Milestone_supervised_learning

September 26, 2021

Data Cleaning and preprocessing

[3]:
```
#1.Load Dataframe
# we will only use four columns from the original dataset
import pandas as pd
df = pd.read_csv ('movies.csv', encoding='latin1')
df=df[['genres', 'id',"imdb_id","overview"]]
```

[5]:
```
df.head()
```

[5]:
```
                                              genres     id    imdb_id  \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, '…    862  tt0114709
1  [{'id': 12, 'name': 'Adventure'}, {'id': 14, '…   8844  tt0113497
2  [{'id': 10749, 'name': 'Romance'}, {'id': 35, …  15602  tt0113228
3  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam…  31357  tt0114885
4                    [{'id': 35, 'name': 'Comedy'}]  11862  tt0113041

                                            overview
0  Led by Woody, Andy's toys live happily in his …
1  When siblings Judy and Peter discover an encha…
2  A family wedding reignites the ancient feud be…
3  Cheated on, mistreated and stepped on, the wom…
4  Just when George Banks has recovered from his …
```

[6]:
```
#drop rows with empty genere
df = df[df.genres != "[]"]
#drop if any of rows with one missing value
df.dropna()
# drop rows with imdb==0
df = df[df.imdb_id != 0]
#create a new column for genres so that genres for each movie is in a list
from ast import literal_eval
df['newgenres'] = df['genres'].apply(literal_eval)
df['newgenres'] =df['newgenres'].apply(lambda x: [i['name'] for i in x] if␣
 ↪isinstance(x, list) else [])
df.head()
```

[6]:
```
                                              genres     id    imdb_id  \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, '…    862  tt0114709
1  [{'id': 12, 'name': 'Adventure'}, {'id': 14, '…   8844  tt0113497
2  [{'id': 10749, 'name': 'Romance'}, {'id': 35, …  15602  tt0113228
3  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam…  31357  tt0114885
4                    [{'id': 35, 'name': 'Comedy'}]  11862  tt0113041

                                            overview  \
0  Led by Woody, Andy's toys live happily in his …
1  When siblings Judy and Peter discover an encha…
2  A family wedding reignites the ancient feud be…
3  Cheated on, mistreated and stepped on, the wom…
4  Just when George Banks has recovered from his …

                      newgenres
0   [Animation, Comedy, Family]
1  [Adventure, Fantasy, Family]
2             [Romance, Comedy]
3      [Comedy, Drama, Romance]
4                       [Comedy]
```

[7]:
```
df.shape
```

[7]:
```
(43024, 5)
```

[8]:
```
genres = []   ## returns list of lists of genres
import json
# extract genres
for i in df['newgenres']:
  genres.append(i)
genres
```

[8]:
```
[['Animation', 'Comedy', 'Family'],
 ['Adventure', 'Fantasy', 'Family'],
 ['Romance', 'Comedy'],
 ['Comedy', 'Drama', 'Romance'],
 ['Comedy'],
 ['Action', 'Crime', 'Drama', 'Thriller'],
 ['Comedy', 'Romance'],
 ['Action', 'Adventure', 'Drama', 'Family'],
 ['Action', 'Adventure', 'Thriller'],
 ['Adventure', 'Action', 'Thriller'],
 ['Comedy', 'Drama', 'Romance'],
 ['Comedy', 'Horror'],
 ['Family', 'Animation', 'Adventure'],
 ['History', 'Drama'],
 ['Action', 'Adventure'],
```

```
['Drama', 'Crime'],
['Drama', 'Romance'],
['Crime', 'Comedy'],
['Crime', 'Comedy', 'Adventure'],
['Action', 'Comedy', 'Crime'],
['Comedy', 'Thriller', 'Crime'],
['Drama', 'Thriller'],
['Action', 'Adventure', 'Crime', 'Thriller'],
['Drama', 'Fantasy', 'Science Fiction', 'Thriller'],
['Drama', 'Romance'],
['Drama'],
['Comedy', 'Drama', 'Family'],
['Drama', 'Romance'],
['Fantasy', 'Science Fiction', 'Adventure'],
['Drama', 'Crime'],
['Drama', 'Crime'],
['Science Fiction', 'Thriller', 'Mystery'],
['Romance', 'Adventure'],
['Fantasy', 'Drama', 'Comedy', 'Family'],
['History', 'Drama', 'Romance'],
['Drama'],
['Adventure', 'History', 'Drama', 'Family'],
['Comedy', 'Family', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Drama'],
['Drama', 'War'],
['Action', 'Crime', 'Drama', 'History'],
['Drama', 'Romance'],
['Action', 'Fantasy'],
['Fantasy', 'Drama', 'Comedy', 'Thriller'],
['Drama', 'Romance'],
['Crime', 'Mystery', 'Thriller'],
['Adventure', 'Animation', 'Drama', 'Family'],
['Drama', 'Romance'],
['Drama', 'Crime', 'Thriller'],
['Action', 'Thriller', 'Drama'],
['Comedy', 'Romance'],
['Drama', 'Foreign'],
['Action', 'Adventure', 'Comedy', 'Family'],
['Drama'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Thriller', 'Drama', 'Mystery'],
['Adventure', 'Family', 'Fantasy'],
['Drama', 'Thriller'],
['Music', 'Drama', 'Family'],
['Comedy'],
```

```
['Comedy', 'Romance'],
['Comedy'],
['Action', 'Science Fiction'],
['Drama', 'Family'],
['Comedy', 'Romance'],
['Comedy'],
['Horror', 'Action', 'Thriller', 'Crime'],
['Action', 'Thriller', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Drama', 'History'],
['Drama', 'Romance'],
['Comedy', 'Family'],
['Horror', 'Science Fiction'],
['Documentary'],
['Drama', 'Thriller'],
['Drama', 'Thriller'],
['Family', 'Drama'],
['Drama', 'Crime'],
['Drama', 'Comedy'],
['Romance', 'Drama'],
['Drama', 'Romance'],
['Action', 'Drama'],
['Adventure', 'Comedy', 'Family'],
['Comedy'],
['Crime', 'Drama', 'Thriller'],
['Drama'],
['Drama', 'Horror', 'Thriller', 'Romance'],
['Comedy', 'Horror', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Action', 'Adventure', 'Drama', 'Thriller'],
['Comedy'],
['Drama'],
['Action', 'Adventure', 'Drama', 'Science Fiction', 'Thriller'],
['Documentary'],
['Drama', 'Thriller'],
['Comedy', 'Crime', 'Drama'],
['Comedy', 'Romance'],
['Drama', 'Mystery', 'Science Fiction', 'Thriller'],
['Comedy'],
['Drama', 'Romance'],
['Drama', 'Comedy', 'Family'],
['Action', 'Comedy', 'Music', 'Family', 'Adventure'],
['Documentary'],
['Crime'],
['Action', 'Drama', 'History', 'War'],
['Crime', 'Drama'],
['Crime', 'Action', 'Comedy', 'Thriller'],
```

```
['Drama', 'Thriller'],
['Drama'],
['Comedy'],
['Documentary', 'Drama', 'Foreign'],
['Comedy', 'Crime', 'Drama'],
['Comedy', 'Romance'],
['Comedy', 'Drama'],
['Action', 'Adventure', 'Comedy', 'Drama', 'Family'],
['Drama'],
['Comedy', 'Romance'],
['Drama', 'Comedy', 'Romance'],
['Drama', 'Romance'],
['Comedy', 'Romance'],
['Fantasy'],
['Drama'],
['Comedy', 'Romance'],
['Comedy', 'Drama'],
['Drama', 'Romance', 'War'],
['Action', 'Thriller', 'Mystery', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Music', 'Documentary'],
['Comedy'],
['Documentary'],
['Comedy'],
['Drama'],
['Drama', 'Romance'],
['Comedy'],
['Drama'],
['Comedy', 'Drama', 'Romance'],
['Action', 'Comedy', 'Crime', 'Thriller'],
['Action', 'Adventure', 'Drama', 'Family'],
['Drama', 'Crime'],
['Romance', 'Drama', 'Comedy'],
['Crime', 'Comedy', 'Drama', 'Thriller'],
['Drama'],
['Adventure'],
['Drama', 'Horror'],
['Action', 'Crime', 'Fantasy'],
['Drama', 'Romance'],
['Drama', 'Thriller'],
['Comedy'],
['Action', 'Comedy'],
['Fantasy', 'Comedy', 'Family'],
['Mystery', 'Crime', 'Drama'],
['Action', 'Adventure', 'Drama', 'Mystery', 'Science Fiction', 'Thriller'],
['Action', 'Thriller', 'Drama'],
['Documentary'],
```

```
['Thriller', 'Action', 'Crime'],
['Drama', 'Mystery', 'Thriller'],
['Action', 'Thriller'],
['Action', 'Comedy', 'Crime', 'Drama'],
['Drama', 'Crime'],
['Action', 'Adventure', 'Drama', 'Romance'],
['Adventure'],
['Action', 'Crime', 'Thriller', 'Drama'],
['Comedy', 'Drama', 'Romance'],
['Adventure', 'Action', 'Drama', 'Science Fiction', 'Thriller'],
['Science Fiction'],
['Comedy', 'Drama'],
['Drama', 'Crime'],
['Drama', 'Comedy'],
['Mystery', 'Horror', 'Thriller'],
['Comedy', 'Drama'],
['Drama', 'Family', 'Romance'],
['Romance', 'Comedy'],
['Action', 'Adventure', 'Science Fiction', 'Family', 'Fantasy'],
['Comedy', 'Drama', 'Romance'],
['Thriller', 'Foreign'],
['Horror', 'Thriller'],
['Crime', 'Drama', 'Mystery', 'Thriller', 'Action'],
['Comedy'],
['Comedy'],
['Fantasy', 'Horror', 'Thriller'],
['Fantasy', 'Comedy', 'Drama'],
['Drama'],
['Drama', 'History', 'Romance'],
['Documentary'],
['Drama'],
['Comedy', 'Drama'],
['Comedy'],
['Science Fiction', 'Horror', 'Action'],
['Drama'],
['Crime', 'Drama', 'Science Fiction', 'Thriller'],
['Drama', 'Romance'],
['Thriller'],
['Drama', 'Fantasy'],
['Drama', 'Romance'],
['Comedy'],
['Action', 'Thriller'],
['Comedy', 'Drama', 'Family'],
['Documentary'],
['Drama', 'Romance'],
['Adventure', 'Action'],
['Drama'],
```

```
['Action', 'Western'],
['Drama'],
['Adventure', 'Action', 'Comedy', 'Crime', 'Family'],
['Drama'],
['Drama', 'War'],
['Drama', 'Romance'],
['Comedy'],
['Drama', 'Thriller'],
['Comedy', 'Drama'],
['Drama', 'Family'],
['Thriller', 'Drama', 'Horror', 'Mystery'],
['Drama', 'Family', 'Romance'],
['Comedy'],
['Romance', 'Drama', 'Comedy'],
['Drama', 'Thriller', 'Crime', 'Mystery', 'Romance'],
['Thriller'],
['Action', 'Adventure', 'Thriller'],
['Crime', 'Comedy', 'Fantasy'],
['Drama', 'Thriller', 'Mystery'],
['Crime', 'Drama', 'Mystery'],
['Comedy'],
['Comedy', 'Drama', 'Romance'],
['Drama', 'Mystery'],
['Thriller', 'Comedy', 'Crime', 'Romance'],
['Comedy', 'Drama', 'History'],
['Comedy', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Action', 'Adventure', 'Drama', 'Family'],
['Romance', 'Animation', 'Family', 'Comedy', 'Adventure'],
['Drama', 'Thriller'],
['Drama', 'Family', 'Fantasy'],
['History', 'Drama'],
['Drama', 'Family', 'Fantasy'],
['Fantasy', 'Animation', 'Science Fiction', 'Family'],
['Drama', 'Action', 'Thriller', 'Crime'],
['Documentary'],
['Drama', 'Fantasy'],
['Comedy', 'Family'],
['Drama', 'Music', 'Romance'],
['Comedy', 'Family', 'Drama'],
['Action', 'Thriller'],
['Comedy', 'Drama', 'Romance'],
['Horror', 'Romance'],
['Drama', 'Romance'],
['Comedy', 'Crime'],
['Comedy', 'Family', 'Romance'],
['Action', 'Drama', 'Mystery', 'Thriller'],
```

