) notices a fire station s', 2),
('After the death of her father, Nina (Smart) is cha', 2),
('In 1961, the "Ark", a Cybertronian spacecraft carr', 2),
('Dr. Thomas Avery (Martin Sheen) is an American oph', 2),
('Frank Vega (Danny Trejo) is a decorated Vietnam Wa', 2),
('Pauline is a disturbed and delusional high school ', 2),
('British art curator Harry Deane (Colin Firth) deci', 2),
('Freddie Quell is a World War II veteran, strugglin', 2),
('George and Linda Gergenblatt are an urban married ', 2),
('Teen siblings Hansel and Gretel Grimm are abducted', 2),
('Two underachieving cousins, Danny (Van Winkle) and', 2),
('In Detroit in 1978, stoic socialite Mickey Dawson ', 2),
('Laurel and Audrey (Zoe Kazan) are identical twins.', 2),
('During a torrential downpour, a woman, later ident', 2),
('An office team go on a two-day team-building semin', 2),
('Alcoholic U.S. Air Marshal Bill Marks boards a Bri', 2),
('According to Dread Central, the producers claim th', 2),
('OsCorp scientist Richard Parker records a video me', 2),
('The film opens in a 2010 firefight in a palm grove', 2),
("In 1988, following his mother's death, a young Pet", 2),
('During a visit to Kazakhstan the daughter of the U', 2),
('Paul, a mechanic, speaks to his daughter Beth on v', 2),
('In Tennessee, a prisoner Colin Evans (Idris Elba) ', 2),
('A futuristic resort, Vice, offers visitors the opp', 2),
('Claire Peterson (Jennifer Lopez) separates from he', 2),
('Nick Wild (Jason Statham) is a recovering gambling', 2),
('Jerry (Ryan Reynolds) is an upbeat man who works a', 2),
('A young girl named Annie Marshall is stalked throu', 2),
('Christian Longo (James Franco), an Oregon man whos', 2),
('Following a nuclear holocaust, the world has becom', 2),
('Riley Andersen is born in Minnesota. Within her mi', 2),
('In 1989, scientist Hank Pym resigns from S.H.I.E.L', 2),
('In Akron, Ohio, 24-year-old Joy Newsome and her 5-', 2),
('Travis Shaw (Benjamin Walker) is a veterinarian, l', 2),
('Above planet Tenemule in the Solana Galaxy, Chairm', 2),
('High school senior Venus "Vee" Delmonico longs to ', 2),
('Laura (Alycia Debnam-Carey) is one of the most pop', 2),
('In a small town near Tampa, Florida, seven-year-ol', 2),
('Three years after murdering his parents, Patrick (', 2),
('Two fugitives from justice, Dale (Teresa Palmer) a', 2),
('Grace (Diana Glenn), her younger sister Lee (Maeve', 2),

```
('Fifteen-year-old Thomas Mollison (Rhys Wakefield) ', 2),
('Raymond Yale (David Roberts) and Carla Smith (Clai', 2),
('Joe Warr (Clive Owen) is a British sports writer w', 2),
('An international architecture student at Universit', 2),
('The Clinic is set in the year 1978 (six years prio', 2),
('Mercenary Martin David (Willem Dafoe) is hired by ', 2),
('Near the rural town of Winton, Queensland, a truck', 2),
("Rabindranath Tagore's elder brother Jyotirindranat", 2),
('Tanisha is a young daughter of a man working for t', 2),
('Bilina (Mim) and Richard (Riaz) belongs to traditi', 2),
('As described in a film magazine,[2] Lord Wheatley ', 2),
('On 26 April 1929, Scotland Yard Detective Frank We', 2),
('George Shuttleworth (Formby) is a chimney sweep fr', 2),
('Grace Perkins (Gracie Fields) is an ordinary worki', 2),
('A World War I American troopship is torpedoed, and', 2),
('In his will, a wealthy match magnate leaves his fo', 2),
('Fanny (Phyllis Calvert) finishes at boarding schoo', 2),
('The story begins with a hangman breaking down when', 2),
('Brothers Cliff and Jim Brandon are a successful wr', 2),
('The story takes place on 14 February 1895. It is a', 2),
('Harry and Christine Garland have a deaf daughter, ', 2),
('Miss Giddens applies for a job as a governess. It ', 2),
('Alexandria December, 1941. Two Italian frogmen are', 2),
('Simon Wells, a middle-aged tourist, is on a boatin', 2),
('An American, Chris Smith, wakes up in an English h', 2),
('Stephen, a married Oxford tutor in his forties, ha', 2),
('Mr. Sebastian is a former Oxford professor, who in', 2),
('Michael "Mick" Marler has risen up through the ran', 2),
('Anthony is a British, black teenager named who goe', 2),
('Jay Mallory is a contract killer living in Montréa', 2),
('The film centres on schoolboy Benjamin Stanfield a', 2),
('Alex Holbeck (Martin Sheen) is recruited as a CIA ', 2),
('Max (Clive Owen) is a promiscuous gay man living i', 2),
('Cremona, 1681 (Language: Italian)\r\nNicolò Bussotti', 2),
('The film follows the lives of three London sisters', 2),
('The film begins at a fashion show, where Chloe Kee', 2),
('Chiefly set in London during the First World War a', 2),
("In this adaptation of Oscar Wilde's play about mis", 2),
('Asylum is set in Britain in the early 1950s. It te', 2),
('In England, Daniel (Matthew Macfadyen) and his wif', 2),
('In a seedy bar in a town ravaged by war, scientist', 2),
('Too insecure to approach the girl of his dreams, D', 2),
('A week in the life of a Soho pimp - Woody - (Rober', 2),
('Brandon Sullivan is an Irish-American New York Cit', 2),
('Set in between Revenge of the Sith and A New Hope,', 2),
('Cheung plays Emily Wang, a former video jockey, wh', 2),
('Joel Rothman (Jonas Chernick) is suffering from in', 2),
('Norman Neale and his best friend Chuck have an ong', 2),
('High school French teacher Sabine reads to her cla', 2),
('Juliette is a Canadian magazine editor who arrives', 2),
('In the prologue, the mother of Sara (Jessica Lownd', 2),
('After their fishing boat sinks, six men in two sma', 2),
('A man attends an erotic show at an underground clu', 2),
('Casey (played by Elma Begovic) returns home from h', 2),
('Mark Berman (Alex Ozerov), an idealistic Jewish Ca', 2),
('In a New York City nursing home, the elderly and i', 2),
('The film follows a nurse, Ye Tong (Kong Lin), who ', 2),
('Still Life is set in Fengjie, a city upstream of t', 2),
('Jin Tian (Wang Likun) is a young woman who recentl', 2),
("Young couple Rana and Guy live in today's Beirut. ", 2),
('Fifty people wake up in a darkened room, arranged ', 2),
('Xiao Nai is a gaming expert, who is also the most ', 2),
('Chun-yau (Bryan Leung) was adopted as a little orp', 2),
('The film is divided into five unrelated segments w', 2),
```

('Half-witted longtime friends Albert (Luis Manzano)', 2),
('The lives of three people take a turn when one of ', 2),
('The Philippines, 1972. Mysterious things are happe', 2),
('The film is set against the backdrop of unrest in ', 2),
('The film starts with a flashback. ACP Bose (Ferdou', 2),
('Amar returns from abroad and is asked by his multi', 2),
('A farmer (Naseeruddin Shah) and a weaver (Om Puri)', 2),
('In India, many illegal activities including supply', 2),
('Joint Commissioner of Police Dev Pratap Singh, a d', 2),
('Anwar, a Muslim, lives a middle-class lifestyle in', 2),
('Dr. Satyajit Chavan is the head doctor at a civil ', 2),
('The movie starts with Akshay Kumar singing the son', 2),
('Manohar moves into a new apartment, 13B, on the 13', 2),
('Arjun Kapoor (Arjun Rampal) is a successful crimin', 2),
('Born into a poor family, Surya grows up with few l', 2),
('A woman, Susan Alves, is in a mental asylum on the', 2),
('Raj (Bobby Deol), Yogi (Sunil Shetty) and Vikram (', 2),
('Singham opens with an honest police officer in Goa', 2),
('Agastya (Akshay Kumar) is a scientist, working on ', 2),
('A young business magnate Aditya Singh (Adhyayan Su', 2),
('Krish Malhotra (Arjun Kapoor), a fresh Engineer fr', 2),
('A young boy named Rana, who hails from the kingdom', 2),
('Shalini, a depressed housewife, attempts suicide, ', 2),
('The film opens with an ancient Chinese mother and ', 2),
('Darshan plays dual role in the movie. This film ha', 2),
('The movie starts with Bharat alias Bachchan (Sudee', 2),
('Some college students from Chennai go on vacation ', 2),
('A student of Sacred Heart Medical College is found', 2),
('Jayaram and Bhavana takes the lead roles as Dr. Ra', 2),
('On 16 September, superstar Sidharth Shankar (Rahma', 2),
('The story revolves around Seetha, a tutor in a Gov', 2),
('The story revolves around a widowed father and son', 2),
('In 2000, George Kalaparambath (Nivin Pauly), a 16-', 2),
('The film Lavender is a remake of the 2006 Korean-H', 2),
('The Silence is based on a true story and revolves ', 2),
('This is the story of Venu, a middle class young ma', 2),
('Pachamuthu (Parthipan) by his own confession is a ', 2),
('The film opens with two people\xa0– a man (Aravind Ak', 2),
('Sivaji, an Indian software architect, returns to I', 2),
('Anbu (Vishal) is a happy-go-lucky engineering stud', 2),
('Prabhu (Uday Kiran) is a college student who does ', 2),
('The Plot opens with water over flowing from the Th', 2),
('Ravi (Vijay), a youth from Thoothukudi, aspires to', 2),
('Ashwin Kumar (Jiiva) is a photojournalist working ', 2),
('Krishna (Vikram) is intellectually disabled and ha', 2),
('The film starts in the 1980s. A young Shiva (Jiiva', 2),
('The Scene begins in a Naxal infested area, where g', 2),
("There was a bomb blast in People's Plaza at Hydera", 2),
('Sakthi (Shira Gaarg), after having separated from ', 2),
('Nayaki is a horror movie . Nayaki is about Gayathr', 2),
('Sachin Joshi is a Kabaddi Player who loved a girl ', 2),
('Raja Chandra Pratap Varma AKA King (Akkineni Nagar', 2),
('Krish (Ram) is a happy-go-lucky guy who lives with', 2),
("Naina (Ileana D'Cruz) is a chivalrous girl. One da", 2),
('The movie begins with a scene where the injured Si', 2),
('Rishi (Prabhas) is a happy-go-lucky son of Bhupath', 2),
('Satyam Master (Rajendra Prasad) is a very famous p', 2),
('Aishwarya, the daughter of a wealthy entrepreneur,', 2),
('The film takes place in suburban Tokyo, and begins', 2),
('Based on real events reported in Japanese newspape', 2),
('While vacationing, Koenma is kidnapped by a pair o', 2),
('Hanamichi Sakuragi is a delinquent and the leader ', 2),
('Best friends Akira and Koichi, known to get into m', 2),
('The film comes accompanied by a short mini-movie f', 2),

```
('Nobody Knows tells the story of four siblings and ', 2),
('Kazuko Yoshiyama, a third-year middle school stude', 2),
('Junior high school teacher Yuko Moriguchi (Takako ', 2),
('Following the events of the TV series, Haruhi Fuji', 2),
('The background of the plot is based on the Stateme', 2),
('In the fictional city of Mitakihara, Japan, a midd', 2),
('A young nursing student, Asuka Ninomiya (Atsuko Ma', 2),
('Nomura Shuhei is a charismatic, wealthy Japanese e', 2),
('Futaba Yoshioka is a 17-year-old high school stude', 2),
('September 1944. Several days before Finland, a co-', 2),
('Dmitry "Dima" Maykov (Grigoriy Dobrygin) is a plai', 2),
('The film is about Janie Sawyer, an American teenag', 2),
('Seoul, 2004. A group of bikers are joy-riding thro', 2),
('The film opens with two bandits breaking into a ra', 1),
('The film is about a family who move to the suburbs', 1),
('Before heading out to a baseball game at a nearby ', 1),
('On a beautiful summer day a father and mother take', 1),
('A thug accosts a girl as she leaves her workplace ', 1),
('A young couple decides to elope after being caught', 1),
('The film opens in a town on the Mexican border. A ', 1),
('A rejected admirer sets up a trap to kill his swee', 1),
('John Wharton, the husband of a true and trusting w', 1),
('An old woman sends a girl begging in the streets o', 1),
('Mary Pickford plays Priscilla an unemployed maid w', 1),
('Hemlock Hoax is a detective who has little respect', 1),
('During the American Civil War a young soldier lose', 1),
('An Irish boy (Olcott) emigrates to America to esca', 1),
('The novel is set in the reign of "old King Henry V', 1),
("The story deals with Tom Brown's efforts to save h", 1),
('Polly Biblett (Mary Pickford), a young lady, tells', 1),
('The film is about a poor married couple living in ', 1),
('Only lasting 15 minutes, it is a light-hearted com', 1),
('The story is of an English manufacturing town {Hud', 1),
('Barney Oldfield races a speeding locomotive to res', 1),
('Bob, a typically devoted husband, is told by his w', 1),
('Cohen (Ford Sterling) and his rival Goldberg (Henr', 1),
('The setting is an early American village, where a ', 1),
('The film begins with the Dreamer, a restless and d', 1),
('A telephone operator is walking out with a handsom', 1),
('In A Busy Day, a wife (played by an energetic Char', 1),
("Graehme, Ned Stewart's father was accused of adult", 1),
('Chaplin plays a waiter who fakes being a Greek Amb', 1),
('The action starts in a park, where a man is trying', 1),
('Chaplin plays a character quite different from the', 1),
('The story involves Chaplin and Chester Conklin wor', 1),
('Mabel and her beau go to an auto race and are join', 1),
('In one of Chaplin\'s "park comedies" for Keystone S', 1),
('James Birch, an English hunter, is accidentally sh', 1),
('Pug, a down-and-out hobo, is talked into pretendin', 1),
('Charlie offers Mabel a ride on his two-seater moto', 1),
("Mabel's Blunder tells the tale of a young woman wh", 1),
('Chaplin, in tramp attire, sits in the park with hi', 1),
('In a hotel lobby a heavily drunk tramp runs into a', 1),
("Chaplin's character attempts to convince a passerb", 1),
('The Masquerader is a comedy short whose plot revol', 1),
('The hero, a janitor played by Chaplin, is fired fr', 1),
('Charlie is in charge of stage "props" and has trou', 1),
('Esra Kincaid (La Reno) takes land by force and, ha', 1),
('A drunk reveller (Chaplin) returns home to a scold', 1),
('Rough-and-tumble Gold Rush-era California: a woman', 1),
('Grace Roberts (played by Lea Leland), marries ranc', 1),
('The Masher played by Charlie Chaplin fights for th', 1),
('Charles Chaplin portrays a womanizing city man who', 1),
('Baron Yamaki (Sessue Hayakawa) is a fisherman who ', 1),
```

