

clustering_report

September 28, 2020

1 Goal

This goal of this assignment was to determine similar movies to recommend using the other ratings and tags. The were data sets provided split. The first task was the merge the data sets to have one dataframe that included `title`, `rating`, and `movieId`. In the notebook `feature_engineering_tags` I reconstructed the `tag` feature to be able to reason about it. In this notebook, I assigned an `polarity` to the tag and merged the polarity with the dataframe. I thought this could help combine movies of similar style of tags without having to vectorize the tag.

I then created a cluster of `polarity` vs `tag`. I created a `KMeans` function to fit the data and used the elbow method to detmine how many clusters were needed for K-Means. The number of clusters was 10.

I relabeld the cluster wiht the movie it matches and created a function that produces movies from the same cluster when asked for a recommendation.

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import nltk
# nltk.download('stopwords')
from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS
from nltk.tokenize import RegexpTokenizer
from nltk.stem.porter import PorterStemmer
import re
import string
from sklearn import preprocessing
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.utils import shuffle
from sklearn import tree
from sklearn.feature_extraction.text import CountVectorizer
from sklearn import metrics
from sklearn.cluster import KMeans
import random
```

First, import the `tags_trimmed` file which is a result of from `feature_engineering_tags`.

```
[2]: movies = pd.read_csv('~Documents/EECS/EECS_731/HW/EECS731_3/data/tags_trimmed.
    ↪csv')
```

```
[3]: moives = movies.drop(columns='Unnamed: 0')
```

Next, I used texblob to assign catagories for polarity.

```
[4]: from textblob import TextBlob
movies[['polarity', 'subjectivity']] = movies['tag'].apply(lambda Text: pd.
    ↪Series(TextBlob(Text).sentiment))
```

```
[5]: movies_rp = movies.loc[:, movies.columns.
    ↪intersection(['movieId', 'rating', 'polarity'])]
movies_rpt = movies.loc[:, movies.columns.
    ↪intersection(['movieId', 'title', 'rating', 'polarity'])]
movies_rpt.set_index('movieId', inplace=True)
# movies_rpt = movies_rpt['polarity'].round(1)
movies_rpt
```

```
[5]:
```

	title	rating	polarity
movieId			
1	Toy Story (1995)	3.920930	0.30
2	Jumanji (1995)	3.431818	-0.10
3	Grumpier Old Men (1995)	3.259615	0.10
5	Father of the Bride Part II (1995)	3.071429	0.00
7	Sabrina (1995)	3.185185	0.00
...
183611	Game Night (2018)	4.000000	0.25
184471	Tomb Raider (2018)	2.500000	-0.40
187593	Deadpool 2 (2018)	3.875000	0.00
187595	Solo: A Star Wars Story (2018)	3.900000	0.00
193565	Gintama: The Movie (2010)	3.500000	0.00

[1554 rows x 3 columns]

I was worried the clusering wouldn't work with positive and negative numbers so I added 1 to each decimal in the polarity column to avoid that.

I dont think K-Means clustering depends on positive or negative numbers, but I wanted to be safe.

```
[6]: movies_rpt['polarity'] = movies_rpt['polarity'] + 1
movies_rpt
```

```
[6]:
```

	title	rating	polarity
movieId			
1	Toy Story (1995)	3.920930	1.30
2	Jumanji (1995)	3.431818	0.90
3	Grumpier Old Men (1995)	3.259615	1.10

5	Father of the Bride Part II (1995)	3.071429	1.00
7	Sabrina (1995)	3.185185	1.00
...
183611	Game Night (2018)	4.000000	1.25
184471	Tomb Raider (2018)	2.500000	0.60
187593	Deadpool 2 (2018)	3.875000	1.00
187595	Solo: A Star Wars Story (2018)	3.900000	1.00
193565	Gintama: The Movie (2010)	3.500000	1.00

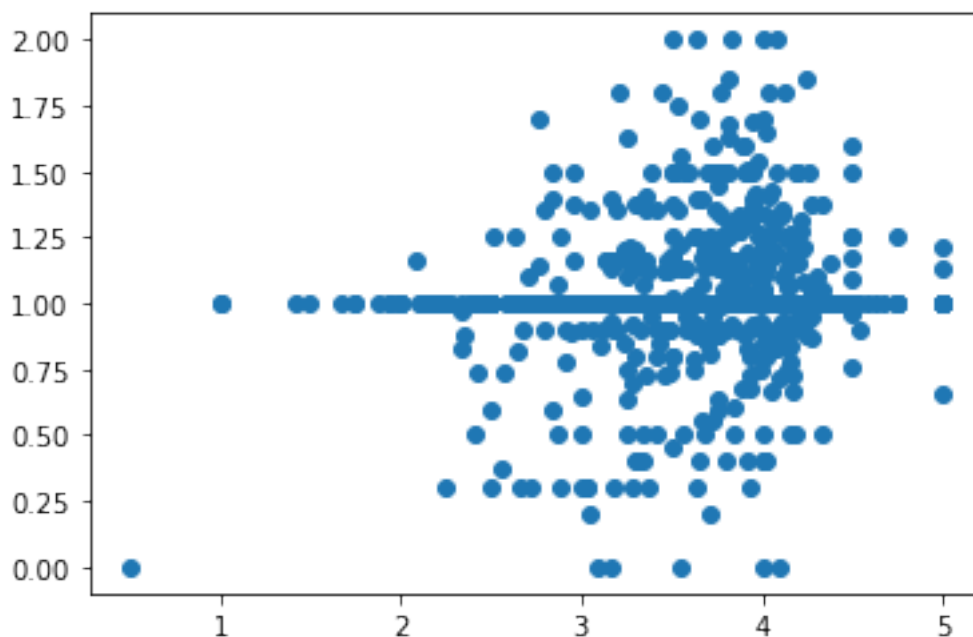
[1554 rows x 3 columns]

1.1 First stab at KMeans Clustering

I did a scatter plot to see how things looked generally.

```
[7]: plt.scatter(movies_rpt['rating'], movies_rpt['polarity'])
```

```
[7]: <matplotlib.collections.PathCollection at 0x11f5466d0>
```



```
[8]: X_rp = movies_rpt[['rating', 'polarity']].values
X_rp
# x_rp is a ndarray
```

```
[8]: array([[3.92093023, 1.3      ],
          [3.43181818, 0.9      ],
          [3.25961538, 1.1      ]],
```

```
...,
[3.875      , 1.          ],
[3.9        , 1.          ],
[3.5        , 1.          ]])
```

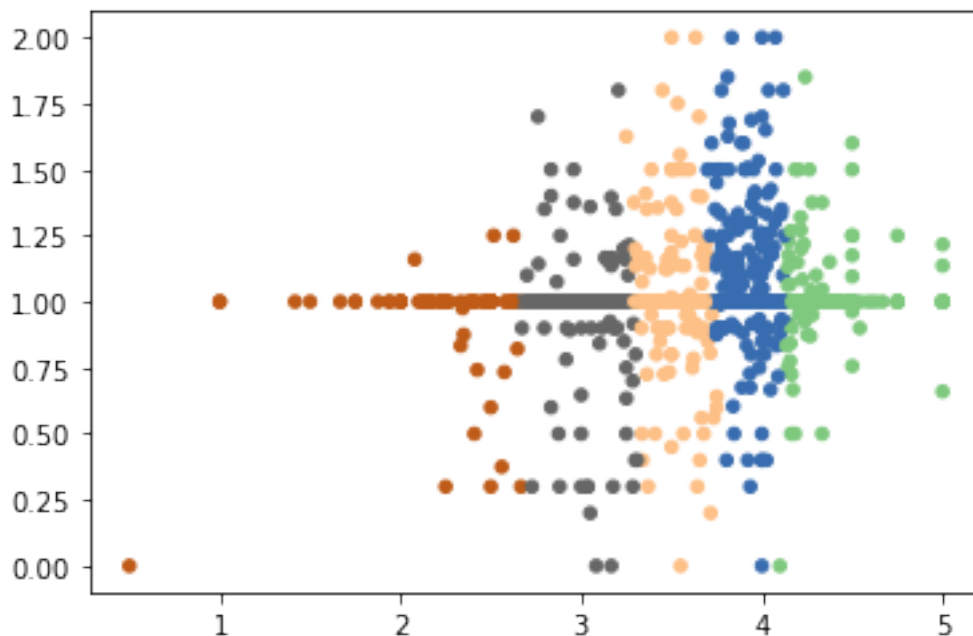
Then, I started with 5 clusters, just to see how it looked.

```
[9]: kmeans_rp = KMeans(n_clusters=5)
      predictions=kmeans_rp.fit_predict(X_rp)
```

```
[10]: def draw_clusters(movies_rpt, predictions, cmap='viridis'):
        fig = plt.figure(figsize=(8,8))
        ax = fig.add_subplot(111)
        plt.xlim(0, 5)
        plt.ylim(0, 5)
        ax.set_xlabel('rating')
        ax.set_ylabel('polarity')
```

```
[11]: clustered = pd.concat([movies_rpt.reset_index(), pd.DataFrame({'group':
    ↳ predictions})], axis=1)
      plt.scatter(clustered['rating'], clustered['polarity'], c=clustered['group'],
    ↳ s=20, cmap='Accent')
```

```
[11]: <matplotlib.collections.PathCollection at 0x11f5823d0>
```



1.2 Elbow Method

Next, I used the elbow method to find the optimal number of clusters.

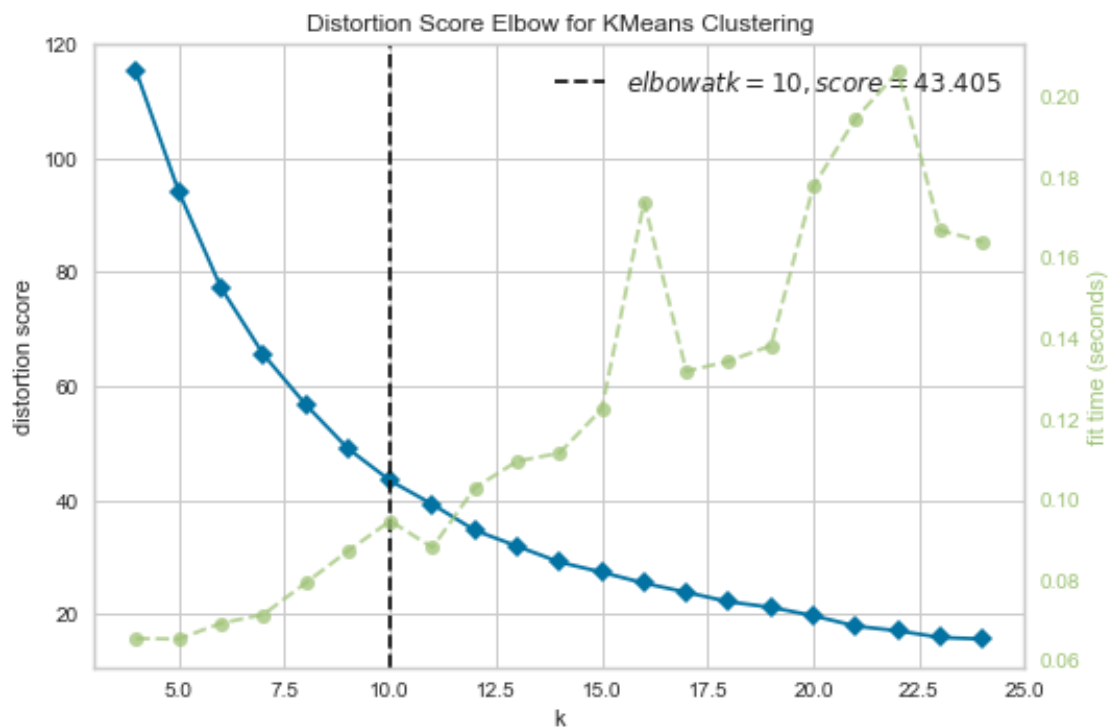
I found the steps to do so at <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>

```
[12]: # must download yellowbrick
from sklearn.cluster import KMeans
# from sklearn.datasets import make_blobs
from yellowbrick.cluster import KElbowVisualizer
```

```
/Users/annarosefritz/opt/anaconda3/lib/python3.8/site-
packages/sklearn/utils/deprecation.py:143: FutureWarning: The
sklearn.metrics.classification module is deprecated in version 0.22 and will be
removed in version 0.24. The corresponding classes / functions should instead be
imported from sklearn.metrics. Anything that cannot be imported from
sklearn.metrics is now part of the private API.
warnings.warn(message, FutureWarning)
```

```
[13]: visualizer = KElbowVisualizer(
      kmeans_rp, k=(4,25)
      #metric='calinski_harabasz', timings=False, locate_elbow=False
    )
```