['Adventure', 'Fantasy', 'Comedy', 'Science Fiction', 'Family'],
['Action', 'Crime', 'Drama', 'Thriller'],
['Adventure', 'Action', 'Science Fiction'],
['Drama', 'Romance'],
['Drama', 'Family', 'Fantasy'],
['Drama'],
['Crime', 'Drama'],
['Drama', 'Romance'],
['Adventure', 'Drama', 'Romance', 'War'],
['Adventure', 'Comedy', 'Family'],
['Action', 'Drama'],
['Crime', 'Drama'],
['Drama', 'Romance'],
['Drama'],
['Comedy', 'Drama', 'History'],
['Drama', 'Horror', 'Science Fiction', 'Romance'],
['Action', 'Comedy', 'Family'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Romance'],
['Fantasy', 'Drama', 'Family'],
['Comedy', 'Romance'],
['Comedy', 'Drama'],
['Crime', 'Drama'],
['Drama', 'Comedy'],
['Drama', 'Thriller'],
['Crime', 'Drama'],
['Action', 'Thriller', 'Crime', 'Drama'],
['Horror'],
['Science Fiction', 'Action', 'Thriller'],
['Drama', 'Romance'],
['Crime', 'Thriller', 'Drama'],
['Comedy', 'Drama', 'Romance'],
['Drama'],
['Drama', 'Romance', 'Thriller'],
['Action', 'Drama', 'Science Fiction', 'Thriller'],
['Thriller', 'Crime', 'Drama'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Thriller', 'Crime'],
['Drama'],
['Comedy', 'Drama'],
['Drama', 'Romance'],
['History', 'Drama'],
['Drama', 'History'],
['Drama', 'History'],
['Action', 'Western'],
['Drama', 'Comedy'],

```
['Drama', 'Comedy'],
['Drama', 'Mystery', 'Romance'],
['Drama', 'Music', 'Mystery'],
['Comedy', 'Drama', 'Mystery'],
['Drama'],
['Comedy', 'Family', 'TV Movie'],
['Horror', 'Drama'],
['Comedy', 'Drama'],
['Animation'],
['Drama', 'Family', 'Fantasy'],
['Action', 'Thriller'],
['Action', 'Adventure', 'Science Fiction'],
['Fantasy', 'Drama', 'Comedy', 'Family'],
['Drama', 'Crime'],
['Crime', 'Thriller'],
['Drama', 'Thriller'],
['Comedy', 'Drama'],
['Crime', 'Comedy'],
['Comedy', 'Drama', 'Romance'],
['Comedy'],
['Drama', 'Romance', 'War'],
['Action', 'Comedy', 'Fantasy', 'Science Fiction'],
['Horror', 'Comedy', 'Thriller'],
['Science Fiction', 'Action', 'Adventure', 'Thriller'],
['Crime', 'Horror', 'Thriller'],
['Drama', 'Romance'],
['Thriller', 'Horror', 'Science Fiction'],
['Comedy'],
['Drama', 'Romance'],
['Thriller', 'Crime'],
['Drama', 'War'],
['Romance', 'Drama'],
['Action', 'Crime', 'Science Fiction', 'Thriller'],
['Comedy', 'Drama', 'Romance'],
['Drama'],
['Comedy', 'Romance'],
['Drama', 'Comedy', 'Romance'],
['Comedy', 'Drama', 'Family'],
['Comedy', 'Mystery'],
['Drama', 'Comedy'],
['Drama', 'Music'],
['Thriller', 'Drama', 'Romance'],
['Action', 'Comedy', 'Thriller'],
['Action', 'Drama', 'Thriller'],
['Drama', 'Thriller', 'Crime', 'Mystery'],
['Comedy', 'Drama', 'Family', 'Romance'],
['Comedy', 'Drama'],
```

```
['Fantasy', 'Action', 'Thriller'],
['Drama'],
['Fantasy', 'Comedy', 'Family'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Romance', 'Drama', 'Crime'],
['Comedy', 'Drama', 'Romance'],
['Action', 'Comedy', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Adventure', 'Drama', 'Family'],
['Documentary', 'Foreign'],
['Family', 'Animation', 'Drama'],
['Drama', 'History'],
['Horror', 'Thriller', 'Mystery', 'Fantasy'],
['Romance', 'Comedy', 'Crime', 'Fantasy'],
['Action', 'Adventure', 'Comedy', 'Drama', 'Western'],
['Drama'],
['Comedy', 'Crime'],
['Comedy', 'Drama'],
['Drama'],
['Crime', 'Drama', 'Thriller'],
['Comedy', 'Family'],
['Drama'],
['Action', 'Adventure', 'Crime', 'Thriller'],
['Action', 'Adventure', 'Crime'],
['Comedy', 'Romance'],
['Thriller', 'Science Fiction', 'Action', 'Crime'],
['Action', 'Thriller'],
['Drama', 'Romance'],
['Fantasy'],
['Drama', 'Action', 'Western'],
['Action', 'Crime', 'Romance', 'Thriller'],
['Comedy', 'Drama'],
['Comedy', 'Drama'],
['Action', 'Comedy', 'Crime'],
['Drama'],
['Drama'],
['Action', 'Crime'],
['Crime', 'Drama', 'Romance'],
['Animation', 'Science Fiction'],
['Action', 'Adventure', 'Thriller'],
['Action', 'Comedy'],
['Action', 'Adventure', 'Crime', 'Drama', 'Thriller'],
['Horror', 'Thriller'],
['Documentary'],
['Romance', 'Adventure', 'Crime', 'Drama'],
['Thriller', 'Drama'],
```

```
['Action', 'Thriller'],
['Comedy'],
['Documentary'],
['Action', 'Fantasy', 'Science Fiction'],
['Horror', 'Mystery', 'Thriller'],
['Drama', 'Romance'],
['Crime', 'Drama'],
['Comedy', 'Family', 'Fantasy'],
['Comedy'],
['Drama', 'Romance'],
['Comedy', 'Crime', 'Music'],
['Comedy', 'Family'],
['Action', 'Comedy'],
['Action', 'Adventure', 'Comedy', 'Drama', 'Western'],
['Comedy', 'Drama', 'Romance'],
['Drama', 'Comedy', 'Fantasy'],
['Comedy'],
['Action', 'Comedy', 'Crime'],
['Action', 'Adventure', 'Drama', 'Family'],
['Mystery', 'Crime', 'Drama', 'Thriller'],
['Drama', 'Action', 'Thriller'],
['Drama'],
['Drama', 'Romance'],
['Horror', 'Science Fiction', 'Thriller'],
['Drama', 'Romance'],
['Drama', 'Crime'],
['Comedy'],
['Drama', 'Comedy'],
['Action', 'Crime', 'Drama', 'Thriller'],
['Action', 'Comedy', 'Drama', 'Western'],
['Comedy', 'Thriller', 'Mystery'],
['Action', 'Adventure', 'Thriller'],
['Comedy', 'Science Fiction', 'Family'],
['Drama', 'Mystery', 'Romance', 'Thriller'],
['Action', 'Comedy', 'Family'],
['Western', 'Action', 'Comedy'],
['Drama'],
['Comedy'],
['Comedy', 'Drama'],
['Crime', 'Action', 'Science Fiction'],
['Documentary'],
['Western', 'Drama', 'Comedy', 'Romance'],
['Comedy'],
['Drama'],
['Drama', 'Comedy', 'Romance'],
['Drama'],
['Comedy'],
```

```
['Comedy', 'Drama'],
['Drama', 'Mystery', 'Romance'],
['Comedy', 'Thriller', 'Mystery', 'Romance', 'Foreign'],
['Comedy', 'Romance'],
['Drama', 'Mystery', 'Thriller'],
['Family', 'Adventure'],
['Crime', 'Drama', 'Thriller'],
['Adventure', 'Action', 'Thriller', 'Crime', 'Mystery'],
['History', 'Action', 'Drama', 'Western'],
['Drama', 'Action', 'Thriller'],
['Family', 'Comedy', 'Crime'],
['Drama', 'Romance'],
['Comedy', 'Drama'],
['Drama', 'Thriller'],
['Action', 'Adventure', 'Crime', 'Thriller'],
['History', 'War', 'Action', 'Drama'],
['Action', 'Comedy', 'War'],
['Comedy'],
['Drama', 'Comedy', 'Romance'],
['Romance', 'Drama'],
['Comedy'],
['Comedy', 'Drama'],
['Comedy', 'Drama'],
['Comedy'],
['Action', 'Drama', 'Thriller', 'Crime', 'Mystery'],
['Drama'],
['Comedy', 'Drama', 'Romance'],
['Drama', 'Music'],
['Comedy'],
['Action', 'Thriller', 'Crime'],
['Adventure', 'Science Fiction'],
['Thriller', 'Crime'],
['Action', 'Crime', 'Drama', 'Thriller'],
['Drama'],
['Action', 'Adventure', 'Drama', 'Family'],
['Adventure', 'Fantasy', 'Action', 'Comedy', 'Family'],
['Comedy', 'Family'],
['Action', 'Comedy', 'Western'],
['Drama', 'Romance'],
['Comedy'],
['Drama', 'Crime'],
['Drama'],
['Comedy', 'Mystery'],
['Drama', 'Action', 'Crime'],
['Action', 'Adventure', 'Drama', 'Thriller'],
['Drama', 'Romance'],
['Thriller', 'Comedy', 'Drama', 'Romance'],
```

```
['Drama', 'Comedy', 'Romance'],
['Drama', 'Romance'],
['Comedy', 'Romance'],
['Comedy', 'Drama', 'Family'],
['Comedy', 'Drama'],
['Adventure'],
['Comedy', 'Drama'],
['Action', 'Drama', 'Science Fiction', 'Thriller'],
['Comedy', 'Drama', 'Family', 'Fantasy', 'Science Fiction'],
['Romance', 'Drama', 'Fantasy'],
['Crime', 'Drama', 'Thriller'],
['Drama'],
['Drama', 'Romance'],
['Drama', 'Romance'],
['Drama'],
['Horror', 'Science Fiction', 'Thriller'],
['Comedy', 'Crime', 'Drama', 'Music', 'Mystery', 'Romance'],
['Comedy'],
['Drama', 'Romance'],
['Comedy', 'War'],
['Action', 'Drama', 'Thriller'],
['Comedy', 'Romance'],
['Action', 'Adventure', 'Crime', 'Science Fiction', 'Thriller'],
['Comedy'],
['Action', 'Crime', 'Drama', 'Thriller'],
['Action', 'Drama', 'Thriller'],
['Drama', 'Romance'],
['Drama'],
['Drama'],
['Drama'],
['Drama', 'History', 'War'],
['Comedy', 'Drama'],
['Drama'],
['Drama'],
['Drama', 'Family', 'Fantasy'],
['Comedy', 'Crime', 'Horror', 'Thriller'],
['Adventure', 'Fantasy', 'Action', 'Thriller', 'Science Fiction'],
['Drama', 'Romance'],
['Comedy', 'Drama'],
['Drama', 'Family'],
['Comedy', 'Drama'],
['Mystery', 'Drama', 'Comedy'],
['Comedy', 'Drama', 'Romance'],
['Drama', 'Thriller'],
['Science Fiction', 'Drama', 'Thriller'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Crime', 'Romance'],
```

```
['Crime', 'Action', 'Mystery', 'Thriller'],
['History', 'Drama', 'Romance'],
['Adventure', 'Comedy', 'Family', 'Fantasy'],
['Action', 'Adventure', 'Thriller', 'Crime'],
['Action', 'Thriller'],
['Drama', 'Music'],
['Comedy', 'Drama', 'Family', 'Romance'],
['Fantasy', 'Animation', 'Family'],
['Action', 'Adventure', 'Comedy'],
['Action', 'Adventure', 'Drama', 'History', 'Western'],
['Drama', 'Action', 'Thriller'],
['Action', 'Thriller', 'Crime', 'Romance'],
['Documentary', 'History'],
['Fantasy', 'Science Fiction', 'Family', 'Animation'],
['Drama', 'Comedy'],
['Thriller', 'Drama', 'Romance'],
['Comedy', 'Drama'],
['Drama', 'Romance'],
['Comedy', 'Romance'],
['Drama', 'Horror', 'Thriller'],
['Comedy', 'Romance'],
['Comedy', 'Drama'],
['Comedy', 'Drama'],
['Comedy', 'Family'],
['Drama'],
['TV Movie', 'Drama', 'Romance', 'Comedy'],
['Drama', 'Romance'],
['Comedy', 'Foreign'],
['Comedy', 'Drama'],
['Romance', 'Comedy', 'Family'],
['Comedy'],
['Drama', 'Family', 'Adventure'],
['Thriller', 'Crime', 'Drama'],
['Comedy', 'Drama', 'History', 'Mystery'],
['History', 'Documentary'],
['Drama', 'Comedy', 'Romance', 'Foreign'],
['Comedy', 'Drama'],
['Romance', 'Drama'],
['Comedy'],
['Comedy', 'Family'],
['Fantasy', 'Drama', 'Thriller', 'Mystery', 'Romance'],
['Animation', 'Family', 'Comedy', 'Adventure', 'Fantasy', 'Romance'],
['Action', 'Thriller', 'Science Fiction'],
['Adventure', 'Drama', 'Western'],
['Thriller', 'Action', 'Drama'],
['Fantasy', 'Action'],
['Crime', 'Drama', 'Thriller'],
```

```
['Fantasy', 'Animation', 'Family'],
['Romance', 'Family', 'Animation', 'Fantasy', 'Music'],
['Animation', 'Family'],
['Romance', 'Comedy'],
['Comedy', 'Drama', 'Fantasy'],
['Adventure', 'Western'],
['Crime', 'Thriller'],
['Drama', 'Foreign', 'Romance'],
['Comedy', 'Romance'],
['Comedy', 'Romance'],
['Horror'],
['Romance', 'Drama'],
['Crime', 'Drama', 'Thriller'],
['Adventure', 'Comedy', 'Drama', 'Family'],
['Animation', 'Science Fiction'],
['Horror', 'Thriller', 'Science Fiction'],
['Comedy', 'Romance'],
['Drama', 'Romance'],
['Drama', 'Mystery', 'Thriller'],
['Comedy', 'Foreign'],
['Animation', 'Comedy', 'Family', 'Adventure'],
['Drama', 'Romance'],
['Comedy', 'Romance'],
['Comedy', 'Family'],
['Adventure', 'Drama', 'Documentary'],
['Drama'],
['Thriller', 'Action', 'Drama'],
['Comedy', 'Drama'],
['Comedy', 'Thriller', 'Crime', 'Romance'],
['Comedy', 'Thriller', 'Crime', 'Drama'],
['Crime', 'Drama', 'Mystery', 'Thriller'],
['Drama', 'Romance'],
['Romance', 'Animation', 'Family', 'Adventure', 'Fantasy'],
['Drama', 'History', 'War', 'Romance'],
['Comedy'],
['Fantasy', 'Science Fiction', 'Family'],
['Comedy', 'Drama'],
['Drama', 'Thriller'],
['Comedy', 'Family'],
['Comedy', 'Drama', 'Romance'],
['Comedy'],
['Drama', 'Horror', 'Mystery', 'Thriller'],
['Adventure', 'Drama', 'Comedy', 'Family'],
['Romance', 'Drama'],
['Comedy'],
['Comedy'],
['Drama', 'Romance'],
```

```
['Drama', 'Thriller', 'Mystery', 'War'],
['Adventure', 'Action', 'Thriller'],
['Drama', 'Comedy'],
['Drama', 'Romance'],
['Comedy'],
['Horror', 'Mystery', 'Thriller'],
['Fantasy'],
['Drama'],
['Drama'],
['Action', 'Comedy'],
['Drama', 'Comedy'],
['Crime', 'Drama', 'Thriller'],
['Romance', 'Comedy', 'Drama'],
['Adventure', 'Animation', 'Family'],
['Drama', 'Thriller', 'Romance'],
['Comedy'],
['Comedy', 'Crime', 'Drama'],
['War', 'Drama', 'Comedy'],
['War', 'Drama', 'Romance'],
['Action', 'Thriller'],
['Drama'],
['Drama'],
['Comedy', 'Science Fiction'],
['Animation', 'Comedy', 'Drama', 'Family', 'Fantasy'],
['Science Fiction'],
['Thriller', 'Action', 'Drama'],
['Comedy', 'Horror', 'Science Fiction'],
['Drama', 'Thriller'],
['Drama', 'Romance'],
['Drama', 'Science Fiction', 'Mystery'],
['Drama'],
['Documentary', 'Foreign'],
['Horror', 'Drama', 'Thriller'],
['Drama'],
['Drama', 'Romance'],
['Action', 'Adventure', 'Comedy', 'Family'],
['Drama', 'Romance'],
['Comedy', 'Romance', 'Drama'],
['Action', 'Adventure', 'Science Fiction', 'Thriller'],
['Drama', 'Foreign'],
['Action', 'Thriller'],
['Crime', 'Mystery', 'Thriller'],
['Crime', 'Drama', 'Romance', 'Thriller'],
['Action', 'Comedy', 'Drama', 'Thriller'],
['Drama', 'Romance'],
['Comedy', 'Drama', 'Family', 'Romance'],
['Drama', 'History'],
```

```
['Drama'],
['Romance', 'Drama', 'Mystery'],
['Action', 'Adventure'],
['Comedy', 'Romance'],
['Drama'],
['Drama', 'Mystery', 'Thriller'],
['Comedy', 'Romance'],
['Animation', 'Comedy', 'Family'],
['Comedy'],
['Adventure', 'Family'],
['Drama', 'Thriller', 'Mystery', 'Romance'],
['Drama'],
['Drama', 'Fantasy', 'Western'],
['War', 'Adventure', 'Drama', 'Romance'],
['Action', 'Crime', 'Drama'],
['Comedy', 'Foreign'],
['Fantasy', 'Comedy', 'Science Fiction'],
['Comedy', 'Fantasy', 'Science Fiction'],
['Documentary'],
['Drama', 'Foreign'],
['Drama', 'Fantasy', 'Horror', 'Thriller'],
['Comedy', 'Drama'],
['Drama', 'Thriller'],
['Comedy', 'Drama', 'Romance'],
['Animation', 'Drama'],
['Drama', 'Mystery', 'Thriller'],
['Action', 'Drama', 'Crime', 'Thriller'],
['Action', 'Adventure', 'Thriller'],
['Comedy'],
['Fantasy', 'Drama', 'Horror', 'Comedy'],
['Action', 'Adventure', 'Drama'],
['Action', 'Science Fiction'],
['Drama'],
['Comedy'],
['Action', 'Animation', 'Science Fiction'],
['Horror', 'Thriller'],
['Action', 'Comedy'],
['Comedy', 'Drama'],
['Family', 'Animation', 'Comedy'],
['Drama', 'Action', 'Crime'],
['Comedy', 'Family'],
['Action', 'Mystery', 'Science Fiction', 'Thriller'],
['Drama', 'War'],
['Drama', 'Comedy', 'War'],
['Horror', 'Comedy', 'Drama', 'Fantasy'],
['Comedy', 'Drama', 'Romance'],
['Drama'],
```

```
['Family', 'Adventure'],
['Documentary'],
['Drama', 'Action'],
['Documentary'],
['Drama', 'History', 'War'],
['Adventure', 'Action'],
['Drama', 'Thriller', 'Crime'],
['Drama', 'Comedy'],
['Drama', 'Romance'],
['Comedy', 'Drama', 'Science Fiction'],
['Drama'],
['Comedy', 'Drama', 'Family'],
['Drama'],
['Music', 'Romance'],
['Drama', 'Romance'],
['Drama'],
['Drama', 'Mystery', 'Horror'],
['Western', 'Drama', 'War'],
['Drama', 'Crime'],
['Comedy', 'Romance'],
['Action', 'Adventure', 'Science Fiction'],
['Drama', 'Romance'],
['Drama', 'Mystery', 'Thriller'],
['Drama', 'Animation', 'Family'],
['Comedy', 'Drama', 'Thriller'],
['Comedy'],
['Action', 'Drama', 'Mystery', 'Thriller'],
['Documentary'],
['Fantasy', 'Comedy', 'Romance', 'Science Fiction'],
['Drama', 'Foreign'],
['Drama', 'Romance', 'War'],
['Comedy', 'Drama'],
['Drama'],
['Comedy'],
['Action', 'Adventure', 'Thriller'],
['Horror', 'Comedy'],
['Drama', 'Mystery', 'Romance'],
['Comedy', 'Drama', 'Family'],
['Drama', 'Fantasy', 'Romance', 'Science Fiction'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Romance'],
['Crime', 'Drama', 'Thriller'],
['Drama'],
['Romance', 'Comedy'],
['Adventure'],
['Action', 'Comedy', 'Thriller'],
['Fantasy', 'Comedy', 'Science Fiction', 'Family'],
```

```
['Family', 'Comedy'],
['Action', 'Mystery', 'Thriller'],
['War', 'Drama'],
['Comedy'],
['Comedy'],
['Adventure', 'Drama'],
['War', 'Drama'],
['Drama', 'Comedy', 'Romance'],
['Drama', 'Foreign'],
['Comedy', 'Foreign'],
['Drama'],
['Thriller'],
['Animation', 'Family', 'Fantasy'],
['Fantasy', 'Comedy', 'Music'],
['Comedy'],
['Romance', 'Drama', 'Comedy'],
['Action', 'Thriller'],
['Comedy'],
['Comedy'],
['Action', 'Drama', 'Thriller'],
['Action', 'Drama', 'Science Fiction', 'Thriller'],
['Comedy', 'Family', 'Fantasy'],
['Drama', 'Comedy', 'Romance'],
['Action', 'Fantasy', 'Thriller'],
['Comedy', 'Family'],
['Drama', 'Horror', 'Thriller'],
['Horror', 'Comedy'],
['Drama'],
['Drama', 'Romance'],
['Drama'],
['Comedy', 'Drama'],
['Action', 'Adventure', 'Science Fiction', 'Thriller'],
['Crime', 'Drama'],
['Drama', 'History'],
['Comedy', 'Drama', 'Romance'],
['Drama', 'Music'],
['Drama'],
['Drama'],
['Drama', 'Crime'],
['Comedy', 'Drama'],
['Action', 'Crime', 'Thriller'],
['Drama', 'Comedy'],
['Drama'],
['Crime', 'Drama', 'Romance', 'Thriller'],
['Comedy', 'Family'],
['Drama'],
['Thriller', 'Drama', 'Music', 'Crime'],
```

```
['Comedy', 'Family'],
['Romance', 'Comedy'],
['Drama'],
['Drama'],
['Drama'],
['Action', 'Crime', 'Drama'],
['Action', 'Adventure', 'Crime'],
['Comedy', 'Drama'],
['Drama'],
['Horror', 'Mystery', 'Thriller'],
['Fantasy', 'Horror', 'Science Fiction'],
['Comedy', 'Family'],
['Drama', 'Thriller'],
['Fantasy', 'Comedy', 'Family'],
['Action', 'Adventure', 'Comedy', 'Thriller'],
['Drama', 'Romance', 'War'],
['Animation', 'Adventure', 'Family'],
['Thriller'],
['Horror', 'Thriller'],
['Drama', 'Comedy', 'Romance'],
['Drama', 'History'],
['Drama', 'Romance'],
['Drama', 'Romance'],
['Drama', 'Romance'],
['Drama', 'History', 'Romance', 'Thriller'],
['Comedy', 'Romance'],
['Comedy', 'Music', 'Romance'],
['Drama', 'Comedy', 'Music', 'Romance'],
['Music', 'Comedy', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Mystery', 'Romance', 'Thriller'],
['Drama', 'Mystery', 'Thriller'],
['Comedy', 'Romance'],
['Drama', 'Mystery', 'Thriller', 'Crime'],
['Comedy', 'Music', 'Romance'],
['Mystery', 'Thriller'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Romance'],
['Comedy', 'Mystery', 'Romance', 'Thriller'],
['Drama', 'Romance'],
['Mystery', 'Crime', 'Thriller'],
['Drama', 'Family', 'Music', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Romance'],
['Drama', 'Comedy', 'Family'],
['Comedy', 'Music', 'Romance'],
['Adventure', 'Family', 'Fantasy'],
```

```
['Drama', 'Romance', 'War'],
['Comedy', 'Drama'],
['Drama'],
['Mystery', 'Drama'],
['Science Fiction', 'Mystery', 'Adventure'],
['Adventure', 'Romance'],
['Drama'],
['Comedy', 'Drama'],
['Drama', 'Mystery'],
['Mystery', 'Thriller', 'War'],
['Thriller', 'Drama', 'Romance'],
['Drama', 'Mystery', 'Thriller'],
['Drama', 'Romance'],
['Crime', 'Drama', 'Mystery', 'Romance', 'Thriller'],
['Comedy', 'Drama', 'Family', 'Romance'],
['Comedy', 'Music'],
['Comedy', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Comedy', 'Music', 'Romance'],
['Comedy', 'Romance'],
['Action', 'Adventure', 'Romance'],
['Action', 'Adventure', 'Drama', 'Romance', 'Western'],
['Drama', 'Mystery'],
['Fantasy', 'Romance', 'Mystery'],
['Adventure', 'Fantasy', 'Drama', 'Mystery'],
['Comedy', 'Music', 'Romance'],
['Comedy', 'War'],
['Comedy', 'Romance'],
['Action', 'Drama', 'Western'],
['Drama'],
['Comedy', 'Crime', 'Mystery'],
['Comedy', 'Drama'],
['Adventure', 'Comedy', 'Family', 'Fantasy', 'Romance'],
['Drama', 'Family', 'Fantasy'],
['Comedy', 'Drama'],
['Comedy', 'Romance'],
['Drama', 'Romance'],
['Drama'],
['Comedy', 'Thriller', 'Mystery', 'Romance'],
['Drama'],
['Comedy', 'Fantasy', 'Romance'],
['Drama', 'Family'],
['Thriller', 'Crime', 'Drama', 'Mystery'],
['Comedy', 'Music', 'Romance'],
['Romance', 'Western'],
['Action', 'Thriller', 'Mystery'],
['War', 'Drama'],
```

```
['Action', 'Adventure', 'Western'],
['Horror'],
['Adventure', 'War', 'Romance'],
['Action', 'Adventure', 'Comedy', 'Crime', 'Drama', 'Romance'],
['Drama', 'Romance'],
['Romance', 'Drama'],
['Drama', 'Comedy', 'Romance'],
['Drama', 'Action', 'Mystery', 'Romance'],
['Comedy'],
['Drama', 'Romance', 'War'],
['Music', 'Mystery'],
['Drama'],
['Action', 'Foreign'],
['Drama'],
['Drama', 'Romance'],
['Comedy', 'Romance', 'Drama'],
['Documentary', 'Family'],
['Action', 'Adventure', 'Drama', 'Family'],
['Drama', 'Romance'],
['Drama', 'Comedy', 'Music'],
['Drama'],
['Action', 'Adventure', 'Thriller'],
['Drama', 'Thriller'],
['Drama', 'Thriller'],
['Drama', 'Romance'],
['Drama'],
['Action', 'Crime', 'Drama', 'Thriller'],
['Action', 'Drama', 'Thriller'],
['Drama', 'Action', 'Crime'],
['Comedy', 'Crime'],
['Comedy', 'Crime', 'Thriller'],
['Comedy'],
['Drama', 'Thriller'],
['Action', 'Adventure', 'Crime', 'Drama', 'Thriller'],
['Action', 'Comedy', 'Drama', 'Family'],
['Crime', 'Drama'],
['Comedy', 'Family'],
['Action', 'Adventure', 'Family', 'Western'],
['Adventure', 'Fantasy', 'Science Fiction', 'Family'],
['Comedy', 'Family', 'Fantasy'],
['Romance', 'Comedy', 'Family', 'Fantasy'],
['Adventure', 'Drama', 'Western'],
['Comedy', 'Family'],
['Adventure'],
['Adventure', 'Comedy', 'Drama', 'Family'],
['Comedy', 'Family'],
['Adventure', 'Family'],
```

```
['Drama', 'Family', 'Comedy'],
['Adventure', 'Drama', 'Science Fiction'],
['Comedy'],
['Fantasy', 'Drama', 'Comedy', 'Family'],
['Family', 'Fantasy', 'Animation', 'Romance'],
['Animation', 'Family', 'Music'],
['Animation', 'Family', 'Music'],
['Animation', 'Family'],
['Animation', 'Drama', 'Family'],
['Adventure'],
['Comedy', 'Family', 'Fantasy'],
['Animation', 'Family'],
['Fantasy', 'Animation', 'Comedy', 'Family'],
['Adventure', 'Fantasy', 'Animation', 'Comedy', 'Family', 'Music'],
['Animation', 'Adventure', 'Family', 'Fantasy'],
['Adventure', 'Animation', 'Drama', 'Family'],
['Comedy', 'Drama', 'Thriller'],
['Drama', 'Family', 'Music', 'Romance'],
['Action', 'Thriller'],
['Horror', 'Thriller', 'Science Fiction'],
['Drama', 'Romance'],
['Documentary'],
['Drama', 'Thriller'],
['Drama'],
['Comedy', 'Drama'],
['Drama', 'Romance'],
['Drama', 'Romance'],
['Comedy', 'Romance'],
['Comedy', 'Drama', 'Romance'],
['Crime', 'Action', 'Mystery', 'Thriller'],
['Adventure'],
['Documentary'],
['Comedy', 'Drama'],
['Drama'],
['Crime', 'Drama'],
['Drama'],
['Action', 'Thriller'],
['Drama', 'Romance'],
['Music', 'Comedy', 'Romance'],
['Drama', 'Romance'],
['Drama', 'Romance'],
['Comedy', 'Romance'],
['Crime', 'Drama', 'Thriller'],
['Drama'],
['Drama'],
…]
```

[9]:
```
#-----------
from collections import Counter
countlists=Counter(str(e) for e in genres)  # returns counts based on lists of␣
 ↪genres
countlists
```