('Ted Ewing (Edward Abeles) invests both his own and', 1),
('Don José, an officer of the law, is seduced by the', 1),
('Based on a summary in a film magazine,[2] Enoch, A', 1),
('John Schuyler (Edward José), a rich Wall Street la', 1),
('Considered a coward by his fiancée and comrades in', 1),
('A farm youth goes to college, pursues the pretty c', 1),
('Luke (Harold Lloyd) works in a shoe store, but has', 1),
('Dora and her father are lost in the kitchen (they ', 1),
('As described in a film magazine,[3] La Valencia (S', 1),
('Mademoiselle Poppea (played by Glaum) is the leadi', 1),
('As the American Civil War begins, Ned Burton leave', 1),
('Charlie is an assistant to Izzy A. Wake, a painter', 1),
('Dr. Edmund B. Stewart is a talented young physicia', 1),
("The narrative hinges on Jason's vow to wreak venge", 1),
('The film opens with the outbreak of a war in the p', 1),
('A sultan agrees to help an evil witch destroy a my', 1),
('Alicia is a poor girl living in the city with her ', 1),
('Upon the death of her parents, little Hulda find h', 1),
('Hoping to improve his financial lot, petty thief H', 1),
('Robbers steal the "Koor-Hal Ruby", killing four in', 1),
('Nadia, the daughter of a nobleman, rejects ancient', 1),
('Mary Beresford (Boland) is the wife of unambitious', 1),
('As described in a film magazine,[3] British noblem', 1),
('The film is about a court case and embezzlement.', 1),
('A father takes his daughter on a trip to Bermuda i', 1),
('As described in a film magazine,[3] Cheyenne Harry', 1),
('As described in a film magazine,[3] Armand Duval (', 1),
('As described in a film magazine,[2] Noisy Jim (Cor', 1),
('As described in a film magazine,[3] Silas Martin (', 1),
('In a slum called Easy Street, the police are faili', 1),
('The setting is the Old West town of Paradise, Neva', 1),
('Polly Biggs (Peggy Hyland) is the eldest of a fami', 1),
('Arbuckle plays a drug store clerk, soda jerk, and ', 1),
('The film begins aboard a steamer crossing the Atla', 1),
('A young man drowns his sorrows in strong drink whe', 1),
('As described in a film magazine,[4] Henri Labordie', 1),
('As described in a film magazine,[5] Barbara Norton', 1),
("The film opens with a sequence in which D'Artagnan", 1),
('As described in Exhibitors Herald, a film magazine', 1),
('As described in a film magazine review,[3] Neil Da', 1),
('Gwendolyn is an 11-year-old girl who is left by he', 1),
('As described in a film magazine,[3] Alexis Caesar ', 1),
('As described in a film magazine,[4] Rebecca Randal', 1),
("A philandering husband's public flirtation with a ", 1),
('Early morning in the Rough family household and Mr', 1),
('As described in a film magazine review,[3] Sebasti', 1),
('At the end of the 19th century in the Far West, a ', 1),
('Lola Montrose (played by Glaum) is a kept woman. T', 1),
('As described in a film magazine,[2] Moctezuma (Hat', 1),
('As described in a film magazine,[3] Susan Sweeney ', 1),
('Set in San Francisco during the early 1900s, the f', 1),
('Con artists Harold (Harold Lloyd) and Snub (Snub P', 1),
('As described in a film magazine,[5] George MacFarl', 1),
('As described in a film magazine,[3] believing he l', 1),
("As described in a film magazine,[3] Cecilia's pare", 1),
('Carroll Wayles (Ford) is a chemist who has discove', 1),
('As described in a film magazine,[3] Eleanor Hamlin', 1),
('As described in a film magazine,[4] Joan (Murray),', 1),
('The plot centers around an inter-racial romance be', 1),
('A young girl suddenly finds herself wealthy, but l', 1),
('The story is about a young woman who is a quarter ', 1),
('As described in a film magazine,[2] the principle ', 1),
('As described in a film magazine,[2] Bess Thurston ', 1),
('As described in a film magazine,[2] determined to ', 1),

```
('Set in the Virginia Hills, Fatty and Buster play r', 1),
('Bebe plays a pretty young thing with several suito', 1),
('As described in a film magazine,[4] disgusted by t', 1),
('Arbuckle plays a drifter who has caught a ride hid', 1),
('The film opens with flashbacks from Tarzan of the ', 1),
('As described in a film magazine,[1] Salome uses he', 1),
('As described in a film magazine,[2] "Kaintuck" Rid', 1),
('After discovering that her grandmother was a gypsy', 1),
('Charlie is in boot camp in the "awkward squad." On', 1),
('Stella Maris (Mary Pickford) was born paralyzed an', 1),
('John and Alice Clayton, Lord and Lady Greystoke (T', 1),
('As described in a film magazine,[4] Yvonne (Vidor)', 1),
('As described in a film magazine,[1] when Maria Val', 1),
('As described in a film magazine,[1] very much in l', 1),
('Nora, An actress (Enid Bennett) is performing in a', 1),
('As described in a film magazine,[2] Doris Standish', 1),
('As described in a film magazine,[2] John Tremble (', 1),
('Lloyd is a serious young middle-class guy on the m', 1),
('As described in a film magazine,[2] the plot of th', 1),
('Barbara, a wealthy female socialite intent on refo', 1),
('The film opens with a quick glimpse into the glamo', 1),
('A police officer finds a baby in a trash can, and ', 1),
('The film tells the story of an attempted Russian i', 1),
('Based upon a description in a film magazine,[5] Le', 1),
('Cabaret singer Olive Martin (played by Baird) appr', 1),
('As described in a film magazine,[3] Blue Jean Bill', 1),
('During World War I, a professional thief known as ', 1),
('As described in a film magazine,[4] Dr. Edward Mea', 1),
('A young woman stands to inherit a fortune, but a c', 1),
('As described in a film magazine,[3] younger son Ja', 1),
('A gang of men frame Harvey Hanford (Houdini) for m', 1),
('As described in a film magazine,[2] ranch owner Ch', 1),
('As described in a film magazine,[2] Jim Montjoy (M', 1),
('As described in a film magazine,[3] New York repor', 1),
('As described in a film magazine,[6] Anthony Hawtho', 1),
('Jason Honeycutt (Harold Goodwin) is a young boy wh', 1),
('As described in a film magazine,[4] Bill (Fairbank', 1),
('As described in a film magazine,[3] Laura Dayne (B', 1),
('The film centers on the relationship between Lady ', 1),
('As described in a film magazine,[4] wealthy clubma', 1),
('As described in a film magazine,[5] Clive (Standin', 1),
('A young man (Fay) goes out to eat breakfeast with ', 1),
('As described in a film magazine,[5] novelest Harri', 1),
('According to a contemporary New York Times article', 1),
('As described in a film magazine,[3] "Toodles" Wald', 1),
('As described in a film magazine,[3] the senior Joh', 1),
('As described in a film magazine,[5] Rosie Nell (Be', 1),
('Based upon a plot summary included in a film revie', 1),
('Robert Clay, a noble America hero of humble means ', 1),
('Charlie works on a farm from 4 a.m. to late at nig', 1),
('Martin Wingrave(Barrymore) is arrested and sent to', 1),
('As described in a film magazine,[2] "True Heart Su', 1),
('As described in a film magazine,[2] Paul Perry (Hu', 1),
('As described in a film magazine,[2] Gudrun Trygavs', 1),
('As described in a film magazine,[3] Dick Armitage ', 1),
('Based upon a plot summary in a newspaper,[3] Amy (', 1),
('As described in a film magazine,[3] Daniel Boone B', 1),
('As described in a film magazine,[4] Lord Angus Cam', 1),
('As described in an adaptation of the film in the O', 1),
('Captain Bob White, an American aviator behind enem', 1),
('As described in a film magazine,[2] old Gordon Rog', 1),
('As described in a film magazine,[3] April Poole (D', 1),
('As described in a film magazine,[2] when Elizabeth', 1),
('Charles Stoddard (played by Hale) is a poor artist', 1),
```