```
[14]: visualizer.fit(X_rp)
      visualizer.show()
```



```
[14]: <matplotlib.axes._subplots.AxesSubplot at 0x11f6a6a00>
```

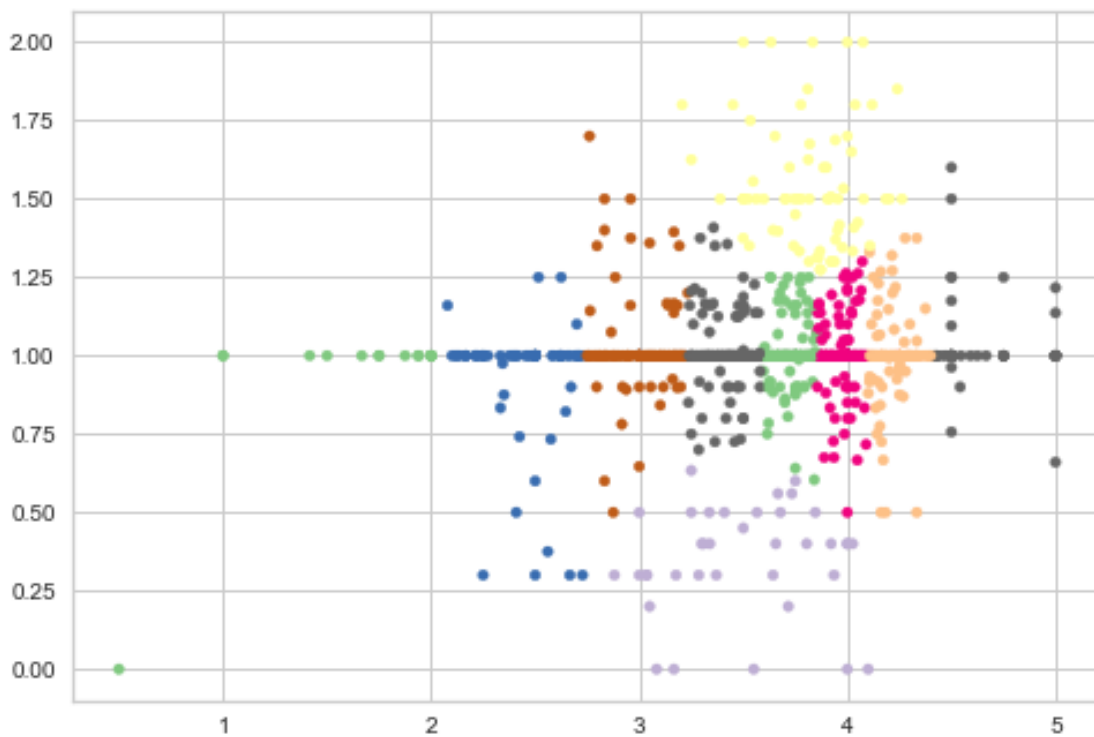
It appeared that the best K for this problem is $K = 10$.

```
[15]: kmeans_rp_2 = KMeans(n_clusters=10)
model = kmeans_rp_2.fit_predict(X_rp)
model2 = kmeans_rp_2.fit(X_rp)
```

For some reason, the cluster graph only works with the function `fit_predict` and then later, the `recommend` function only works with the `fit` function.

```
[16]: clustered = pd.concat([movies_rpt.reset_index(), pd.DataFrame({'group':
    ↳model})], axis=1)
plt.scatter(clustered['rating'], clustered['polarity'], c=clustered['group'],
    ↳s=20, cmap='Accent')
```

```
[16]: <matplotlib.collections.PathCollection at 0x120d213a0>
```



2 Making Recommendations

2.1 Finding movie cluster belongs

In order to predict similar movies, we need to relate the cluster to the title. `movie_rpt` is the original data frame. We take each index and row (so every element) and pass it through the model. To the row I appended the title.

Now, we have the titles sorted by where they cluster element was.

```
[17]: model2.labels_
```

```
[17]: array([6, 7, 7, ..., 0, 0, 7], dtype=int32)
```

Here, I could see which elements belong in each cluster, but it doesn't give me the title.

I found how to do this here: <https://stackoverflow.com/questions/32232067/cluster-points-after-kmeans-clustering-scikit-learn>

```
[18]: n_clusters = 10
labels_num = {i: np.where(model2.labels_ == i)[0] for i in range(model2.
    ↳ n_clusters)}
labels_num
```

```
[18]: {0: array([ 5,  6,  8, 11, 18, 22, 30, 33, 38, 42, 45,
          52, 53, 54, 56, 57, 62, 66, 67, 68, 79, 86,
          91, 102, 108, 113, 116, 118, 120, 122, 127, 129, 142,
         144, 147, 148, 165, 167, 173, 184, 185, 194, 198, 201,
         202, 206, 212, 217, 221, 223, 225, 228, 231, 235, 245,
         266, 271, 272, 275, 280, 281, 284, 304, 311, 312, 317,
         327, 329, 334, 337, 346, 349, 353, 354, 356, 360, 362,
         366, 369, 371, 374, 377, 389, 396, 397, 401, 403, 404,
         412, 414, 415, 416, 418, 420, 422, 424, 434, 446, 450,
         451, 458, 462, 463, 464, 466, 467, 468, 469, 480, 485,
         491, 492, 497, 503, 504, 506, 528, 529, 534, 535, 541,
         548, 551, 561, 563, 572, 579, 581, 583, 587, 590, 593,
         599, 603, 613, 615, 616, 622, 623, 626, 632, 640, 644,
         648, 650, 651, 652, 659, 660, 662, 667, 668, 671, 680,
         695, 698, 704, 711, 715, 723, 724, 729, 731, 732, 733,
         743, 746, 747, 753, 755, 757, 762, 766, 767, 770, 773,
         774, 775, 778, 794, 801, 807, 813, 814, 818, 825, 836,
         848, 857, 860, 863, 864, 865, 868, 876, 877, 881, 889,
         894, 899, 900, 906, 907, 914, 920, 922, 923, 929, 930,
         932, 936, 940, 943, 952, 966, 972, 982, 987, 988, 992,
         997, 1000, 1003, 1011, 1013, 1027, 1028, 1032, 1034, 1040, 1041,
        1048, 1049, 1051, 1053, 1058, 1060, 1061, 1070, 1074, 1080, 1082,
        1088, 1090, 1092, 1093, 1095, 1099, 1101, 1104, 1120, 1122, 1125,
        1134, 1136, 1143, 1147, 1154, 1161, 1162, 1172, 1180, 1190, 1197,
        1199, 1211, 1214, 1217, 1218, 1219, 1225, 1234, 1238, 1242, 1243,
```

```

1244, 1263, 1266, 1268, 1269, 1276, 1277, 1279, 1280, 1284, 1287,
1295, 1298, 1303, 1304, 1308, 1309, 1314, 1315, 1321, 1327, 1328,
1333, 1334, 1337, 1342, 1350, 1352, 1356, 1357, 1361, 1363, 1367,
1368, 1373, 1380, 1383, 1392, 1399, 1406, 1411, 1417, 1422, 1424,
1430, 1431, 1433, 1440, 1441, 1446, 1454, 1455, 1461, 1470, 1474,
1491, 1499, 1513, 1523, 1538, 1544, 1546, 1547, 1551, 1552]),
1: array([ 383,  478,  479,  554,  557,  691,  781,  792,  809,  827,  883,
          918, 1016, 1063, 1237, 1262]),
2: array([ 59,  83, 114, 158, 180, 182, 188, 224, 251, 254, 269,
          279, 293, 297, 302, 315, 344, 395, 406, 425, 441, 442,
          520, 530, 542, 544, 545, 594, 617, 670, 688, 697, 699,
          707, 709, 712, 714, 728, 761, 777, 780, 824, 829, 831,
          838, 847, 873, 878, 898, 905, 908, 916, 928, 962, 970,
          976, 981, 1023, 1030, 1036, 1064, 1067, 1097, 1109, 1118, 1124,
         1131, 1150, 1152, 1175, 1209, 1213, 1249, 1251, 1258, 1265, 1275,
         1293, 1329, 1339, 1343, 1354, 1370, 1372, 1381, 1384, 1394, 1408,
         1428, 1451, 1483, 1484, 1526, 1528, 1532, 1534, 1542]),
3: array([  3,  4, 10, 15, 23, 25, 43, 44, 48, 69, 70,
          71, 74, 82, 87, 93, 96, 103, 105, 106, 117, 123,
          140, 154, 168, 175, 177, 179, 236, 241, 242, 247, 273,
          375, 379, 380, 388, 399, 402, 405, 410, 427, 428, 436,
          449, 481, 488, 490, 500, 501, 505, 512, 513, 525, 527,
          539, 540, 549, 556, 559, 564, 569, 571, 580, 582, 585,
          586, 588, 592, 597, 598, 601, 604, 606, 611, 618, 649,
          657, 661, 678, 681, 682, 683, 684, 686, 689, 700, 703,
          710, 734, 736, 742, 764, 772, 779, 787, 795, 800, 806,
          811, 819, 822, 830, 841, 844, 855, 869, 887, 891, 901,
          902, 903, 909, 910, 913, 959, 977, 985, 1006, 1076, 1100,
         1146, 1155, 1156, 1168, 1184, 1203, 1210, 1221, 1224, 1226, 1240,
         1241, 1260, 1267, 1281, 1286, 1294, 1301, 1324, 1344, 1347, 1391,
         1405, 1429, 1437, 1438, 1447, 1459, 1462, 1482, 1489, 1495, 1515,
         1516, 1522, 1530]),
4: array([  7, 13, 14, 16, 21, 26, 27, 29, 37, 46, 49,
          50, 60, 65, 72, 75, 76, 77, 84, 94, 97, 111,
          126, 134, 143, 151, 157, 163, 169, 170, 172, 178, 183,
          186, 189, 192, 193, 196, 197, 199, 200, 205, 207, 208,
          209, 211, 216, 218, 219, 220, 222, 226, 227, 229, 230,
          232, 234, 239, 250, 256, 257, 260, 263, 265, 268, 274,
          276, 278, 282, 285, 286, 287, 288, 289, 290, 291, 292,
          294, 295, 296, 298, 299, 300, 301, 303, 305, 306, 307,
          308, 309, 310, 313, 314, 316, 318, 319, 320, 321, 322,
          323, 324, 325, 326, 330, 331, 333, 335, 336, 338, 339,
          341, 342, 343, 345, 347, 348, 350, 351, 352, 355, 358,
          359, 363, 368, 370, 372, 382, 385, 390, 391, 393, 400,
          413, 429, 432, 437, 440, 443, 444, 447, 452, 453, 454,
          455, 456, 459, 461, 465, 483, 484, 487, 493, 496, 514,
          517, 519, 521, 522, 523, 526, 532, 536, 537, 538, 547,

```



```

567, 568, 573, 576, 577, 584, 595, 607, 609, 612, 619,
624, 629, 630, 631, 633, 637, 638, 642, 647, 655, 658,
663, 665, 666, 669, 676, 685, 692, 694, 696, 706, 708,
713, 717, 720, 721, 730, 737, 739, 741, 745, 748, 749,
750, 752, 756, 763, 776, 782, 783, 786, 788, 790, 793,
799, 803, 804, 820, 833, 850, 852, 859, 866, 870, 871,
874, 879, 880, 884, 888, 890, 895, 897, 915, 917, 925,
926, 931, 937, 938, 939, 941, 944, 945, 948, 950, 956,
957, 958, 961, 965, 968, 974, 975, 983, 999, 1005, 1007,
1009, 1010, 1012, 1017, 1018, 1019, 1022, 1035, 1037, 1050, 1056,
1059, 1062, 1065, 1068, 1071, 1075, 1083, 1085, 1087, 1091, 1094,
1102, 1103, 1105, 1106, 1108, 1114, 1115, 1116, 1117, 1119, 1121,
1129, 1130, 1132, 1133, 1135, 1140, 1141, 1142, 1148, 1149, 1151,
1158, 1159, 1160, 1163, 1164, 1171, 1173, 1174, 1176, 1177, 1178,
1179, 1183, 1186, 1187, 1188, 1189, 1191, 1195, 1196, 1205, 1207,
1220, 1228, 1229, 1230, 1231, 1236, 1239, 1245, 1247, 1250, 1253,
1254, 1255, 1257, 1261, 1270, 1273, 1278, 1288, 1289, 1291, 1292,
1296, 1300, 1305, 1306, 1323, 1340, 1353, 1355, 1359, 1369, 1376,
1377, 1378, 1382, 1386, 1400, 1403, 1404, 1409, 1410, 1412, 1418,
1421, 1423, 1432, 1435, 1442, 1444, 1445, 1453, 1456, 1458, 1468,
1469, 1472, 1475, 1480, 1485, 1487, 1494, 1496, 1501, 1504, 1512,
1524, 1531, 1539, 1540, 1549]),
5: array([ 80, 146, 204, 214, 328, 332, 634, 639, 693, 718, 738,
740, 760, 823, 867, 964, 971, 991, 1096, 1098, 1139, 1212,
1248, 1471, 1486, 1510, 1511, 1527, 1533]),
6: array([ 0, 17, 36, 47, 98, 107, 153, 187, 191, 195, 253,
262, 340, 419, 426, 438, 510, 627, 643, 664, 687, 702,
705, 735, 861, 886, 933, 942, 986, 1111, 1123, 1200, 1223,
1232, 1235, 1274, 1290, 1311, 1318, 1322, 1332, 1351, 1358, 1371,
1375, 1389, 1395, 1414, 1415, 1425, 1426, 1439, 1450, 1457, 1479,
1492, 1493, 1503, 1505, 1506, 1507, 1509, 1514, 1519, 1521, 1525,
1536, 1537]),
7: array([ 1, 2, 9, 12, 20, 24, 28, 32, 34, 35, 40,
41, 51, 55, 58, 63, 64, 73, 78, 81, 85, 88,
89, 90, 92, 95, 99, 100, 101, 104, 109, 110, 112,
115, 119, 121, 124, 125, 128, 130, 131, 132, 135, 136,
137, 138, 139, 141, 145, 149, 150, 152, 155, 156, 159,
162, 164, 166, 171, 174, 176, 181, 190, 203, 210, 213,
215, 233, 237, 238, 240, 243, 244, 246, 248, 249, 252,
255, 258, 259, 261, 264, 267, 270, 277, 283, 357, 361,
364, 365, 367, 373, 376, 378, 384, 386, 392, 394, 407,
408, 409, 411, 417, 421, 423, 430, 433, 435, 439, 445,
448, 457, 460, 474, 477, 482, 486, 489, 494, 495, 498,
499, 502, 509, 511, 515, 516, 518, 524, 533, 543, 546,
550, 555, 558, 562, 566, 570, 578, 591, 596, 600, 602,
608, 610, 621, 625, 628, 635, 641, 645, 653, 654, 656,
672, 673, 675, 677, 679, 701, 716, 719, 722, 725, 726,

```