[9]:
```
Counter({"['Animation', 'Comedy', 'Family']": 72,
         "['Adventure', 'Fantasy', 'Family']": 16,
         "['Romance', 'Comedy']": 213,
         "['Comedy', 'Drama', 'Romance']": 593,
         "['Comedy']": 3621,
         "['Action', 'Crime', 'Drama', 'Thriller']": 97,
         "['Comedy', 'Romance']": 930,
         "['Action', 'Adventure', 'Drama', 'Family']": 22,
         "['Action', 'Adventure', 'Thriller']": 59,
         "['Adventure', 'Action', 'Thriller']": 37,
         "['Comedy', 'Horror']": 184,
         "['Family', 'Animation', 'Adventure']": 13,
         "['History', 'Drama']": 95,
         "['Action', 'Adventure']": 93,
         "['Drama', 'Crime']": 174,
         "['Drama', 'Romance']": 1301,
         "['Crime', 'Comedy']": 53,
         "['Crime', 'Comedy', 'Adventure']": 3,
         "['Action', 'Comedy', 'Crime']": 86,
         "['Comedy', 'Thriller', 'Crime']": 5,
         "['Drama', 'Thriller']": 497,
         "['Action', 'Adventure', 'Crime', 'Thriller']": 18,
         "['Drama', 'Fantasy', 'Science Fiction', 'Thriller']": 5,
         "['Drama']": 5000,
         "['Comedy', 'Drama', 'Family']": 79,
         "['Fantasy', 'Science Fiction', 'Adventure']": 2,
         "['Science Fiction', 'Thriller', 'Mystery']": 1,
         "['Romance', 'Adventure']": 4,
         "['Fantasy', 'Drama', 'Comedy', 'Family']": 4,
         "['History', 'Drama', 'Romance']": 7,
         "['Adventure', 'History', 'Drama', 'Family']": 2,
         "['Comedy', 'Family', 'Romance']": 27,
         "['Drama', 'War']": 204,
         "['Action', 'Crime', 'Drama', 'History']": 1,
         "['Action', 'Fantasy']": 16,
         "['Fantasy', 'Drama', 'Comedy', 'Thriller']": 1,
         "['Crime', 'Mystery', 'Thriller']": 39,
         "['Adventure', 'Animation', 'Drama', 'Family']": 4,
         "['Drama', 'Crime', 'Thriller']": 39,
         "['Action', 'Thriller', 'Drama']": 10,
         "['Drama', 'Foreign']": 283,
```

```
"['Action', 'Adventure', 'Comedy', 'Family']": 15,
"['Thriller', 'Drama', 'Mystery']": 32,
"['Adventure', 'Family', 'Fantasy']": 26,
"['Music', 'Drama', 'Family']": 1,
"['Action', 'Science Fiction']": 100,
"['Drama', 'Family']": 197,
"['Horror', 'Action', 'Thriller', 'Crime']": 2,
"['Action', 'Thriller', 'Romance']": 7,
"['Drama', 'History']": 267,
"['Comedy', 'Family']": 192,
"['Horror', 'Science Fiction']": 222,
"['Documentary']": 2723,
"['Family', 'Drama']": 46,
"['Drama', 'Comedy']": 532,
"['Romance', 'Drama']": 343,
"['Action', 'Drama']": 175,
"['Adventure', 'Comedy', 'Family']": 19,
"['Crime', 'Drama', 'Thriller']": 225,
"['Drama', 'Horror', 'Thriller', 'Romance']": 3,
"['Comedy', 'Horror', 'Romance']": 3,
"['Action', 'Adventure', 'Drama', 'Thriller']": 40,
"['Action', 'Adventure', 'Drama', 'Science Fiction', 'Thriller']": 2,
"['Comedy', 'Crime', 'Drama']": 56,
"['Drama', 'Mystery', 'Science Fiction', 'Thriller']": 8,
"['Drama', 'Comedy', 'Family']": 15,
"['Action', 'Comedy', 'Music', 'Family', 'Adventure']": 1,
"['Crime']": 134,
"['Action', 'Drama', 'History', 'War']": 16,
"['Crime', 'Drama']": 430,
"['Crime', 'Action', 'Comedy', 'Thriller']": 1,
"['Documentary', 'Drama', 'Foreign']": 2,
"['Comedy', 'Drama']": 1135,
"['Action', 'Adventure', 'Comedy', 'Drama', 'Family']": 1,
"['Drama', 'Comedy', 'Romance']": 203,
"['Fantasy']": 51,
"['Drama', 'Romance', 'War']": 52,
"['Action', 'Thriller', 'Mystery', 'Romance']": 1,
"['Music', 'Documentary']": 107,
"['Action', 'Comedy', 'Crime', 'Thriller']": 17,
"['Romance', 'Drama', 'Comedy']": 45,
"['Crime', 'Comedy', 'Drama', 'Thriller']": 2,
"['Adventure']": 110,
"['Drama', 'Horror']": 101,
"['Action', 'Crime', 'Fantasy']": 2,
"['Action', 'Comedy']": 242,
"['Fantasy', 'Comedy', 'Family']": 14,
"['Mystery', 'Crime', 'Drama']": 11,
```

```
        "['Action', 'Adventure', 'Drama', 'Mystery', 'Science Fiction',
'Thriller']": 3,
        "['Thriller', 'Action', 'Crime']": 29,
        "['Drama', 'Mystery', 'Thriller']": 112,
        "['Action', 'Thriller']": 301,
        "['Action', 'Comedy', 'Crime', 'Drama']": 7,
        "['Action', 'Adventure', 'Drama', 'Romance']": 13,
        "['Action', 'Crime', 'Thriller', 'Drama']": 3,
        "['Adventure', 'Action', 'Drama', 'Science Fiction', 'Thriller']": 1,
        "['Science Fiction']": 198,
        "['Mystery', 'Horror', 'Thriller']": 19,
        "['Drama', 'Family', 'Romance']": 23,
        "['Action', 'Adventure', 'Science Fiction', 'Family', 'Fantasy']": 1,
        "['Thriller', 'Foreign']": 17,
        "['Horror', 'Thriller']": 528,
        "['Crime', 'Drama', 'Mystery', 'Thriller', 'Action']": 1,
        "['Fantasy', 'Horror', 'Thriller']": 18,
        "['Fantasy', 'Comedy', 'Drama']": 9,
        "['Drama', 'History', 'Romance']": 50,
        "['Science Fiction', 'Horror', 'Action']": 3,
        "['Crime', 'Drama', 'Science Fiction', 'Thriller']": 4,
        "['Thriller']": 465,
        "['Drama', 'Fantasy']": 70,
        "['Adventure', 'Action']": 27,
        "['Action', 'Western']": 114,
        "['Adventure', 'Action', 'Comedy', 'Crime', 'Family']": 1,
        "['Thriller', 'Drama', 'Horror', 'Mystery']": 9,
        "['Drama', 'Thriller', 'Crime', 'Mystery', 'Romance']": 2,
        "['Crime', 'Comedy', 'Fantasy']": 1,
        "['Drama', 'Thriller', 'Mystery']": 37,
        "['Crime', 'Drama', 'Mystery']": 56,
        "['Drama', 'Mystery']": 74,
        "['Thriller', 'Comedy', 'Crime', 'Romance']": 1,
        "['Comedy', 'Drama', 'History']": 15,
        "['Romance', 'Animation', 'Family', 'Comedy', 'Adventure']": 1,
        "['Drama', 'Family', 'Fantasy']": 15,
        "['Fantasy', 'Animation', 'Science Fiction', 'Family']": 7,
        "['Drama', 'Action', 'Thriller', 'Crime']": 64,
        "['Drama', 'Music', 'Romance']": 86,
        "['Comedy', 'Family', 'Drama']": 9,
        "['Horror', 'Romance']": 6,
        "['Comedy', 'Crime']": 144,
        "['Action', 'Drama', 'Mystery', 'Thriller']": 14,
        "['Adventure', 'Fantasy', 'Comedy', 'Science Fiction', 'Family']": 2,
        "['Adventure', 'Action', 'Science Fiction']": 21,
        "['Adventure', 'Drama', 'Romance', 'War']": 3,
        "['Drama', 'Horror', 'Science Fiction', 'Romance']": 1,
```

```
"['Action', 'Comedy', 'Family']": 11,
"['Fantasy', 'Drama', 'Family']": 8,
"['Action', 'Thriller', 'Crime', 'Drama']": 12,
"['Horror']": 974,
"['Science Fiction', 'Action', 'Thriller']": 10,
"['Crime', 'Thriller', 'Drama']": 17,
"['Drama', 'Romance', 'Thriller']": 43,
"['Action', 'Drama', 'Science Fiction', 'Thriller']": 5,
"['Thriller', 'Crime', 'Drama']": 58,
"['Thriller', 'Crime']": 49,
"['Drama', 'Mystery', 'Romance']": 18,
"['Drama', 'Music', 'Mystery']": 2,
"['Comedy', 'Drama', 'Mystery']": 17,
"['Comedy', 'Family', 'TV Movie']": 1,
"['Horror', 'Drama']": 36,
"['Animation']": 242,
"['Action', 'Adventure', 'Science Fiction']": 55,
"['Crime', 'Thriller']": 98,
"['Action', 'Comedy', 'Fantasy', 'Science Fiction']": 3,
"['Horror', 'Comedy', 'Thriller']": 8,
"['Science Fiction', 'Action', 'Adventure', 'Thriller']": 10,
"['Crime', 'Horror', 'Thriller']": 24,
"['Thriller', 'Horror', 'Science Fiction']": 15,
"['Action', 'Crime', 'Science Fiction', 'Thriller']": 3,
"['Comedy', 'Mystery']": 25,
"['Drama', 'Music']": 143,
"['Thriller', 'Drama', 'Romance']": 20,
"['Action', 'Comedy', 'Thriller']": 36,
"['Action', 'Drama', 'Thriller']": 130,
"['Drama', 'Thriller', 'Crime', 'Mystery']": 22,
"['Comedy', 'Drama', 'Family', 'Romance']": 19,
"['Fantasy', 'Action', 'Thriller']": 4,
"['Romance', 'Drama', 'Crime']": 14,
"['Action', 'Comedy', 'Romance']": 28,
"['Adventure', 'Drama', 'Family']": 24,
"['Documentary', 'Foreign']": 121,
"['Family', 'Animation', 'Drama']": 1,
"['Horror', 'Thriller', 'Mystery', 'Fantasy']": 2,
"['Romance', 'Comedy', 'Crime', 'Fantasy']": 1,
"['Action', 'Adventure', 'Comedy', 'Drama', 'Western']": 3,
"['Action', 'Adventure', 'Crime']": 14,
"['Thriller', 'Science Fiction', 'Action', 'Crime']": 1,
"['Drama', 'Action', 'Western']": 9,
"['Action', 'Crime', 'Romance', 'Thriller']": 1,
"['Action', 'Crime']": 74,
"['Crime', 'Drama', 'Romance']": 45,
"['Animation', 'Science Fiction']": 31,
```

```
"['Action', 'Adventure', 'Crime', 'Drama', 'Thriller']": 15,
"['Romance', 'Adventure', 'Crime', 'Drama']": 1,
"['Thriller', 'Drama']": 180,
"['Action', 'Fantasy', 'Science Fiction']": 11,
"['Horror', 'Mystery', 'Thriller']": 130,
"['Comedy', 'Family', 'Fantasy']": 25,
"['Comedy', 'Crime', 'Music']": 1,
"['Drama', 'Comedy', 'Fantasy']": 3,
"['Mystery', 'Crime', 'Drama', 'Thriller']": 5,
"['Drama', 'Action', 'Thriller']": 53,
"['Horror', 'Science Fiction', 'Thriller']": 78,
"['Action', 'Comedy', 'Drama', 'Western']": 2,
"['Comedy', 'Thriller', 'Mystery']": 8,
"['Comedy', 'Science Fiction', 'Family']": 2,
"['Drama', 'Mystery', 'Romance', 'Thriller']": 14,
"['Western', 'Action', 'Comedy']": 2,
"['Crime', 'Action', 'Science Fiction']": 2,
"['Western', 'Drama', 'Comedy', 'Romance']": 1,
"['Comedy', 'Thriller', 'Mystery', 'Romance', 'Foreign']": 1,
"['Family', 'Adventure']": 20,
"['Adventure', 'Action', 'Thriller', 'Crime', 'Mystery']": 2,
"['History', 'Action', 'Drama', 'Western']": 1,
"['Family', 'Comedy', 'Crime']": 5,
"['History', 'War', 'Action', 'Drama']": 1,
"['Action', 'Comedy', 'War']": 4,
"['Action', 'Drama', 'Thriller', 'Crime', 'Mystery']": 2,
"['Action', 'Thriller', 'Crime']": 51,
"['Adventure', 'Science Fiction']": 18,
"['Adventure', 'Fantasy', 'Action', 'Comedy', 'Family']": 3,
"['Action', 'Comedy', 'Western']": 21,
"['Drama', 'Action', 'Crime']": 51,
"['Thriller', 'Comedy', 'Drama', 'Romance']": 2,
"['Comedy', 'Drama', 'Family', 'Fantasy', 'Science Fiction']": 2,
"['Romance', 'Drama', 'Fantasy']": 5,
"['Comedy', 'Crime', 'Drama', 'Music', 'Mystery', 'Romance']": 1,
"['Comedy', 'War']": 21,
"['Action', 'Adventure', 'Crime', 'Science Fiction', 'Thriller']": 3,
"['Drama', 'History', 'War']": 77,
"['Comedy', 'Crime', 'Horror', 'Thriller']": 3,
"['Adventure', 'Fantasy', 'Action', 'Thriller', 'Science Fiction']": 2,
"['Mystery', 'Drama', 'Comedy']": 2,
"['Science Fiction', 'Drama', 'Thriller']": 6,
"['Comedy', 'Crime', 'Romance']": 15,
"['Crime', 'Action', 'Mystery', 'Thriller']": 2,
"['Adventure', 'Comedy', 'Family', 'Fantasy']": 6,
"['Action', 'Adventure', 'Thriller', 'Crime']": 2,
"['Fantasy', 'Animation', 'Family']": 12,
```

```
          "['Action', 'Adventure', 'Comedy']": 61,
          "['Action', 'Adventure', 'Drama', 'History', 'Western']": 2,
          "['Action', 'Thriller', 'Crime', 'Romance']": 2,
          "['Documentary', 'History']": 70,
          "['Fantasy', 'Science Fiction', 'Family', 'Animation']": 1,
          "['Drama', 'Horror', 'Thriller']": 104,
          "['TV Movie', 'Drama', 'Romance', 'Comedy']": 1,
          "['Comedy', 'Foreign']": 76,
          "['Romance', 'Comedy', 'Family']": 7,
          "['Drama', 'Family', 'Adventure']": 4,
          "['Comedy', 'Drama', 'History', 'Mystery']": 1,
          "['History', 'Documentary']": 32,
          "['Drama', 'Comedy', 'Romance', 'Foreign']": 49,
          "['Fantasy', 'Drama', 'Thriller', 'Mystery', 'Romance']": 2,
          "['Animation', 'Family', 'Comedy', 'Adventure', 'Fantasy', 'Romance']":
1,
          "['Action', 'Thriller', 'Science Fiction']": 34,
          "['Adventure', 'Drama', 'Western']": 9,
          "['Thriller', 'Action', 'Drama']": 20,
          "['Fantasy', 'Action']": 9,
          "['Romance', 'Family', 'Animation', 'Fantasy', 'Music']": 1,
          "['Animation', 'Family']": 173,
          "['Comedy', 'Drama', 'Fantasy']": 26,
          "['Adventure', 'Western']": 16,
          "['Drama', 'Foreign', 'Romance']": 77,
          "['Adventure', 'Comedy', 'Drama', 'Family']": 4,
          "['Horror', 'Thriller', 'Science Fiction']": 14,
          "['Animation', 'Comedy', 'Family', 'Adventure']": 8,
          "['Adventure', 'Drama', 'Documentary']": 2,
          "['Comedy', 'Thriller', 'Crime', 'Romance']": 1,
          "['Comedy', 'Thriller', 'Crime', 'Drama']": 1,
          "['Crime', 'Drama', 'Mystery', 'Thriller']": 82,
          "['Romance', 'Animation', 'Family', 'Adventure', 'Fantasy']": 1,
          "['Drama', 'History', 'War', 'Romance']": 1,
          "['Fantasy', 'Science Fiction', 'Family']": 6,
          "['Drama', 'Horror', 'Mystery', 'Thriller']": 29,
          "['Adventure', 'Drama', 'Comedy', 'Family']": 3,
          "['Drama', 'Thriller', 'Mystery', 'War']": 1,
          "['Romance', 'Comedy', 'Drama']": 101,
          "['Adventure', 'Animation', 'Family']": 39,
          "['Drama', 'Thriller', 'Romance']": 34,
          "['War', 'Drama', 'Comedy']": 3,
          "['War', 'Drama', 'Romance']": 24,
          "['Comedy', 'Science Fiction']": 60,
          "['Animation', 'Comedy', 'Drama', 'Family', 'Fantasy']": 1,
          "['Comedy', 'Horror', 'Science Fiction']": 30,
          "['Drama', 'Science Fiction', 'Mystery']": 6,
```

```
"['Horror', 'Drama', 'Thriller']": 18,
"['Comedy', 'Romance', 'Drama']": 49,
"['Action', 'Adventure', 'Science Fiction', 'Thriller']": 17,
"['Crime', 'Drama', 'Romance', 'Thriller']": 19,
"['Action', 'Comedy', 'Drama', 'Thriller']": 10,
"['Romance', 'Drama', 'Mystery']": 9,
"['Adventure', 'Family']": 41,
"['Drama', 'Thriller', 'Mystery', 'Romance']": 6,
"['Drama', 'Fantasy', 'Western']": 1,
"['War', 'Adventure', 'Drama', 'Romance']": 4,
"['Action', 'Crime', 'Drama']": 88,
"['Fantasy', 'Comedy', 'Science Fiction']": 7,
"['Comedy', 'Fantasy', 'Science Fiction']": 15,
"['Drama', 'Fantasy', 'Horror', 'Thriller']": 6,
"['Animation', 'Drama']": 32,
"['Action', 'Drama', 'Crime', 'Thriller']": 8,
"['Fantasy', 'Drama', 'Horror', 'Comedy']": 1,
"['Action', 'Adventure', 'Drama']": 74,
"['Action', 'Animation', 'Science Fiction']": 21,
"['Family', 'Animation', 'Comedy']": 5,
"['Action', 'Mystery', 'Science Fiction', 'Thriller']": 2,
"['Drama', 'Comedy', 'War']": 6,
"['Horror', 'Comedy', 'Drama', 'Fantasy']": 2,
"['Drama', 'Action']": 88,
"['Drama', 'Thriller', 'Crime']": 65,
"['Comedy', 'Drama', 'Science Fiction']": 15,
"['Music', 'Romance']": 26,
"['Drama', 'Mystery', 'Horror']": 8,
"['Western', 'Drama', 'War']": 1,
"['Drama', 'Animation', 'Family']": 1,
"['Comedy', 'Drama', 'Thriller']": 24,
"['Fantasy', 'Comedy', 'Romance', 'Science Fiction']": 2,
"['Horror', 'Comedy']": 80,
"['Drama', 'Fantasy', 'Romance', 'Science Fiction']": 2,
"['Fantasy', 'Comedy', 'Science Fiction', 'Family']": 5,
"['Family', 'Comedy']": 40,
"['Action', 'Mystery', 'Thriller']": 11,
"['War', 'Drama']": 104,
"['Adventure', 'Drama']": 88,
"['Animation', 'Family', 'Fantasy']": 18,
"['Fantasy', 'Comedy', 'Music']": 2,
"['Action', 'Fantasy', 'Thriller']": 4,
"['Action', 'Crime', 'Thriller']": 139,
"['Thriller', 'Drama', 'Music', 'Crime']": 1,
"['Fantasy', 'Horror', 'Science Fiction']": 7,
"['Action', 'Adventure', 'Comedy', 'Thriller']": 11,
"['Animation', 'Adventure', 'Family']": 18,
```

```
"['Drama', 'History', 'Romance', 'Thriller']": 3,
"['Comedy', 'Music', 'Romance']": 101,
"['Drama', 'Comedy', 'Music', 'Romance']": 8,
"['Music', 'Comedy', 'Romance']": 10,
"['Mystery', 'Romance', 'Thriller']": 3,
"['Drama', 'Mystery', 'Thriller', 'Crime']": 7,
"['Mystery', 'Thriller']": 109,
"['Comedy', 'Mystery', 'Romance', 'Thriller']": 5,
"['Mystery', 'Crime', 'Thriller']": 5,
"['Drama', 'Family', 'Music', 'Romance']": 6,
"['Mystery', 'Drama']": 41,
"['Science Fiction', 'Mystery', 'Adventure']": 2,
"['Adventure', 'Romance']": 11,
"['Mystery', 'Thriller', 'War']": 1,
"['Crime', 'Drama', 'Mystery', 'Romance', 'Thriller']": 5,
"['Comedy', 'Music']": 107,
"['Action', 'Adventure', 'Romance']": 8,
"['Action', 'Adventure', 'Drama', 'Romance', 'Western']": 3,
"['Fantasy', 'Romance', 'Mystery']": 1,
"['Adventure', 'Fantasy', 'Drama', 'Mystery']": 1,
"['Action', 'Drama', 'Western']": 23,
"['Comedy', 'Crime', 'Mystery']": 17,
"['Adventure', 'Comedy', 'Family', 'Fantasy', 'Romance']": 1,
"['Comedy', 'Thriller', 'Mystery', 'Romance']": 3,
"['Comedy', 'Fantasy', 'Romance']": 23,
"['Thriller', 'Crime', 'Drama', 'Mystery']": 10,
"['Romance', 'Western']": 14,
"['Action', 'Thriller', 'Mystery']": 7,
"['Action', 'Adventure', 'Western']": 14,
"['Adventure', 'War', 'Romance']": 1,
"['Action', 'Adventure', 'Comedy', 'Crime', 'Drama', 'Romance']": 1,
"['Drama', 'Action', 'Mystery', 'Romance']": 1,
"['Music', 'Mystery']": 1,
"['Action', 'Foreign']": 42,
"['Documentary', 'Family']": 20,
"['Drama', 'Comedy', 'Music']": 17,
"['Comedy', 'Crime', 'Thriller']": 19,
"['Action', 'Comedy', 'Drama', 'Family']": 4,
"['Action', 'Adventure', 'Family', 'Western']": 1,
"['Adventure', 'Fantasy', 'Science Fiction', 'Family']": 3,
"['Romance', 'Comedy', 'Family', 'Fantasy']": 1,
"['Drama', 'Family', 'Comedy']": 5,
"['Adventure', 'Drama', 'Science Fiction']": 6,
"['Family', 'Fantasy', 'Animation', 'Romance']": 1,