('Buster, a particularly untalented golfer plays gol', 1),
('As described in a film magazine,[9] Grace Goodrigh', 1),
('As described in a film publication,[4] Betty Palme', 1),
('As described in a film publication,[4] the proud, ', 1),
('As summarized in a film publication,[2] Laurie Dev', 1),
('As described in a film magazine,[3] Rex Rossmore (', 1),
('The action in Haunted Spooks centres around Harold', 1),
('As described in a film magazine,[4] Princeton Hadl', 1),
('The film revolves around a young woman who sleepwa', 1),
('As described in a film publication,[2] Huckleberry', 1),
('Mary (Seymour) is the daughter of a French man and', 1),
('As described in a film magazine,[3] in a small tow', 1),
('As described in a film magazine,[2] Peter Lane (Tu', 1),
('As described in a film magazine,[1] Judy (Minter),', 1),
('As described in a film magazine,[3] granddaughter ', 1),
('As described in a film magazine,[4] Thomas Bevan (', 1),
('While at an amusement park, trying vainly to forge', 1),
('The film opens in the Ozarks where a distraught Po', 1),
('As described in film publications,[2][3] Janie (Gi', 1),
('As described in a film publication,[2] a youth (Ar', 1),
('As described in a film magazine,[3] it is his love', 1),
('Nicholas Van Alstyne is the richest man in New Yor', 1),
('A period newspaper gives the following description', 1),
('As summarized in a film publication,[5] Marjorie B', 1),
('As described in a film magazine,[4] Deane Kendall ', 1),
('As described in a film magazine,[2] Princess Kalor', 1),
("As described in a film magazine,[4] David Markely'", 1),
('The main charter Karl Breitman played by Henry B. ', 1),
('Amanda Afflick (Mary Pickford) is a poor laundry w', 1),
('Based upon a review in a film publication,[2] Sari', 1),
('Frumpy wife Beth devotes herself to bettering her ', 1),
('The film opens with Sylvia Landry (Evelyn Preer), ', 1),
("As described in a film magazine,[4] Mary's (Bennet", 1),
('Socialite Anatol Spencer (Reid), finding his relat', 1),
("Mrs. O'Brien (Herring) is eager to be accepted as ", 1),
('As summarized in a film publication,[2] Buck (Jone', 1),
('At the end of class, poorly paid college professor', 1),
('The plot of Experience was summarized in the Augus', 1),
('Mrs. Mallory (Williams) persuades Mary Maddock (Ay', 1),
('Madariaga "The Centaur" (Pomeroy Cannon), a harsh ', 1),
('Buster Keaton is walking by and peers through a ba', 1),
('Keaton plays a teller at a successful bank. Unbekn', 1),
('As described in a film publication,[3] idealistic ', 1),
('As described in a film magazine,[3] raised aboard ', 1),
('Based upon a summary in a film publication,[2] Ang', 1),
('As described in a film magazine,[3] John Trott (Hu', 1),
('A hapless hero (Laurel), who after being thrown ou', 1),
('As described in a film publication,[5] several yea', 1),
('Harold works in an office on a tall building next ', 1),
('As described in a film magazine,[3] the Brandeis o', 1),
('Based upon a summary in a film publication,[4] Cha', 1),
('Just before the French Revolution, Henriette takes', 1),
('The film is set up as a series of humorous tricks ', 1),
('As summarized in a film publication,[3] David Mars', 1),
('"The Boy" (Lloyd) is an idle playboy and heir to $', 1),
('Captain John Nelson (Hobart Bosworth) and his crew', 1),
('Max Linder returns home drunk after his bachelor p', 1),
('As described in a film publication,[4] the story i', 1),
('While discussing heredity, a man recalls the story', 1),
('The Sky Pilot (Bowers) arrives in a small rough-an', 1),
('As described in a film magazine,[2] easy going ran', 1),
('David Kinemon, youngest son of West Virginia tenan', 1),
('As described in a film magazine,[3] after her drun', 1),
('As described in a film magazine,[4] Josephine Gers', 1),

Loading [MathJax]/extensions/Safe.js

('Captain Fitzgerald (Alec B. Francis), a retired gu', 1),
('The film is a contemporary (1920s, though the book', 1),
('Buster wants to marry a girl, but her father disap', 1),
('The Sick-Little-Well-Girl (played by Mildred Davis', 1),
('An Indian maid and American girl (both played by F', 1),
('Keaton plays a botany student who is accidentally ', 1),
('In Sicily, Count Martinello is assassinated by Car', 1),
('The silent drama tells the story of a man who name', 1),
('As described in a studio publication,[2] Bert Lyon', 1),
('The opening scenes are set in Scandinavia, where a', 1),
('Chaplin plays a laborer on a house construction si', 1),
('As described in a film magazine,[3] the board of d', 1),
('As described in a film publication,[3] Margaret "P', 1),
('Based upon a review in a film publication,[2] Cyri', 1),
('Leon de Severac is fed up with his daughter Jacque', 1),
('Spoiled son Andrew Kane (Joseph Striker) competes ', 1),
('As reviewed in a film magazine,[3] a rich man\'s so', 1),
('After fifteen years, Joshua Judd (Charles Ogle) te', 1),
('Ambrose Applejohn is bored with his life in Cornwa', 1),
('At the ball celebrating the wedding of Henry of Na', 1),
('A young man (Keaton) has a series of encounters in', 1),
('Lee Clavering (Tearle), a playwright in New York, ', 1),
('Glenn Kilbourne (Richard Dix) returns from the war', 1),
('A university student named John Browning (Charles ', 1),
('Sue Graham (Normand) is a small town girl who trav', 1),
('Julia (Bessie Love) is a "small-town heartbreaker"', 1),
('The Canfield and McKay families have been feuding ', 1),
('The Pilgrim (Chaplin), an escaped convict, steals ', 1),
('The film opens in 1922 with Harold Lloyd (the char', 1),
('In 1906, Wilse Dilling (Lon Chaney), a crippled ga', 1),
('The film is a spoof of Rudolph Valentino\u2019s hit 192', 1),
('Benadict Hisston is a foreign agent, part of a con', 1),
('The title is also the moniker of a renowned safe-c', 1),
('Remember "Mem" Steddon (Eleanor Boardman) marries ', 1),
('Three plots in three different historical periods\u2014', 1),
('Harold Van Pelham (Lloyd) is a rich businessman wh', 1),
('Jamil (Ramon Novarro) is a soldier in the Bedouin ', 1),
('Cecilie Brunner (Murray) was once a good natured w', 1),
('Crippled by the war, Oliver Bashforth (Richard Bar', 1),
('Kerry Harlan (La Rocque) is unable to work because', 1),
('John McTeague is a miner working in Placer County,', 1),
('Paul Beaumont (Lon Chaney) is a scientist who labo', 1),
('Toodie and Budge are identified as the two best ch', 1),
('American millionaire Samuel C. Adams brings his da', 1),
('Gritzko (John Gilbert) is a Russian nobleman and T', 1),
('Episodic in nature (effectively three short films ', 1),
('The film presents an idealized image of the constr', 1),
('A family from Poland has been left homeless in the', 1),
('The shop girl Tessie McGuire is invited by her bos', 1),
('The Duke of Chartres is in love with Princess Henr', 1),
('Wealthy Rollo Treadway (Buster Keaton) suddenly de', 1),
('Madame L\'Enigme (Laurette Taylor) is a fortune-tel', 1),
('Jean Leonnec (Ramon Novarro) and Marise La Noue (E', 1),
('In Renaissance Florence, a Florentine trader meets', 1),
('A movie theater projectionist and janitor (Buster ', 1),
('A dancer known as Lou Lorraine feels her life is g', 1),
('John Woolfolk and his wife are riding down a count', 1),
('Through a series of confusions, Dudley Ainsworth (', 1),
('In the United States in 1917, James "Jim" Apperson', 1),
('Matrimony is not only a good thing, but also a goo', 1),
('The King of Illyris (Lewis Stone) marries a neighb', 1),
('During the First World War Captain Alan Trent - on', 1),
('American millionaire Frederick Harmon (played by L', 1),
('Although he wins the championship by a knockout, p', 1),