```

727, 744, 751, 758, 759, 765, 784, 791, 796, 798, 802,
805, 808, 810, 815, 816, 821, 828, 832, 834, 837, 840,
843, 845, 846, 849, 854, 856, 858, 862, 872, 875, 882,
885, 892, 896, 904, 911, 912, 919, 921, 924, 934, 946,
947, 949, 951, 954, 955, 963, 967, 969, 973, 978, 979,
980, 984, 989, 993, 994, 995, 998, 1001, 1002, 1004, 1008,
1014, 1015, 1020, 1021, 1024, 1039, 1042, 1043, 1044, 1045, 1046,
1047, 1052, 1055, 1057, 1069, 1072, 1073, 1077, 1078, 1079, 1081,
1084, 1107, 1112, 1113, 1126, 1127, 1128, 1137, 1144, 1145, 1153,
1157, 1165, 1166, 1167, 1169, 1170, 1181, 1185, 1193, 1194, 1198,
1202, 1208, 1215, 1216, 1227, 1233, 1246, 1252, 1256, 1259, 1264,
1271, 1272, 1282, 1283, 1285, 1297, 1299, 1302, 1307, 1310, 1312,
1313, 1316, 1317, 1319, 1320, 1325, 1326, 1330, 1331, 1335, 1336,
1338, 1341, 1345, 1348, 1349, 1360, 1366, 1374, 1379, 1385, 1388,
1390, 1393, 1396, 1398, 1401, 1416, 1420, 1427, 1434, 1443, 1449,
1452, 1460, 1464, 1465, 1466, 1473, 1476, 1477, 1488, 1490, 1497,
1498, 1500, 1508, 1518, 1520, 1529, 1535, 1541, 1543, 1553]),
8: array([ 61, 133, 161, 560, 574, 575, 674, 769, 812, 817, 893,
          953, 960, 1025, 1086, 1089, 1346, 1364, 1387, 1407, 1448, 1463,
          1478, 1545, 1548]),
9: array([ 19, 31, 39, 160, 381, 387, 398, 431, 470, 471, 472,
          473, 475, 476, 507, 508, 531, 552, 553, 565, 589, 605,
          614, 620, 636, 646, 690, 754, 768, 771, 785, 789, 797,
          826, 835, 839, 842, 851, 853, 927, 935, 990, 996, 1026,
          1029, 1031, 1033, 1038, 1054, 1066, 1110, 1138, 1182, 1192, 1201,
          1204, 1206, 1222, 1362, 1365, 1397, 1402, 1413, 1419, 1436, 1467,
          1481, 1502, 1517, 1550])}]

```

With this, I saw that some clusters have way more values than others. So, it's not a great model because the data is not evenly distributed. But, it works.

I couldn't figure out how to map the values to the titles so I used a different way instead.

This function uses `title` rather than numbers. It gave me an array where the titles of each movie are assigned to the proper cluster based on the rating and the polarity.

```

[19]: movie_to_cluster = [[] for i in range(10)]

for x, y in movies_rpt.iterrows():
    model_input = y[['rating', 'polarity']].to_numpy().reshape(1, -1)
    point = model2.predict(model_input)[0]
    movie_to_cluster[point].append(y['title'])

```