"['Animation', 'Family', 'Music']": 7,
"['Animation', 'Drama', 'Family']": 11,
"['Fantasy', 'Animation', 'Comedy', 'Family']": 3,
```

```
        "['Adventure', 'Fantasy', 'Animation', 'Comedy', 'Family', 'Music']":
1,
        "['Animation', 'Adventure', 'Family', 'Fantasy']": 6,
        "['Mystery']": 70,
        "['Adventure', 'Drama', 'Romance']": 21,
        "['War', 'Drama', 'Romance', 'Action']": 1,
        "['Family', 'Fantasy']": 30,
        "['Comedy', 'Romance', 'Science Fiction']": 6,
        "['Adventure', 'Comedy']": 55,
        "['Drama', 'War', 'Action']": 4,
        "['Thriller', 'Mystery']": 55,
        "['Romance', 'Crime', 'Drama', 'Thriller']": 1,
        "['Science Fiction', 'Adventure', 'Family', 'Fantasy']": 1,
        "['Action', 'Romance', 'War']": 1,
        "['Action', 'Adventure', 'Drama', 'Western']": 19,
        "['Mystery', 'Comedy', 'Crime']": 8,
        "['Comedy', 'Fantasy', 'Family']": 7,
        "['Adventure', 'Action', 'Thriller', 'Science Fiction']": 12,
        "['Science Fiction', 'Action']": 25,
        "['Adventure', 'Comedy', 'Fantasy']": 10,
        "['Mystery', 'Crime', 'Drama', 'History', 'Romance']": 1,
        "['Animation', 'Drama', 'Fantasy']": 9,
        "['Action']": 278,
        "['Comedy', 'Science Fiction', 'Fantasy']": 1,
        "['Western']": 318,
        "['Comedy', 'Romance', 'War']": 5,
        "['Mystery', 'Documentary', 'Crime']": 1,
        "['Comedy', 'Crime', 'Drama', 'Romance']": 7,
        "['Documentary', 'Music']": 185,
        "['Adventure', 'Family', 'Fantasy', 'Comedy', 'Romance']": 1,
        "['Horror', 'Action', 'Thriller', 'Science Fiction']": 13,
        "['Adventure', 'Drama', 'History', 'War']": 6,
        "['Science Fiction', 'Drama']": 28,
        "['Drama', 'Fantasy', 'Romance']": 24,
        "['Fantasy', 'Horror', 'Comedy']": 7,
        "['Action', 'Drama', 'History']": 24,
        "['Music', 'Comedy', 'Action', 'Crime']": 2,
        "['Adventure', 'Animation', 'Comedy', 'Science Fiction', 'Family']": 3,
        "['War', 'Drama', 'History', 'Action', 'Romance']": 1,
        "['Drama', 'History', 'Music']": 4,
        "['Drama', 'Science Fiction', 'Thriller']": 27,
        "['Action', 'Drama', 'History', 'War', 'Adventure']": 1,
        "['Fantasy', 'Drama']": 36,
        "['War']": 37,
        "['Action', 'Comedy', 'Horror', 'Science Fiction']": 6,
        "['Horror', 'Comedy', 'Fantasy']": 5,
        "['Adventure', 'Drama', 'History', 'Thriller', 'War']": 1,
```

```
        "['Action', 'Drama', 'Music', 'Thriller']": 1,
        "['Romance', 'Fantasy', 'Drama', 'Comedy']": 1,
        "['Music', 'Drama']": 57,
        "['Adventure', 'Comedy', 'Science Fiction', 'Family']": 1,
        "['Crime', 'Drama', 'Comedy']": 8,
        "['Science Fiction', 'Animation']": 13,
        "['Adventure', 'Action', 'Fantasy']": 9,
        "['Drama', 'Comedy', 'History', 'Romance']": 2,
        "['Thriller', 'Comedy', 'Drama']": 3,
        "['Science Fiction', 'Drama', 'Fantasy', 'Romance']": 2,
        "['Action', 'Adventure', 'Drama', 'History']": 11,
        "['Music']": 109,
        "['War', 'Drama', 'History']": 38,
        "['Adventure', 'Science Fiction', 'Action']": 13,
        "['History', 'Drama', 'Western', 'Crime']": 1,
        "['Thriller', 'Drama', 'Science Fiction', 'Action']": 1,
        "['Drama', 'Action', 'Comedy', 'Crime']": 9,
        "['Science Fiction', 'Action', 'Horror']": 7,
        "['Horror', 'Mystery']": 53,
        "['Fantasy', 'Drama', 'Horror', 'Thriller']": 2,
        "['Crime', 'Drama', 'Horror', 'Mystery', 'Thriller']": 6,
        "['Science Fiction', 'Horror', 'Comedy']": 4,
        "['Thriller', 'Horror']": 152,
        "['Romance', 'Horror']": 1,
        "['Fantasy', 'Horror']": 40,
        "['Action', 'Drama', 'Horror', 'Mystery', 'Thriller']": 1,
        "['Documentary', 'Drama']": 85,
        "['Fantasy', 'Drama', 'Comedy', 'Science Fiction', 'Romance',
'Family']": 2,
        "['Science Fiction', 'Adventure', 'Mystery']": 4,
        "['Science Fiction', 'Adventure']": 8,
        "['Crime', 'Action', 'Adventure', 'Drama', 'Western']": 1,
        "['Western', 'Adventure']": 5,
        "['Romance']": 97,
        "['Comedy', 'Drama', 'Family', 'Music', 'Romance']": 4,
        "['Action', 'Science Fiction', 'Thriller']": 49,
        "['Horror', 'Thriller', 'Adventure']": 1,
        "['Animation', 'Comedy']": 66,
        "['Crime', 'Horror', 'Mystery']": 10,
        "['Action', 'Adventure', 'Drama', 'History', 'Romance', 'War']": 3,
        "['Drama', 'Romance', 'Family', 'Comedy']": 1,
        "['Romance', 'Crime', 'Drama']": 19,
        "['History', 'Drama', 'Music']": 3,
        "['Action', 'Drama', 'Mystery', 'Thriller', 'Crime']": 3,
        "['Drama', 'Thriller', 'Documentary', 'Romance', 'Foreign']": 1,
        "['Crime', 'Drama', 'TV Movie']": 7,
        "['Action', 'Adventure', 'Comedy', 'Family', 'Romance']": 1,
```

```
        "['Action', 'Adventure', 'Comedy', 'Crime', 'Drama', 'Thriller']": 2,
        "['Drama', 'Action', 'Comedy', 'Thriller']": 6,
        "['Action', 'Adventure', 'Comedy', 'Crime', 'Thriller']": 7,
        "['Music', 'Drama', 'Comedy']": 10,
        "['Music', 'Comedy']": 34,
        "['Comedy', 'Foreign', 'Romance']": 8,
        "['Action', 'Comedy', 'Drama', 'Foreign', 'Romance']": 2,
        "['Thriller', 'Action', 'Romance', 'Science Fiction', 'Adventure']": 1,
        "['Action', 'Crime', 'Drama', 'Mystery', 'Thriller']": 18,
        "['Animation', 'Comedy', 'Family', 'Music']": 1,
        "['Action', 'Adventure', 'Family', 'Fantasy', 'Science Fiction',
'Thriller']": 1,
        "['Adventure', 'Horror', 'Thriller']": 8,
        "['Action', 'Comedy', 'Thriller', 'Romance']": 6,
        "['Science Fiction', 'Action', 'Drama', 'Thriller']": 3,
        "['Science Fiction', 'Comedy', 'Crime']": 2,
        "['Action', 'Crime', 'Drama', 'Romance', 'Thriller']": 2,
        "['Fantasy', 'Family', 'Action']": 1,
        "['Science Fiction', 'Drama', 'Comedy']": 3,
        "['Music', 'Comedy', 'Drama', 'Romance']": 5,
        "['Comedy', 'Drama', 'Romance', 'Fantasy']": 2,
        "['Comedy', 'Thriller']": 35,
        "['Animation', 'Family', 'Fantasy', 'Adventure', 'Comedy', 'Romance']":
1,
        "['Science Fiction', 'Horror', 'Drama', 'Thriller']": 1,
        "['Action', 'Adventure', 'Comedy', 'Science Fiction']": 8,
        "['Family']": 91,
        "['Adventure', 'Fantasy', 'Action']": 18,
        "['Adventure', 'Comedy', 'Family', 'Romance']": 2,
        "['Horror', 'Science Fiction', 'Mystery']": 2,
        "['Action', 'Adventure', 'Fantasy', 'Horror', 'Science Fiction',
'Thriller']": 3,
        "['Science Fiction', 'Action', 'Adventure']": 12,
        "['Adventure', 'Fantasy', 'Action', 'Science Fiction']": 6,
        "['Adventure', 'Drama', 'History']": 20,
        "['Horror', 'Thriller', 'Mystery']": 38,
        "['Comedy', 'Family', 'Science Fiction']": 16,
        "['Thriller', 'Science Fiction', 'Mystery', 'Romance']": 1,
        "['Fantasy', 'Drama', 'Mystery', 'Family']": 1,
        "['Fantasy', 'Drama', 'Comedy', 'Thriller', 'Science Fiction',
'Romance']": 1,
        "['Action', 'Adventure', 'Mystery', 'Thriller']": 2,
        "['Drama', 'Music', 'Romance', 'Foreign']": 3,
        "['Comedy', 'Drama', 'Fantasy', 'Romance']": 12,
        "['Action', 'Thriller', 'Adventure', 'Crime']": 2,
        "['Comedy', 'Thriller', 'Crime', 'Action']": 2,
        "['Drama', 'History', 'Mystery']": 3,
```

```
"['Comedy', 'Horror', 'Thriller']": 17,
"['Drama', 'Adventure']": 29,
"['Drama', 'Romance', 'Comedy']": 40,
"['Mystery', 'Science Fiction']": 6,
"['Action', 'Adventure', 'Family', 'Science Fiction']": 1,
"['Thriller', 'Fantasy', 'Crime', 'Drama']": 1,
"['Crime', 'Comedy', 'Romance']": 4,
"['Drama', 'Action', 'Romance']": 16,
"['Action', 'Drama', 'Foreign', 'History', 'Thriller']": 1,
"['Adventure', 'Fantasy', 'Drama', 'Comedy', 'Music']": 1,
"['Adventure', 'Action', 'Horror', 'Science Fiction']": 1,
"['Thriller', 'Comedy']": 11,
"['Drama', 'Crime', 'Mystery']": 11,
"['Mystery', 'Horror', 'Fantasy']": 2,
"['Music', 'Action', 'Comedy', 'Crime']": 1,
"['Romance', 'Action', 'Drama', 'Thriller']": 1,
"['Mystery', 'Thriller', 'Drama', 'Romance', 'Crime']": 1,
"['Drama', 'Crime', 'Mystery', 'Thriller']": 5,
"['Thriller', 'Crime', 'Action']": 4,
"['Music', 'Comedy', 'Fantasy', 'Adventure']": 1,
"['Drama', 'Romance', 'Foreign']": 65,
"['Adventure', 'Documentary']": 18,
"['Action', 'Comedy', 'Crime', 'Drama', 'Thriller']": 5,
"['Crime', 'Action', 'Drama']": 10,
"['Drama', 'Action', 'Comedy', 'Thriller', 'Crime']": 4,
"['Adventure', 'Family', 'Science Fiction']": 2,
"['Comedy', 'Crime', 'Mystery', 'Thriller']": 6,
"['Adventure', 'Fantasy']": 24,
"['Crime', 'Action', 'Comedy']": 10,
"['Drama', 'Foreign', 'History']": 8,
"['Romance', 'Drama', 'TV Movie']": 1,
"['Crime', 'Drama', 'History', 'Romance']": 1,
"['Action', 'Drama', 'Romance']": 31,
"['Fantasy', 'Animation', 'Drama', 'Romance', 'Family']": 1,
"['Adventure', 'Drama', 'Comedy']": 4,
"['Action', 'Adventure', 'Comedy', 'Western']": 4,
"['Action', 'Adventure', 'Comedy', 'Romance']": 9,
"['Action', 'Comedy', 'Romance', 'Science Fiction', 'Thriller']": 3,
"['Animation', 'Family', 'Adventure']": 12,
"['Mystery', 'Science Fiction', 'Thriller']": 7,
"['Romance', 'Comedy', 'Crime']": 7,
"['Romance', 'Comedy', 'Crime', 'Drama']": 5,
"['Action', 'Thriller', 'Science Fiction', 'Adventure']": 3,
"['Comedy', 'Adventure', 'Fantasy', 'Science Fiction', 'Action']": 1,
"['Mystery', 'Drama', 'Thriller']": 18,
"['Science Fiction', 'Horror']": 45,
"['Action', 'Drama', 'Romance', 'War']": 2,
```

```
"['Drama', 'Western']": 41,
"['Adventure', 'Drama', 'History', 'Romance']": 4,
"['Crime', 'Drama', 'Music']": 4,
"['Adventure', 'Comedy', 'History', 'Romance']": 2,
"['Drama', 'Family', 'Music']": 5,
"['Comedy', 'Crime', 'Science Fiction', 'Thriller']": 1,
"['Horror', 'Thriller', 'Comedy']": 7,
"['Horror', 'Mystery', 'Science Fiction']": 13,
"['Thriller', 'Comedy', 'Horror']": 7,
"['Thriller', 'Horror', 'Action']": 6,
"['Horror', 'Thriller', 'Fantasy']": 3,
"['Thriller', 'Horror', 'Mystery']": 27,
"['Adventure', 'Action', 'Comedy', 'Thriller', 'Crime']": 4,
"['Comedy', 'Horror', 'Fantasy']": 3,
"['Drama', 'Science Fiction']": 71,
"['Adventure', 'Comedy', 'Family', 'Science Fiction']": 6,
"['Comedy', 'Family', 'Western']": 3,
"['Action', 'Science Fiction', 'Adventure']": 11,
"['Mystery', 'Horror', 'Science Fiction']": 2,
"['Adventure', 'Family', 'Science Fiction', 'Action']": 1,
"['Comedy', 'Action', 'Adventure']": 12,
"['Family', 'Science Fiction', 'Adventure']": 4,
"['Comedy', 'Animation', 'Family']": 8,
"['Comedy', 'Drama', 'Family', 'Music']": 5,
"['Fantasy', 'Comedy', 'Family', 'Adventure']": 2,
"['Romance', 'Comedy', 'Family', 'Action', 'Adventure']": 1,
"['Action', 'Comedy', 'Family', 'Western']": 1,
"['Action', 'Adventure', 'Crime', 'Drama', 'Mystery', 'Thriller']": 6,
"['Fantasy', 'Comedy', 'Romance']": 14,
"['Fantasy', 'Drama', 'Mystery']": 5,
"['Family', 'Animation']": 53,
"['Comedy', 'Family', 'Fantasy', 'Drama']": 1,
"['Drama', 'Music', 'Family']": 1,
"['Adventure', 'Animation', 'Comedy', 'Family']": 24,
"['Animation', 'Music', 'Family', 'Adventure', 'Fantasy']": 1,
"['Action', 'Adventure', 'Comedy', 'Family', 'Music']": 1,
"['Adventure', 'Animation', 'Crime', 'Family', 'Fantasy']": 1,
"['Fantasy', 'Family', 'Animation', 'Adventure']": 2,
"['Family', 'Adventure', 'Animation', 'Romance']": 1,
"['Action', 'Adventure', 'Science Fiction', 'Family']": 1,
"['Fantasy', 'Animation', 'Romance', 'Family']": 2,
"['Fantasy', 'Mystery', 'Thriller']": 4,
"['Family', 'Music']": 6,
"['Fantasy', 'Drama', 'Animation', 'Adventure']": 1,
"['Drama', 'Romance', 'Science Fiction']": 12,
"['Action', 'Drama', 'Horror', 'Science Fiction']": 3,
"['Drama', 'Animation', 'Family', 'Comedy', 'Fantasy']": 1,
```

```
"['Crime', 'Mystery']": 24,
"['Thriller', 'Comedy', 'Crime', 'Adventure']": 1,
"['Comedy', 'Romance', 'Fantasy']": 5,
"['Adventure', 'Fantasy', 'Comedy', 'Family', 'Music', 'Romance']": 1,
"['Comedy', 'Family', 'Romance', 'Science Fiction']": 1,
"['Adventure', 'Animation', 'Drama']": 2,
"['Fantasy', 'Animation', 'Drama', 'Family']": 1,
"['Drama', 'Family', 'Animation', 'Adventure']": 1,
"['Crime', 'Drama', 'Horror', 'Thriller']": 21,
"['Horror', 'Drama', 'Mystery']": 5,
"['Drama', 'Family', 'Fantasy', 'Adventure']": 2,
"['Horror', 'Action']": 13,
"['Adventure', 'Action', 'Drama', 'Fantasy', 'Mystery']": 1,
"['Fantasy', 'Comedy']": 26,
"['Adventure', 'Crime', 'Drama', 'Mystery', 'Thriller']": 2,
"['Comedy', 'Mystery', 'Thriller']": 10,
"['Drama', 'Comedy', 'Animation']": 2,
"['Adventure', 'Drama', 'Action']": 13,
"['Crime', 'Drama', 'History', 'Thriller']": 2,
"['Drama', 'Thriller', 'Crime', 'Romance']": 11,
"['Thriller', 'War']": 7,
"['Adventure', 'Crime']": 3,
"['Thriller', 'Science Fiction', 'Mystery']": 8,
"['Western', 'Action', 'Drama', 'History']": 1,
"['Comedy', 'Drama', 'Foreign']": 27,
"['Drama', 'Comedy', 'Thriller', 'Mystery']": 2,
"['Comedy', 'Drama', 'Music', 'Romance']": 20,
"['Drama', 'Comedy', 'Romance', 'Family']": 7,
"['Comedy', 'Adventure']": 27,
"['Fantasy', 'Comedy', 'Science Fiction', 'Romance']": 12,
"['Action', 'Thriller', 'Crime', 'Adventure']": 1,
"['Comedy', 'Crime', 'Drama', 'Thriller']": 11,
"['Science Fiction', 'Mystery', 'Horror']": 1,
"['Fantasy', 'Drama', 'Romance']": 9,
"['TV Movie', 'Drama']": 65,
"['Thriller', 'Science Fiction']": 29,
"['Drama', 'Fantasy', 'Comedy']": 5,
"['Fantasy', 'Drama', 'Comedy']": 6,
"['Action', 'War', 'Science Fiction']": 1,
"['Thriller', 'Comedy', 'Fantasy', 'Horror']": 1,
"['Action', 'Drama', 'Fantasy', 'Horror', 'Thriller']": 1,
"['Drama', 'Crime', 'Romance']": 4,
"['History', 'Drama', 'War']": 8,
"['Horror', 'Comedy', 'Crime']": 2,
"['Action', 'Drama', 'Horror', 'Science Fiction', 'Thriller']": 4,
"['Action', 'Adventure', 'Drama', 'Horror']": 1,
"['Action', 'Horror']": 25,
```

```
        "['Fantasy', 'Horror', 'Science Fiction', 'Thriller']": 3,
        "['Adventure', 'Comedy', 'Drama', 'Family', 'Fantasy']": 3,
        "['Comedy', 'Family', 'Drama', 'Fantasy']": 1,
        "['Adventure', 'Animation', 'Drama', 'Family', 'Music']": 1,
        "['Romance', 'History']": 2,
        "['Family', 'Fantasy', 'Adventure', 'Science Fiction', 'Comedy']": 1,
        "['Action', 'Adventure', 'Thriller', 'War']": 4,
        "['Comedy', 'Thriller', 'Crime', 'Mystery']": 5,
        "['Action', 'Adventure', 'Crime', 'Drama', 'Family', 'Mystery',
'Thriller']": 1,
        "['Adventure', 'Drama', 'Action', 'Romance', 'Family']": 1,
        "['Action', 'Adventure', 'Family', 'Fantasy']": 8,
        "['Adventure', 'Fantasy', 'Drama', 'Action', 'Science Fiction']": 3,
        "['Comedy', 'Drama', 'War']": 19,
        "['Drama', 'Action', 'Western', 'Romance']": 2,
        "['Horror', 'Action', 'Science Fiction']": 7,
        "['Adventure', 'Science Fiction', 'Comedy']": 1,
        "['Horror', 'Fantasy']": 15,
        "['Drama', 'Family', 'Fantasy', 'Science Fiction']": 1,
        "['Drama', 'Horror', 'Science Fiction']": 22,
        "['Thriller', 'Action', 'Comedy', 'Crime']": 4,
        "['Horror', 'Music']": 8,
        "['Romance', 'Drama', 'History']": 15,
        "['Romance', 'Comedy', 'Thriller']": 1,
        "['Action', 'Comedy', 'Drama', 'War']": 1,
        "['Comedy', 'Western']": 45,
        "['Action', 'Comedy', 'Horror', 'Crime']": 1,
        "['Documentary', 'Drama', 'History', 'War']": 1,
        "['Animation', 'Fantasy', 'Science Fiction']": 5,
        "['Comedy', 'Drama', 'Family', 'Science Fiction']": 2,
        "['Thriller', 'Crime', 'Mystery']": 19,
        "['Action', 'Adventure', 'Drama', 'Horror', 'Science Fiction',
'Thriller', 'Western']": 1,
        "['Adventure', 'Action', 'Romance', 'Science Fiction']": 1,
        "['Science Fiction', 'Adventure', 'Drama', 'Action']": 1,
        "['Adventure', 'Science Fiction', 'Mystery']": 1,
        "['Action', 'Crime', 'Mystery', 'Thriller']": 8,
        "['Science Fiction', 'Thriller']": 56,
        "['Science Fiction', 'Comedy', 'Family']": 2,
        "['Comedy', 'Horror', 'Thriller', 'Western']": 1,
        "['Action', 'Drama', 'Foreign']": 24,
        "['Animation', 'Family', 'Comedy']": 23,
        "['Crime', 'Comedy', 'Thriller']": 2,
        "['Comedy', 'Action']": 39,
        "['Thriller', 'Comedy', 'Crime', 'Mystery']": 2,
        "['TV Movie', 'History', 'Documentary', 'War']": 1,
        "['Crime', 'Romance', 'Comedy', 'Drama', 'Music']": 1,
```

```
          "['Action', 'Thriller', 'Science Fiction', 'Horror']": 2,
          "['Science Fiction', 'Action', 'Fantasy']": 4,
          "['Comedy', 'Family', 'Fantasy', 'Science Fiction']": 4,
          "['Fantasy', 'Drama', 'Comedy', 'Romance']": 4,