('The plot concerns Beaumont, a horse breeder with a', 1),
('A young woman marries the wastrel son of a British', 1),
('In London, Lady Margaret Windermere is busy discou', 1),
('Annie Rooney is a young girl who spends her days w', 1),
('Following the death of the owner of the Hunt ranch', 1),
('Cathleen Gillis (May McAvoy) falls in love with Ja', 1),
('Zara (Aileen Pringle) is a gypsy rogue who joins w', 1),
('Tim Kelly (Jackie Coogan) and Max Ginsberg (Max Da', 1),
('The film opens with the debut of the new season at', 1),
('Tim Kelly (Jackie Coogan) is a kid who runs away f', 1),
('The film takes a behind-the-scenes look at the rom', 1),
('Because she married a circus performer, Judge Fost', 1),
('The film opens with a foreword:\r\nThe story begins ', 1),
('Jimmy Shannon (Buster Keaton) is the junior partne', 1),
('At 40, businesswoman Jane Vale (Pauline Frederick)', 1),
('Jennie Hagan (Swanson) is a waitress who dreams of', 1),
('Singer Nora (Mae Busch) left her husband for new f', 1),
('Richard Gaylor Jr. (Richard Dix) is a modern Lotha', 1),
('Jan (Lon Chaney) is a Swedish farmer and Glory (No', 1),
('Tex Hartwell (Buck Jones) rescues an old cobbler f', 1),
('Dick Bannister is the new field boss of the Ford L', 1),
("Alfred's father wants him to make a man of himself", 1),
('Old Bill (Sydney Chaplin), a jovial Limey sergeant', 1),
('Beverly Calhoun (Davies) impersonates the Prince o', 1),
('The Blackbird (Lon Chaney) is a thief who uses a s', 1),
('Several struggling bohemians try to survive in the', 1),
('Comedy about a small-town girl unhappy with her fa', 1),
("The film stars actor Rod La Rocque as 'Jerry Clegg", 1),
('May (Claire Windsor) is married to Roger (Conrad N', 1),
("In the prologue, Don José, warned of his wife's in", 1),
('In the house of late father, Ella Cinders (Colleen', 1),
('The film concerns a rich young Frenchman (Conrad N', 1),
('Burlesque chorus girl Orchid Murphy (Gloria Swanso', 1),
('The film is a romantic melodrama[1] about two chil', 1),
('Millionaire J. Harold Manners (Harold Lloyd) finds', 1),
('A deceiver leads the fast set in Paris and is invo', 1),
('Western & Atlantic Railroad train engineer Johnnie', 1),
('Following a series of robberies of the K & A Railr', 1),
('The film tells the story of Jack, a spy for the Co', 1),
('Julian Perryam (Lloyd Hughes) gets thrown out of O', 1),
('Jewel Courage (Borden) rejects a suitor (Hamilton)', 1),
('After a bully helps him out of a jam, Samuel "Kid"', 1),
('A telephone line repairman named Joe Graham (Spenc', 1),
('In the Latin Quarter of Paris, sculptor Margaret D', 1),
('As a young boy growing up in a Spanish family with', 1),
('Polly Brewster (Daniels), a penniless Hollywood ex', 1),
('Early in the 19th century, the USS Constitution is', 1),
('Isadore Goldberg, an enterprising Russian Jew, com', 1),
('At the beginning of the story, Ahab (John Barrymor', 1),
('Paul Giraud is happily married to Suzanna, and tog', 1),
('Sam Bisbee (W.C. Fields) is a small-town glazier w', 1),
('Paul Bergot (Harry Langdon) is a Belgian immigrant', 1),
('The story opens in Paris at a masquerade ball wher', 1),
('The wealthy matriarch Doña Bernarda Brull (Mattox)', 1),
("The film tells of Harry (Langdon) a ne'er-do-well ", 1),
('As a child, Barbara is orphaned when her settler p', 1),
('"The little fur-bearing animals that are sacrifice', 1),
('Jean Waddington (Esther Ralston) and Ted Larrabee ', 1),
('Newlywed Carl (Ralph Forbes) goes war where he ', 1),
('Marion Bright enrolls in college to pursue a hands', 1),
("A baby boy is found abandoned in a Hell's Kitchen ", 1),
('Captain Robert Kelly (McCoy) while attached to the', 1),
('Father Adrien (Iván Petrovich), a monk at the Trap', 1),
('The story tells of Harry Howells (Langdon), a rece', 1),

```
('The Hickorys are a respected family in Hickoryvill', 1),
('When a successful prizefighter known as "Killer" A', 1),
('A girl travelling by train to meet her boyfriend m', 1),
('In a cultured and peaceful home on the outskirts o', 1),
('The silent tells the story of Harry Shelby (Langdo', 1),
('The film depicts a young woman (Swanson) granted t', 1),
('Tina (Marion Davies) is the drudge of the Red Mill', 1),
('Serafina (Marceline Day) is captured by Don Baltha', 1),
('During World War I an entertainer named Greg Lee (', 1),
('Cock Robin (John Gilbert) is a sideshow barker in ', 1),
('A vacationing Woman from the City (Margaret Living', 1),
('In the story, which opens in the early 1900s, Jann', 1),
('The film starts with the loading of a ship called ', 1),
('Jack Powell and David Armstrong are rivals in the ', 1),
('Captain Winslow takes a revolutionary new clipper ', 1),
('The plot concerns four orphans (Janet Gaynor, Nanc', 1),
('Two young men, "Speed" Doolittle (Arthur Lake) and', 1),
('The film tells the story of a woman (Dorothy Macka', 1),
('A young doctor (William Collier Jr) is accused by ', 1),
('A group of American coeds/flappers arrives at the ', 1),
('The story takes place in old Mexico, where a maske', 1),
('The story centers on an Asian woman named Onoto (L', 1),
('Born on the Fourth of July, 1900, John Sims (James', 1),
('Marianne (Greta Garbo) is a poor French country gi', 1),
('An American tramp steamer docks in New York harbor', 1),
('A flapper masquerades as her straight-laced cousin', 1),
('After finding out her father and his estate is in ', 1),
('Wealthy divorced American Willoughby Quimby has be', 1),
('Set in the fictitious European kingdom of Balanca,', 1),
('Mother Bernle is a widow in Bavaria with four sons', 1),
('Benny Horowitz (John Gilbert), a reformed gangster', 1),
('A Viennese ingénue named Toni LeBrun, (Corinne Gri', 1),
('The film is a semi-historical narrative and depict', 1),
('This drama features a romance between two differen', 1),
('Patience Winslow (Esther Ralston) is an impulsive ', 1),
('A young man tries to enlist in the United States A', 1),
('Dolly "Angel Face" Morgan is a parolee out to flee', 1),
('In 1928 Hollywood, director Leo Andreyev (William ', 1),
('Tito (Lon Chaney), a traveling circus clown, finds', 1),
('Betty Dallas (Crawford) is a passenger on a stagec', 1),
('In World War I, four young men from various walks ', 1),
('Seven young English aviators are billeted at the B', 1),
('Judge Ross, on the Federal Bench, rules in favor o', 1),
('Hannerl (Philbin) is a young woman growing up in O', 1),
("Count D'Estrange tries to save his nephew Hubert f", 1),
("In 1690's England, King James II sentences his pol", 1),
('Don Wilson, a famous blackface comedian, is prepar', 1),
('Leslie Hatton, a poor farmer, becomes a captain an', 1),
('Fannie Brand (Fanny Brice), an industrious girl wh', 1),
('Rosalind McGill (White) is a cloak room girl falls', 1),
('The film opens after the great flood, with Noah an', 1),
('A man is put on trial for the murder on his best f', 1),
('"Dangerous Diana" Medford (Crawford) is outwardly ', 1),
('Tasia (Dolores del Río), a beautiful dancer lower ', 1),
('A free-spirited young girl has three middle-aged a', 1),
('A smoking, drinking, jazz listening, young prostit', 1),
('Doug Caswell (Arthur Rankin) falls for Irene Gordo', 1),
("George (George O'Brien) is a sailor and smooth-tal", 1),
('Dixie Dugan, a Brooklyn cutie, goes to the offices', 1),
('Young Peggy Pepper (Marion Davies) wants to be in ', 1),
('After years of hopeful struggle, Al Stone (Jolson)', 1),
('A self-centered polo player (Haines) has to redeem', 1),
('Jerry McGuire (William Collier Jr.) is a dress des', 1),
('Everybody in New York City "is in such a hurry tha', 1),
```

```
('William "Steamboat Bill" Canfield is the owner and', 1),
('The young brother of two police officers borrows o', 1),
('A spirited young woman (Gaynor) tries to steal mon', 1),
('Molly Kelly (Viola Dana) intends to marry a millio', 1),
('The Boys (Stan and Oliver) are trying to attend a ', 1),
('In Vienna in 1914, Prince Nicki is the scion of a ', 1),
('Anna (Jacqueline Gadsden) cannot bring herself to ', 1),
('Arrogant and wise-cracking Brice Wayne (William Ha', 1),
('Louis Hozenozzle and 2d Lieut. Rodney Ramsbottom, ', 1),
('An impoverished young woman named Letty Mason (Lil', 1),
('Diana Merrick (Greta Garbo), Neville (John Gilbert', 1),
('Marian is sent to prison for a crime she did not c', 1),
('Brooks (William Norton Bailey), a publisher and hi', 1),
('The scene opens where the passenger train, hauled ', 1),
('Lash is the head coal stoker on a steam ship whose', 1),
('In the French colony of Cochin-China, young French', 1),
('In turn-of-the-century Vienna, simple peasant girl', 1),
('A spoiled rich girl falls for a poor chauffeur. Th', 1),
('Norma Besant, daughter of a Southern doctor, is an', 1),
('A gang of thieves rob an African diamond company o', 1),
("In 1874, Disraeli's ambitious foreign policy, aime", 1),
('Young David Carroll takes over the publication of ', 1),
('Coal miner Hagon Derk (Charles Bickford) is senten', 1),
('Naomi Iverson learns that her father has assumed t', 1),
('Tired of being hit and hurt, football star Marty R', 1),
('Jason Robards Sr. (Carvel Emerson) and George Fawc', 1),
('Stephen Ghent, a mineowner, falls in love with Rut', 1),
('Judy Page is a young society girl who falls in lov', 1),
('Sharecroppers Zeke and Spunk Johnson sell their fa', 1),
('A Southern belle (Loy) must work in a gambling hou', 1),
('Although being engaged against her will with a wea', 1),
('A poor songwriter from the south, Barry Holmes, tr', 1),
('Disgusted that his fiancee, Diane (Jetta Goudal) h', 1),
("Bored and lonely living on her husband's rubber pl", 1),
('Ann Carter, an inexperienced young woman, accepts ', 1),
('Set on the Maine coast, a young sloop skipper Bram', 1),
('Saddled with the care of a younger brother and una', 1),
('Bobby Murray is the middleweight champion, managed', 1),
('A wealthy man dresses up as Napoleon for a fancy d', 1),
('Heiress Billie Brown (Crawford), is engaged to mar', 1),
 ...]
```

In [11]:
```python
movie_df[movie_df["Plot"].str.contains("At night, a carnivorous")]
```

Out[11]:

	Title	Genre	Plot
15637	prey	action	At night, a carnivorous, shape-shifting alien ...
15638	prey	action	At night, a carnivorous, shape-shifting alien ...
15639	prey	action	At night, a carnivorous, shape-shifting alien ...
15640	prey	action	At night, a carnivorous, shape-shifting alien ...
15641	prey	action	At night, a carnivorous, shape-shifting alien ...
15642	prey	action	At night, a carnivorous, shape-shifting alien ...
15643	prey	action	At night, a carnivorous, shape-shifting alien ...
15644	prey	action	At night, a carnivorous, shape-shifting alien ...
15645	prey	action	At night, a carnivorous, shape-shifting alien ...

In [12]:
```python
print("Rows before: {}".format(movie_df.shape[0]))
movie_df = movie_df.drop_duplicates(subset=['Plot'])
print("...fter: {}".format(movie_df.shape[0]))
```

Loading [MathJax]/extensions/Safe.js

```
Rows before: 17303
Rows after: 16584
```

In [13]: `movie_df[movie_df["Plot"].str.contains("At night, a carnivorous")]`

Out[13]:

	Title	Genre	Plot
15637	prey	action	At night, a carnivorous, shape-shifting alien ...

In [14]: `Counter(movie_df["Title"]).most_common()`

```
Out[14]: [('angel', 4),
         ('romance', 3),
         ('godzilla', 3),
         ('darling', 3),
         ('i love you', 3),
         ('love', 2),
         ('the valiant', 2),
         ('welcome home', 2),
         ('oh, what a night', 2),
         ('i, the jury', 2),
         ('another time, another place', 2),
         ('tarzan, the ape man', 2),
         ('see no evil', 2),
         ('the karate kid', 2),
         ('mother', 2),
         ('war', 2),
         ('london', 2),
         ('the importance of being earnest', 2),
         ('the great train robbery', 1),
         ('the suburbanite', 1),
         ('how brown saw the baseball game', 1),
         ('the adventures of dollie', 1),
         ('the black viper', 1),
         ('a calamitous elopement', 1),
         ('the fight for freedom', 1),
         ('at the altar', 1),
         ("a drunkard's reformation", 1),
         ('the golden louis', 1),
         ('an arcadian maid', 1),
         ('hemlock hoax, the detective', 1),
         ('the house with closed shutters', 1),
         ('a lad from old ireland', 1),
         ('the black arrow: a tale of the two roses', 1),
         ('brown of harvard', 1),
         ('sweet memories', 1),
         ('the musketeers of pig alley', 1),
         ('petticoat camp', 1),
         ('put yourself in his place', 1),
         ("barney oldfield's race for a life", 1),
         ("bob's baby", 1),
         ('cohen saves the flag', 1),
         ('the quakeress', 1),
         ('the restless spirit', 1),
         ('the telephone girl and the lady', 1),
         ('a busy day', 1),
         ('the call of the north', 1),
         ('caught in a cabaret', 1),
         ('caught in the rain', 1),
         ('cruel, cruel love', 1),
         ('dough and dynamite', 1),
         ('the escape', 1),
         ('gentlemen of nerve', 1),
         ('getting acquainted', 1),
         ("his father's rifle", 1),
         ('the knockout', 1),
         ('mabel at the wheel', 1),
         ("mabel's blunder", 1),
         ("mabel's married life", 1),
         ("mabel's strange predicament", 1),
         ('making a living', 1),
         ('the masquerader', 1),
         ('the new janitor', 1),
         ('the property man', 1),
         ('rose of the rancho', 1),
```