```

[20]: movie_to_cluster

```

```

[20]: [['American President, The (1995)',
        'Nixon (1995)',
        'Sense and Sensibility (1995)',

```

'Leaving Las Vegas (1995)',
 'Dead Man Walking (1995)',
 'Richard III (1995)',
 "Mr. Holland's Opus (1995)",
 'Bottle Rocket (1996)',
 'Anne Frank Remembered (1995)',
 'Apollo 13 (1995)',
 'Crimson Tide (1995)',
 'Circle of Friends (1995)',
 'Clerks (1994)',
 'Don Juan DeMarco (1995)',
 'Eat Drink Man Woman (Yin shi nan nu) (1994)',
 'Ed Wood (1994)',
 'Immortal Beloved (1994)',
 'Little Women (1994)',
 'Little Princess, A (1995)',
 'Madness of King George, The (1994)',
 'Three Colors: Blue (Trois couleurs: Bleu) (1993)',
 "What's Eating Gilbert Grape (1993)",
 'Clear and Present Danger (1994)',
 'When a Man Loves a Woman (1994)',
 "Carlito's Way (1993)",
 'In the Line of Fire (1993)',
 'Jurassic Park (1993)',
 'Much Ado About Nothing (1993)',
 'Philadelphia (1993)',
 'Remains of the Day, The (1993)',
 'Searching for Bobby Fischer (1993)',
 'Shadowlands (1993)',
 'Aladdin (1992)',
 'Dances with Wolves (1990)',
 'Snow White and the Seven Dwarfs (1937)',
 'Beauty and the Beast (1991)',
 'Rock, The (1996)',
 'Wallace & Gromit: A Close Shave (1995)',
 'Time to Kill, A (1996)',
 'American in Paris, An (1951)',
 "Breakfast at Tiffany's (1961)",
 'Charade (1963)',
 'Sabrina (1954)',
 'Wizard of Oz, The (1939)',
 'Gone with the Wind (1939)',
 '2001: A Space Odyssey (1968)',
 'Spellbound (1945)',
 'Mark of Zorro, The (1940)',
 'My Man Godfrey (1936)',
 'Thin Man, The (1934)',

'Around the World in 80 Days (1956)',
 'Bringing Up Baby (1938)',
 'Night of the Living Dead (1968)',
 'Meet John Doe (1941)',
 'Mary Poppins (1964)',
 'E.T. the Extra-Terrestrial (1982)',
 'On Golden Pond (1981)',
 'Return of the Pink Panther, The (1975)',
 'When We Were Kings (1996)',
 'Grifters, The (1990)',
 'English Patient, The (1996)',
 'Strictly Ballroom (1992)',
 'Blues Brothers, The (1980)',
 'Annie Hall (1977)',
 'Right Stuff, The (1983)',
 'Terminator, The (1984)',
 'Day the Earth Stood Still, The (1951)',
 'Better Off Dead... (1985)',
 'Deer Hunter, The (1978)',
 'Arsenic and Old Lace (1944)',
 'Fantasia (1940)',
 'Heathers (1989)',
 'Gandhi (1982)',
 'Room with a View, A (1986)',
 'Forbidden Planet (1956)',
 'When Harry Met Sally... (1989)',
 'Cape Fear (1991)',
 'Omen, The (1976)',
 'Star Trek: First Contact (1996)',
 'Sling Blade (1996)',
 'Crucible, The (1996)',
 'Star Trek II: The Wrath of Khan (1982)',
 'Last of the Mohicans, The (1992)',
 'Waiting for Guffman (1996)',
 'Donnie Brasco (1997)',
 'Grosse Pointe Blank (1997)',
 'Shall We Dance? (Shall We Dansu?) (1996)',
 'Brassed Off (1996)',
 'Hunt for Red October, The (1990)',
 'Game, The (1997)',
 'Ice Storm, The (1997)',
 'Mrs. Brown (a.k.a. Her Majesty, Mrs. Brown) (1997)',
 'Washington Square (1997)',
 'Stripes (1981)',
 'Witness (1985)',
 'Truman Show, The (1998)',
 'Dark City (1998)',

'Out of Sight (1998)',
 'Pi (1998)',
 "There's Something About Mary (1998)",
 'Tom Jones (1963)',
 'French Connection, The (1971)',
 'Rocky (1976)',
 'Kramer vs. Kramer (1979)',
 'Out of Africa (1985)',
 'Last Emperor, The (1987)',
 'Rain Man (1988)',
 'Driving Miss Daisy (1989)',
 'Exorcist, The (1973)',
 'Last Temptation of Christ, The (1988)',
 'Roger & Me (1989)',
 'Purple Rose of Cairo, The (1985)',
 'Jungle Book, The (1967)',
 'Outsiders, The (1983)',
 'Indiana Jones and the Temple of Doom (1984)',
 'Dead Zone, The (1983)',
 'Few Good Men, A (1992)',
 'Edward Scissorhands (1990)',
 'Mighty, The (1998)',
 'Children of a Lesser God (1986)',
 'Elizabeth (1998)',
 'Simple Plan, A (1998)',
 'Shakespeare in Love (1998)',
 'Hilary and Jackie (1998)',
 'Name of the Rose, The (Name der Rose, Der) (1986)',
 'Planet of the Apes (1968)',
 'Following (1998)',
 "Cookie's Fortune (1999)",
 'eXistenZ (1999)',
 'Superman (1978)',
 'Frankenstein (1931)',
 'Invasion of the Body Snatchers (1956)',
 'South Park: Bigger, Longer and Uncut (1999)',
 'Ghostbusters (a.k.a. Ghost Busters) (1984)',
 'Sixth Sense, The (1999)',
 'Airplane! (1980)',
 'Big (1988)',
 'Deliverance (1972)',
 'Excalibur (1981)',
 "Ferris Bueller's Day Off (1986)",
 'South Pacific (1958)',
 'Time Bandits (1981)',
 'American Movie (1999)',
 'Trading Places (1983)',

'Dead Again (1991)',
 'Dogma (1999)',
 'Commitments, The (1991)',
 'Shop Around the Corner, The (1940)',
 'Natural, The (1984)',
 'Fisher King, The (1991)',
 'Grand Illusion (La grande illusion) (1937)',
 'Great Santini, The (1979)',
 'Magnolia (1999)',
 'Cry in the Dark, A (1988)',
 'Wonder Boys (2000)',
 'Hoosiers (a.k.a. Best Shot) (1986)',
 'Animal House (1978)',
 'Modern Times (1936)',
 'High Fidelity (2000)',
 'Me Myself I (2000)',
 'Parenthood (1989)',
 'Auntie Mame (1958)',
 'Caddyshack (1980)',
 'Virgin Suicides, The (1999)',
 'Limelight (1952)',
 'Hamlet (1990)',
 'Serpico (1973)',
 'X-Men (2000)',
 'Eyes of Tammy Faye, The (2000)',
 'Almost Famous (2000)',
 'Best in Show (2000)',
 'Billy Elliot (2000)',
 'Crouching Tiger, Hidden Dragon (Wo hu cang long) (2000)',
 'Wall Street (1987)',
 'Finding Forrester (2000)',
 'O Brother, Where Art Thou? (2000)',
 'State and Main (2000)',
 'Traffic (2000)',
 'Evil Dead, The (1981)',
 'Reversal of Fortune (1990)',
 'Bridget Jones's Diary (2001)',
 'Shrek (2001)',
 'Divided We Fall (Musíme si pomáhat) (2000)',
 'Gentlemen Prefer Blondes (1953)',
 'Princess and the Warrior, The (Krieger und die Kaiserin, Der) (2000)',
 'Accused, The (1988)',
 'Bill & Ted's Excellent Adventure (1989)',
 'Session 9 (2001)',
 'Fiddler on the Roof (1971)',
 'Man Who Wasn't There, The (2001)',
 'Now, Voyager (1942)',

"Ocean's Eleven (2001)",
 'And Then There Were None (1945)',
 'Royal Tenenbaums, The (2001)',
 'Truly, Madly, Deeply (1991)',
 'The Count of Monte Cristo (2002)',
 'Monsoon Wedding (2001)',
 'Frailty (2001)',
 'About a Boy (2002)',
 'Minority Report (2002)',
 'Rabbit-Proof Fence (2002)',
 'Z (1969)',
 'Kid Stays in the Picture, The (2002)',
 'Igby Goes Down (2002)',
 'Heaven (2002)',
 'Bowling for Columbine (2002)',
 'Punch-Drunk Love (2002)',
 'Frida (2002)',
 'Roger Dodger (2002)',
 'Far from Heaven (2002)',
 'Quiet American, The (2002)',
 '25th Hour (2002)',
 'Chicago (2002)',
 'Amen. (2002)',
 'Pillow Talk (1959)',
 'Nowhere in Africa (Nirgendwo in Afrika) (2001)',
 'Raising Victor Vargas (2002)',
 'Man Without a Past, The (Mies vailla menneisyyttä) (2002)',
 'Better Luck Tomorrow (2002)',
 'Mighty Wind, A (2003)',
 'Identity (2003)',
 'X2: X-Men United (2003)',
 "Man on the Train (Homme du train, L') (2002)",
 'Italian Job, The (2003)',
 'Whale Rider (2002)',
 'Sinbad: Legend of the Seven Seas (2003)',
 'Pirates of the Caribbean: The Curse of the Black Pearl (2003)',
 'Dirty Pretty Things (2002)',
 'Seabiscuit (2003)',
 'American Splendor (2003)',
 'Sure Thing, The (1985)',
 'Matchstick Men (2003)',
 'Dummy (2002)',
 'Millennium Actress (Sennen joyû) (2001)',
 'Hello, Dolly! (1969)',
 'Triplets of Belleville, The (Les triplettes de Belleville) (2003)',
 'Seven Brides for Seven Brothers (1954)',
 'Kiss of the Spider Woman (1985)',

'In Cold Blood (1967)',
 'Mystic River (2003)',
 'Unvanquished, The (Aparajito) (1957)',
 'Elf (2003)',
 'Desk Set (1957)',
 'Ordet (Word, The) (1955)',
 'Passion of Joan of Arc, The (Passion de Jeanne d'Arc, La) (1928)",
 'Hannah and Her Sisters (1986)',
 'Presumed Innocent (1990)',
 'Midnight Clear, A (1992)',
 'Follow the Fleet (1936)',
 'National Velvet (1944)',
 'In America (2002)',
 'Big Fish (2003)',
 'House of Sand and Fog (2003)',
 'The Butterfly Effect (2004)',
 'Thirteen (2003)',
 'Kill Bill: Vol. 2 (2004)',
 'Mean Girls (2004)',
 'Miracle Worker, The (1962)',
 'Postman Always Rings Twice, The (1946)',
 'Pierrot le fou (1965)',
 'Dark Days (2000)',
 'How to Steal a Million (1966)',
 'Maltese Falcon, The (a.k.a. Dangerous Female) (1931)',
 'Oliver Twist (1948)',
 'Saved! (2004)',
 'Dear Frankie (2004)',
 'Tokyo Godfathers (2003)',
 'Spider-Man 2 (2004)',
 'Before Sunset (2004)',
 'Anchorman: The Legend of Ron Burgundy (2004)',
 'Garden State (2004)',
 'Primer (2004)',
 'Sideways (2004)',
 'Ray (2004)',
 'Incredibles, The (2004)',
 'Finding Neverland (2004)',
 'Porco Rosso (Crimson Pig) (Kurenai no buta) (1992)',
 'Animatrix, The (2003)',
 'Cat Returns, The (Neko no ongaeshi) (2002)',
 'Twilight Samurai, The (Tasogare Seibei) (2002)',
 'Mean Creek (2004)',
 'Born into Brothels (2004)',
 'Million Dollar Baby (2004)',
 'Hotel Rwanda (2004)',
 'Woodsman, The (2004)',

'Merchant of Venice, The (2004)',
'Sin City (2005)',
'Crash (2004)',
'Hustle & Flow (2005)',
"Devil's Rejects, The (2005)",
'Junebug (2005)',
'Grizzly Man (2005)',
'Lord of War (2005)',
'Everything Is Illuminated (2005)',
'Capote (2005)',
'Good Night, and Good Luck. (2005)',
'Jarhead (2005)',
'Walk the Line (2005)',
'Breakfast on Pluto (2005)',
'Munich (2005)',
'Lucky Number Slevin (2006)',
'Stranger than Fiction (2006)',
"Pan's Labyrinth (Laberinto del fauno, El) (2006)",
'Good Year, A (2006)',
'Blood Diamond (2006)',
'Fracture (2007)',
'Paprika (Paprika) (2006)',
'Bourne Ultimatum, The (2007)',
'Superbad (2007)',
'American Gangster (2007)',
'Bank Job, The (2008)',
'Forbidden Kingdom, The (2008)',
'Pineapple Express (2008)',
'Coraline (2009)',
'Star Trek (2009)',
'Hangover, The (2009)',
'Zombieland (2009)',
'Town Called Panic, A (Panique au village) (2009)',
'Scott Pilgrim vs. the World (2010)',
'127 Hours (2010)',
'Black Swan (2010)',
'True Grit (2010)',
'Warrior (2011)',
'Margin Call (2011)',
'Moonrise Kingdom (2012)',
'Life of Pi (2012)',
'Hobbit: An Unexpected Journey, The (2012)',
'Croods, The (2013)',
'Short Term 12 (2013)',
'Inside Llewyn Davis (2013)',
'John Wick (2014)',
'Big Eyes (2014)',

'Kung Fury (2015)',
 'Silence (2016)',
 'The Lego Batman Movie (2017)',
 'Blade Runner 2049 (2017)',
 'Jumanji: Welcome to the Jungle (2017)',
 'The Shape of Water (2017)',
 'Deadpool 2 (2018)',
 'Solo: A Star Wars Story (2018)'],
 ['Jaws 3-D (1983)',
 'Poltergeist II: The Other Side (1986)',
 'Poltergeist III (1988)',
 'Rocky V (1990)',
 'Karate Kid, Part III, The (1989)',
 'Brandon Teena Story, The (1998)',
 'My Demon Lover (1987)',
 "Losin' It (1983)",
 'Ice Castles (1978)',
 'Big Top Pee-Wee (1988)',
 'Oh, God! Book II (1980)',
 'Jonah: A VeggieTales Movie (2002)',
 "Walk, Don't Run (1966)",
 'Angie (1994)',
 'Being Julia (2004)',
 'Begotten (1990)'],
 ['Hoop Dreams (1994)',
 'Shawshank Redemption, The (1994)',
 'In the Name of the Father (1993)',
 'World of Apu, The (Apu Sansar) (1959)',
 'Godfather, The (1972)',
 'Philadelphia Story, The (1940)',
 'It Happened One Night (1934)',
 'His Girl Friday (1940)',
 'Secrets & Lies (1996)',
 'Shall We Dance (1937)',
 'Streetcar Named Desire, A (1951)',
 'Paths of Glory (1957)',
 'Lawrence of Arabia (1962)',
 "Once Upon a Time in the West (C'era una volta il West) (1968)",
 'Ran (1985)',
 'Harold and Maude (1971)',
 'Raise the Red Lantern (Da hong deng long gao gao gua) (1991)',
 'Kolya (Kolja) (1996)',
 'Ponette (1996)',
 'Wings of the Dove, The (1997)',
 'Last Days of Disco, The (1998)',
 'Beyond Silence (Jenseits der Stille) (1996)',
 'Lifeboat (1944)',

'Impostors, The (1998)',
 'Nights of Cabiria (Notti di Cabiria, Le) (1957)',
 'Central Station (Central do Brasil) (1998)',
 'Celebration, The (Festen) (1998)',
 'Buena Vista Social Club (1999)',
 'Oscar and Lucinda (a.k.a. Oscar & Lucinda) (1997)',
 'Last Picture Show, The (1971)',
 'Man Bites Dog (C'est arrivé près de chez vous) (1992)",
 'Born Yesterday (1950)',
 'Taking of Pelham One Two Three, The (1974)',
 'Double Indemnity (1944)',
 'Guess Who's Coming to Dinner (1967)",
 'Hustler, The (1961)',
 'Place in the Sun, A (1951)',
 'What Ever Happened to Baby Jane? (1962)',