          "['Thriller', 'Comedy', 'Crime']": 9,
          "['Comedy', 'Drama', 'Fantasy', 'Mystery']": 2,
          "['Drama', 'Fantasy', 'Horror', 'Mystery', 'Thriller']": 2,
          "['Action', 'Adventure', 'Fantasy', 'Science Fiction']": 16,
          "['Comedy', 'Action', 'Adventure', 'Fantasy', 'Science Fiction']": 1,
          "['Comedy', 'Horror', 'Music', 'Science Fiction']": 2,
          "['Comedy', 'Documentary']": 67,
          "['Adventure', 'Fantasy', 'Horror', 'Science Fiction']": 3,
          "['Action', 'War', 'Drama']": 6,
          "['Adventure', 'Drama', 'Thriller']": 9,
          "['Adventure', 'Comedy', 'Crime', 'Science Fiction']": 1,
          "['Drama', 'Thriller', 'Mystery', 'Music', 'Romance']": 1,
          "['Drama', 'Comedy', 'Foreign']": 40,
          "['Animation', 'Comedy', 'Music']": 4,
          "['Action', 'Adventure', 'Comedy', 'Science Fiction', 'Western']": 1,
          "['Thriller', 'Drama', 'Crime', 'Romance']": 2,
          "['Horror', 'Comedy', 'Action', 'Science Fiction', 'Thriller']": 1,
          "['Comedy', 'Fantasy']": 49,
          "['Horror', 'Thriller', 'Fantasy', 'Mystery']": 1,
          "['Adventure', 'Fantasy', 'Action', 'Comedy', 'Science Fiction']": 1,
          "['Romance', 'Action', 'Comedy']": 4,
          "['Adventure', 'Drama', 'Action', 'History']": 5,
          "['Horror', 'Comedy', 'Music']": 1,
          "['Action', 'Comedy', 'Mystery']": 1,
          "['Crime', 'Thriller', 'Mystery']": 8,
          "['Adventure', 'Animation', 'Family', 'Fantasy', 'Science Fiction']":
2,
          "['Mystery', 'Thriller', 'Drama']": 9,
          "['Romance', 'Crime', 'Thriller', 'Drama']": 1,
          "['Thriller', 'Romance']": 19,
          "['Comedy', 'Fantasy', 'Horror']": 9,
          "['Fantasy', 'Horror', 'Drama']": 2,
          "['Thriller', 'Comedy', 'Horror', 'Mystery']": 1,
          "['Comedy', 'Horror', 'Music']": 4,
          "['Comedy', 'Thriller', 'Horror']": 2,
          "['Action', 'Drama', 'Horror', 'Thriller']": 6,
          "['Adventure', 'Comedy', 'Romance']": 11,
          "['Comedy', 'Adventure', 'Romance']": 3,
          "['Fantasy', 'Drama', 'Comedy', 'Romance', 'Family']": 1,
          "['Adventure', 'Animation', 'Family', 'Fantasy']": 13,
          "['Action', 'Adventure', 'Crime', 'Romance', 'Thriller']": 1,
          "['Thriller', 'Action', 'Science Fiction', 'Crime']": 1,
          "['Animation', 'Thriller']": 3,
```

```
"['Action', 'Thriller', 'War']": 8,
"['Comedy', 'Drama', 'Music']": 38,
"['Drama', 'Mystery', 'War']": 3,
"['Animation', 'Fantasy', 'Music']": 1,
"['Drama', 'Adventure', 'Thriller']": 1,
"['Fantasy', 'Adventure']": 11,
"['Animation', 'Drama', 'Family', 'Fantasy']": 3,
"['Adventure', 'Action', 'Comedy']": 16,
"['Drama', 'Action', 'Comedy']": 14,
"['Action', 'War', 'Western']": 2,
"['Comedy', 'Drama', 'Foreign', 'Music']": 1,
"['Mystery', 'Horror']": 22,
"['Action', 'Adventure', 'Animation', 'Family']": 10,
"['Thriller', 'Drama', 'Horror', 'Science Fiction']": 1,
"['Comedy', 'Family', 'Music']": 2,
"['Drama', 'Comedy', 'Crime', 'Fantasy']": 1,
"['Adventure', 'Comedy', 'Drama', 'Romance']": 6,
"['Mystery', 'Science Fiction', 'Drama']": 2,
"['Music', 'Romance', 'War']": 1,
"['Action', 'Adventure', 'War']": 7,
"['Action', 'Thriller', 'Adventure']": 7,
"['Romance', 'Adventure', 'Drama']": 4,
"['Comedy', 'Family', 'Adventure', 'Crime']": 1,
"['Drama', 'History', 'Documentary']": 1,
"['Family', 'Fantasy', 'Science Fiction', 'Adventure', 'Comedy']": 1,
"['Fantasy', 'Animation', 'Comedy', 'Crime', 'Family']": 1,
"['Adventure', 'Fantasy', 'Animation']": 1,
"['Romance', 'Drama', 'Science Fiction']": 1,
"['Thriller', 'Mystery', 'Drama']": 3,
"['Science Fiction', 'Comedy', 'Horror']": 5,
"['Action', 'Adventure', 'Comedy', 'Drama']": 7,
"['Comedy', 'Drama', 'Family', 'War']": 3,
"['Adventure', 'Western', 'History']": 1,
"['Fantasy', 'Comedy', 'Adventure']": 4,
"['Adventure', 'Drama', 'Action', 'History', 'War']": 3,
"['Mystery', 'Thriller', 'Crime', 'Drama']": 4,
"['History', 'Action', 'Drama', 'Adventure', 'War']": 1,
"['Adventure', 'Comedy', 'Fantasy', 'Science Fiction']": 3,
"['Drama', 'Fantasy', 'Thriller', 'Mystery']": 1,
"['History', 'Drama', 'Action']": 6,
"['Romance', 'Drama', 'Thriller', 'Horror', 'Mystery']": 1,
"['Adventure', 'Comedy', 'Crime']": 8,
"['Drama', 'Romance', 'History']": 10,
"['Action', 'Fantasy', 'Horror', 'Mystery']": 1,
"['Drama', 'War', 'Romance', 'Western']": 1,
"['War', 'Action', 'Thriller']": 5,
"['Fantasy', 'Drama', 'Crime']": 1,
```

```
        "['Action', 'Drama', 'War']": 38,
        "['Animation', 'Fantasy', 'Family', 'Comedy']": 1,
        "['Drama', 'Action', 'Adventure', 'Crime', 'Thriller']": 1,
        "['Thriller', 'Drama', 'Crime']": 24,
        "['Drama', 'History', 'Thriller']": 12,
        "['Crime', 'Thriller', 'Horror', 'Drama']": 2,
        "['Horror', 'Drama', 'Thriller', 'Science Fiction']": 1,
        "['Action', 'Adventure', 'Crime', 'Drama', 'Mystery', 'Romance',
'Thriller']": 1,
        "['Comedy', 'Mystery', 'Romance']": 3,
        "['Comedy', 'Drama', 'Romance', 'Thriller']": 9,
        "['Thriller', 'Action', 'Drama', 'Music', 'Romance']": 1,
        "['Adventure', 'Drama', 'Romance', 'Western']": 2,
        "['Action', 'Comedy', 'Horror']": 17,
        "['Comedy', 'Crime', 'Drama', 'Horror']": 2,
        "['Adventure', 'Fantasy', 'Drama', 'Science Fiction', 'Romance']": 1,
        "['Action', 'Comedy', 'Romance', 'Thriller']": 5,
        "['Drama', 'Adventure', 'Romance', 'Thriller']": 1,
        "['Drama', 'Action', 'Thriller', 'Crime', 'Mystery']": 5,
        "['Science Fiction', 'Drama', 'Fantasy']": 4,
        "['Thriller', 'Science Fiction', 'Action']": 7,
        "['War', 'Comedy', 'Drama']": 5,
        "['Music', 'Romance', 'Comedy']": 5,
        "['Comedy', 'Drama', 'Romance', 'Science Fiction']": 2,
        "['Drama', 'Action', 'History', 'War']": 3,
        "['Crime', 'Horror']": 9,
        "['Drama', 'Action', 'Thriller', 'Mystery']": 7,
        "['Comedy', 'Romance', 'Foreign']": 9,
        "['War', 'Action']": 10,
        "['Drama', 'Western', 'Romance']": 2,
        "['Action', 'Drama', 'Romance', 'Thriller']": 7,
        "['Action', 'History', 'War']": 8,
        "['Drama', 'History', 'Thriller', 'War']": 2,
        "['Science Fiction', 'Drama', 'Romance']": 4,
        "['Comedy', 'Music', 'Drama']": 2,
        "['Drama', 'Thriller', 'History']": 5,
        "['Comedy', 'Fantasy', 'Romance', 'Science Fiction']": 2,
        "['Crime', 'Family', 'Comedy']": 1,
        "['Adventure', 'Comedy', 'Family', 'Music']": 1,
        "['Animation', 'Science Fiction', 'Family']": 1,
        "['Adventure', 'Fantasy', 'Action', 'Science Fiction', 'Family']": 4,
        "['Drama', 'Action', 'History']": 11,
        "['War', 'Action', 'Adventure', 'Drama']": 2,
        "['Drama', 'Music', 'TV Movie']": 3,
        "['Drama', 'Thriller', 'Crime', 'Adventure']": 1,
        "['Thriller', 'Action', 'Adventure', 'Crime']": 3,
        "['Drama', 'Action', 'Comedy', 'Music', 'Romance']": 1,
```

```
        "['Science Fiction', 'Action', 'Adventure', 'Comedy', 'Family']": 1,
        "['Science Fiction', 'Adventure', 'Action', 'Comedy', 'Family']": 1,
        "['Action', 'Adventure', 'Comedy', 'Family', 'Fantasy', 'Science
Fiction']": 3,
        "['Drama', 'Fantasy', 'Horror']": 21,
        "['Thriller', 'Adventure', 'Drama']": 4,
        "['Adventure', 'Drama', 'Action', 'Thriller', 'Science Fiction',
'Foreign']": 1,
        "['Fantasy', 'Drama', 'Comedy', 'Science Fiction', 'Romance']": 6,
        "['Adventure', 'Drama', 'Fantasy', 'Romance']": 1,
        "['Comedy', 'Drama', 'Romance', 'Music']": 2,
        "['Adventure', 'Fantasy', 'Comedy', 'Family']": 2,
        "['Adventure', 'Drama', 'Action', 'Romance']": 5,
        "['Drama', 'Science Fiction', 'Adventure', 'Mystery']": 1,
        "['Action', 'Comedy', 'Drama']": 28,
        "['Action', 'Adventure', 'Drama', 'History', 'War']": 9,
        "['Drama', 'Crime', 'Horror']": 2,
        "['Action', 'Adventure', 'War', 'Thriller']": 1,
        "['Crime', 'Drama', 'History']": 7,
        "['Horror', 'Thriller', 'Romance']": 3,
        "['Action', 'Comedy', 'Drama', 'Romance']": 9,
        "['Action', 'Drama', 'Thriller', 'War']": 6,
        "['Drama', 'Horror', 'Romance']": 8,
        "['Science Fiction', 'Comedy', 'Family', 'Romance']": 1,
        "['Comedy', 'Drama', 'Foreign', 'Romance']": 13,
        "['Action', 'Horror', 'Science Fiction']": 35,
        "['Action', 'Horror', 'Science Fiction', 'Thriller']": 11,
        "['Action', 'Drama', 'Adventure']": 3,
        "['Romance', 'Drama', 'Adventure']": 2,
        "['Drama', 'Romance', 'Music']": 4,
        "['Drama', 'Adventure', 'Fantasy']": 1,
        "['Animation', 'Drama', 'Family', 'Science Fiction', 'Mystery']": 1,
        "['Action', 'Science Fiction', 'War']": 5,
        "['Comedy', 'Fantasy', 'Music', 'Romance']": 1,
        "['Adventure', 'Family', 'Comedy', 'Western']": 1,
        "['Drama', 'Action', 'Music', 'Romance']": 2,
        "['Adventure', 'Family', 'Fantasy', 'Romance']": 3,
        "['Drama', 'Comedy', 'Crime']": 20,
        "['Adventure', 'Action', 'Comedy', 'Western']": 1,
        "['Adventure', 'Comedy', 'Drama']": 20,
        "['Action', 'Adventure', 'Thriller', 'Science Fiction']": 1,
        "['Action', 'Romance', 'Western']": 4,
        "['Action', 'War']": 17,
        "['Drama', 'Romance', 'Western']": 9,
        "['Adventure', 'Horror']": 14,
        "['Action', 'Adventure', 'Drama', 'War']": 14,
        "['Fantasy', 'Drama', 'Mystery', 'Romance']": 1,
```

```
"['Comedy', 'Crime', 'Action']": 6,
"['Western', 'Comedy']": 20,
"['Adventure', 'Family', 'Romance']": 2,
"['Action', 'Adventure', 'Family']": 10,
"['Drama', 'Fantasy', 'Horror', 'Science Fiction']": 2,
"['Music', 'Drama', 'Romance']": 16,
"['Drama', 'Horror', 'Science Fiction', 'Thriller']": 8,
"['Animation', 'Science Fiction', 'Fantasy', 'Adventure']": 1,
"['Horror', 'Comedy', 'Science Fiction']": 8,
"['Science Fiction', 'Action', 'Comedy', 'Horror']": 1,
"['Comedy', 'Horror', 'Action']": 3,
"['Comedy', 'Horror', 'Action', 'Science Fiction']": 1,
"['Action', 'Adventure', 'Romance', 'Science Fiction']": 1,
"['Crime', 'Science Fiction', 'Thriller']": 1,
"['Fantasy', 'Drama', 'Horror', 'Action', 'Thriller']": 1,
"['Action', 'Adventure', 'Comedy', 'Crime', 'Drama']": 3,
"['Drama', 'Family', 'Foreign', 'Music']": 1,
"['Adventure', 'Action', 'Comedy', 'Thriller']": 7,
"['Action', 'Comedy', 'Mystery', 'Thriller']": 4,
"['Animation', 'Music']": 6,
"['Adventure', 'Action', 'Thriller', 'Crime']": 4,
"['Action', 'Adventure', 'Comedy', 'Fantasy']": 6,
"['Animation', 'Action', 'Science Fiction', 'Family', 'Adventure']": 1,
"['Drama', 'History', 'War', 'Action']": 3,
"['Action', 'Adventure', 'Animation', 'Comedy', 'Family']": 2,
"['Crime', 'Thriller', 'Action']": 4,
"['Action', 'Adventure', 'Comedy', 'Crime']": 6,
"['Adventure', 'Horror', 'Science Fiction', 'Thriller']": 1,
"['Action', 'Adventure', 'Drama', 'Fantasy']": 5,
"['Horror', 'Foreign']": 31,
"['Music', 'Family', 'Animation']": 3,
"['Romance', 'Comedy', 'Music']": 11,
"['Action', 'Adventure', 'Crime', 'Mystery']": 2,
"['Adventure', 'Fantasy', 'Animation', 'Science Fiction', 'Family']":
4,
"['Adventure', 'Action', 'Western']": 3,
"['Action', 'Adventure', 'Family', 'Fantasy', 'Science Fiction']": 3,
"['History', 'Comedy']": 5,
"['Animation', 'Drama', 'Family', 'Science Fiction']": 2,
"['Action', 'Comedy', 'Drama', 'Romance', 'Western']": 2,
"['Adventure', 'Comedy', 'War']": 3,
"['Thriller', 'Horror', 'War']": 1,
"['Drama', 'Horror', 'Thriller', 'Crime']": 10,
"['Adventure', 'Horror', 'Action', 'Science Fiction']": 5,
"['Action', 'Comedy', 'Drama', 'Crime']": 1,
"['Crime', 'Action', 'Adventure']": 4,
"['Drama', 'Crime', 'Music']": 1,
```

```
        "['Horror', 'Thriller', 'War']": 3,
        "['Drama', 'Fantasy', 'Science Fiction']": 9,
        "['Adventure', 'Horror', 'Science Fiction']": 6,
        "['Music', 'Horror', 'Romance']": 1,
        "['Thriller', 'Crime', 'Science Fiction']": 2,
        "['Action', 'Drama', 'Thriller', 'Crime']": 4,
        "['Thriller', 'Adventure', 'Fantasy', 'Science Fiction', 'Romance']":
1,
        "['Horror', 'Fantasy', 'Animation', 'Family']": 1,
        "['Comedy', 'Crime', 'Thriller', 'Mystery', 'Horror']": 1,
        "['Mystery', 'Thriller', 'Horror']": 17,
        "['Family', 'Horror']": 1,
        "['Action', 'Adventure', 'Comedy', 'Crime', 'Romance', 'Thriller']": 1,
        "['Thriller', 'Action', 'Science Fiction']": 13,
        "['Adventure', 'Drama', 'Action', 'Thriller', 'War']": 1,
        "['Family', 'Comedy', 'Fantasy']": 5,
        "['Science Fiction', 'Thriller', 'Drama']": 4,
        "['Action', 'Adventure', 'Drama', 'Romance', 'Thriller']": 1,
        "['Adventure', 'Animation', 'Comedy', 'Family', 'Fantasy']": 7,
        "['Horror', 'Drama', 'Thriller', 'Mystery']": 2,
        "['Thriller', 'Western', 'Mystery']": 1,
        "['Drama', 'War', 'Western']": 2,
        "['Adventure', 'Drama', 'Action', 'Comedy', 'Romance']": 1,
        "['Drama', 'Family', 'Romance', 'Music']": 1,
        "['Thriller', 'Action', 'Adventure', 'Drama', 'Crime']": 1,
        "['Comedy', 'Romance', 'Family']": 6,
        "['Drama', 'Action', 'Crime', 'Mystery']": 1,
        "['Fantasy', 'Adventure', 'Animation', 'Comedy', 'Family', 'Music']":
1,
        "['Horror', 'Thriller', 'Crime', 'Mystery']": 2,
        "['Action', 'Crime', 'Drama', 'Thriller', 'Western']": 2,
        "['Action', 'Adventure', 'Comedy', 'Drama', 'Family', 'Thriller']": 1,
        "['Action', 'Comedy', 'Science Fiction']": 27,
        "['Adventure', 'Thriller']": 7,
        "['Action', 'Fantasy', 'Science Fiction', 'Adventure', 'Thriller']": 1,
        "['Horror', 'Comedy', 'Action', 'Fantasy', 'Science Fiction']": 1,
        "['Comedy', 'Fantasy', 'Horror', 'Romance']": 3,
        "['Crime', 'Drama', 'Thriller', 'Horror']": 2,
        "['Science Fiction', 'Animation', 'Comedy', 'Family']": 1,
        "['Adventure', 'Fantasy', 'Animation', 'Action', 'Comedy']": 1,
        "['Action', 'Adventure', 'Comedy', 'Thriller', 'Crime']": 1,
        "['Action', 'Comedy', 'Crime', 'Romance']": 3,
        "['Music', 'Romance', 'Family']": 2,
        "['Comedy', 'Drama', 'Music', 'Romance', 'Science Fiction']": 1,
        "['Drama', 'Horror', 'Thriller', 'Science Fiction', 'Foreign']": 1,
        "['Thriller', 'Action', 'Crime', 'Drama']": 9,
        "['Action', 'Horror', 'Thriller', 'Science Fiction']": 3,
```

```
            "['Comedy', 'Family', 'Music', 'Romance']": 2,
            "['Crime', 'Comedy', 'Drama']": 7,
            "['History', 'Comedy', 'Drama']": 3,
            "['Action', 'Comedy', 'Family', 'Adventure']": 1,
            "['Crime', 'Mystery', 'Thriller', 'Action']": 1,
            "['Adventure', 'Fantasy', 'Animation', 'Action', 'Family']": 2,
            "['Mystery', 'Adventure', 'Comedy', 'Romance', 'Drama']": 1,
            "['Action', 'Adventure', 'Horror', 'Thriller']": 5,
            "['Drama', 'Fantasy', 'Comedy', 'Family']": 1,
            "['Fantasy', 'Drama', 'Comedy', 'Science Fiction', 'Romance', 'Family',
    'Foreign']": 2,
            "['Fantasy', 'Action', 'Adventure']": 12,
            "['Adventure', 'Drama', 'Fantasy', 'Science Fiction']": 2,
            "['Family', 'Fantasy', 'Music']": 4,
            "['Drama', 'Action', 'Thriller', 'Romance']": 9,
            "['Adventure', 'Drama', 'Romance', 'Action']": 1,
            "['Documentary', 'Music', 'Foreign']": 5,
            "['History', 'Romance', 'War']": 1,
            "['Drama', 'Music', 'Romance', 'War']": 2,
            "['War', 'Western']": 7,
            "['Action', 'Adventure', 'Drama', 'War', 'Romance']": 1,
            "['War', 'Action', 'Drama']": 11,
            "['Action', 'Drama', 'History', 'Thriller', 'War']": 3,
            "['Romance', 'Drama', 'Music']": 17,
            "['Drama', 'Action', 'War']": 10,
            "['War', 'Drama', 'Romance', 'Foreign']": 1,
            "['Animation', 'Family', 'Adventure', 'Science Fiction']": 2,
            "['Adventure', 'Fantasy', 'Action', 'Thriller']": 3,
            "['Comedy', 'Family', 'Romance', 'Fantasy']": 1,
            "['Drama', 'Science Fiction', 'Adventure']": 2,
            "['Mystery', 'Drama', 'Romance', 'Crime']": 1,
            "['Horror', 'Comedy', 'Foreign']": 3,
            "['Comedy', 'Drama', 'Science Fiction', 'Thriller']": 1,
            "['Adventure', 'Action', 'Animation', 'Fantasy', 'Science Fiction',
    'Thriller']": 1,
            "['Documentary', 'Adventure']": 7,
            "['Animation', 'Horror', 'Thriller']": 1,
            "['Thriller', 'Horror', 'Drama']": 7,
            …})
```