Loading [MathJax]/extensions/Safe.js

```
('the rounders', 1),
('salomy jane', 1),
('should a woman divorce?', 1),
('those love pangs', 1),
("tillie's punctured romance", 1),
('the wrath of the gods', 1),
('after five', 1),
('carmen', 1),
('enoch arden', 1),
('a fool there was', 1),
('four feathers', 1),
('fresh from the farm', 1),
('giving them fits', 1),
('love, loot and crash', 1),
('the soul of broadway', 1),
('the toast of death', 1),
('the warrens of virginia', 1),
('work', 1),
('the abandonment', 1),
('the bondman', 1),
('civilization', 1),
('a daughter of the gods', 1),
('divorce and the daughter', 1),
('hulda from holland', 1),
('lights of new york', 1),
('manhattan madness', 1),
("the soul's cycle", 1),
('the stepping stone', 1),
('under two flags', 1),
("youth's endearing charm", 1),
('all aboard', 1),
('bucking broadway', 1),
('camille', 1),
("cheyenne's pal", 1),
('the devil-stone', 1),
('easy street', 1),
('golden rule kate', 1),
('great expectations', 1),
('her right to live', 1),
('his wedding night', 1),
('the immigrant', 1),
('in again, out again', 1),
('the jury of fate', 1),
('the man without a country', 1),
('a modern musketeer', 1),
('oh doctor!', 1),
("peggy, the will o' the wisp", 1),
('the poor little rich girl', 1),
('reaching for the moon', 1),
('rebecca of sunnybrook farm', 1),
('a reckless romeo', 1),
('the rough house', 1),
('souls in pawn', 1),
('straight shooting', 1),
('a strange transgressor', 1),
('the woman god forgot', 1),
('all woman', 1),
('amarilly of clothes-line alley', 1),
('are crooks dishonest?', 1),
('believe me, xantippe', 1),
("the bride's awakening", 1),
('cecilia of the pink roses', 1),
('the cook', 1),
('the craving', 1),
```

```
('the deciding kiss', 1),
('face value', 1),
('the forbidden city', 1),
('fuss and feathers', 1),
('the goddess of lost lake', 1),
('he comes up smiling', 1),
('hell bent', 1),
('the legion of death', 1),
('moonshine', 1),
('the non-stop kid', 1),
('old wives for new', 1),
('out west', 1),
('the romance of tarzan', 1),
('salomé', 1),
('the scarlet drop', 1),
('set free', 1),
('shoulder arms', 1),
('stella maris', 1),
('tarzan of the apes', 1),
('till i come back to you', 1),
('under the yoke', 1),
("we can't have everything", 1),
('when do we eat?', 1),
('which woman?', 1),
('the whispering chorus', 1),
('ask father', 1),
('better times', 1),
('bolshevism on trial', 1),
('bumping into broadway', 1),
('daddy-long-legs', 1),
('dangerous hours', 1),
("don't change your husband", 1),
('the echo of youth', 1),
('the exquisite thief', 1),
('the false faces', 1),
('for better, for worse', 1),
('from hand to mouth', 1),
('the girl who stayed at home', 1),
('the grim game', 1),
("a gun fightin' gentleman", 1),
('happy though married', 1),
('the haunted bedroom', 1),
('hawthorne of the u.s.a.', 1),
("heart o' the hills", 1),
('his majesty, the american', 1),
('the law of men', 1),
('male and female', 1),
("a man's fight", 1),
('the miracle of love', 1),
('pistols for breakfast', 1),
('the probation wife', 1),
('ravished armenia', 1),
('the roaring road', 1),
('a romance of happy valley', 1),
('scarlet days', 1),
('a society exile', 1),
('soldiers of fortune', 1),
('sunnyside', 1),
('the test of honor', 1),
('true heart susie', 1),
('the turn in the road', 1),
('the unpainted woman', 1),
('the virtuous thief', 1),
('what every woman learns', 1),
```

```
('when the clouds roll by', 1),
('the white heather', 1),
('the world and its woman', 1),
('yankee doodle in berlin', 1),
("you're fired", 1),
('april folly', 1),
('blind youth', 1),
('a child for sale', 1),
('convict 13', 1),
("the devil's pass key", 1),
('the false road', 1),
('the family honor', 1),
('the girl in number 29', 1),
('hairpins', 1),
('haunted spooks', 1),
("her husband's friend", 1),
('high and dizzy', 1),
('huckleberry finn', 1),
('the idol dancer', 1),
('in the heart of a fool', 1),
('the jack-knife man', 1),
("judy of rogue's harbor", 1),
("lady rose's daughter", 1),
('the love flower', 1),
('number, please?', 1),
('pollyanna', 1),
('remodeling her husband', 1),
('the round-up', 1),
('the saphead', 1),
('sex', 1),
('shore acres', 1),
('silk hosiery', 1),
('silk husbands and calico wives', 1),
('the slim princess', 1),
('something to think about', 1),
('a splendid hazard', 1),
('suds', 1),
('the virgin of stamboul', 1),
('why change your wife?', 1),
('within our gates', 1),
('the woman in the suitcase', 1),
('the affairs of anatol', 1),
('among those present', 1),
('the big punch', 1),
('the blot', 1),
('the boat', 1),
('experience', 1),
('forbidden fruit', 1),
('forever', 1),
('the four horsemen of the apocalypse', 1),
('the goat', 1),
('the haunted house', 1),
('heedless moths', 1),
('i do', 1),
('the lotus eater', 1),
('the love light', 1),
('love never dies', 1),
('the lucky dog', 1),
("mother o' mine", 1),
('never weaken', 1),
('no woman knows', 1),
('the nut', 1),
('orphans of the storm', 1),
('the playhouse', 1),
```

```
('roads of destiny', 1),
('a sailor-made man', 1),
('the sea lion', 1),
('seven years bad luck', 1),
('sheltered daughters', 1),
('the silver lining', 1),
('the sky pilot', 1),
('sure fire', 1),
("tol'able david", 1),
('uncharted seas', 1),
("woman's place", 1),
('beyond the rocks', 1),
('a blind bargain', 1),
('daydreams', 1),
('dr. jack', 1),
('dusk to dawn', 1),
('the electric house', 1),
('fair lady', 1),
('foolish wives', 1),
('the loaded door', 1),
('moran of the lady letty', 1),
('pay day', 1),
('peacock alley', 1),
("peg o' my heart", 1),
("ridin' wild", 1),
('trifling women', 1),
('wildness of youth', 1),
('the woman he married', 1),
('the young rajah', 1),
("ambrose applejohn's adventure", 1),
('ashes of vengeance', 1),
('the balloonatic', 1),
('black oxen', 1),
('the call of the canyon', 1),
('the daring years', 1),
('the extra girl', 1),
('gentle julia', 1),
('our hospitality', 1),
('the pilgrim', 1),
('safety last!', 1),
('the shock', 1),
('the shriek of araby', 1),
('the silent command', 1),
('slippy mcgee', 1),
('souls for sale', 1),
('three ages', 1),
('why worry?', 1),
('the arab', 1),
('circe, the enchantress', 1),
('the enchanted cottage', 1),
('feet of clay', 1),
('greed', 1),
('he who gets slapped', 1),
("helen's babies", 1),
('her night of romance', 1),
('his hour', 1),
('hot water', 1),
('the iron horse', 1),
("isn't life wonderful", 1),
('manhandled', 1),
('monsieur beaucaire', 1),
('the navigator', 1),
('one night in rome', 1),
('the red lily', 1),
```

```
('romola', 1),
('sherlock, jr.', 1),
('the shooting of dan mcgrew', 1),
('wild oranges', 1),
('the yankee consul', 1),
('the big parade', 1),
('cheaper to marry', 1),
('confessions of a queen', 1),
('the dark angel', 1),
('fifty-fifty', 1),
("hogan's alley", 1),
('kentucky pride', 1),
('the lady', 1),
("lady windermere's fan", 1),
('little annie rooney', 1),
('the lucky horseshoe', 1),
('the mad whirl', 1),
('the mystic', 1),
('old clothes', 1),
('phantom of the opera', 1),
('the rag man', 1),
('sally, irene and mary', 1),
('sally of the sawdust', 1),
('the salvation hunters', 1),
('seven chances', 1),
('smouldering fires', 1),
('stage struck', 1),
('time, the comedian', 1),
('too many kisses', 1),
('the tower of lies', 1),
('the trail rider', 1),
('wolf blood', 1),
('battling butler', 1),
("the better 'ole", 1),
('beverly of graustark', 1),
('the blackbird', 1),
('la bohème', 1),
('the boy friend', 1),
('cruise of the jasper b', 1),
('dance madness', 1),
('don juan', 1),
('ella cinders', 1),
('the exquisite sinner', 1),
('fine manners', 1),
('flesh and the devil', 1),
("for heaven's sake", 1),
('the gay deceiver', 1),
('the general', 1),
('the great k & a train robbery', 1),
('hands up!', 1),
('high steppers', 1),
('the joy girl', 1),
('kid boots', 1),
('looking for trouble', 1),
('the magician', 1),
('mare nostrum', 1),
("miss brewster's millions", 1),
('old ironsides', 1),
('private izzy murphy', 1),
('the sea beast', 1),
('so this is paris', 1),
("so's your old man", 1),
('the strong man', 1),
('the temptress', 1),
```

```
('torrent', 1),
('tramp, tramp, tramp', 1),
('the winning of barbara worth', 1),
('adam and evil', 1),
('children of divorce', 1),
('the enemy', 1),
('the fair co-ed', 1),
('for the love of mike', 1),
('foreign devils', 1),
('the garden of allah', 1),
('his first flame', 1),
('the kid brother', 1),
('knockout reilly', 1),
('a little journey', 1),
('london after midnight', 1),
('long pants', 1),
('the love of sunya', 1),
('the red mill', 1),
('resurrection', 1),
('the road to romance', 1),
('rookies', 1),
('the show', 1),
('sunrise: a song of two humans', 1),
('the way of all flesh', 1),
('why girls love sailors', 1),
('wings', 1),
('the yankee clipper', 1),
('4 devils', 1),
('the air circus', 1),
('the barker', 1),
('beware of bachelors', 1),
('the cardboard lover', 1),
('the cavalier', 1),
('the crimson city', 1),
('the crowd', 1),
('the divine woman', 1),
('the docks of new york', 1),
("don't marry", 1),
('doomsday', 1),
('drums of love', 1),
('dry martini', 1),
('forbidden hours', 1),
('four sons', 1),
('four walls', 1),
('the garden of eden', 1),
('glorious betsy', 1),
('the godless girl', 1),
('half a bride', 1),
('heart trouble', 1),
('a lady of chance', 1),
('the last command', 1),
('laugh, clown, laugh', 1),
('the law of the range', 1),
('the legion of the condemned', 1),
('lilac time', 1),
('the lion and the mouse', 1),
('love me and the world is mine', 1),
('the magnificent flirt', 1),
('the man who laughs', 1),
('the matinee idol', 1),
('the mating call', 1),
('my man', 1),
('naughty baby', 1),
("noah's ark", 1),
```

```
('on trial', 1),
('our dancing daughters', 1),
('the red dance', 1),
('red hair', 1),
('sadie thompson', 1),
('say it with sables', 1),
('sharp shooters', 1),
('show girl', 1),
('show people', 1),
('the singing fool', 1),
('the smart set', 1),
('so this is love?', 1),
('speedy', 1),
('steamboat bill, jr.', 1),
('stop that man!', 1),
('street angel', 1),
('that certain thing', 1),
('we faw down', 1),
('the wedding march', 1),
('west of zanzibar', 1),
('west point', 1),
('wife savers', 1),
('the wind', 1),
('a woman of affairs', 1),
('acquitted', 1),
('the aviator', 1),
('berth marks', 1),
('captain lash', 1),
('careers', 1),
('the case of lena smith', 1),
('children of the ritz', 1),
('coquette', 1),
('desert nights', 1),
('disraeli', 1),
('drag', 1),
('dynamite', 1),
('fancy baggage', 1),
('the forward pass', 1),
('the gamblers', 1),
('the great divide', 1),
('half marriage', 1),
('hallelujah!', 1),
('hardboiled rose', 1),
('his glorious night', 1),
('jazz heaven', 1),
('lady of the pavements', 1),
('the letter', 1),
('the locked door', 1),
('masked emotions', 1),
("new year's eve", 1),
('night parade', 1),
('one hysterical night', 1),
('our modern maidens', 1),
('the pagan', 1),
('queen kelly', 1),
('redskin', 1),
('the river', 1),
('seven faces', 1),
('seven footprints to satan', 1),
('show boat', 1),
('the sky hawk', 1),
("smilin' guns", 1),
('smiling irish eyes', 1),
('spite marriage', 1),
```

```
('the squall', 1),
('stark mad', 1),
('street girl', 1),
('synthetic sin', 1),
('this is heaven', 1),
('the trespasser', 1),
('under the southern cross', 1),
('the very idea', 1),
('wedding rings', 1),
('welcome danger', 1),
('where east is east', 1),
('why be good?', 1),
('wild orchids', 1),
('the wolf song', 1),
('woman trap', 1),
('the younger generation', 1),
('abraham lincoln', 1),
('animal crackers', 1),
('anna christie', 1),
('beau bandit', 1),
('behind the make-up', 1),
('the benson murder case', 1),
('the big house', 1),
('the big trail', 1),
('borrowed wives', 1),
('the boudoir diplomat', 1),
('the case of sergeant grischa', 1),
('check and double check', 1),
('courage', 1),
('the devil to pay!', 1),
("the devil's holiday", 1),
('the divorcee', 1),
('doughboys', 1),
('dumbbells in ermine', 1),
('extravagance', 1),
('the fall guy', 1),
('feet first', 1),
('the florodora girl', 1),
('for the defense', 1),
('general crack', 1),
('girl of the port', 1),
('the girl said no', 1),
('the green goddess', 1),
('grumpy', 1),
('half shot at sunrise', 1),
('he knew women', 1),
('hell harbor', 1),
("hell's heroes", 1),
('hide-out', 1),
('hit the deck', 1),
('hook, line and sinker', 1),
('hot curves', 1),
('isle of escape', 1),
("journey's end", 1),
('a lady surrenders', 1),
('a lady to love', 1),
('the lash', 1),
('laughter', 1),
('lawful larceny', 1),
("lovin' the ladies", 1),
('madam satan', 1),
('mamba', 1),
("the man from blankley's", 1),
('a man from wyoming', 1),
```

```
('the man hunter', 1),
('the matrimonial bed', 1),
('montana moon', 1),
('morocco', 1),
('mothers cry', 1),
("near the rainbow's end", 1),
('the office wife', 1),
('oh sailor behave', 1),
('oklahoma cyclone', 1),
('on your back', 1),
("one night at susie's", 1),
('the other tomorrow', 1),
('our blushing brides', 1),
('outside the law', 1),
('outward bound', 1),
('part time wife', 1),
('the pay-off', 1),
('playing around', 1),
("river's end", 1),
('road to paradise', 1),
('rough romance', 1),
('rough waters', 1),
('the runaway bride', 1),
('the sap from syracuse', 1),
('sarah and son', 1),
('the sea god', 1),
('second choice', 1),
('second wife', 1),
("she's my weakness", 1),
('the silver horde', 1),
('sin takes a holiday', 1),
('the social lion', 1),
('son of the gods', 1),
('soup to nuts', 1),
('street of chance', 1),
('such men are dangerous', 1),
('sweet mama', 1),
('the texan', 1),
('those three french girls', 1),
('tom sawyer', 1),
('troopers three', 1),
('the way of all men', 1),
('what men want', 1),
('wide open', 1),
('young eagles', 1),
('young man of manhattan', 1),
('24 hours', 1),
('alexander hamilton', 1),
('ambassador bill', 1),
("annabelle's affairs", 1),
('are these our children?', 1),
('arrowsmith', 1),
('the avenger', 1),
('bachelor apartment', 1),
('the bachelor father', 1),
('bad girl', 1),
('bad sister', 1),
('beau ideal', 1),
('behind office doors', 1),
('beyond victory', 1),
('the big shot', 1),
('blonde crazy', 1),
('body and soul', 1),
('border law', 1),
```

```
('born to love', 1),
('bought', 1),
('broadminded', 1),
('captain applejack', 1),
('caught cheating', 1),
('caught plastered', 1),
('chances', 1),
('the common law', 1),
('confessions of a co-ed', 1),
('consolation marriage', 1),
('cracked nuts', 1),
('daughter of the dragon', 1),
('devotion', 1),
('dirigible', 1),
("don't bet on women", 1),
('the easiest way', 1),
('east lynne', 1),
('east of borneo', 1),
("everything's rosie", 1),
('the false madonna', 1),
('fighting caravans', 1),
('the fighting marshal', 1),
('five star final', 1),
('flying high', 1),
('a free soul', 1),
('freighters of destiny', 1),
('friends and lovers', 1),
('the front page', 1),
("gentleman's fate", 1),
('girls about town', 1),
('goldie', 1),
('his woman', 1),
('a holy terror', 1),
('honor among lovers', 1),
('a house divided', 1),
('illicit', 1),
('inspiration', 1),
('it pays to advertise', 1),
('just a gigolo', 1),
('kept husbands', 1),
('ladies of the big house', 1),
('the lady refuses', 1),
('the last flight', 1),
('laugh and get rich', 1),
('laughing sinners', 1),
('local boy makes good', 1),
('lonely wives', 1),
('the magnificent lie', 1),
('the maltese falcon', 1),
('the man in possession', 1),
('man of the world', 1),
('mata hari', 1),
('men in her life', 1),
('millie', 1),
('the miracle woman', 1),
('monkey business', 1),
('mr. lemon of orange', 1),
('my sin', 1),
('the naughty flirt', 1),
('nice women', 1),
('night nurse', 1),
("other men's women", 1),
('the painted desert', 1),
('pardon us', 1),
```