 'Two Family House (2000)',
 'With a Friend Like Harry... (Harry, un ami qui vous veut du bien) (2000)',
 'Hope and Glory (1987)',
 'More (1998)',
 'Crossing Delancey (1988)',
 'Running on Empty (1988)',
 'Jesus of Montreal (Jésus de Montréal) (1989)',
 'Marat/Sade (1966)',
 'Witness for the Prosecution (1957)',
 'Going Places (Valseuses, Les) (1974)',
 '84 Charing Cross Road (1987)',
 'The Importance of Being Earnest (1952)',
 'Red Beard (Akahige) (1965)',
 'Lavender Hill Mob, The (1951)',
 'Gallipoli (1981)',
 'He Loves Me... He Loves Me Not (À la folie... pas du tout) (2002)',
 'Lady Jane (1986)',
 'Europa Europa (Hitlerjunge Salomon) (1990)',
 'Awful Truth, The (1937)',
 'You Only Live Once (1937)',
 'I Capture the Castle (2003)',
 'Magdalene Sisters, The (2002)',
 'Babette's Feast (Babettes gæstebud) (1987)",
 'Come and See (Idi i smotri) (1985)',
 'Wedding Banquet, The (Xi yan) (1993)',
 'Woman Under the Influence, A (1974)',
 'Adam's Rib (1949)",
 'Fog of War: Eleven Lessons from the Life of Robert S. McNamara, The (2003)',
 'Ox-Bow Incident, The (1943)',
 'Wit (2001)',
 'Oklahoma! (1955)',
 'Infernal Affairs (Mou gaan dou) (2002)',

'Jetée, La (1962)',
 'Into the Woods (1991)',
 'Control Room (2004)',
 'Andalusian Dog, An (Chien andalou, Un) (1929)',
 'Battle of Algiers, The (La battaglia di Algeri) (1966)',
 'Neon Genesis Evangelion: The End of Evangelion (Shin seiki Evangelion Gekijō-
 ban: Air/Magokoro wo, kimi ni) (1997)',
 'Return, The (Vozvrashcheniye) (2003)',
 'Wild Parrots of Telegraph Hill, The (2003)',
 'Match Factory Girl, The (Tulitikkutehtaan tyttö) (1990)',
 'Voices of a Distant Star (Hoshi no koe) (2003)',
 'Reefer Madness: The Movie Musical (2005)',
 'Deliver Us from Evil (2006)',
 'Tekkonkinkreet (Tekkon kinkurīto) (2006)',
 'Elite Squad (Tropa de Elite) (2007)',
 'Funny Games U.S. (2007)',
 'Son of Rambow (2007)',
 'FLCL (2000)',
 'Brothers Bloom, The (2008)',
 'Inside Job (2010)',
 'Hands Over the City (Le mani sulla città) (1963)',
 'Frank (2014)',
 'Babadook, The (2014)',
 'World of Tomorrow (2015)',
 'Who Killed Chea Vichea? (2010)',
 'Paterson',
 'Captain Fantastic (2016)',
 'Black Mirror: White Christmas (2014)'],
 ['Father of the Bride Part II (1995)',
 'Sabrina (1995)',
 'Copycat (1995)',
 'Dangerous Minds (1995)',
 'Restoration (1995)',
 'How to Make an American Quilt (1995)',
 'Batman Forever (1995)',
 'Congo (1995)',
 'Net, The (1995)',
 'Mary Shelley's Frankenstein (Frankenstein) (1994)",
 'Miracle on 34th Street (1994)',
 'My Family (1995)',
 'Natural Born Killers (1994)',
 'Santa Clause, The (1994)',
 'Virtuosity (1995)',
 'Corrina, Corrina (1994)',
 'It Could Happen to You (1994)',
 'Addams Family Values (1993)',
 'Airheads (1994)',

'Black Beauty (1994)',
'M. Butterfly (1993)',
'Renaissance Man (1994)',
'Home Alone (1990)',
'Jack and Sarah (1995)',
'Arrival, The (1996)',
'First Wives Club, The (1996)',
'Chain Reaction (1996)',
'Tin Cup (1996)',
'Farewell to Arms, A (1932)',
'Love Bug, The (1969)',
'Parent Trap, The (1961)',
'Pete's Dragon (1977)',
'Private Benjamin (1980)',
'101 Dalmatians (1996)',
'Batman Returns (1992)',
'Grease (1978)',
'Scream (1996)',
'Liar Liar (1997)',
'Romy and Michele's High School Reunion (1997)',
'Lost World: Jurassic Park, The (1997)',
'G.I. Jane (1997)',
'Midnight in the Garden of Good and Evil (1997)',
'Scream 2 (1997)',
'Sphere (1998)',
'Armageddon (1998)',
'Mask of Zorro, The (1998)',
'Honey, I Shrunk the Kids (1989)',
'Parent Trap, The (1998)',
'Something Wicked This Way Comes (1983)',
'Splash (1984)',
'Lord of the Rings, The (1978)',
'Pretty in Pink (1986)',
'St. Elmo's Fire (1985)',
'Married to the Mob (1988)',
'My Blue Heaven (1990)',
'Siege, The (1998)',
'Waterboy, The (1998)',
'Star Trek: Insurrection (1998)',
'Karate Kid, Part II, The (1986)',
'You've Got Mail (1998)',
'Crocodile Dundee (1986)',
'8MM (1999)',
'Christine (1983)',
'Never Been Kissed (1999)',
'Pushing Tin (1999)',
'Dick Tracy (1990)',

'Star Wars: Episode I - The Phantom Menace (1999)',
'Superman II (1980)',
'War of the Worlds, The (1953)',
'Big Daddy (1999)',
'Arachnophobia (1990)',
'Blair Witch Project, The (1999)',
'Ghostbusters II (1989)',
'Mystery Men (1999)',
'Radio Days (1987)',
'Pelican Brief, The (1993)',
'Meatballs (1979)',
'Scrooged (1988)',
'Fatal Attraction (1987)',
'Terrorist, The (a.k.a. Malli) (Theeviravaathi) (1998)',
'Sister Act (1992)',
'Alive (1993)',
'Death Becomes Her (1992)',
'Far and Away (1992)',
'White Men Can't Jump (1992)',
'Cutting Edge, The (1992)',
'Volunteers (1985)',
'...And Justice for All (1979)',
'Lord of the Flies (1963)',
'Big Kahuna, The (2000)',
'Pee-wee's Big Adventure (1985)',
'Soapdish (1991)',
'How the Grinch Stole Christmas (a.k.a. The Grinch) (2000)',
'Miss Congeniality (2000)',
'Glass Menagerie, The (1987)',
'Avalon (1990)',
'Revenge of the Nerds (1984)',
'Spy Kids (2001)',
'Love Story (1970)',
'Throw Momma from the Train (1987)',
'Cannonball Run, The (1981)',
'Legally Blonde (2001)',
'My Stepmother Is an Alien (1988)',
'Jurassic Park III (2001)',
'Turner & Hooch (1989)',
'Serendipity (2001)',
'Bill & Ted's Bogus Journey (1991)',
'Panic Room (2002)',
'Enigma (2001)',
'Lovely & Amazing (2001)',
'Pumpkin (2002)',
'Men in Black II (a.k.a. MIIB) (a.k.a. MIB 2) (2002)',
'Signs (2002)',

'Good Girl, The (2002)',
 'Down and Out in Beverly Hills (1986)',
 'Shanghai Knights (2003)',
 'Green Card (1990)',
 'Phone Booth (2002)',
 'Down with Love (2003)',
 'Pieces of April (2003)',
 'Flying Down to Rio (1933)',
 '13 Going on 30 (2004)',
 'Rose Red (2002)',
 'Babylon 5: In the Beginning (1998)',
 'After the Thin Man (1936)',
 'Educating Rita (1983)',
 'Dresser, The (1983)',
 'Freaks (1932)',
 'Bachelor and the Bobby-Soxer, The (1947)',
 'Village, The (2004)',
 'Funny Thing Happened on the Way to the Forum, A (1966)',
 'Vera Drake (2004)',
 'Saw (2004)',
 'Duel (1971)',
 '"Lemony Snicket's A Series of Unfortunate Events (2004)"',
 'Charlie and the Chocolate Factory (2005)',
 'Meet the Fockers (2004)',
 'Ballad of Jack and Rose, The (2005)',
 'Fantastic Four (2005)',
 'In Her Shoes (2005)',
 'Da Vinci Code, The (2006)',
 'Marie Antoinette (2006)',
 'Vicky Cristina Barcelona (2008)',
 'X-Men Origins: Wolverine (2009)',
 'Restrepo (2010)',
 'Upside Down: The Creation Records Story (2010)',
 'Friends with Benefits (2011)',
 'Thousand Words, A (2012)',
 'Upstream Color (2013)',
 'Invincible Iron Man, The (2007)',
 'A Million Ways to Die in the West (2014)',
 'Maze Runner, The (2014)',
 '"The Rabbi's Cat (Le chat du rabbin) (2011)"',
 'Self/less (2015)',
 'Suicide Squad (2016)',
 'Beasts of No Nation (2015)',
 'Eye in the Sky (2016)'],
 ['Casino (1995)',
 'Persuasion (1995)',
 'City of Lost Children, The (Cité des enfants perdus, La) (1995)',
]

'Twelve Monkeys (a.k.a. 12 Monkeys) (1995)',
 'Cry, the Beloved Country (1995)',
 'Seven (a.k.a. Se7en) (1995)',
 'Usual Suspects, The (1995)',
 'Postman, The (Postino, Il) (1994)',
 'Taxi Driver (1976)',
 'Crumb (1994)',
 'Umbrellas of Cherbourg, The (Parapluies de Cherbourg, Les) (1964)',
 'Before Sunrise (1995)',
 'Heavenly Creatures (1994)',
 'Star Wars: Episode IV - A New Hope (1977)',
 'Murder in the First (1995)',
 'Once Were Warriors (1994)',
 'Léon: The Professional (a.k.a. The Professional) (Léon) (1994)',
 'Pulp Fiction (1994)',
 'To Live (Huozhe) (1994)',
 'Forrest Gump (1994)',
 'Wonderful, Horrible Life of Leni Riefenstahl, The (Macht der Bilder: Leni Riefenstahl, Die) (1993)',
 'Fugitive, The (1993)',
 "Schindler's List (1993)",
 'Blade Runner (1982)',
 'Terminator 2: Judgment Day (1991)',
 ' Fargo (1996)',
 'Song of the Little Road (Pathar Panchali) (1955)',
 'Wallace & Gromit: The Best of Aardman Animation (1996)',
 'Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1964)',
 'Trainspotting (1996)',
 'Lone Star (1996)',
 'Emma (1996)',
 "Singin' in the Rain (1952)",
 'Vertigo (1958)',
 'Gaslight (1944)',
 'Apartment, The (1960)',
 'Some Like It Hot (1959)',
 'Maltese Falcon, The (1941)',
 'My Fair Lady (1964)',
 'Roman Holiday (1953)',
 'Meet Me in St. Louis (1944)',
 'Citizen Kane (1941)',
 'All About Eve (1950)',
 'Women, The (1939)',
 'Rebecca (1940)',
 'Notorious (1946)',
 'Adventures of Robin Hood, The (1938)',
 'Ghost and Mrs. Muir, The (1947)',
 'Lost Horizon (1937)',

'Top Hat (1935)',
 'Giant (1956)',
 "It's a Wonderful Life (1946)",
 'Mr. Smith Goes to Washington (1939)',
 'Penny Serenade (1941)',
 '39 Steps, The (1935)',
 'African Queen, The (1951)',
 'Cat on a Hot Tin Roof (1958)',
 'Big Night (1996)',
 'Sound of Music, The (1965)',
 'Fish Called Wanda, A (1988)',
 "Monty Python's Life of Brian (1979)",
 'Bonnie and Clyde (1967)',
 'Platoon (1986)',
 "Sophie's Choice (1982)",
 'Rebel Without a Cause (1955)',
 'Monty Python and the Holy Grail (1975)',
 'Wallace & Gromit: The Wrong Trousers (1993)',
 'Enchanted April (1992)',
 'My Left Foot (1989)',
 'Thin Blue Line, The (1988)',
 "One Flew Over the Cuckoo's Nest (1975)",
 'Star Wars: Episode V - The Empire Strikes Back (1980)',
 'Princess Bride, The (1987)',
 'Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark)
 (1981)',
 'Aliens (1986)',
 'Good, the Bad and the Ugly, The (Buono, il brutto, il cattivo, Il) (1966)',
 '12 Angry Men (1957)',
 'Clockwork Orange, A (1971)',
 'To Kill a Mockingbird (1962)',
 'Apocalypse Now (1979)',
 'Star Wars: Episode VI - Return of the Jedi (1983)',
 'Third Man, The (1949)',
 'Goodfellas (1990)',
 'Alien (1979)',
 'Psycho (1960)',
 'Godfather: Part II, The (1974)',
 'Full Metal Jacket (1987)',
 'Grand Day Out with Wallace and Gromit, A (1989)',
 'Henry V (1989)',
 'Amadeus (1984)',
 'Raging Bull (1980)',
 'Boot, Das (Boat, The) (1981)',
 'Sting, The (1973)',
 'Seventh Seal, The (Sjunde inseglet, Det) (1957)',
 'Glory (1989)',

'Rosencrantz and Guildenstern Are Dead (1990)',
'Manhattan (1979)',
"Miller's Crossing (1990)",
'Dead Poets Society (1989)',
'Graduate, The (1967)',
'Femme Nikita, La (Nikita) (1990)',
'Bridge on the River Kwai, The (1957)',
'Chinatown (1974)',
'Shining, The (1980)',
'Stand by Me (1986)',
'Great Escape, The (1963)',
'Unforgiven (1992)',
'Manchurian Candidate, The (1962)',
'Back to the Future (1985)',
'Patton (1970)',
'Cool Hand Luke (1967)',
'Cyrano de Bergerac (1990)',
'Young Frankenstein (1974)',
'Great Dictator, The (1940)',
'High Noon (1952)',
'Big Sleep, The (1946)',
'This Is Spinal Tap (1984)',
'Indiana Jones and the Last Crusade (1989)',
'Being There (1979)',
'Killing Fields, The (1984)',
'Man Who Would Be King, The (1975)',
'Butch Cassidy and the Sundance Kid (1969)',
'Cape Fear (1962)',
'Breaking the Waves (1996)',
'Shine (1996)',
'Paradise Lost: The Child Murders at Robin Hood Hills (1996)',
'Jaws (1975)',
'Raising Arizona (1987)',
'Hamlet (1996)',
'Whole Wide World, The (1996)',
'Hearts and Minds (1996)',
'Anna Karenina (1997)',
'L.A. Confidential (1997)',
'Sweet Hereafter, The (1997)',
'Big Lebowski, The (1998)',
'Everest (1998)',
'Spanish Prisoner, The (1997)',
'Children of Heaven, The (Bacheha-Ye Aseman) (1997)',
'Dream for an Insomniac (1996)',
'Picnic at Hanging Rock (1975)',
'Mutiny on the Bounty (1935)',
'Lost Weekend, The (1945)',

"Gentleman's Agreement (1947)",
 "All the King's Men (1949)",
 'On the Waterfront (1954)',
 'In the Heat of the Night (1967)',
 'Midnight Cowboy (1969)',
 'Chariots of Fire (1981)',
 'Seven Samurai (Shichinin no samurai) (1954)',
 'Dangerous Liaisons (1988)',
 'Saving Private Ryan (1998)',
 'Out of the Past (1947)',
 'Blue Velvet (1986)',
 "Rosemary's Baby (1968)",
 'Strangers on a Train (1951)',
 'Untouchables, The (1987)',
 'Shadow of a Doubt (1943)',
 'Suspicion (1941)',
 'Lady Vanishes, The (1938)',
 'Say Anything... (1989)',
 'Producers, The (1968)',
 'Elephant Man, The (1980)',