[10]:
```
unique = set()

for l in genres:
    unique = unique.union(set(l))
unique  # unique genres  # some of genres are not genres but production names␣
 ↪therefore we drop them
unique
```

[10]:
```
{'Action',
 'Adventure',
 'Animation',
 'Aniplex',
 'BROSTA TV',
 'Carousel Productions',
 'Comedy',
 'Crime',
 'Documentary',
 'Drama',
 'Family',
 'Fantasy',
 'Foreign',
 'GoHands',
 'History',
 'Horror',
 'Mardock Scramble Production Committee',
 'Music',
 'Mystery',
 'Odyssey Media',
 'Pulser Productions',
 'Rogue State',
 'Romance',
 'Science Fiction',
 'Sentai Filmworks',
 'TV Movie',
 'Telescene Film Group Productions',
 'The Cartel',
 'Thriller',
 'Vision View Entertainment',
 'War',
 'Western'}
```

[11]:
```
productionname=["Aniplex","BROSTA TV","Carousel Productions","GoHands","Mardock␣
 ↪Scramble Production Committee","Odyssey Media","Pulser Productions",
                "Rogue State","Sentai Filmworks","Telescene Film Group␣
 ↪Productions","The Cartel","Vision View Entertainment"]

df=df.assign(Cleaned_genres=[[x for x in y if x not in productionname] for y in␣
 ↪df.newgenres])
#------------
genres = []   ## returns list of lists of genres
import json
# extract genres
for i in df['Cleaned_genres']:
    genres.append(i)
```

```
unique = set()

for genre in genres:
    unique = unique.union(set(genre))
#---------------------------------
# remove samples with 0 genre tags
unique
```

[11]:
```
{'Action',
 'Adventure',
 'Animation',
 'Comedy',
 'Crime',
 'Documentary',
 'Drama',
 'Family',
 'Fantasy',
 'Foreign',
 'History',
 'Horror',
 'Music',
 'Mystery',
 'Romance',
 'Science Fiction',
 'TV Movie',
 'Thriller',
 'War',
 'Western'}
```

[12]:
```
len(unique)
```

[12]:
```
20
```

[13]:
```
df.head()
```

[13]:

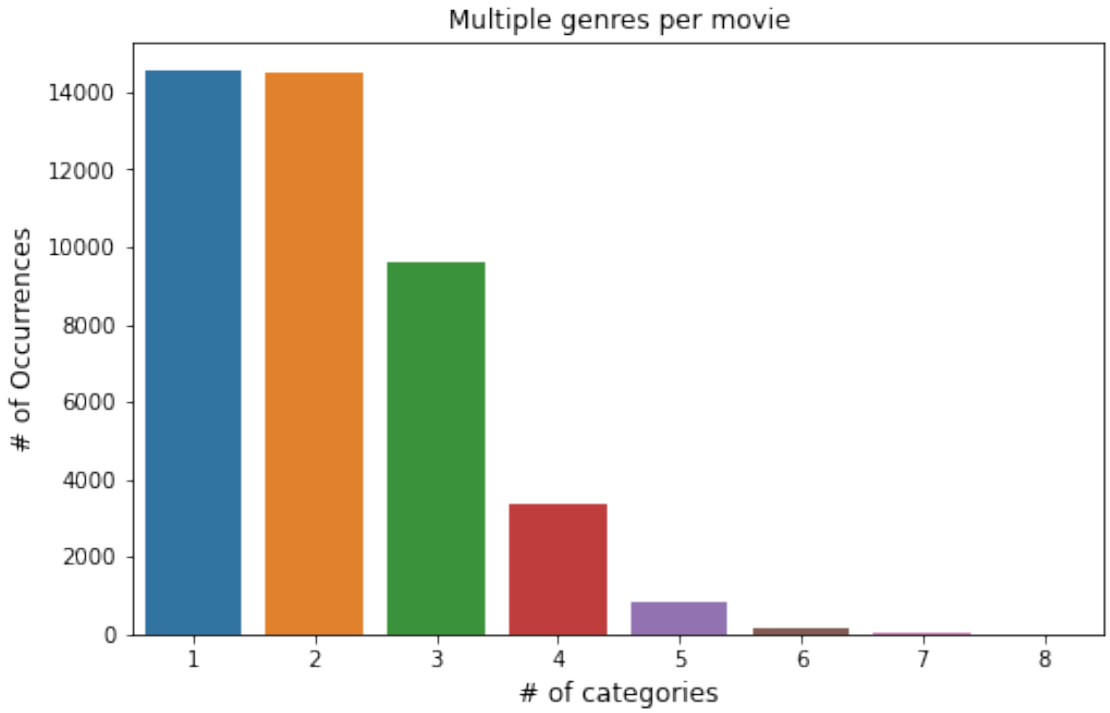
	genres	id	imdb_id	overview
0	[{'id': 16, 'name': 'Animation'}, {'id': 35, '…	862	tt0114709	Led by Woody, Andy's toys live happily in his …
1	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '…	8844	tt0113497	When siblings Judy and Peter discover an encha…
2	[{'id': 10749, 'name': 'Romance'}, {'id': 35, …	15602	tt0113228	A family wedding reignites the ancient feud be…
3	[{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam…	31357	tt0114885	Cheated on, mistreated and stepped on, the wom…
4	[{'id': 35, 'name': 'Comedy'}]	11862	tt0113041	

```
4  Just when George Banks has recovered from his …

                      newgenres                Cleaned_genres
0   [Animation, Comedy, Family]   [Animation, Comedy, Family]
1  [Adventure, Fantasy, Family]  [Adventure, Fantasy, Family]
2             [Romance, Comedy]             [Romance, Comedy]
3      [Comedy, Drama, Romance]      [Comedy, Drama, Romance]
4                       [Comedy]                       [Comedy]
```

[14]:
```
df.loc[df['Cleaned_genres'] == "[]"]
df=df[df['Cleaned_genres'].map(lambda d: len(d)) > 0]
```

[15]:
```
import seaborn as sns
import matplotlib.pyplot as plt
rowsums=df.Cleaned_genres.map(len)


x=rowsums.value_counts()#plot
plt.figure(figsize=(8,5))
ax = sns.barplot(x.index, x.values)
plt.title("Multiple genres per movie")
plt.ylabel('# of Occurrences', fontsize=12)
plt.xlabel('# of categories', fontsize=12)
x
```

[15]:
```
1    14559
2    14480
3     9585
4     3379
5      834
6      157
7       24
8        3
Name: Cleaned_genres, dtype: int64
```



2. Check if data is balanced or imbalanced

[16]:
```
import itertools, collections
# return the number of genres
c = collections.Counter(itertools.chain(*genres))
counters=dict(c)
counters
```

[16]:
```
{'Animation': 1935,
 'Comedy': 13182,
 'Family': 2770,
 'Adventure': 3496,
 'Fantasy': 2313,
 'Romance': 6735,
 'Drama': 20265,
 'Action': 6596,
 'Crime': 4307,
 'Thriller': 7624,
 'Horror': 4673,
 'History': 1398,
 'Science Fiction': 3049,
 'Mystery': 2467,
 'War': 1323,
 'Foreign': 1622,
```

```
 'Music': 1598,
 'Documentary': 3932,
 'Western': 1042,
 'TV Movie': 767}
```

[17]:
```
by_count = c.most_common()
by_count
```

[17]:
```
[('Drama', 20265),
 ('Comedy', 13182),
 ('Thriller', 7624),
 ('Romance', 6735),
 ('Action', 6596),
 ('Horror', 4673),
 ('Crime', 4307),
 ('Documentary', 3932),
 ('Adventure', 3496),
 ('Science Fiction', 3049),
 ('Family', 2770),
 ('Mystery', 2467),
 ('Fantasy', 2313),
 ('Animation', 1935),
 ('Foreign', 1622),
 ('Music', 1598),
 ('History', 1398),
 ('War', 1323),
 ('Western', 1042),
 ('TV Movie', 767)]
```

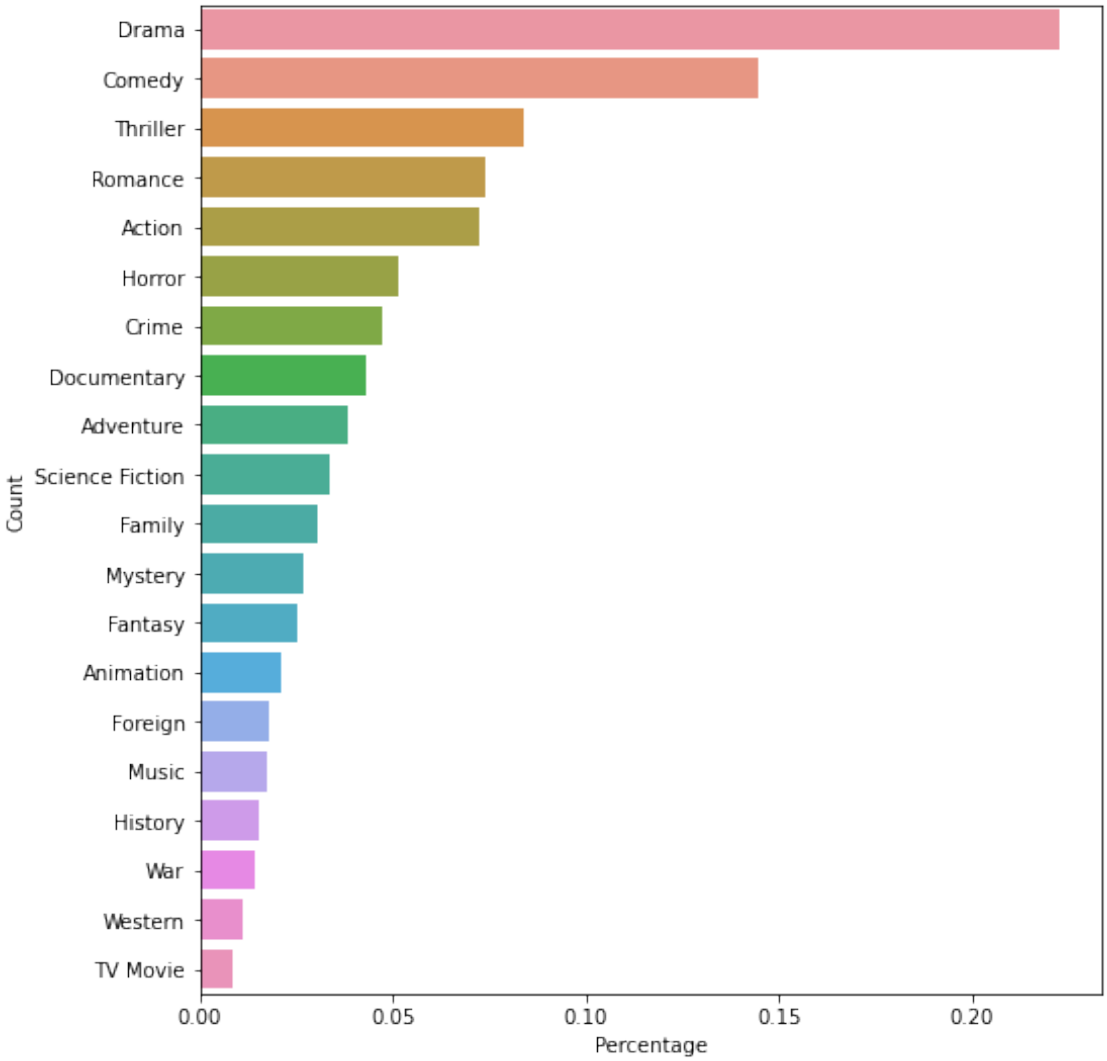
[18]:
```
total = sum(c.values())
by_freq = [(k, v/total) for k, v in by_count]
```

[19]:
```
by_freq   # percentage of genres and there is an imbalance
```

[19]:
```
[('Drama', 0.2224625112521132),
 ('Comedy', 0.1447076646101829),
 ('Thriller', 0.0836937668781698),
 ('Romance', 0.07393461698904429),
 ('Action', 0.07240872066217315),
 ('Horror', 0.05129865852855292),
 ('Crime', 0.04728083079017279),
 ('Documentary', 0.0431642040090456),
 ('Adventure', 0.03837793927152172),
 ('Science Fiction', 0.03347092014841812),
 ('Family', 0.03040814982325949),
 ('Mystery', 0.027081915384108722),
 ('Fantasy', 0.02539135398599249),
```

```
('Animation', 0.021241794190616288),
('Foreign', 0.01780578303730213),
('Music', 0.01754231892330999),
('History', 0.015346784640042153),
('War', 0.014523459283816717),
('Western', 0.011438733615825411),
('TV Movie', 0.00841987397633214)]
```

[20]:
```
import matplotlib.pyplot as plt
import seaborn as sns
df1 = pd.DataFrame(by_freq, columns=["Genre", "Percentage"])
#g = by_freq.nlargest(columns="Count", n = 50)
plt.figure(figsize=(8,9))
ax = sns.barplot(data=df1, x= "Percentage", y = "Genre")
ax.set(ylabel = 'Count')
plt.show()
```



3. Data preprocessing for Overview and Transform text to a vector

[21]:
```
#Converting Text to Features
from sklearn.preprocessing import MultiLabelBinarizer
```

[22]:
```
df.head()
```

[22]:
```
                                              genres     id    imdb_id  \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, '…    862  tt0114709
1  [{'id': 12, 'name': 'Adventure'}, {'id': 14, '…   8844  tt0113497
2  [{'id': 10749, 'name': 'Romance'}, {'id': 35, …  15602  tt0113228
3  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam…  31357  tt0114885
4                    [{'id': 35, 'name': 'Comedy'}]  11862  tt0113041

                                            overview  \
0  Led by Woody, Andy's toys live happily in his …
1  When siblings Judy and Peter discover an encha…
2  A family wedding reignites the ancient feud be…
3  Cheated on, mistreated and stepped on, the wom…
4  Just when George Banks has recovered from his …

                      newgenres                Cleaned_genres
0   [Animation, Comedy, Family]   [Animation, Comedy, Family]
1  [Adventure, Fantasy, Family]  [Adventure, Fantasy, Family]
2             [Romance, Comedy]             [Romance, Comedy]
3      [Comedy, Drama, Romance]      [Comedy, Drama, Romance]
4                       [Comedy]                       [Comedy]
```

[23]:
```
#df['overview'].dtypes
df['overview']=df['overview'].astype(str) # convert df to string
```

[24]:
```
df['overview'].dtypes
```

[24]:
```
dtype('O')
```

[25]:
```
df['overview'][0]
```

[25]:
```
"Led by Woody, Andy's toys live happily in his room until Andy's birthday brings
Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody
plots against Buzz. But when circumstances separate Buzz and Woody from their
owner, the duo eventually learns to put aside their differences."
```

[26]:
```
df
```

[26]:

```
                                                  genres      id    imdb_id  \
0      [{'id': 16, 'name': 'Animation'}, {'id': 35, '…     862  tt0114709
1      [{'id': 12, 'name': 'Adventure'}, {'id': 14, '…    8844  tt0113497
2      [{'id': 10749, 'name': 'Romance'}, {'id': 35, …   15602  tt0113228
3      [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam…   31357  tt0114885
4                          [{'id': 35, 'name': 'Comedy'}]   11862  tt0113041
…                                                    …     …        …
45459           [{'id': 878, 'name': 'Science Fiction'}]  222848  tt0112613
45460  [{'id': 18, 'name': 'Drama'}, {'id': 28, 'name…   30840  tt0102797
45461  [{'id': 18, 'name': 'Drama'}, {'id': 10751, 'n…  439050  tt6209470
45462                        [{'id': 18, 'name': 'Drama'}]  111109  tt2028550
45463  [{'id': 28, 'name': 'Action'}, {'id': 18, 'nam…   67758  tt0303758

                                                 overview  \
0      Led by Woody, Andy's toys live happily in his …
1      When siblings Judy and Peter discover an encha…
2      A family wedding reignites the ancient feud be…
3      Cheated on, mistreated and stepped on, the wom…
4      Just when George Banks has recovered from his …
…                                                    …
45459  It's the year 3000 AD. The world's most danger…
45460  Yet another version of the classic epic, with …
45461        Rising and falling between a man and woman.
45462  An artist struggles to finish his work while a…
45463  When one of her hits goes wrong, a professiona…

                          newgenres                Cleaned_genres
0       [Animation, Comedy, Family]   [Animation, Comedy, Family]
1      [Adventure, Fantasy, Family]  [Adventure, Fantasy, Family]
2                 [Romance, Comedy]             [Romance, Comedy]
3          [Comedy, Drama, Romance]      [Comedy, Drama, Romance]
4                           [Comedy]                       [Comedy]
…                                 …                             …
45459              [Science Fiction]             [Science Fiction]
45460      [Drama, Action, Romance]      [Drama, Action, Romance]
45461                [Drama, Family]               [Drama, Family]
45462                        [Drama]                       [Drama]
45463     [Action, Drama, Thriller]     [Action, Drama, Thriller]

[43021 rows x 6 columns]
```

[27]:

```
%pip install nltk
import nltk
from nltk.corpus import stopwords


nltk.download('stopwords')
```

```
Collecting nltk
  Downloading nltk-3.6.3-py3-none-any.whl (1.5 MB)
     |                | 1.5 MB 18.4 MB/s eta 0:00:01
Requirement already satisfied: joblib in /opt/conda/lib/python3.7/site-
packages (from nltk) (1.0.1)
Collecting regex
  Downloading
regex-2021.8.28-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (745
kB)
     |                | 745 kB 20.4 MB/s eta 0:00:01
Requirement already satisfied: click in /opt/conda/lib/python3.7/site-
packages (from nltk) (7.1.2)
Requirement already satisfied: tqdm in /opt/conda/lib/python3.7/site-packages
(from nltk) (4.58.0)
Installing collected packages: regex, nltk
Successfully installed nltk-3.6.3 regex-2021.8.28
Note: you may need to restart the kernel to use updated packages.

[nltk_data] Downloading package stopwords to /home/jovyan/nltk_data…
[nltk_data]   Unzipping corpora/stopwords.zip.
```