('parlor, bedroom and bath', 1),
("peach o'reno", 1),
('platinum blonde', 1),
('possessed', 1),
('reducing', 1),
('the road to singapore', 1),
('the royal bed', 1),
('the runaround', 1),
('safe in hell', 1),
('salvation nell', 1),
('scandal sheet', 1),
('the sea ghost', 1),
('secret service', 1),
('seed', 1),
('the sin of madelon claudet', 1),
('the sin ship', 1),
('sky raiders', 1),
('son of india', 1),
('strangers may kiss', 1),
('street scene', 1),
('strictly dishonorable', 1),
('suicide fleet', 1),
('susan lenox (her fall and rise)', 1),
('sweepstakes', 1),
('tabu', 1),
('tarnished lady', 1),
('this modern age', 1),
('three who loved', 1),
('the tip-off', 1),
('tonight or never', 1),
('transgression', 1),
('traveling husbands', 1),
('the unholy garden', 1),
('up for murder', 1),
('the viking', 1),
('west of broadway', 1),
('the woman between', 1),
('women of all nations', 1),
('working girls', 1),
('the yellow ticket', 1),
("young donovan's kid", 1),
('20,000 years in sing sing', 1),
('after the ball', 1),
('after tomorrow', 1),
('the age of consent', 1),
('the all american', 1),
('almost married', 1),
('american madness', 1),
('the animal kingdom', 1),
('are you listening?', 1),
('as you desire me', 1),
('big city blues', 1),
('blessed event', 1),
('blonde venus', 1),
('broadway to cheyenne', 1),
('broken lullaby', 1),
('business and pleasure', 1),
('but the flesh is weak', 1),
('the cabin in the cotton', 1),
('call her savage', 1),
('cock of the air', 1),
('come on danger!', 1),
('the conquerors', 1),
('cowboy counsellor', 1),

```
('the crash', 1),
('the crusader', 1),
('cynara', 1),
('dancers in the dark', 1),
('daring danger', 1),
('devil and the deep', 1),
('the devil is driving', 1),
("devil's lottery", 1),
('disorderly conduct', 1),
('downstairs', 1),
('evenings for sale', 1),
('faithless', 1),
('fast life', 1),
('flowers and trees', 1),
('frisco jenny', 1),
('grand hotel', 1),
('the greeks had a word for them', 1),
('the heart of new york', 1),
("hell's highway", 1),
("hell's house", 1),
('heritage of the desert', 1),
('high pressure', 1),
('horse feathers', 1),
('hot saturday', 1),
('huddle', 1),
('i am a fugitive from a chain gang', 1),
('island of lost souls', 1),
("it's tough to be famous", 1),
('klondike', 1),
('kongo', 1),
('ladies of the jury', 1),
('lady with a past', 1),
('the last mile', 1),
('lawyer man', 1),
('life begins', 1),
('love in high gear', 1),
('madame butterfly', 1),
('make me a star', 1),
('manhattan tower', 1),
('the man called back', 1),
('man wanted', 1),
('the man who played god', 1),
('the match king', 1),
('me and my gal', 1),
('men of chance', 1),
('million dollar legs', 1),
('misleading lady', 1),
('the monster walks', 1),
('movie crazy', 1),
('mr. robinson crusoe', 1),
('my pal, the king', 1),
('new morals for old', 1),
('night world', 1),
('no man of her own', 1),
('no more orchids', 1),
('once in a lifetime', 1),
('one way passage', 1),
('out of singapore', 1),
('the painted woman', 1),
('the passionate plumber', 1),
('the phantom president', 1),
('police court', 1),
('polly of the circus', 1),
('rackety rax', 1),
```

```
('rain', 1),
('rasputin and the empress', 1),
('red dust', 1),
('red-headed woman', 1),
('the rich are always with us', 1),
('ride him, cowboy', 1),
("ridin' for justice", 1),
('rockabye', 1),
('scarlet dawn', 1),
('silver dollar', 1),
('sinners in the sun', 1),
('sky bride', 1),
('sky devils', 1),
('skyscraper souls', 1),
("smilin' through", 1),
('speak easily', 1),
("state's attorney", 1),
('the stoker', 1),
('the strange case of clara deane', 1),
('strange interlude', 1),
('a successful calamity', 1),
('symphony of six million', 1),
('tarzan the ape man', 1),
('taxi!', 1),
('the tenderfoot', 1),
('texas cyclone', 1),
('they call it sin', 1),
('they just had to get married', 1),
('they never come back', 1),
('thirteen women', 1),
('this is the night', 1),
('tiger shark', 1),
('tom brown of culver', 1),
('two against the world', 1),
('two-fisted law', 1),
('unashamed', 1),
('union depot', 1),
('uptown new york', 1),
('virtue', 1),
('washington merry-go-round', 1),
('westward passage', 1),
('the wet parade', 1),
('what price hollywood?', 1),
('while paris sleeps', 1),
('white zombie', 1),
('you said a mouthful', 1),
('after tonight', 1),
('aggie appleby, maker of men', 1),
('air hostess', 1),
("ann carver's profession", 1),
('ann vickers', 1),
('baby face', 1),
('the barbarian', 1),
('beauty for sale', 1),
('before dawn', 1),
('blondie johnson', 1),
('brief moment', 1),
('broadway bad', 1),
('bureau of missing persons', 1),
('by candlelight', 1),
('chance at heaven', 1),
('child of manhattan', 1),
('christopher bean', 1),
('christopher strong', 1),
```

```
('the cohens and kellys in trouble', 1),
('college coach', 1),
('convention city', 1),
('counsellor at law', 1),
('damaged lives', 1),
('deluge', 1),
("the devil's brother", 1),
('diplomaniacs', 1),
("don't bet on love", 1),
("dora's dunking doughnuts", 1),
('duck soup', 1),
('elmer, the great', 1),
('the emperor jones', 1),
("employees' entrance", 1),
('eskimo', 1),
('ex-lady', 1),
('fast workers', 1),
('flying devils', 1),
('forgotten', 1),
('golden harvest', 1),
('goodbye love', 1),
('havana widows', 1),
('headline shooter', 1),
('heroes for sale', 1),
('his private secretary', 1),
('i cover the waterfront', 1),
('i have lived', 1),
('i loved a woman', 1),
('if i were free', 1),
('infernal machine', 1),
('king of the arena', 1),
("the king's vacation", 1),
('lady for a day', 1),
('laughing at life', 1),
('laughter in hell', 1),
('the life of jimmy dolan', 1),
('looking forward', 1),
('the mad game', 1),
('the man from monterey', 1),
('man of the forest', 1),
('meet the baron', 1),
('midnight mary', 1),
('the midnight patrol', 1),
('mr. broadway', 1),
('murders in the zoo', 1),
('mystery of the wax museum', 1),
('night flight', 1),
('night of terror', 1),
('no other woman', 1),
('only yesterday', 1),
('our betters', 1),
('parachute jumper', 1),
('parole girl', 1),
('the past of mary holmes', 1),
('picture brides', 1),
('picture snatcher', 1),
('pleasure cruise', 1),
('the prizefighter and the lady', 1),
('queen christina', 1),
('sagebrush trail', 1),
("saturday's millions", 1),
('the secret of madame blanche', 1),
('shanghai madness', 1),
('she had to say yes', 1),
```

```
('the silk express', 1),
('skyway', 1),
('so this is africa', 1),
('the solitaire man', 1),
('the son of kong', 1),
('sons of the desert', 1),
('stage mother', 1),
('the story of temple drake', 1),
('supernatural', 1),
('sweepings', 1),
('three-cornered moon', 1),
('three little pigs', 1),
('the thrill hunter', 1),
('tillie and gus', 1),
('to the last man', 1),
('topaze', 1),
('turn back the clock', 1),
("the warrior's husband", 1),
('what! no beer?', 1),
('when ladies meet', 1),
('wild boys of the road', 1),
('the working man', 1),
('zoo in budapest', 1),
('affairs of a gentleman', 1),
('as the earth turns', 1),
('beggars in ermine', 1),
('the captain hates the sea', 1),
("the cat's-paw", 1),
('cheating cheaters', 1),
('cockeyed cavaliers', 1),
('college rhythm', 1),
('come on marines!', 1),
('crimson romance', 1),
('cross country cruise', 1),
('cross streets', 1),
('double door', 1),
('the dover road', 1),
('easy to love', 1),
('eight girls in a boat', 1),
('embarrassing moments', 1),
...]
```

In [15]:
```python
movie_df[movie_df["Title"] == "godzilla"]
```

Out[15]:

	Title	Genre	Plot
10485	godzilla	action	In the 1960s, a marine iguana nest is exposed ...
13045	godzilla	action	In 1954, Godzilla, an ancient alpha predator, ...
20496	godzilla	action	When the Japanese freighter Eiko-maru is destr...