 'Life Is Beautiful (La Vita è bella) (1997)',
 'American History X (1998)',
 'Little Voice (1998)',
 'Last Days, The (1998)',
 'Office Space (1999)',
 'Haunting, The (1963)',
 'King and I, The (1956)',
 'Matrix, The (1999)',
 'Mildred Pierce (1945)',
 'Run Lola Run (Lola rennt) (1998)',
 'Killing, The (1956)',
 '400 Blows, The (Les quatre cents coups) (1959)',
 'Iron Giant, The (1999)',
 'Christmas Story, A (1983)',
 'Sanjuro (Tsubaki Sanjûrô) (1962)',
 'Lady Eve, The (1941)',
 'Palm Beach Story, The (1942)',
 'Gilda (1946)',
 'Indochine (1992)',
 'Fight Club (1999)',
 'Straight Story, The (1999)',
 'Princess Mononoke (Mononoke-hime) (1997)',
 'Yojimbo (1961)',
 '42 Up (1998)',
 'Grapes of Wrath, The (1940)',
 'Places in the Heart (1984)',
 'Flawless (1999)',

'End of the Affair, The (1999)',
 'Green Mile, The (1999)',
 'Titus (1999)',
 'Howards End (1992)',
 'Circus, The (1928)',
 'Kid, The (1921)',
 'Splendor in the Grass (1961)',
 'Creature Comforts (1989)',
 'Good Earth, The (1937)',
 'Inherit the Wind (1960)',
 'Midnight Express (1978)',
 'Network (1976)',
 'Odd Couple, The (1968)',
 'Guys and Dolls (1955)',
 'Gold Rush, The (1925)',
 'White Christmas (1954)',
 'Blood Simple (1984)',
 'Conversation, The (1974)',
 'Anatomy of a Murder (1959)',
 'Breaker Morant (1980)',
 'Official Story, The (La historia oficial) (1985)',
 'The Spiral Staircase (1945)',
 'Dancer in the Dark (2000)',
 'You Can Count on Me (2000)',
 'Friendly Persuasion (1956)',
 'In the Mood For Love (Fa yeung nin wa) (2000)',
 'Widow of St. Pierre, The (Veuve de Saint-Pierre, La) (2000)',
 "Long Night's Journey Into Day (2000)",
 "Bishop's Wife, The (1947)",
 'Elmer Gantry (1960)',
 'Manhunter (1986)',
 'Memento (2000)',
 'Scarface (1983)',
 'Days of Wine and Roses (1962)',
 'Faust (1926)',
 'Vanishing, The (Spoonloos) (1988)',
 'Little Foxes, The (1941)',
 'Silkwood (1983)',
 'Donnie Darko (2001)',
 'Blue Angel, The (Blaue Engel, Der) (1930)',
 'Lord of the Rings: The Fellowship of the Ring, The (2001)',
 'Beautiful Mind, A (2001)',
 'Yentl (1983)',
 'Bad and the Beautiful, The (1952)',
 'Sleuth (1972)',
 'Promises (2001)',
 'Rashomon (Rashômon) (1950)',

'Rome, Open City (a.k.a. Open City) (Roma, città aperta) (1945)',
 'Believer, The (2001)',
 'Requiem for a Heavyweight (1962)',
 'Yearling, The (1946)',
 'Spirited Away (Sen to Chihiro no kamikakushi) (2001)',
 'Real Women Have Curves (2002)',
 'Grave of the Fireflies (Hotaru no haka) (1988)',
 'World of Henry Orient, The (1964)',
 'Talk to Her (Hable con Ella) (2002)',
 'Zulu (1964)',
 'Lord of the Rings: The Two Towers, The (2002)',
 'My Neighbor Totoro (Tonari no Totoro) (1988)',
 'Nicholas Nickleby (2002)',
 'Pianist, The (2002)',
 'City of God (Cidade de Deus) (2002)',
 'Dogfight (1991)',
 'Missing (1982)',
 'Monty Python Live at the Hollywood Bowl (1982)',
 'Night of the Shooting Stars (Notte di San Lorenzo, La) (1982)',
 'All the Real Girls (2003)',
 'Red Badge of Courage, The (1951)',
 'Stone Reader (2002)',
 'Talk of the Town, The (1942)',
 'Born Free (1966)',
 'Stevie (2002)',
 'Spellbound (2002)',
 'Laputa: Castle in the Sky (Tenkû no shiro Rapyuta) (1986)',
 'Cinemanía (2002)',
 'Gigantic (A Tale of Two Johns) (2002)',
 'Finding Nemo (2003)',
 'Capturing the Friedmans (2003)',
 'Shenandoah (1965)',
 'Barton Fink (1991)',
 'Long, Hot Summer, The (1958)',
 '28 Days Later (2002)',
 'One, Two, Three (1961)',
 'Secret Lives of Dentists, The (2002)',
 'Lost in Translation (2003)',
 'Luther (2003)',
 'Lagaan: Once Upon a Time in India (2001)',
 'All the President's Men (1976)',
 'Monty Python's The Meaning of Life (1983)',
 'Ginger Snaps (2000)',
 'Yankee Doodle Dandy (1942)',
 'Kill Bill: Vol. 1 (2003)',
 'Billabong Odyssey (2003)',
 'Tupac: Resurrection (2003)',

"Devil's Playground (2002)",
 'Tale of Two Cities, A (1935)',
 'Night of the Hunter, The (1955)',
 'Roberta (1935)',
 'Swing Time (1936)',
 'Dark Victory (1939)',
 'Birdman of Alcatraz (1962)',
 'Red River (1948)',
 'Passage to India, A (1984)',
 'Black Orpheus (Orfeu Negro) (1959)',
 'Hero (Ying xiong) (2002)',
 'Horse Feathers (1932)',
 'Night at the Opera, A (1935)',
 'My Flesh and Blood (2003)',
 'I Was a Male War Bride (1949)',
 'Dark Passage (1947)',
 'They Drive by Night (1940)',
 'Strada, La (1954)',
 'Touching the Void (2003)',
 'Eternal Sunshine of the Spotless Mind (2004)',
 'Dogville (2003)',
 "I'm Not Scared (Io non ho paura) (2003)",
 'Watch on the Rhine (1943)',
 'Three Faces of Eve, The (1957)',
 'Woman of the Year (1942)',
 'Pat and Mike (1952)',
 'Cat People (1942)',
 'Camelot (1967)',
 'Throne of Blood (Kumonosu jô) (1957)',
 '11\09"01 - September 11 (2002)',
 'High and Low (Tengoku to jigoku) (1963)',
 'Zelig (1983)',
 'Shame (Skammen) (1968)',
 'Broadway Danny Rose (1984)',
 'Weather Underground, The (2002)',
 'Spring, Summer, Fall, Winter... and Spring (Bom yeoreum gaeul gyeoul geurigo
 bom) (2003)',
 'Dark Water (Honogurai mizu no soko kara) (2002)',
 'Captain Blood (1935)',
 'Americanization of Emily, The (1964)',
 'Anne of the Thousand Days (1969)',
 'Hiroshima Mon Amour (1959)',
 'Day of the Jackal, The (1973)',
 'Safety Last! (1923)',
 'Caine Mutiny, The (1954)',
 'Black Narcissus (1947)',
 'Divorcee, The (1930)',

'Johnny Belinda (1948)',
 'Maria Full of Grace (Maria, Llena eres de gracia) (2004)',
 'Bill Cosby, Himself (1983)',
 "Alice Doesn't Live Here Anymore (1974)",
 'Shaun of the Dead (2004)',
 'Come Back, Little Sheba (1952)',
 'Tarnation (2003)',
 'The Machinist (2004)',
 'Guerrilla: The Taking of Patty Hearst (2004)',
 "That's Entertainment (1974)",
 'Greed (1924)',
 'Stage Door (1937)',
 'Letter, The (1940)',
 'Random Harvest (1942)',
 'Letter to Three Wives, A (1949)',
 "Kiki's Delivery Service (Majo no takkyûbin) (1989)",
 'Old Boy (2003)',
 'Story of the Weeping Camel, The (Geschichte vom weinenden Kamel, Die) (2003)',
 'Riding Giants (2004)',
 'I Remember Mama (1948)',
 'Nobody Knows (Dare mo shiranai) (2004)',
 'Zelary (2003)',
 'Three Musketeers, The (1973)',
 'Enron: The Smartest Guys in the Room (2005)',
 'Cinderella Man (2005)',
 'Georgy Girl (1966)',
 'Serenity (2005)',
 'Kiss Kiss Bang Bang (2005)',
 'Lady Vengeance (Sympathy for Lady Vengeance) (Chinjeolhan geumjassi) (2005)',
 'Departed, The (2006)',
 'Prestige, The (2006)',
 'Hot Fuzz (2007)',
 '3:10 to Yuma (2007)',
 'There Will Be Blood (2007)',
 'Cat Soup (Nekojiru-so) (2001)',
 'In Bruges (2008)',
 'Dark Knight, The (2008)',
 'WALL·E (2008)',
 "Dr. Horrible's Sing-Along Blog (2008)",
 'Inglourious Basterds (2009)',
 'Moon (2009)',
 'Neon Genesis Evangelion: Death & Rebirth (Shin seiki Evangelion Gekijô-ban: Shito shinsei) (1997)',
 'Up (2009)',
 'Hurt Locker, The (2008)',
 'Mary and Max (2009)',

'Shutter Island (2010)',
 'Inception (2010)',
 'Rare Exports: A Christmas Tale (Rare Exports) (2010)',
 'Tucker & Dale vs Evil (2010)',
 'Starsuckers (2009)',
 'Dark Knight Rises, The (2012)',
 'Intouchables (2011)',
 'Argo (2012)',
 'Zero Dark Thirty (2012)',
 'It's Such a Beautiful Day (2012)",
 'History of Future Folk, The (2012)',
 'Prisoners (2013)',
 'Dallas Buyers Club (2013)',
 'Wolf of Wall Street, The (2013)',
 'Interstellar (2014)',
 'Whiplash (2014)',
 'Guardians of the Galaxy (2014)',
 'Wild Tales (2014)',
 'Paddington (2014)',
 'Kingsman: The Secret Service (2015)',
 'Avengers: Infinity War - Part I (2018)',
 'The Martian (2015)',
 'Just Eat It: A Food Waste Story (2014)',
 'Everybody Wants Some (2016)',
 'John Wick: Chapter Two (2017)',
 'Logan (2017)',
 'Game Night (2018)'],
 ['Three Colors: White (Trzy kolory: Bialy) (1994)',
 'Silence of the Lambs, The (1991)',
 'Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)',
 'Ninotchka (1939)',
 'Treasure of the Sierra Madre, The (1948)',
 'M (1931)',
 'Fistful of Dollars, A (Per un pugno di dollari) (1964)',
 'Bad Seed, The (1956)',
 'City Lights (1931)',
 'Misery (1990)',
 'Blazing Saddles (1974)',
 'Eraserhead (1977)',
 'Requiem for a Dream (2000)',
 'Score, The (2001)',
 'Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)",
 'Patch of Blue, A (1965)',
 'Irreversible (Irréversible) (2002)',
 'Lilya 4-Ever (Lilja 4-ever) (2002)',
 'Battle Royale (Batoru rowaiaru) (2000)',
 'Salo, or The 120 Days of Sodom (Salò o le 120 giornate di Sodoma) (1976)',

'Good bye, Lenin! (2003)',
 'Notebook, The (2004)',
 "Love Me If You Dare (Jeux d'enfants) (2003)",
 'Captain Phillips (2013)',
 'Gone Girl (2014)',
 'A Pigeon Sat on a Branch Reflecting on Existence (2014)',
 'The Hateful Eight (2015)',
 '10 Cloverfield Lane (2016)',
 'Sausage Party (2016)'],
 ['Toy Story (1995)',
 'Babe (1995)',
 'Braveheart (1995)',
 'Kids (1995)',
 'Lion King, The (1994)',
 'Blue Sky (1994)',
 'Primal Fear (1996)',
 'Rear Window (1954)',
 'North by Northwest (1959)',
 'Casablanca (1942)',
 "William Shakespeare's Romeo + Juliet (1996)",
 'Reservoir Dogs (1992)',
 'Akira (1988)',
 'Gattaca (1997)',
 'Good Will Hunting (1997)',
 'Man in the Iron Mask, The (1998)',
 'Watership Down (1978)',
 'Year of Living Dangerously, The (1982)',
 'Insider, The (1999)',
 'Toy Story 2 (1999)',
 'Hard-Boiled (Lat sau san taam) (1992)',
 'Erin Brockovich (2000)',
 'Do the Right Thing (1989)',
 'Gladiator (2000)',
 "Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the
 Philosopher's Stone) (2001)",
 'Blade II (2002)',
 'Harry Potter and the Chamber of Secrets (2002)',
 'Catch Me If You Can (2002)',
 'Cowboy Bebop: The Movie (Cowboy Bebop: Tengoku no Tobira) (2001)',
 'Jezebel (1938)',
 'Lord of the Rings: The Return of the King, The (2003)',
 'Harry Potter and the Prisoner of Azkaban (2004)',
 'Prime of Miss Jean Brodie, The (1969)',
 'Murder on the Orient Express (1974)',
 'Country Girl, The (1954)',
 'Layer Cake (2004)',
 "Howl's Moving Castle (Hauru no ugoku shiro) (2004)",

'40-Year-Old Virgin, The (2005)',
 'Corpse Bride (2005)',
 'Wallace & Gromit in The Curse of the Were-Rabbit (2005)',
 'Harry Potter and the Goblet of Fire (2005)',
 'Illusionist, The (2006)',
 'Ratatouille (2007)',
 'Gone Baby Gone (2007)',
 'Juno (2007)',
 'Meet Dave (2008)',
 'Zack and Miri Make a Porno (2008)',
 '(500) Days of Summer (2009)',
 'District 9 (2009)',
 'Social Network, The (2010)',
 'Town, The (2010)',
 'Avengers, The (2012)',
 'Skyfall (2012)',
 'Django Unchained (2012)',
 'The Lego Movie (2014)',
 'Big Hero 6 (2014)',
 'The Imitation Game (2014)',
 'Mad Max: Fury Road (2015)',
 'Thor: Ragnarok (2017)',
 'Guardians of the Galaxy 2 (2017)',
 'Doctor Strange (2016)',
 'A Story of Children and Film (2013)',
 'The Hunger Games: Mockingjay - Part 2 (2015)',
 'The Revenant (2015)',
 'Room (2015)',
 'Big Short, The (2015)',
 'Arrival (2016)',
 'La La Land (2016)'],
 ['Jumanji (1995)',
 'Grumpier Old Men (1995)',
 'Get Shorty (1995)',
 'Othello (1995)',
 'Clueless (1995)',
 'To Die For (1995)',
 'Mighty Aphrodite (1995)',
 'In the Bleak Midwinter (1995)',
 'Happy Gilmore (1996)',
 'Muppet Treasure Island (1996)',
 'Up Close and Personal (1996)',
 'Basketball Diaries, The (1995)',
 'Billy Madison (1995)',
 'Dolores Claiborne (1995)',
 'Forget Paris (1995)',
 'I.Q. (1994)',

'Just Cause (1995)',
 'Nell (1994)',
 'Quiz Show (1994)',
 'Stargate (1994)',
 'Star Trek: Generations (1994)',
 'While You Were Sleeping (1995)',
 'Muriel's Wedding (1994)',
 'Adventures of Priscilla, Queen of the Desert, The (1994)',
 'Client, The (1994)',
 'Four Weddings and a Funeral (1994)',
 'Paper, The (1994)',
 'Speed (1994)',
 'True Lies (1994)',
 'Age of Innocence, The (1993)',
 'Dave (1993)',
 'Firm, The (1993)',
 'Hudsucker Proxy, The (1994)',
 'What's Love Got to Do with It? (1993)',
 'Mrs. Doubtfire (1993)',
 'Radioland Murders (1994)',
 'Romper Stomper (1992)',
 'Rudy (1993)',
 'Secret Garden, The (1993)',