[27]:
```
True
```

[28]:
```
from nltk.stem import WordNetLemmatizer
import re
```

[ ]:
```

```

[29]:
```
%pip install spacy
```

```
Collecting spacy
  Downloading
spacy-3.1.3-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (5.9 MB)
     |                | 5.9 MB 17.7 MB/s eta 0:00:01
Requirement already satisfied: setuptools in
/opt/conda/lib/python3.7/site-packages (from spacy) (49.6.0.post20210108)
Collecting murmurhash<1.1.0,>=0.28.0
  Downloading murmurhash-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (20 kB)
Collecting typer<0.5.0,>=0.3.0
  Downloading typer-0.4.0-py3-none-any.whl (27 kB)
Collecting srsly<3.0.0,>=2.4.1
  Downloading srsly-2.4.1-cp37-cp37m-manylinux2014_x86_64.whl (456 kB)
     |                | 456 kB 99.7 MB/s eta 0:00:01
Requirement already satisfied: typing-extensions<4.0.0.0,>=3.7.4 in
/opt/conda/lib/python3.7/site-packages (from spacy) (3.7.4.3)
Collecting catalogue<2.1.0,>=2.0.6
  Downloading catalogue-2.0.6-py3-none-any.whl (17 kB)
Collecting spacy-legacy<3.1.0,>=3.0.8
```

```
  Downloading spacy_legacy-3.0.8-py2.py3-none-any.whl (14 kB)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/opt/conda/lib/python3.7/site-packages (from spacy) (4.58.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/opt/conda/lib/python3.7/site-packages (from spacy) (2.25.1)
Requirement already satisfied: numpy>=1.15.0 in /opt/conda/lib/python3.7/site-
packages (from spacy) (1.20.1)
Collecting thinc<8.1.0,>=8.0.9
  Downloading
thinc-8.0.10-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (623 kB)
     |                  | 623 kB 89.2 MB/s eta 0:00:01
Collecting preshed<3.1.0,>=3.0.2
  Downloading preshed-3.0.5-cp37-cp37m-manylinux2014_x86_64.whl (126 kB)
     |                  | 126 kB 107.0 MB/s eta 0:00:01
Collecting cymem<2.1.0,>=2.0.2
  Downloading cymem-2.0.5-cp37-cp37m-manylinux2014_x86_64.whl (35 kB)
Requirement already satisfied: packaging>=20.0 in /opt/conda/lib/python3.7/site-
packages (from spacy) (20.9)
Collecting pydantic!=1.8,!=1.8.1,<1.9.0,>=1.7.4
  Downloading pydantic-1.8.2-cp37-cp37m-manylinux2014_x86_64.whl (10.1 MB)
     |                  | 10.1 MB 100.8 MB/s eta 0:00:01
Collecting wasabi<1.1.0,>=0.8.1
  Downloading wasabi-0.8.2-py3-none-any.whl (23 kB)
Collecting blis<0.8.0,>=0.4.0
  Downloading blis-0.7.4-cp37-cp37m-manylinux2014_x86_64.whl (9.8 MB)
     |                  | 9.8 MB 91.2 MB/s eta 0:00:01
|                  | 9.5 MB 91.2 MB/s eta 0:00:01
Collecting pathy>=0.3.5
  Downloading pathy-0.6.0-py3-none-any.whl (42 kB)
     |                  | 42 kB 10.0 MB/s eta 0:00:01
Requirement already satisfied: jinja2 in /opt/conda/lib/python3.7/site-
packages (from spacy) (2.11.3)
Requirement already satisfied: zipp>=0.5 in /opt/conda/lib/python3.7/site-
packages (from catalogue<2.1.0,>=2.0.6->spacy) (3.4.0)
Requirement already satisfied: pyparsing>=2.0.2 in
/opt/conda/lib/python3.7/site-packages (from packaging>=20.0->spacy) (2.4.7)
Collecting smart-open<6.0.0,>=5.0.0
  Downloading smart_open-5.2.1-py3-none-any.whl (58 kB)
     |                  | 58 kB 36.4 MB/s  eta 0:00:01
Requirement already satisfied: idna<3,>=2.5 in
/opt/conda/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy)
(2.10)
Requirement already satisfied: chardet<5,>=3.0.2 in
/opt/conda/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy)
(4.0.0)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy)
(2020.12.5)
```

```
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/opt/conda/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy)
(1.26.3)
Requirement already satisfied: click<9.0.0,>=7.1.1 in
/opt/conda/lib/python3.7/site-packages (from typer<0.5.0,>=0.3.0->spacy) (7.1.2)
Requirement already satisfied: MarkupSafe>=0.23 in
/opt/conda/lib/python3.7/site-packages (from jinja2->spacy) (1.1.1)
Installing collected packages: murmurhash, cymem, catalogue, wasabi, typer,
srsly, smart-open, pydantic, preshed, blis, thinc, spacy-legacy, pathy, spacy
Successfully installed blis-0.7.4 catalogue-2.0.6 cymem-2.0.5 murmurhash-1.0.5
pathy-0.6.0 preshed-3.0.5 pydantic-1.8.2 smart-open-5.2.1 spacy-3.1.3 spacy-
legacy-3.0.8 srsly-2.4.1 thinc-8.0.10 typer-0.4.0 wasabi-0.8.2
Note: you may need to restart the kernel to use updated packages.
```

[30]:
```
import nltk
nltk.download('punkt')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package punkt to /home/jovyan/nltk_data…
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package wordnet to /home/jovyan/nltk_data…
[nltk_data]   Unzipping corpora/wordnet.zip.
```

[30]:
```
True
```

[31]:
```
import nltk
import pandas as pd

w_tokenizer = nltk.tokenize.WhitespaceTokenizer()
lemmatizer = nltk.stem.WordNetLemmatizer()

def lemmatize_text(text):
    return [lemmatizer.lemmatize(w) for w in w_tokenizer.tokenize(text)]
```

[32]:
```
def Clean(text):

    #remove all punctuation and special characters
    text = re.sub('[^a-zA-Z0-9]', ' ', text)
    # remove everything except alphabets
    #text = re.sub("[^a-zA-Z]"," ",text)
    # remove whitespaces
    text = ' '.join(text.split())
    # convert text to lowercase
    text = text.lower()
    #remove single characters
    text=' '.join([w for w in text.split() if len(w)>1])
    #Lemmatizer
```

```
    from nltk.stem import WordNetLemmatizer
    lemmatizer = WordNetLemmatizer()


    word_list = nltk.word_tokenize(text)
    text = ' '.join([lemmatizer.lemmatize(w) for w in word_list if w not in␣
↪stopwords.words('english')])

    return text
```

[33]:
```
df["newoverview"]=df['overview'].apply(lambda x : Clean(x))

df["newoverview"]
```

[33]:
```
0        led woody andy toy live happily room andy birt…
1        sibling judy peter discover enchanted board ga…
2        family wedding reignites ancient feud next doo…
3        cheated mistreated stepped woman holding breat…
4        george bank recovered daughter wedding receive…
                              …
45459    year 3000 ad world dangerous woman banished re…
45460    yet another version classic epic enough variat…
45461                              rising falling man woman
45462    artist struggle finish work storyline cult pla…
45463    one hit go wrong professional assassin end sui…
Name: newoverview, Length: 43021, dtype: object
```

[34]:
```
multilabel_binarizer = MultiLabelBinarizer()
#https://towardsai.net/p/latest/predicting-genres-from-movie-dialogue
```

[35]:
```
multilabel_binarizer.fit(df['Cleaned_genres'])
```

[35]:
```
MultiLabelBinarizer()
```

[36]:
```
# transform target variable
y = multilabel_binarizer.transform(df['Cleaned_genres'])
```

[37]:
```
y[0]
```

[37]:
```
array([0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

[38]:
```
multilabel_binarizer.classes_
```

[38]:
```
array(['Action', 'Adventure', 'Animation', 'Comedy', 'Crime',
       'Documentary', 'Drama', 'Family', 'Fantasy', 'Foreign', 'History',
       'Horror', 'Music', 'Mystery', 'Romance', 'Science Fiction',
       'TV Movie', 'Thriller', 'War', 'Western'], dtype=object)
```

[39]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
#tfidf_vectorizer = TfidfVectorizer(max_df=0.8, max_features=1000,ngram_range =␣
↪(1,2))
tfidf_vectorizer = TfidfVectorizer(min_df=10,max_features=10000,ngram_range =␣
↪(1,2))
```

[40]:
```
# split dataset into training and validation set
from sklearn.model_selection import train_test_split
X_train, X_test, ytrain, ytest = train_test_split(df['newoverview'], y,␣
↪test_size=0.2, random_state=9)
```

[41]:
```
# create TF-IDF features
X_train = tfidf_vectorizer.fit_transform(X_train)
X_test = tfidf_vectorizer.transform(X_test)

#X_train_vec = tfidf_vectorizer.fit_transform(xtrain)
#X_valid_vec = tfidf_vectorizer.transform(xval)
```

[42]:
```
X_train
```

[42]:
```
<34416x10000 sparse matrix of type '<class 'numpy.float64'>'
        with 932166 stored elements in Compressed Sparse Row format>
```

[ ]:
```

```

Model Building

[43]:
```
#1 Random forest calssifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import f1_score
from sklearn.metrics import roc_auc_score
from sklearn.metrics import average_precision_score
from sklearn.metrics import recall_score
```

[44]:
```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import f1_score
rfClassifier = RandomForestClassifier(n_jobs=-1)
rfClassifier.fit(X_train, ytrain)
rfPreds = rfClassifier.predict(X_test)
f1_score(ytest, rfPreds, average="micro")
```

[44]:
```
0.356000162002349
```

[45]:
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,rfPreds)
```

[45]:
```
0.16025566531086577
```

[46]:
```
import sklearn.metrics as metrics

metrics.hamming_loss(ytest, rfPreds) #hyperparameter tuning
```

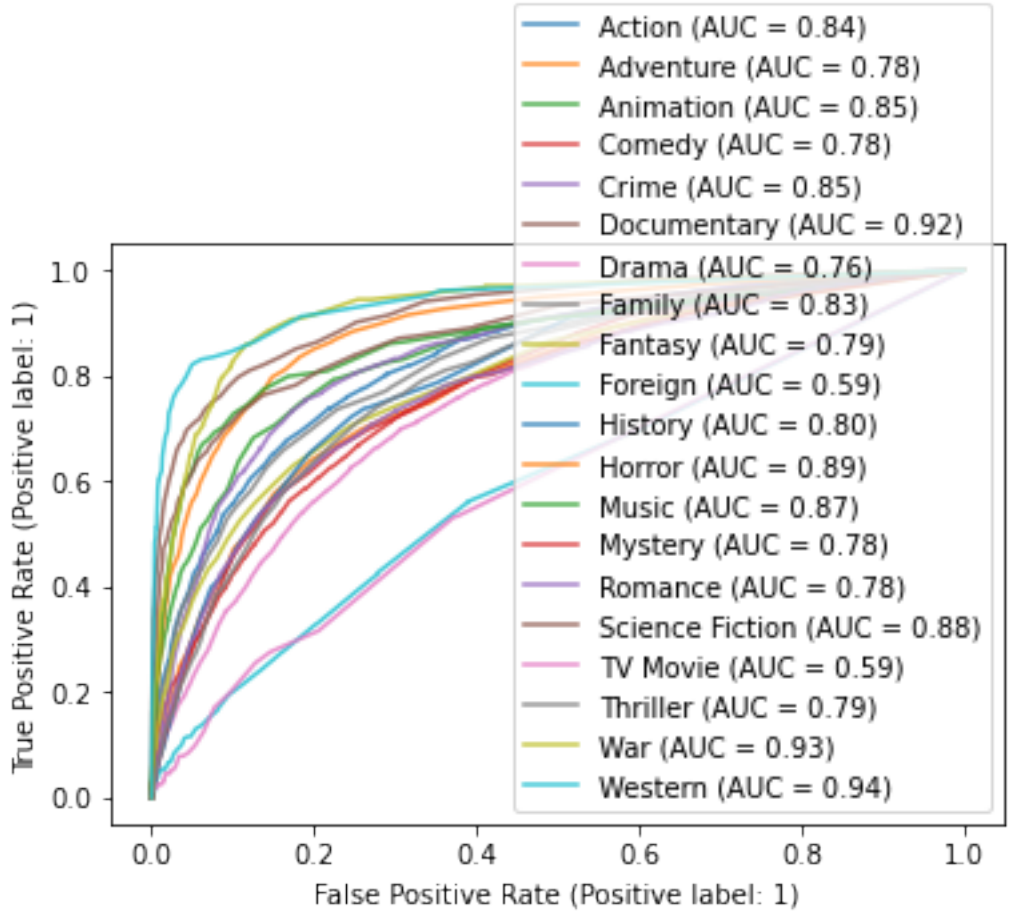
[46]: 0.09239395700174317

[47]:
```
from sklearn.metrics import classification_report

label_names = multilabel_binarizer.classes_

print(classification_report(ytest, rfPreds,target_names=label_names))
```

```
                precision    recall  f1-score   support

        Action       0.83      0.04      0.08      1305
     Adventure       1.00      0.01      0.01       706
     Animation       0.81      0.03      0.07       378
        Comedy       0.71      0.24      0.36      2660
         Crime       0.56      0.04      0.08       842
   Documentary       0.90      0.39      0.54       812
         Drama       0.63      0.73      0.68      4017
        Family       0.50      0.00      0.01       540
       Fantasy       0.67      0.00      0.01       465
       Foreign       1.00      0.00      0.01       328
       History       0.00      0.00      0.00       291
        Horror       0.77      0.10      0.17       908
         Music       0.58      0.02      0.04       309
       Mystery       0.33      0.00      0.00       478
       Romance       0.68      0.10      0.18      1339
Science Fiction       0.86      0.12      0.21       582
      TV Movie       0.00      0.00      0.00       152
      Thriller       0.65      0.07      0.12      1494
           War       1.00      0.02      0.04       271
       Western       1.00      0.03      0.06       207

     micro avg       0.67      0.24      0.36     18084
     macro avg       0.67      0.10      0.13     18084
  weighted avg       0.69      0.24      0.28     18084
   samples avg       0.47      0.30      0.35     18084
```

[48]:
```
labelPlots ={}
from sklearn import metrics
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):
```

```
  rfClassifier = RandomForestClassifier(n_jobs=-1)
  rfClassifier.fit(X_train, ytrain[:,i])
  ax = plt.gca()
 # labelPlots[i]= plot_roc_curve(rfClassifier, X_test, ytest[:,i], name=␣
↪('label_'+str(i)), ax=ax, alpha=0.8)
  #labelPlots[i]= plot_roc_curve(rfClassifier, X_test, ytest[:,i], name=␣
↪('label_'+str(i)), ax=ax, alpha=0.8)
  labelPlots[i]= plot_roc_curve(rfClassifier, X_test, ytest[:,i], name=␣
↪(label_names[i]), ax=ax, alpha=0.8)

plt.show()
```



[50]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
```

[51]:
```
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
```

```
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label)


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[7289   11]
 [1252   53]]
Confusion matrix for label Adventure:
[[7899    0]
 [ 701    5]]
Confusion matrix for label Animation:
[[8224    3]
 [ 365   13]]
Confusion matrix for label Comedy:
[[5683  262]
 [2027  633]]
Confusion matrix for label Crime:
[[7735   28]
 [ 806   36]]
Confusion matrix for label Documentary:
[[7756   37]
 [ 494  318]]
Confusion matrix for label Drama:
[[2884 1704]
 [1101 2916]]
Confusion matrix for label Family:
[[8063    2]
 [ 538    2]]
Confusion matrix for label Fantasy:
[[8139    1]
 [ 463    2]]
Confusion matrix for label Foreign:
[[8277    0]
 [ 327    1]]
Confusion matrix for label History:
[[8314    0]
 [ 291    0]]
Confusion matrix for label Horror:
[[7671   26]
 [ 819   89]]
Confusion matrix for label Music:
```

```
[[8291    5]
 [ 302    7]]
Confusion matrix for label Mystery:
[[8125    2]
 [ 477    1]]
Confusion matrix for label Romance:
[[7201   65]
 [1203  136]]
Confusion matrix for label Science Fiction:
[[8011   12]
 [ 511   71]]
Confusion matrix for label TV Movie:
[[8453    0]
 [ 152    0]]
Confusion matrix for label Thriller:
[[7057   54]
 [1394  100]]
Confusion matrix for label War:
[[8334    0]
 [ 265    6]]
Confusion matrix for label Western:
[[8398    0]
 [ 201    6]]
```

[52]:
```
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize="true")


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[0.99849315 0.00150685]
 [0.95938697 0.04061303]]
Confusion matrix for label Adventure:
[[1.         0.        ]
 [0.99291785 0.00708215]]
Confusion matrix for label Animation:
[[9.99635347e-01 3.64652972e-04]
 [9.65608466e-01 3.43915344e-02]]
Confusion matrix for label Comedy:
[[0.95592935 0.04407065]
```

```
 [0.76203008 0.23796992]]
Confusion matrix for label Crime:
[[0.99639315 0.00360685]
 [0.95724466 0.04275534]]
Confusion matrix for label Documentary:
[[0.99525215 0.00474785]
 [0.60837438 0.39162562]]
Confusion matrix for label Drama:
[[0.62859634 0.37140366]
 [0.27408514 0.72591486]]
Confusion matrix for label Family:
[[9.99752015e-01 2.47985121e-04]
 [9.96296296e-01 3.70370370e-03]]
Confusion matrix for label Fantasy:
[[9.99877150e-01 1.22850123e-04]
 [9.95698925e-01 4.30107527e-03]]
Confusion matrix for label Foreign:
[[1.         0.        ]
 [0.99695122 0.00304878]]
Confusion matrix for label History:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Horror:
[[0.99662206 0.00337794]
 [0.90198238 0.09801762]]
Confusion matrix for label Music:
[[9.99397300e-01 6.02700096e-04]
 [9.77346278e-01 2.26537217e-02]]
Confusion matrix for label Mystery:
[[9.99753907e-01 2.46093269e-04]
 [9.97907950e-01 2.09205021e-03]]
Confusion matrix for label Romance:
[[0.99105423 0.00894577]
 [0.89843167 0.10156833]]
Confusion matrix for label Science Fiction:
[[0.9985043  0.0014957 ]
 [0.87800687 0.12199313]]
Confusion matrix for label TV Movie:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Thriller:
[[0.99240613 0.00759387]
 [0.9330656  0.0669344 ]]
Confusion matrix for label War:
[[1.         0.        ]
 [0.97785978 0.02214022]]
Confusion matrix for label Western:
[[1.         0.        ]
```

```
[0.97101449 0.02898551]]
```

[53]:
```
#2. Decision Tree Classifier
from sklearn.tree import DecisionTreeClassifier
dtClassifier = DecisionTreeClassifier()
dtClassifier.fit(X_train, ytrain)
rfPreds = dtClassifier.predict(X_test)
f1_score(ytest, rfPreds, average="micro")
```

[53]: 0.3706929754392202

[ ]:

[54]:
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,rfPreds)
```

[54]: 0.11946542707728065

[55]:
```
import sklearn.metrics as metrics

metrics.hamming_loss(ytest, rfPreds) #hyperparameter tuning
```

[55]: 0.12342242882045322

[56]:
```
from sklearn.metrics import classification_report

label_names = multilabel_binarizer.classes_

print(classification_report(ytest, rfPreds,target_names=label_names))
```

```
                precision    recall  f1-score   support

        Action       0.37      0.26      0.30      1305
     Adventure       0.20      0.14      0.16       706
     Animation       0.22      0.17      0.19       378
        Comedy       0.44      0.42      0.43      2660
         Crime       0.29      0.27      0.28       842
   Documentary       0.53      0.48      0.50       812
         Drama       0.57      0.59      0.58      4017
        Family       0.23      0.17      0.20       540
       Fantasy       0.20      0.14      0.17       465
       Foreign       0.06      0.03      0.04       328
       History       0.10      0.06      0.08       291
        Horror       0.38      0.35      0.36       908
         Music       0.34      0.23      0.27       309
       Mystery       0.16      0.11      0.13       478
       Romance       0.31      0.26      0.28      1339
Science Fiction       0.33      0.26      0.29       582
```

```
       TV Movie       0.01      0.01      0.01       152
       Thriller       0.33      0.28      0.30      1494
            War       0.33      0.23      0.27       271
        Western       0.30      0.19      0.23       207

      micro avg       0.40      0.35      0.37     18084
      macro avg       0.28      0.23      0.25     18084
   weighted avg       0.38      0.35      0.36     18084
    samples avg       0.44      0.39      0.38     18084
```

[57]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label)


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[6716  584]
 [ 967  338]]
Confusion matrix for label Adventure:
[[7523  376]
 [ 610   96]]
Confusion matrix for label Animation:
[[7999  228]
 [ 313   65]]
Confusion matrix for label Comedy:
[[4511 1434]
 [1534 1126]]
Confusion matrix for label Crime:
[[7203  560]
 [ 616  226]]
Confusion matrix for label Documentary:
[[7440  353]
 [ 420  392]]
Confusion matrix for label Drama:
[[2774 1814]
 [1639 2378]]
Confusion matrix for label Family:
```

```
[[7764  301]
 [ 448   92]]
Confusion matrix for label Fantasy:
[[7880  260]
 [ 398   67]]
Confusion matrix for label Foreign:
[[8094  183]
 [ 317   11]]
Confusion matrix for label History:
[[8158  156]
 [ 273   18]]
Confusion matrix for label Horror:
[[7177  520]
 [ 594  314]]
Confusion matrix for label Music:
[[8158  138]
 [ 239   70]]
Confusion matrix for label Mystery:
[[7848  279]
 [ 426   52]]
Confusion matrix for label Romance:
[[6513  753]
 [ 993  346]]
Confusion matrix for label Science Fiction:
[[7711  312]
 [ 431  151]]
Confusion matrix for label TV Movie:
[[8362   91]
 [ 151    1]]
Confusion matrix for label Thriller:
[[6258  853]
 [1082  412]]
Confusion matrix for label War:
[[8211  123]
 [ 210   61]]
Confusion matrix for label Western:
[[8303   95]
 [ 167   40]]
```

[58]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
```

```
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize="true")