In [16]:
```python
movie_df = movie_df[["Title", "Plot", "Genre"]]
movie_df.columns = ["title", "text", "label_name"]
movie_df
```

	title	text	label_name
0	the great train robbery	The film opens with two bandits breaking into ...	action
1	the suburbanite	The film is about a family who move to the sub...	comedy
2	how brown saw the baseball game	Before heading out to a baseball game at a nea...	comedy
5	the adventures of dollie	On a beautiful summer day a father and mother ...	drama
6	the black viper	A thug accosts a girl as she leaves her workpl...	drama
...	...	...	...
21223	selam	The film opens with a Senegalese boy named Kha...	drama
21224	particle (film)	Zeynep lost her job at weaving factory, and he...	drama
21225	çalgı çengi ikimiz	Two musicians, Salih and Gürkan, described the...	comedy
21226	olanlar oldu	Zafer, a sailor living with his mother Döndü i...	comedy
21227	istanbul kırmızısı	The writer Orhan Şahin returns to İstanbul aft...	romance

16584 rows × 3 columns

In [17]:
```python
Counter(movie_df["label_name"]).most_common()
```

Out[17]:
```
[('drama', 5579),
 ('comedy', 4160),
 ('action', 3463),
 ('thriller', 1925),
 ('romance', 884),
 ('animation', 573)]
```

In [18]:
```python
labels = list(set(movie_df["label_name"]))
labels
```

Out[18]:
```
['thriller', 'animation', 'action', 'romance', 'comedy', 'drama']
```

In [19]:
```python
id2label = {k:v for k,v in enumerate(labels)}
label2id = {v:k for k,v in enumerate(labels)}
label2id
```

Out[19]:
```
{'thriller': 0,
 'animation': 1,
 'action': 2,
 'romance': 3,
 'comedy': 4,
 'drama': 5}
```

In [20]:
```python
movie_df["label"] = movie_df["label_name"].apply(lambda x: label2id[x])
movie_df
```

Loading [MathJax]/extensions/Safe.js

	title	text	label_name	label
0	the great train robbery	The film opens with two bandits breaking into ...	action	2
1	the suburbanite	The film is about a family who move to the sub...	comedy	4
2	how brown saw the baseball game	Before heading out to a baseball game at a nea...	comedy	4
5	the adventures of dollie	On a beautiful summer day a father and mother ...	drama	5
6	the black viper	A thug accosts a girl as she leaves her workpl...	drama	5
...	...	...	...	...
21223	selam	The film opens with a Senegalese boy named Kha...	drama	5
21224	particle (film)	Zeynep lost her job at weaving factory, and he...	drama	5
21225	çalgı çengi ikimiz	Two musicians, Salih and Gürkan, described the...	comedy	4
21226	olanlar oldu	Zafer, a sailor living with his mother Döndü i...	comedy	4
21227	istanbul kırmızısı	The writer Orhan Şahin returns to İstanbul aft...	romance	3

16584 rows × 4 columns

Creating train, dev, test split

In [21]:
```python
from sklearn.model_selection import train_test_split

df_train, df_tmp = train_test_split(
        movie_df,stratify=movie_df["label"], test_size=0.2)

df_val, df_test = train_test_split(
        df_tmp,stratify=df_tmp["label"], test_size=0.5)
```

In [22]:
```python
from datasets import Features, ClassLabel, Value, Dataset, DatasetDict

ds_features = Features({"text": Value("string"), "label": ClassLabel(names=labels)})

dataset = DatasetDict({
    "train": Dataset.from_pandas(df_train.reset_index(drop=True),features=ds_features),
    "valid": Dataset.from_pandas(df_val.reset_index(drop=True),features=ds_features),
    "test": Dataset.from_pandas(df_test.reset_index(drop=True),features=ds_features)})

dataset
```

Out[22]:
```
DatasetDict({
    train: Dataset({
        features: ['text', 'label'],
        num_rows: 13267
    })
    valid: Dataset({
        features: ['text', 'label'],
        num_rows: 1658
    })
    test: Dataset({
        features: ['text', 'label'],
        num_rows: 1659
    })
})
```

Count Baseline

Loading [MathJax]/extensions/Safe.js

To start off with train a simple count baseline, to see how well we perform if we only look at the words of the training dataset.

In [23]:
```python
from nltk.corpus import stopwords
```

In [24]:
```python
stopwords_en = stopwords.words("english")
stopwords_en
```

```
Out[24]:  ['i',
           'me',
           'my',
           'myself',
           'we',
           'our',
           'ours',
           'ourselves',
           'you',
           "you're",
           "you've",
           "you'll",
           "you'd",
           'your',
           'yours',
           'yourself',
           'yourselves',
           'he',
           'him',
           'his',
           'himself',
           'she',
           "she's",
           'her',
           'hers',
           'herself',
           'it',
           "it's",
           'its',
           'itself',
           'they',
           'them',
           'their',
           'theirs',
           'themselves',
           'what',
           'which',
           'who',
           'whom',
           'this',
           'that',
           "that'll",
           'these',
           'those',
           'am',
           'is',
           'are',
           'was',
           'were',
           'be',
           'been',
           'being',
           'have',
           'has',
           'had',
           'having',
           'do',
           'does',
           'did',
           'doing',
           'a',
           'an',
           'the',
           'and',
```

```
'but',
'if',
'or',
'because',
'as',
'until',
'while',
'of',
'at',
'by',
'for',
'with',
'about',
'against',
'between',
'into',
'through',
'during',
'before',
'after',
'above',
'below',
'to',
'from',
'up',
'down',
'in',
'out',
'on',
'off',
'over',
'under',
'again',
'further',
'then',
'once',
'here',
'there',
'when',
'where',
'why',
'how',
'all',
'any',
'both',
'each',
'few',
'more',
'most',
'other',
'some',
'such',
'no',
'nor',
'not',
'only',
'own',
'same',
'so',
'than',
'too',
'very',
's',
't',
```

```
        'can',
        'will',
        'just',
        'don',
        "don't",
        'should',
        "should've",
        'now',
        'd',
        'll',
        'm',
        'o',
        're',
        've',
        'y',
        'ain',
        'aren',
        "aren't",
        'couldn',
        "couldn't",
        'didn',
        "didn't",
        'doesn',
        "doesn't",
        'hadn',
        "hadn't",
        'hasn',
        "hasn't",
        'haven',
        "haven't",
        'isn',
        "isn't",
        'ma',
        'mightn',
        "mightn't",
        'mustn',
        "mustn't",
        'needn',
        "needn't",
        'shan',
        "shan't",
        'shouldn',
        "shouldn't",
        'wasn',
        "wasn't",
        'weren',
        "weren't",
        'won',
        "won't",
        'wouldn',
        "wouldn't"]
```

In [25]:
```python
from nltk.tokenize import word_tokenize

word_tokenize("This is just a test.")
```

Out[25]: `['This', 'is', 'just', 'a', 'test', '.']`

In [26]:
```python
most_common_words = {}

for l in labels:
    tokenized_text = df_train[df_train["label_name"] == l]["text"].str.lower().apply(
        lambda x: [e for e in word_tokenize(x) if e not in stopwords_en and len(e) > 2]
```

```
        most_common_words[l] = Counter(tokenized_text.explode()).most_common()
```

```
In [27]: for k,v in most_common_words.items():
             print(k,"\n",v[:15])
             print("="*120)
```

```
thriller
 [('one', 1684), ('house', 1640), ('back', 1357), ('finds', 1312), ('police', 1259), ('t
ells', 1246), ('find', 1156), ('two', 1130), ('killed', 1120), ('later', 1119), ('man',
1105), ('death', 1061), ('home', 1051), ('goes', 977), ('car', 969)]
========================================================================================
==============================
animation
 [('tom', 1404), ('jerry', 991), ('bugs', 856), ('back', 701), ('one', 458), ('however',
390), ('time', 368), ('two', 360), ('away', 348), ('find', 347), ('tries', 320), ('cat',
305), ('tells', 305), ('takes', 286), ('new', 286)]
========================================================================================
==============================
action
 [('one', 2508), ('back', 1976), ('men', 1878), ('two', 1722), ('killed', 1656), ('fathe
r', 1577), ('man', 1474), ('tells', 1433), ('kill', 1421), ('later', 1397), ('police', 1
383), ('also', 1361), ('find', 1323), ('time', 1256), ('new', 1240)]
========================================================================================
==============================
romance
 [('love', 1191), ('father', 710), ('one', 497), ('family', 483), ('get', 456), ('tell
s', 439), ('life', 435), ('day', 417), ('back', 408), ('time', 392), ('gets', 386), ('mo
ther', 378), ('marriage', 369), ('comes', 359), ('friend', 357)]
========================================================================================
==============================
comedy
 [('one', 2357), ('back', 2143), ('get', 2085), ('new', 2047), ('two', 1913), ('father',
1729), ('home', 1711), ('tells', 1702), ('also', 1509), ('love', 1501), ('man', 1496),
('house', 1495), ('time', 1465), ('find', 1458), ('money', 1445)]
========================================================================================
==============================
drama
 [('one', 3299), ('father', 3277), ('home', 2572), ('tells', 2490), ('mother', 2469),
('family', 2427), ('life', 2419), ('back', 2356), ('man', 2239), ('love', 2234), ('new',
2218), ('two', 2199), ('wife', 2082), ('also', 2000), ('son', 1974)]
========================================================================================
==============================
```

```
In [28]: most_common_words_10k = {}
         for k,v in most_common_words.items():
             most_common_words_10k[k] = [word for word,count in v][:10000]

         most_common_words_10k["romance"][:3], len(most_common_words_10k["romance"])
```

```
Out[28]: (['love', 'father', 'one'], 10000)
```

```
In [29]: idx = 3
         df_val.iloc[idx]
```

```
Out[29]: title                                          raksha
         text            Rajeev (Jagapati Babu) owner of Raksha Constru...
         label_name                                    thriller
         label                                                0
         Name: 20164, dtype: object
```

```
In [30]: for genre, words in most_common_words_10k.items():
             print(genre)
```

Loading [MathJax]/extensions/Safe.js

```
        print(len(list(set(word_tokenize(df_val.iloc[idx].text)) & set(words))))
        print("="*10)
```

```
thriller
90
==========
animation
88
==========
action
87
==========
romance
89
==========
comedy
88
==========
drama
89
==========
```

In [31]:
```python
import numpy as np
from random import randint

def count_classifier(row):
    len_intersection = []

    for l in labels:
        len_intersection.append(len(list(set(word_tokenize(row)) & set(most_common_words

    winner = list(np.argwhere(len_intersection == np.amax(len_intersection)).flatten())

    if len(winner) == 1:
        pred = winner[0]
    else:
        pred = winner[randint(0,len(winner)-1)]

    return pred
```

In [32]:
```python
df_val["count_clf_pred"] = df_val["text"].apply(count_classifier)
df_val
```

Loading [MathJax]/extensions/Safe.js

Out[32]:

	title	text	label_name	label	count_clf_pred
18779	maleyali jotheyali	Preetham is the only son of a wealthy but supe...	romance	3	4
2976	ding dong williams	Hollywood's Sunrise Studios is producing a fil...	comedy	4	4
3335	on our merry way	Oliver Pease (Burgess Meredith) has deceived h...	comedy	4	4
20164	raksha	Rajeev (Jagapati Babu) owner of Raksha Constru...	thriller	0	0
10935	amy's orgasm	Amy is a single 29 year old Jewish woman. She ...	comedy	4	4
...	...	...	...	...	...
892	annabelle's affairs	After they are separated shortly after their m...	comedy	4	2
7374	honky tonk freeway	In a small Florida tourist town named Ticlaw, ...	comedy	4	4
19757	sawaari	The film starts with a murder mystery. The mai...	thriller	0	4
15511	dracula ad 1972	A century later, a new generation of Britons a...	thriller	0	0
6021	the navy vs. the night monsters	The dull, workaday life at the small American ...	action	2	2