 'Short Cuts (1993)',
 'Six Degrees of Separation (1993)',
 'Sleepless in Seattle (1993)',
 'So I Married an Axe Murderer (1993)',
 'Nightmare Before Christmas, The (1993)',
 'Three Musketeers, The (1993)',
 'War Room, The (1993)',
 'Welcome to the Dollhouse (1995)',
 'Ghost (1990)',
 'Batman (1989)',
 'Pinocchio (1940)',
 'Pretty Woman (1990)',
 'Aristocats, The (1970)',
 'Courage Under Fire (1996)',
 'Mission: Impossible (1996)',
 'Mystery Science Theater 3000: The Movie (1996)',
 'Truth About Cats & Dogs, The (1996)',
 'Cold Comfort Farm (1995)',
 'Twister (1996)',
 'Independence Day (a.k.a. ID4) (1996)',
 'Very Brady Sequel, A (1996)',
 'Ransom (1996)',
 'Twelfth Night (1996)',
 'Gay Divorcee, The (1934)',

'My Favorite Year (1982)',
 'Foreign Correspondent (1940)',
 'Father of the Bride (1950)',
 'Gigi (1958)',
 'Beat the Devil (1953)',
 'Fly Away Home (1996)',
 'Michael Collins (1996)',
 'Chamber, The (1996)',
 'Cinderella (1950)',
 'Sword in the Stone, The (1963)',
 'Dumbo (1941)',
 'Alice in Wonderland (1951)',
 'Fox and the Hound, The (1981)',
 'That Thing You Do! (1996)',
 'Crossfire (1947)',
 'Victor/Victoria (1982)',
 'Candidate, The (1972)',
 'Dirty Dancing (1987)',
 'Doors, The (1991)',
 'Top Gun (1986)',
 'People vs. Larry Flynt, The (1996)',
 'Bob Roberts (1992)',
 'Passion Fish (1992)',
 'Field of Dreams (1989)',
 'Birds, The (1963)',
 'Carrie (1976)',
 'Nosferatu (Nosferatu, eine Symphonie des Grauens) (1922)',
 'Mirror Has Two Faces, The (1996)',
 'Preacher's Wife, The (1996)",
 'Star Trek VI: The Undiscovered Country (1991)',
 'Star Trek IV: The Voyage Home (1986)',
 'Jerry Maguire (1996)',
 'Sneakers (1992)',
 'Evita (1996)',
 'Benny & Joon (1993)',
 'My Best Friend's Wedding (1997)",
 'Face/Off (1997)',
 'Men in Black (a.k.a. MIB) (1997)',
 'Air Force One (1997)',
 'The Devil's Advocate (1997)",
 'Rainmaker, The (1997)',
 'Sliding Doors (1998)',
 'Titanic (1997)',
 'Great Expectations (1998)',
 'Wedding Singer, The (1998)',
 'Primary Colors (1998)',
 'X-Files: Fight the Future, The (1998)',

"Buffalo '66 (a.k.a. Buffalo 66) (1998)",
 'West Side Story (1961)',
 'Oliver! (1968)',
 'Friday the 13th Part VI: Jason Lives (1986)',
 'Poltergeist (1982)',
 'Back to the Future Part II (1989)',
 'Godfather: Part III, The (1990)',
 'Negotiator, The (1998)',
 'Tender Mercies (1983)',
 'And the Band Played On (1993)',
 'Lady and the Tramp (1955)',
 '101 Dalmatians (One Hundred and One Dalmatians) (1961)',
 'L.A. Story (1991)',
 'Addams Family, The (1991)',
 'Secret of NIMH, The (1982)',
 'Blade (1998)',
 'Next Stop Wonderland (1998)',
 'Stage Fright (1950)',
 'Broadcast News (1987)',
 'My Cousin Vinny (1992)',
 "Bug's Life, A (1998)",
 'Very Bad Things (1998)',
 'Prince of Egypt, The (1998)',
 'Karate Kid, The (1984)',
 "Christmas Vacation (National Lampoon's Christmas Vacation) (1989)",
 'Ruthless People (1986)',
 'Simply Irresistible (1999)',
 'Other Sister, The (1999)',
 '10 Things I Hate About You (1999)',
 'Thing from Another World, The (1951)',
 'Trekkies (1997)',
 'American Pie (1999)',
 'Eyes Wide Shut (1999)',
 'Spartacus (1960)',
 'Mission, The (1986)',
 "Hard Day's Night, A (1964)",
 'Risky Business (1983)',
 'Brief Encounter (1946)',
 'Sydney (Hard Eight) (1996)',
 'Who Framed Roger Rabbit? (1988)',
 "They Shoot Horses, Don't They? (1969)",
 'Stand and Deliver (1988)',
 'Moonstruck (1987)',
 'Sleepy Hollow (1999)',
 'Man on the Moon (1999)',
 'Galaxy Quest (1999)',
 'Hurricane, The (1999)',

'Girl, Interrupted (1999)',
 'Fast Times at Ridgemont High (1982)',
 'JFK (1991)',
 'Hook (1991)',
 'My Life (1993)',
 'Return to Me (2000)',
 'Predator (1987)',
 'Prince of Tides, The (1991)',
 'Keeping the Faith (2000)',
 'Coming Home (1978)',
 'Steel Magnolias (1989)',
 'Invisible Man, The (1933)',
 'Meet the Parents (2000)',
 'Unbreakable (2000)',
 'Caveman's Valentine, The (2001)",
 'Alfie (1966)',
 "River's Edge (1986)",
 'Enemy at the Gates (2001)',
 'Luzhin Defence, The (2000)',
 'Norma Rae (1979)',
 "Himalaya (Himalaya - l'enfance d'un chef) (1999)",
 'Mississippi Burning (1988)',
 'Seven Year Itch, The (1955)',
 'Tootsie (1982)',
 'Final Fantasy: The Spirits Within (2001)',
 'Coming to America (1988)',
 'Short Circuit (1986)',
 'Tetsuo, the Ironman (Tetsuo) (1988)',
 'Gross Anatomy (a.k.a. A Cut Above) (1989)',
 'Major League (1989)',
 'Ghost World (2001)',
 'UHF (1989)',
 'Black Robe (1991)',
 'Little Man Tate (1991)',
 'Zoolander (2001)',
 "Coal Miner's Daughter (1980)",
 'K-PAX (2001)',
 "Devil's Backbone, The (Espinazo del diablo, El) (2001)",
 'Defiant Ones, The (1958)',
 "Monster's Ball (2001)",
 'Bad News Bears, The (1976)',
 'Return of the Secaucus 7 (1980)',
 'Spider-Man (2002)',
 'Insomnia (2002)',
 'Road to Perdition (2002)',
 'Possession (2002)',
 'One Hour Photo (2002)',

'Red Dragon (2002)',
 'Pride of the Yankees, The (1942)',
 'Grey Zone, The (2001)',
 'Standing in the Shadows of Motown (2002)',
 'Love Liza (2002)',
 'Confessions of a Dangerous Mind (2002)',
 'Alice Adams (1935)',
 'Imitation of Life (1959)',
 'May (2002)',
 'Kiss Me Goodbye (1982)',
 'Black Stallion, The (1979)',
 'Old School (2003)',
 'Spider (2002)',
 'Bend It Like Beckham (2002)',
 'Ringu (Ring) (1998)',
 'Salaam Bombay! (1988)',
 'Sweet Charity (1969)',
 'In a Lonely Place (1950)',
 'Ghosts of the Abyss (2003)',
 'Holes (2003)',
 'Winged Migration (Peuple migrateur, Le) (2001)',
 'Fahrenheit 451 (1966)',
 'Decade Under the Influence, A (2003)',
 'Blue Car (2002)',
 'Owning Mahowny (2003)',
 'Shape of Things, The (2003)',
 'Bruce Almighty (2003)',
 'Silk Stockings (1957)',
 'Murder on a Sunday Morning (Un coupable idéal) (2001)',
 'Music Box (1989)',
 'Mississippi Masala (1991)',
 'Anastasia (1956)',
 'Gospel According to St. Matthew, The (Vangelo secondo Matteo, Il) (1964)',
 'Wait Until Dark (1967)',
 'Kind Hearts and Coronets (1949)',
 'Red Shoes, The (1948)',
 'Pink Panther, The (1963)',
 'Waco: The Rules of Engagement (1997)',
 'Ikiru (1952)',
 'Paper Moon (1973)',
 'Secondhand Lions (2003)',
 'My Life Without Me (2003)',
 'Regarding Henry (1991)',
 'School of Rock (2003)',
 'Station Agent, The (2003)',
 'Elephant (2003)',
 'Eye, The (Gin gwai) (Jian gui) (2002)',

"You'll Never Get Rich (1941)",
 'Shattered Glass (2003)',
 'Father of the Bride (1991)',
 'Macbeth (a.k.a. Tragedy of Macbeth, The) (1971)',
 'Hunchback of Notre Dame, The (1939)',
 'Pygmalion (1938)',
 'Monster (2003)',
 'Cold Mountain (2003)',
 'Jane Eyre (1970)',
 'Diary of a Country Priest (Journal d'un curé de campagne) (1951)",
 'Punisher, The (2004)',
 'Paper Clips (2004)',
 'Chaplin (1992)',
 'Wages of Fear, The (Salaire de la peur, Le) (1953)',
 'Another Thin Man (1939)',
 'Thin Man Goes Home, The (1945)',
 'Shadow of the Thin Man (1941)',
 'Song of the Thin Man (1947)',
 'Stray Dog (Nora inu) (1949)',
 'Bad Day at Black Rock (1955)',
 'Sansho the Bailiff (Sanshō dayū) (1954)',
 'College (1927)',
 'Make Way for Tomorrow (1937)',
 'Shrek 2 (2004)',
 'Napoleon Dynamite (2004)',
 'Super Size Me (2004)',
 'Our Hospitality (1923)',
 'Fahrenheit 9/11 (2004)',
 'Danny Deckchair (2003)',
 'Mask (1985)',
 'House of Flying Daggers (Shi mian mai fu) (2004)',
 'Fury (1936)',
 'Lady from Shanghai, The (1947)',
 'Au Hasard Balthazar (1966)',
 'Gia (1998)',
 'Millions (2004)',
 'Spanglish (2004)',
 'Life Aquatic with Steve Zissou, The (2004)',
 'Aviator, The (2004)',
 'In Good Company (2004)',
 'Kingdom of Heaven (2005)',
 'Star Wars: Episode III - Revenge of the Sith (2005)',
 'Wedding Crashers (2005)',
 'Broken Flowers (2005)',
 'Duma (2005)',
 'Constant Gardener, The (2005)',
 'Proof (2005)',

'Why We Fight (2005)',
'Waiting... (2005)',
'History of Violence, A (2005)',
'Oliver Twist (2005)',
'Squid and the Whale, The (2005)',
'North Country (2005)',
'Syriana (2005)',
'Pride & Prejudice (2005)',
'Chronicles of Narnia: The Lion, the Witch and the Wardrobe, The (2005)',
'King Kong (2005)',
'Producers, The (2005)',
'Looking for Comedy in the Muslim World (2005)',
'X-Men: The Last Stand (2006)',
'Babel (2006)',
'Night at the Museum (2006)',
'Meet the Robinsons (2007)',
'Fido (2006)',
'I Am Legend (2007)',
'Definitely, Maybe (2008)',
'Chronicles of Narnia: Prince Caspian, The (2008)',
'Get Smart (2008)',
'Step Brothers (2008)',
'Burn After Reading (2008)',
'Changeling (2008)',
'Boy in the Striped Pajamas, The (Boy in the Striped Pyjamas, The) (2008)',
'I Love You, Man (2009)',
'Haunted World of El Superbeasto, The (2009)',
'Avatar (2009)',
'Sintel (2010)',
'Tron: Legacy (2010)',
'The Hunger Games (2012)',
'Beasts of the Southern Wild (2012)',
'Taken 2 (2012)',
'Movie 43 (2013)',
'Man of Steel (2013)',
'Grown Ups 2 (2013)',
'Gravity (2013)',
'Hobbit: The Desolation of Smaug, The (2013)',
'Anchorman 2: The Legend Continues (2013)',
'Snowpiercer (2013)',
'Two Days, One Night (Deux jours, une nuit) (2014)',
'Angel's Egg (Tenshi no tamago) (1985)',
'The Hobbit: The Battle of the Five Armies (2014)',
'Good Copy Bad Copy (2007)',
'The Interview (2014)',
'The DUFF (2015)',
'The Man from U.N.C.L.E. (2015)',

'Sicario (2015)',
 'Rabbits (2002)',
 "Don't Breathe (2016)",
 'It Comes at Night (2017)',
 'Dunkirk (2017)',
 'Gintama: The Movie (2010)'],
 ['Houseguest (1994)',
 'Sliver (1993)',
 'Mulholland Falls (1996)',
 'Patch Adams (1998)',
 'Village of the Damned (1960)',
 'Children of the Damned (1963)',
 'Talented Mr. Ripley, The (1999)',
 "Dude, Where's My Car? (2000)",
 'Evolution (2001)',
 'A.I. Artificial Intelligence (2001)',
 'Butterflies Are Free (1972)',
 'Final Destination 2 (2003)',
 'Daredevil (2003)',
 'Hulk (2003)',
 '21 Grams (2003)',
 'WarGames (1983)',
 'Lady in the Water (2006)',
 'Bug (2007)',
 'Hancock (2008)',
 'Terminator Salvation (2009)',
 'Ted (2012)',
 'Now You See Me (2013)',
 'Divergent (2014)',
 'Mother! (2017)',
 'The Greatest Showman (2017)'],
 ['It Takes Two (1995)',
 'Mary Reilly (1996)',
 'Boomerang (1992)',
 'Space Jam (1996)',
 'Grease 2 (1982)',
 'In Love and War (1996)',
 'Jungle2Jungle (a.k.a. Jungle 2 Jungle) (1997)',
 'Postman, The (1997)',
 'Friday the 13th Part 2 (1981)',
 'Friday the 13th Part 3: 3D (1982)',
 'Friday the 13th Part IV: The Final Chapter (1984)',
 'Friday the 13th Part V: A New Beginning (1985)',
 'Halloween II (1981)',
 'Halloween III: Season of the Witch (1982)',
 'Cujo (1983)',
 'Children of the Corn (1984)',

'Night at the Roxbury, A (1998)',
'Rocky III (1982)',
'Rocky IV (1985)',
'Crocodile Dundee II (1988)',
'Superman III (1983)',
'Haunting, The (1999)',
'Cat's Eye (1985)",
'Astronaut's Wife, The (1999)",
'Home Alone 2: Lost in New York (1992)',
'Taming of the Shrew, The (1967)',
'Scream 3 (2000)',
'Supergirl (1984)',
'Punchline (1988)',
'House of Mirth, The (2000)',
'Series 7: The Contenders (2001)',
'Greatest Story Ever Told, The (1965)',
'Necessary Roughness (1991)',
'Big Business (1988)',
'Twins (1988)',
'Look Who's Talking (1989)",
'America's Sweethearts (2001)",
'Vanishing, The (1993)',
'SpaceCamp (1986)',
'Staying Alive (1983)',
'Personal Velocity (2002)',
'House of 1000 Corpses (2003)',
'Legal Eagles (1986)',
'Legally Blonde 2: Red, White & Blonde (2003)',
'League of Extraordinary Gentlemen, The (a.k.a. LXG) (2003)',
'Northfork (2003)',
'Lara Croft Tomb Raider: The Cradle of Life (2003)',
'Double Life, A (1947)',
'Once Bitten (1985)',
'Sleeping with the Enemy (1991)',
'Shot in the Dark, A (1964)',
'Passion of the Christ, The (2004)',
'Lust for Life (1956)',
'Purple Rain (1984)',
'Hunting of the President, The (2004)',
'Hello Again (1987)',
'Sands of Iwo Jima (1949)',
'Night and Day (1946)',
'Invisible, The (2007)',
'Fantastic Four: Rise of the Silver Surfer (2007)',
'Twilight (2008)',
'Observe and Report (2009)',
'Transformers: Revenge of the Fallen (2009)',