1658 rows × 5 columns

In [33]:
```python
# From here: https://github.com/nlp-with-transformers/notebooks/blob/main/02_classificat

from sklearn.metrics import classification_report
from sklearn.metrics import ConfusionMatrixDisplay, confusion_matrix
import matplotlib.pyplot as plt

def plot_confusion_matrix(y_preds, y_true, labels):
    cm = confusion_matrix(y_true, y_preds, normalize="true")
    fig, ax = plt.subplots(figsize=(6, 6))
    disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=labels)
    disp.plot(cmap="Blues", values_format=".2f", ax=ax, colorbar=False)
    plt.title("Normalized confusion matrix")
    plt.show()
```
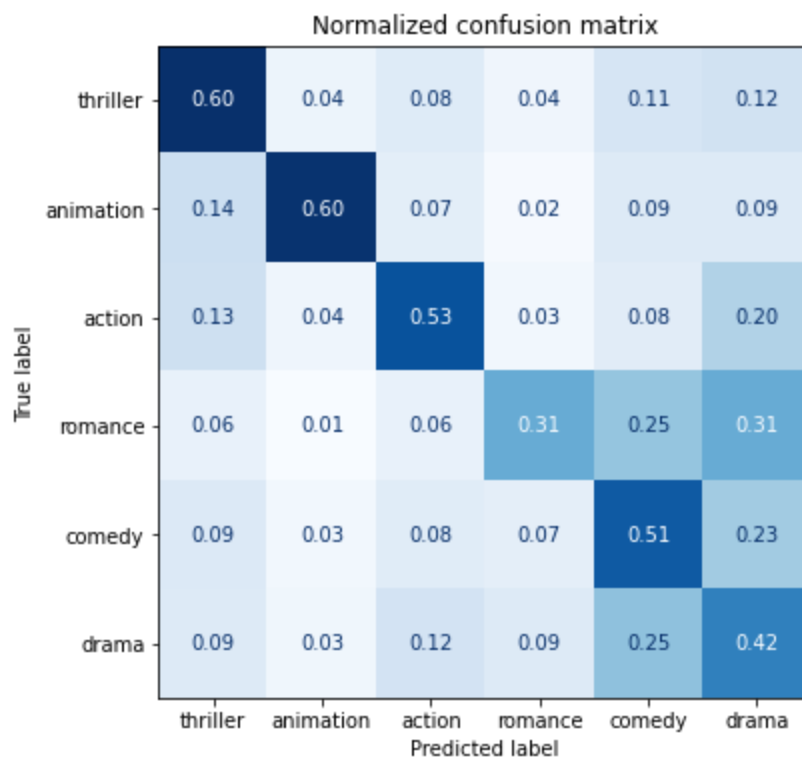
In [34]:
```python
print(classification_report(list(df_val["label"]),df_val["count_clf_pred"], target_names
```

```
              precision    recall  f1-score   support

    thriller       0.44      0.60      0.51       192
   animation       0.40      0.60      0.48        57
      action       0.59      0.53      0.56       346
     romance       0.23      0.31      0.27        89
      comedy       0.49      0.51      0.50       416
       drama       0.52      0.42      0.46       558

    accuracy                           0.49      1658
   macro avg       0.45      0.49      0.46      1658
weighted avg       0.50      0.49      0.49      1658
```

In [35]:
```python
plot_confusion_matrix(df_val["count_clf_pred"], list(df_val["label"]), labels)
```

Loading [MathJax]/extensions/Safe.js

Normalized confusion matrix

	thriller	animation	action	romance	comedy	drama
thriller	0.60	0.04	0.08	0.04	0.11	0.12
animation	0.14	0.60	0.07	0.02	0.09	0.09
action	0.13	0.04	0.53	0.03	0.08	0.20
romance	0.06	0.01	0.06	0.31	0.25	0.31
comedy	0.09	0.03	0.08	0.07	0.51	0.23
drama	0.09	0.03	0.12	0.09	0.25	0.42

True label / Predicted label

Transformer Model

Now we are training a more complex model, namely one based on the Transformer archicture. Some the code is based on the excellent NLP with Transformers book, in particular chapter 2: https://github.com/nlp-with-transformers/notebooks/blob/main/02_classification.ipynb.

```python
In [36]: from transformers import AutoTokenizer

model_ckpt = "distilbert-base-uncased"
tokenizer = AutoTokenizer.from_pretrained(model_ckpt)
```

```python
In [37]: def tokenize(batch):
    return tokenizer(batch["text"], padding=True, truncation=True)
```

```python
In [38]: dataset_encoded = dataset.map(tokenize, batched=True, batch_size=None)
```

```
  0%|          | 0/1 [00:00<?, ?ba/s]
  0%|          | 0/1 [00:00<?, ?ba/s]
  0%|          | 0/1 [00:00<?, ?ba/s]
```

```python
In [39]: print(dataset_encoded["train"].column_names)
```

```
['text', 'label', 'input_ids', 'attention_mask']
```

```python
In [40]: import torch
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
```

```python
In [41]: from transformers import AutoModelForSequenceClassification, AutoConfig

num_labels = len(labels)

config = (AutoConfig
          .from_pretrained(model_ckpt, num_labels=num_labels,
                           label2id=label2id, id2label=id2label))
```

ModelForSequenceClassification

Loading [MathJax]/extensions/Safe.js

```
        .from_pretrained(model_ckpt, config=config)
        .to(device))
```

In [42]:
```python
from sklearn.metrics import accuracy_score, f1_score

def compute_metrics(pred):
    labels = pred.label_ids
    preds = pred.predictions.argmax(-1)
    f1 = f1_score(labels, preds, average="weighted")
    acc = accuracy_score(labels, preds)
    return {"accuracy": acc, "f1": f1}
```

In [43]:
```python
from transformers import Trainer, TrainingArguments

batch_size = 16
logging_steps = len(dataset_encoded["train"]) // batch_size
model_name = f"{model_ckpt}-finetuned"
training_args = TrainingArguments(output_dir=model_name,
                                  num_train_epochs=4,
                                  learning_rate=2e-5,
                                  per_device_train_batch_size=batch_size,
                                  per_device_eval_batch_size=batch_size,
                                  weight_decay=0.01,
                                  evaluation_strategy="epoch",
                                  disable_tqdm=False,
                                  logging_steps=logging_steps,
#                                   push_to_hub=True,
                                  log_level="error",
                                  save_total_limit=1)
```

In [44]:
```python
from transformers import Trainer

trainer = Trainer(model=model, args=training_args,
                  compute_metrics=compute_metrics,
                  train_dataset=dataset_encoded["train"],
                  eval_dataset=dataset_encoded["valid"],
                  tokenizer=tokenizer)
trainer.train();
```

Loading [MathJax]/extensions/Safe.js

Epoch	Training Loss	Validation Loss	Accuracy	F1
1	1.049600	0.871576	0.673703	0.672098
2	0.732500	0.818706	0.700241	0.698558
3	0.546600	0.845013	0.711098	0.710802
4	0.408600	0.900630	0.712907	0.711442

In [45]:
```python
preds_output = trainer.predict(dataset_encoded["valid"])
```
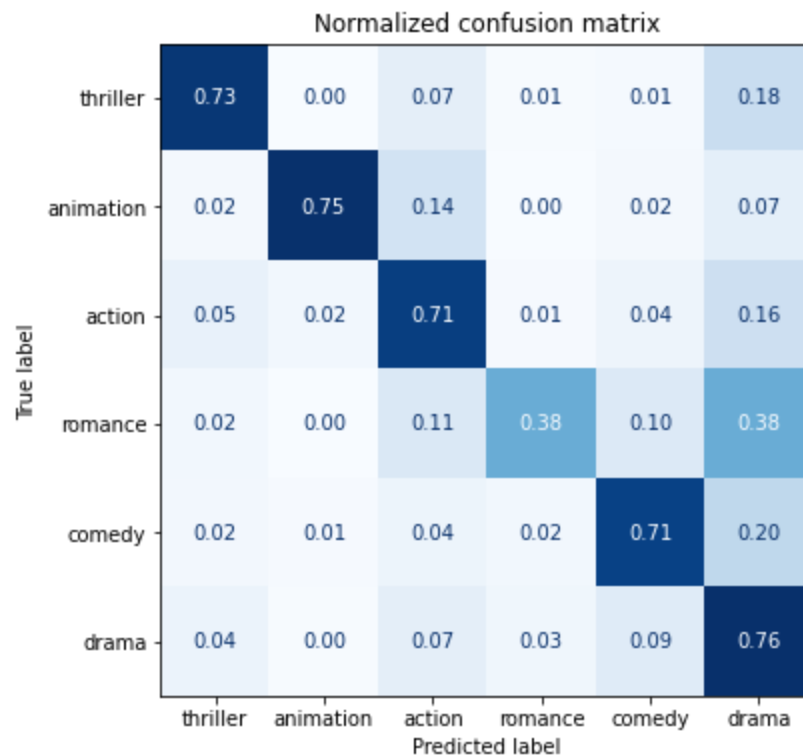
[104/104 00:11]

In [46]:
```python
preds_output.metrics
```

Out[46]:
```
{'test_loss': 0.9006300568580627,
 'test_accuracy': 0.7129071170084439,
 'test_f1': 0.711441744064859,
 'test_runtime': 11.189,
 'test_samples_per_second': 148.182,
 'test_steps_per_second': 9.295}
```

In [47]:
```python
y_preds = np.argmax(preds_output.predictions, axis=1)
y_valid = np.array(df_val["label"])
```

In [48]:
```python
plot_confusion_matrix(y_preds, y_valid, labels)
```

Normalized confusion matrix

	thriller	animation	action	romance	comedy	drama
thriller	0.73	0.00	0.07	0.01	0.01	0.18
animation	0.02	0.75	0.14	0.00	0.02	0.07
action	0.05	0.02	0.71	0.01	0.04	0.16
romance	0.02	0.00	0.11	0.38	0.10	0.38
comedy	0.02	0.01	0.04	0.02	0.71	0.20
drama	0.04	0.00	0.07	0.03	0.09	0.76

In [49]:
```python
print(classification_report(y_valid,y_preds, target_names=labels))
```

```
              precision    recall  f1-score   support

    thriller       0.72      0.73      0.73       192
   animation       0.77      0.75      0.76        57
      action       0.73      0.71      0.72       346
     romance       0.54      0.38      0.45        89
      comedy       0.79      0.71      0.75       416
       drama       0.67      0.76      0.71       558

    accuracy                           0.71      1658
   macro avg       0.70      0.67      0.69      1658
weighted avg       0.71      0.71      0.71      1658
```

In [61]: `df_train[df_train["text"].str.contains("Jane Foster")]`

Out[61]:

	title	text	label_name	label
12603	thor	In 965 AD, Odin, king of Asgard, wages war aga...	action	2
12961	thor: the dark world	Eons ago, Bor, father of Odin, clashes with th...	action	2

Inference

In [51]:
```python
text = "Thor enlists the help of Valkyrie, Korg and ex-girlfriend Jane Foster to fight G

encoding = tokenizer(text,  padding=True, truncation=True, return_tensors="pt")
encoding = {k: v.to(trainer.model.device) for k,v in encoding.items()}

outputs = trainer.model(**encoding)
```

In [52]:
```python
logits = outputs.logits
logits.shape
```

Out[52]: `torch.Size([1, 6])`

In [53]:
```python
predictions = torch.nn.functional.softmax(logits.squeeze().cpu(), dim=-1)
predictions
```

Out[53]:
```
tensor([0.0344, 0.0048, 0.9169, 0.0028, 0.0301, 0.0111],
        grad_fn=<SoftmaxBackward0>)
```

In [54]: `id2label[int(predictions.argmax())]`

Out[54]: `'action'`

Loading [MathJax]/extensions/Safe.js