```
'Gentlemen Broncos (2009)',
'Green Lantern (2011)',
'What If (2013)',
'Godzilla (2014)',
'Taken 3 (2015)',
'Batman v Superman: Dawn of Justice (2016)',
'Tomb Raider (2018)']]
```

2.2 Finally, make recommendations

Before writing the function to make movie recommendations, I first played with the `predict` function to make sure I was using it correctly. Below is a sample for how to use the `predict` function.

```
[21]: Y = [[1,1]]
      Y2 = model2.predict(Y)
      Y2
```

```
[21]: array([1], dtype=int32)
```

The `recommend` function recommends a movie. It take the movie you provide, and give back a movie that is in that same cluster. It randomly selects movies from the cluster with the `random` function.

```
[22]: def recommend(input):
      # here we find the row the movie is at
      movie_row = movies_rpt.loc[movies_rpt.title == input]
      # then we need the parameter from that row to predict
      model_input = movie_row[['rating', 'polarity']]
      cluster = model2.predict(model_input)[0]
      #randomly recommend something from that sample
      print(random.sample(movie_to_cluster[cluster], 2))
```

```
[23]: recommend('Father of the Bride Part II (1995)')
```

```
['After the Thin Man (1936)', 'White Men Can't Jump (1992)']
```

```
[24]: recommend('Tomb Raider (2018)')
```

```
['Green Lantern (2011)', 'Transformers: Revenge of the Fallen (2009)']
```

3 Trail and ERROR

Here, I discuss some of the things I tried that I couldn't get to work. And so it goes.

3.1 Pivot Table

I tried to make a pivot table, but there were too many NaN values. Not sure what to do. I think I have the wrong index, columns, and values, for the pivot table.

```
[25]: movie_rp_pivot = movies.pivot(index='polarity', columns= 'title',
    ↪values='rating')
```

```
[26]: movie_rp_pivot.head(5)
```

```
[26]: title      (500) Days of Summer (2009)  ...And Justice for All (1979)  \
polarity
-1.000                                NaN                                NaN
-0.800                                NaN                                NaN
-0.700                                NaN                                NaN
-0.700                                NaN                                NaN
-0.625                                NaN                                NaN

title      10 Cloverfield Lane (2016)  10 Things I Hate About You (1999)  \
polarity
-1.000                                NaN                                NaN
-0.800                                NaN                                NaN
-0.700                                NaN                                NaN
-0.700                                NaN                                NaN
-0.625                                NaN                                NaN

title      101 Dalmatians (1996)  \
polarity
-1.000                                NaN
-0.800                                NaN
-0.700                                NaN
-0.700                                NaN
-0.625                                NaN

title      101 Dalmatians (One Hundred and One Dalmatians) (1961)  \
polarity
-1.000                                NaN
-0.800                                NaN
-0.700                                NaN
-0.700                                NaN
-0.625                                NaN

title      11'09"01 - September 11 (2002)  12 Angry Men (1957)  \
polarity
-1.000                                NaN                                NaN
-0.800                                NaN                                NaN
-0.700                                NaN                                NaN
-0.700                                NaN                                NaN
-0.625                                NaN                                NaN

title      127 Hours (2010)  13 Going on 30 (2004)  ...  \
polarity                                                    ...
```

-1.000	NaN	NaN	...
-0.800	NaN	NaN	...
-0.700	NaN	NaN	...
-0.700	NaN	NaN	...
-0.625	NaN	NaN	...

title	Young Frankenstein (1974)	Z (1969)	\
polarity			
-1.000	NaN	NaN	
-0.800	NaN	NaN	
-0.700	NaN	NaN	
-0.700	NaN	NaN	
-0.625	NaN	NaN	

title	Zack and Miri Make a Porno (2008)	Zelary (2003)	Zelig (1983)	\
polarity				
-1.000		NaN	NaN	NaN
-0.800		NaN	NaN	NaN
-0.700		NaN	NaN	NaN
-0.700		NaN	NaN	NaN
-0.625		NaN	NaN	NaN

title	Zero Dark Thirty (2012)	Zombieland (2009)	Zoolander (2001)	\
polarity				
-1.000	NaN	NaN	NaN	
-0.800	NaN	NaN	NaN	
-0.700	NaN	NaN	NaN	
-0.700	NaN	NaN	NaN	
-0.625	NaN	NaN	NaN	

title	Zulu (1964)	eXistenZ (1999)
polarity		
-1.000	NaN	NaN
-0.800	NaN	NaN
-0.700	NaN	NaN
-0.700	NaN	NaN
-0.625	NaN	NaN

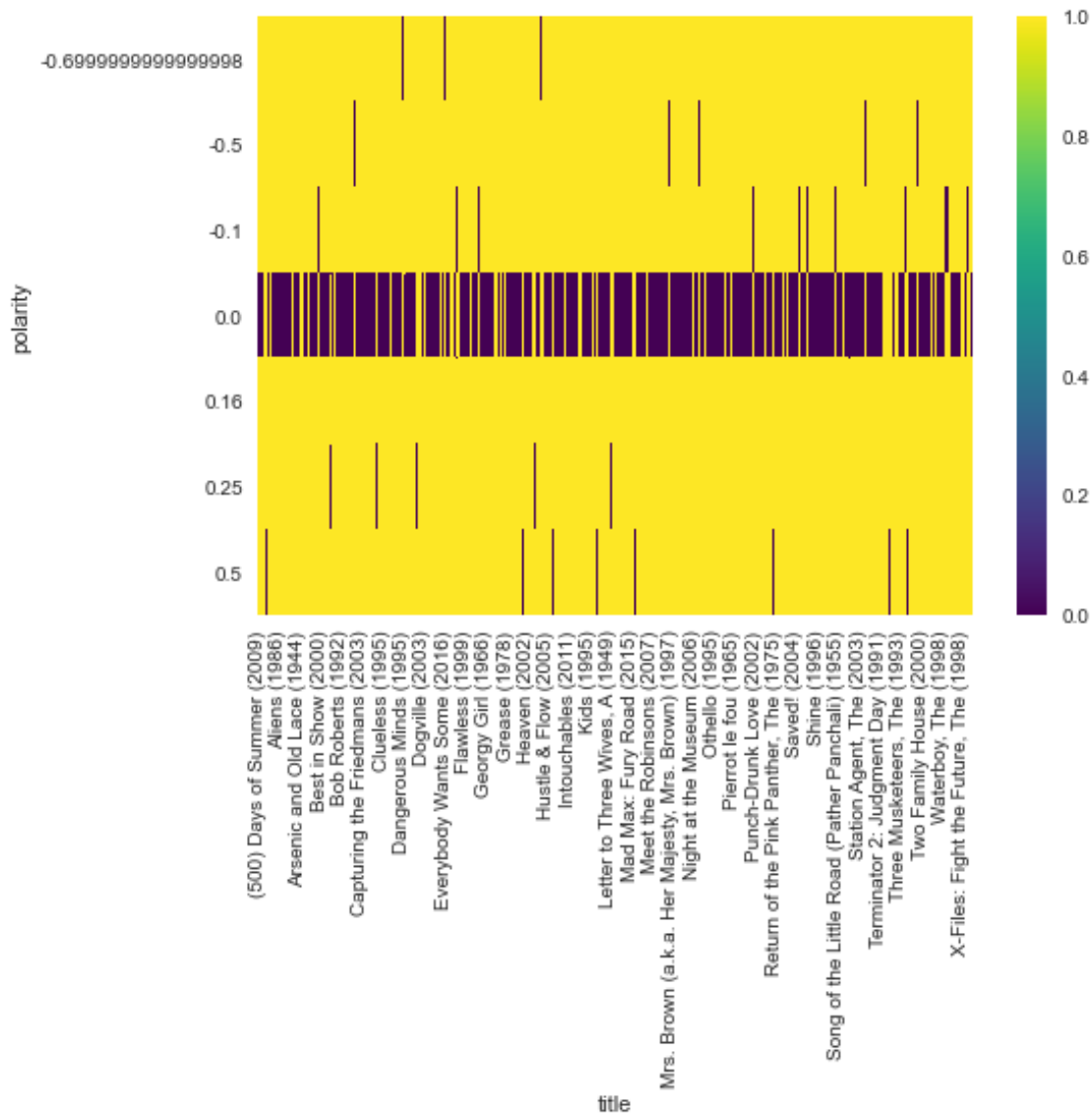
[5 rows x 1554 columns]

There are a lot of NaN values here. This is referred to as “low-dense”. So, I used the `thresh` command so that the column must have 10 or more values to be kept. This didn’t help much.

```
[27]: polarity = {-1, -0.5, 0 ,0.5, 1}
```

```
[28]: movie_rp_pivot= movie_rp_pivot.dropna(thresh=10)
```

```
[29]: import seaborn as sns
sns.heatmap(movie_rp_pivot.isnull(), cmap='viridis')
plt.show()
```



Pivot Table Findings I know I did this wrong because reading the table is unclear. Instead, would try again with different features as part of the table.

3.2 Bag of Words

Bag of words turns text into numerical feature vector. It describes the word occurrence.

Here I created a bag of words from <https://www.freecodecamp.org/news/an-introduction-to-bag-of-words-and-how-to-code-it-in-python-for-nlp-282e87a9da04/>

I performed vectorization with help from <http://brandonrose.org/clustering#K-means-clustering> and <https://pythonprogramminglanguage.com/kmeans-text-clustering/>

```
[30]: from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(movies['tag'])
print(X.toarray())
X
```

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

```
[30]: <1554x1742 sparse matrix of type '<class 'numpy.int64'>'
      with 5210 stored elements in Compressed Sparse Row format>
```

```
[31]: from sklearn.cluster import KMeans
num_clusters = 10
km_bag = KMeans(n_clusters=num_clusters)
model_bag = km_bag.fit(X)

clusters = km_bag.labels_.tolist()
y_pred = km_bag.fit_predict(X)
y_pred
```

```
[31]: array([1, 1, 1, ..., 1, 1, 1], dtype=int32)
```

```
X3 = (movies['rating'].values,X) print(X3.toarray())
```

```
[32]: order_centroids = km_bag.cluster_centers_.argsort()[:, :-1]
terms = vectorizer.get_feature_names()
for i in range(num_clusters):
    print("Cluster %d:" % i),
    for ind in order_centroids[i, :10]:
        print(' %s' % terms[ind]),
    print
```

Cluster 0:

```
in
queue
netflix
movie
black
dialogue
```


robert
samurai
up
hayao
Cluster 1:
comedy
and
atmospheric
funny
ending
disney
dark
war
superhero
religion
Cluster 2:
dark
twist
ending
comedy
thought
psychological
psychology
violence
provoking
challenging
Cluster 3:
crime
organized
tarantino
corruption
jean
romance
love
murder
men
heist
Cluster 4:
visually
appealing
atmospheric
provoking
thought
dreamlike
surreal
hallucinatory
beautiful
stunning

Cluster 5:

violence
dialogue
soundtrack
film
cult
great
tarantino
nonlinear
noir
crime

Cluster 6:

sci
space
fi
classic
epic
action
adventure
wars
star
goodie

Cluster 7:

comedy
dark
black
coen
quirky
brothers
drugs
humor
friendship
short

Cluster 8:

space
travel
sci
fi
time
aliens
robots
scifi
effects
cult

Cluster 9:

bad
acting
plot

```
fi  
evans  
sequel  
female  
marvel  
christian  
post
```

```
[33]: kmeans_bag = KMeans(n_clusters=10)  
model = kmeans_rp_2.fit_predict(X_rp)  
model2 = kmeans_rp_2.fit(X_rp)
```

I used the same `KMeans` techniques here and if I was going to proceed I would need to map each title to the cluster it belongs to based on tag. This is something to do as part of future work.