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

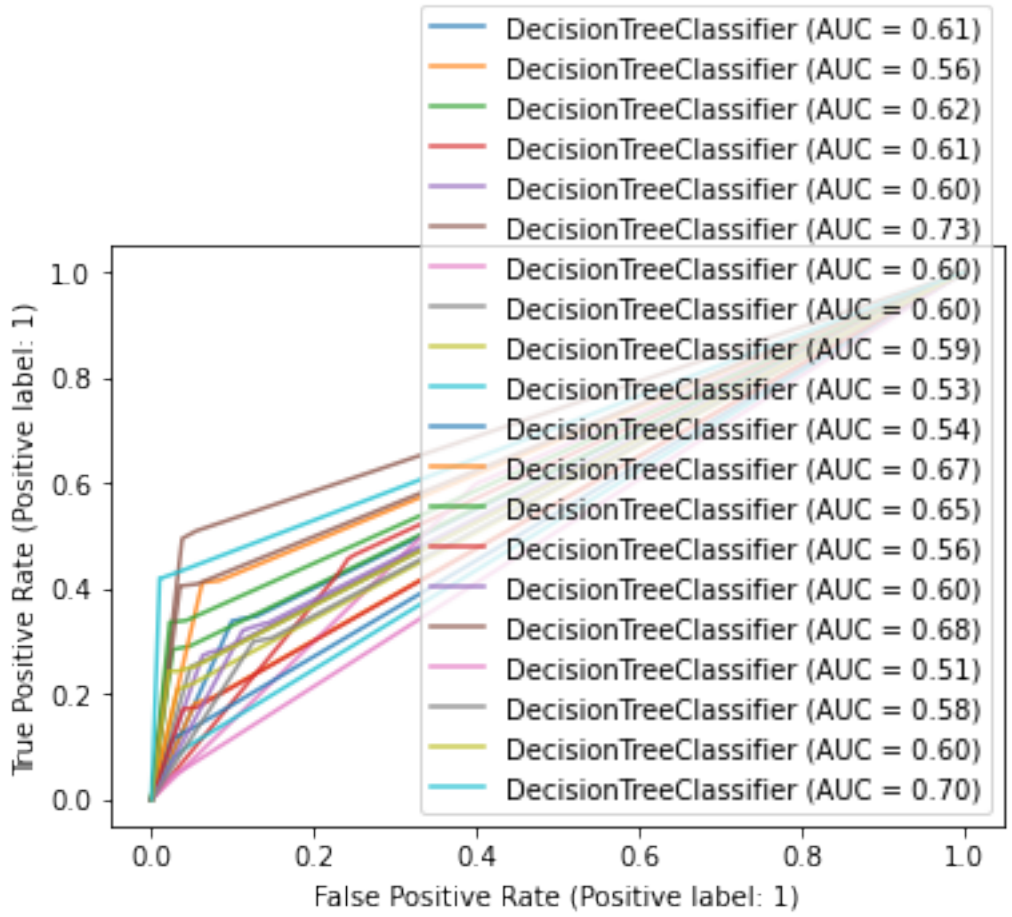
```
Confusion matrix for label Action:
[[0.92       0.08      ]
 [0.74099617 0.25900383]]
Confusion matrix for label Adventure:
[[0.95239904 0.04760096]
 [0.86402266 0.13597734]]
Confusion matrix for label Animation:
[[0.97228637 0.02771363]
 [0.82804233 0.17195767]]
Confusion matrix for label Comedy:
[[0.7587889  0.2412111 ]
 [0.57669173 0.42330827]]
Confusion matrix for label Crime:
[[0.92786294 0.07213706]
 [0.73159145 0.26840855]]
Confusion matrix for label Documentary:
[[0.95470294 0.04529706]
 [0.51724138 0.48275862]]
Confusion matrix for label Drama:
[[0.60462075 0.39537925]
 [0.40801593 0.59198407]]
Confusion matrix for label Family:
[[0.96267824 0.03732176]
 [0.82962963 0.17037037]]
Confusion matrix for label Fantasy:
[[0.96805897 0.03194103]
 [0.85591398 0.14408602]]
Confusion matrix for label Foreign:
[[0.97789054 0.02210946]
 [0.96646341 0.03353659]]
Confusion matrix for label History:
[[0.98123647 0.01876353]
 [0.93814433 0.06185567]]
Confusion matrix for label Horror:
[[0.93244121 0.06755879]
 [0.65418502 0.34581498]]
Confusion matrix for label Music:
[[0.98336548 0.01663452]
 [0.77346278 0.22653722]]
```

```
Confusion matrix for label Mystery:
[[0.96566999 0.03433001]
 [0.89121339 0.10878661]]
Confusion matrix for label Romance:
[[0.89636664 0.10363336]
 [0.74159821 0.25840179]]
Confusion matrix for label Science Fiction:
[[0.9611118  0.0388882 ]
 [0.74054983 0.25945017]]
Confusion matrix for label TV Movie:
[[0.98923459 0.01076541]
 [0.99342105 0.00657895]]
Confusion matrix for label Thriller:
[[0.880045   0.119955  ]
 [0.72423025 0.27576975]]
Confusion matrix for label War:
[[0.98524118 0.01475882]
 [0.77490775 0.22509225]]
Confusion matrix for label Western:
[[0.98868778 0.01131222]
 [0.80676329 0.19323671]]
```

[ ]:
```
labelPlots ={}
from sklearn import metrics
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):

  dtClassifier = DecisionTreeClassifier()
  dtClassifier.fit(X_train,  ytrain[:,i])
  ax = plt.gca()

  labelPlots[i]= plot_roc_curve(dtClassifier, X_test, ytest[:,i], name=␣
 ↪(label_names[i]), ax=ax, alpha=0.8)

plt.show()
```

[60]:
```
labelPlots ={}
from sklearn import metrics
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):

  dtClassifier = DecisionTreeClassifier()
  dtClassifier.fit(X_train,  ytrain[:,i])
```

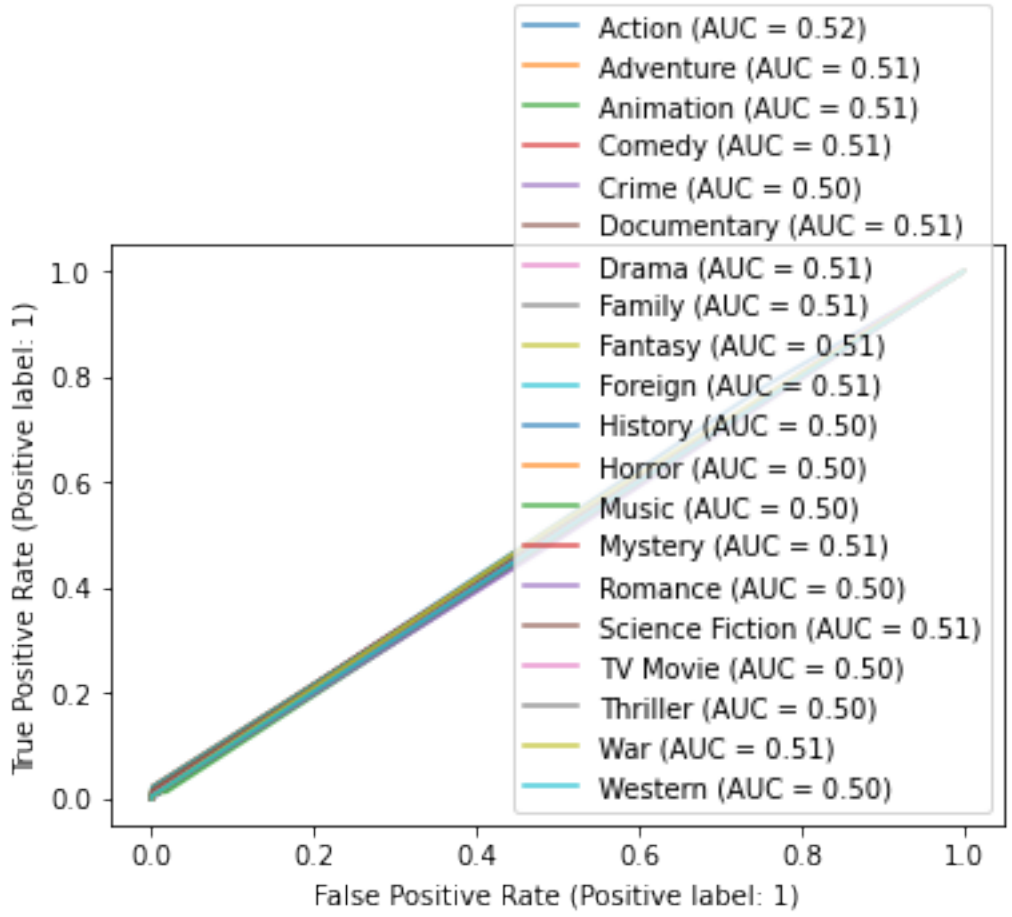
```
  ax = plt.gca()
 a=0.8)
  labelPlots[i]= plot_roc_curve(dtClassifier, X_test, ytest[:,i], ax=ax,␣
 ↪alpha=0.8)

plt.show()
```



[61]:
```
#3. KNN CLassifier

from sklearn.neighbors import KNeighborsClassifier


knnClf = KNeighborsClassifier()

knnClf.fit(X_train, ytrain)
rfPreds = knnClf.predict(X_test)
f1_score(ytest, rfPreds, average="micro")
```

[61]: 0.22356275303643727

[62]:
```
labelPlots ={}
from sklearn import metrics
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):

    knnClf = KNeighborsClassifier()

    knnClf.fit(X_train, ytrain[:,i])
    ax = plt.gca()

    labelPlots[i]= plot_roc_curve(knnClf, X_test, ytest[:,i], name=␣
 ↪(label_names[i]), ax=ax, alpha=0.8)


plt.show()
```



[64]:
```
import sklearn.metrics as metrics

metrics.hamming_loss(ytest, rfPreds) #hyperparameter tuning
```

[64]: 0.11143521208599651

[65]:
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,rfPreds)
```

[65]: 0.08692620569436374

[66]:
```
from sklearn.metrics import classification_report

label_names = multilabel_binarizer.classes_

print(classification_report(ytest, rfPreds,target_names=label_names))
```

```
                precision    recall  f1-score   support

        Action       0.33      0.00      0.00      1305
     Adventure       0.50      0.00      0.00       706
     Animation       1.00      0.01      0.01       378
        Comedy       0.32      0.29      0.31      2660
         Crime       1.00      0.00      0.00       842
   Documentary       1.00      0.00      0.01       812
         Drama       0.47      0.49      0.48      4017
        Family       0.50      0.00      0.00       540
       Fantasy       1.00      0.00      0.00       465
       Foreign       0.00      0.00      0.00       328
       History       0.00      0.00      0.00       291
        Horror       0.00      0.00      0.00       908
         Music       0.00      0.00      0.00       309
       Mystery       0.00      0.00      0.00       478
       Romance       1.00      0.00      0.00      1339
Science Fiction       1.00      0.01      0.01       582
       TV Movie       0.00      0.00      0.00       152
       Thriller       0.00      0.00      0.00      1494
            War       0.00      0.00      0.00       271
       Western       0.00      0.00      0.00       207

     micro avg       0.42      0.15      0.22     18084
     macro avg       0.41      0.04      0.04     18084
  weighted avg       0.45      0.15      0.15     18084
   samples avg       0.32      0.18      0.22     18084
```

[67]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize='true')


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[9.99452055e-01 5.47945205e-04]
 [9.98467433e-01 1.53256705e-03]]
Confusion matrix for label Adventure:
[[9.99873402e-01 1.26598304e-04]
 [9.98583569e-01 1.41643059e-03]]
Confusion matrix for label Animation:
[[1.         0.        ]
 [0.99470899 0.00529101]]
Confusion matrix for label Comedy:
[[0.72632464 0.27367536]
 [0.70864662 0.29135338]]
Confusion matrix for label Crime:
[[1.         0.        ]
 [0.99881235 0.00118765]]
Confusion matrix for label Documentary:
[[1.         0.        ]
 [0.99630542 0.00369458]]
Confusion matrix for label Drama:
[[0.51569311 0.48430689]
 [0.50933532 0.49066468]]
Confusion matrix for label Family:
[[9.99876007e-01 1.23992560e-04]
 [9.98148148e-01 1.85185185e-03]]
Confusion matrix for label Fantasy:
[[1.         0.        ]
 [0.99784946 0.00215054]]
Confusion matrix for label Foreign:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label History:
[[1. 0.]
```

```
 [1. 0.]]
Confusion matrix for label Horror:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Music:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Mystery:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Romance:
[[1.00000000e+00 0.00000000e+00]
 [9.99253174e-01 7.46825990e-04]]
Confusion matrix for label Science Fiction:
[[1.         0.        ]
 [0.99484536 0.00515464]]
Confusion matrix for label TV Movie:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Thriller:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label War:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Western:
[[1. 0.]
 [1. 0.]]
```

[68]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label)


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[7296    4]
 [1303    2]]
Confusion matrix for label Adventure:
```

```
[[7898    1]
 [ 705    1]]
Confusion matrix for label Animation:
[[8227    0]
 [ 376    2]]
Confusion matrix for label Comedy:
[[4318 1627]
 [1885  775]]
Confusion matrix for label Crime:
[[7763    0]
 [ 841    1]]
Confusion matrix for label Documentary:
[[7793    0]
 [ 809    3]]
Confusion matrix for label Drama:
[[2366 2222]
 [2046 1971]]
Confusion matrix for label Family:
[[8064    1]
 [ 539    1]]
Confusion matrix for label Fantasy:
[[8140    0]
 [ 464    1]]
Confusion matrix for label Foreign:
[[8277    0]
 [ 328    0]]
Confusion matrix for label History:
[[8314    0]
 [ 291    0]]
Confusion matrix for label Horror:
[[7697    0]
 [ 908    0]]
Confusion matrix for label Music:
[[8296    0]
 [ 309    0]]
Confusion matrix for label Mystery:
[[8127    0]
 [ 478    0]]
Confusion matrix for label Romance:
[[7266    0]
 [1338    1]]
Confusion matrix for label Science Fiction:
[[8023    0]
 [ 579    3]]
Confusion matrix for label TV Movie:
[[8453    0]
 [ 152    0]]
Confusion matrix for label Thriller:
```

```
[[7111    0]
 [1494    0]]
Confusion matrix for label War:
[[8334    0]
 [ 271    0]]
Confusion matrix for label Western:
[[8398    0]
 [ 207    0]]
```

[69]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rfPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize="true")


for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[9.99452055e-01 5.47945205e-04]
 [9.98467433e-01 1.53256705e-03]]
Confusion matrix for label Adventure:
[[9.99873402e-01 1.26598304e-04]
 [9.98583569e-01 1.41643059e-03]]
Confusion matrix for label Animation:
[[1.         0.        ]
 [0.99470899 0.00529101]]
Confusion matrix for label Comedy:
[[0.72632464 0.27367536]
 [0.70864662 0.29135338]]
Confusion matrix for label Crime:
[[1.         0.        ]
 [0.99881235 0.00118765]]
Confusion matrix for label Documentary:
[[1.         0.        ]
 [0.99630542 0.00369458]]
Confusion matrix for label Drama:
[[0.51569311 0.48430689]
 [0.50933532 0.49066468]]
Confusion matrix for label Family:
[[9.99876007e-01 1.23992560e-04]
 [9.98148148e-01 1.85185185e-03]]
```

```
Confusion matrix for label Fantasy:
[[1.         0.        ]
 [0.99784946 0.00215054]]
Confusion matrix for label Foreign:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label History:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Horror:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Music:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Mystery:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Romance:
[[1.00000000e+00 0.00000000e+00]
 [9.99253174e-01 7.46825990e-04]]
Confusion matrix for label Science Fiction:
[[1.         0.        ]
 [0.99484536 0.00515464]]
Confusion matrix for label TV Movie:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Thriller:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label War:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Western:
[[1. 0.]
 [1. 0.]]
```

[70]:
```
from functools import partial
def print_evaluation_scores(ytest, rfPreds):
    f1_score_macro = partial(f1_score,average="macro")
    f1_score_micro = partial(f1_score,average="micro")
    f1_score_weighted = partial(f1_score,average="weighted")

    average_precision_score_macro =␣
 ↪partial(average_precision_score,average="macro")
    average_precision_score_micro =␣
 ↪partial(average_precision_score,average="micro")
```

```
    average_precision_score_weighted =␣
↪partial(average_precision_score,average="weighted")

    scores =␣
↪[accuracy_score,f1_score_macro,f1_score_micro,f1_score_weighted,average_precision_score_mac
            average_precision_score_micro,average_precision_score_weighted]
    for score in scores:
        print(score,score(ytest,rfPreds))
print_evaluation_scores(ytest, rfPreds)
```

```
<function accuracy_score at 0x7f8de781b7a0> 0.08692620569436374
functools.partial(<function f1_score at 0x7f8de78213b0>, average='macro')
0.041610839898434074
functools.partial(<function f1_score at 0x7f8de78213b0>, average='micro')
0.22356275303643727
functools.partial(<function f1_score at 0x7f8de78213b0>, average='weighted')
0.15334737525691033
functools.partial(<function average_precision_score at 0x7f8de78cb710>,
average='macro') 0.10628649921578719
functools.partial(<function average_precision_score at 0x7f8de78cb710>,
average='micro') 0.15275061042691684
functools.partial(<function average_precision_score at 0x7f8de78cb710>,
average='weighted') 0.21539980031640457
```

Onevsrest

[71]:
```
#4KNN

from sklearn.neighbors import KNeighborsClassifier
from sklearn.multiclass import OneVsRestClassifier

knnClf = KNeighborsClassifier()
knnClf = OneVsRestClassifier(knnClf)
knnClf.fit(X_train, ytrain)
knnPredictions = knnClf.predict(X_test)
f1_score(ytest, knnPredictions, average="micro")
```

[71]:
```
0.22356275303643727
```

[72]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = knnPredictions[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
↪y_true=y_true_label,normalize='true')
```

```
for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[9.99452055e-01 5.47945205e-04]
 [9.98467433e-01 1.53256705e-03]]
Confusion matrix for label Adventure:
[[9.99873402e-01 1.26598304e-04]
 [9.98583569e-01 1.41643059e-03]]
Confusion matrix for label Animation:
[[1.         0.        ]
 [0.99470899 0.00529101]]
Confusion matrix for label Comedy:
[[0.72632464 0.27367536]
 [0.70864662 0.29135338]]
Confusion matrix for label Crime:
[[1.         0.        ]
 [0.99881235 0.00118765]]
Confusion matrix for label Documentary:
[[1.         0.        ]
 [0.99630542 0.00369458]]
Confusion matrix for label Drama:
[[0.51569311 0.48430689]
 [0.50933532 0.49066468]]
Confusion matrix for label Family:
[[9.99876007e-01 1.23992560e-04]
 [9.98148148e-01 1.85185185e-03]]
Confusion matrix for label Fantasy:
[[1.         0.        ]
 [0.99784946 0.00215054]]
Confusion matrix for label Foreign:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label History:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Horror:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Music:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Mystery:
[[1. 0.]
```

```
 [1. 0.]]
Confusion matrix for label Romance:
[[1.00000000e+00 0.00000000e+00]
 [9.99253174e-01 7.46825990e-04]]
Confusion matrix for label Science Fiction:
[[1.         0.        ]
 [0.99484536 0.00515464]]
Confusion matrix for label TV Movie:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Thriller:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label War:
[[1. 0.]
 [1. 0.]]
Confusion matrix for label Western:
[[1. 0.]
 [1. 0.]]
```

[73]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = knnPredictions[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label)

for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

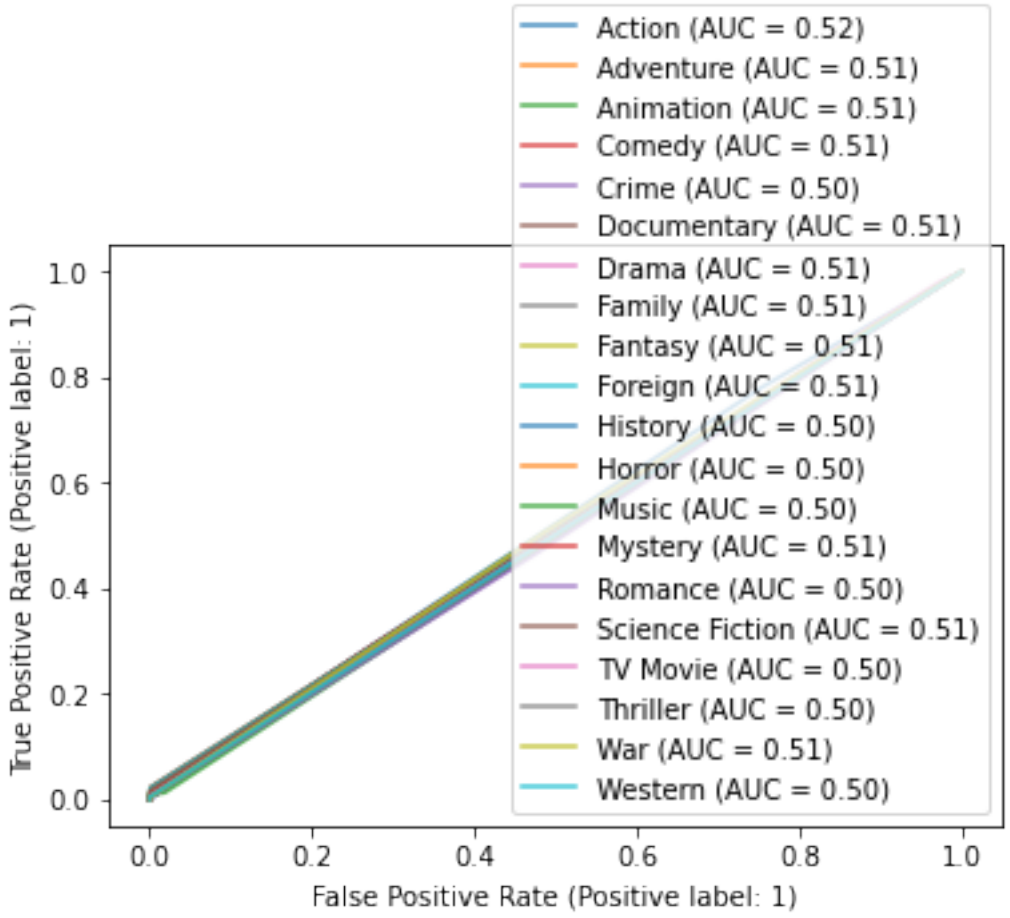
```
Confusion matrix for label Action:
[[7296    4]
 [1303    2]]
Confusion matrix for label Adventure:
[[7898    1]
 [ 705    1]]
Confusion matrix for label Animation:
[[8227    0]
 [ 376    2]]
Confusion matrix for label Comedy:
[[4318 1627]
 [1885  775]]
Confusion matrix for label Crime:
[[7763    0]
```

```
 [ 841    1]]
Confusion matrix for label Documentary:
[[7793    0]
 [ 809    3]]
Confusion matrix for label Drama:
[[2366 2222]
 [2046 1971]]
Confusion matrix for label Family:
[[8064    1]
 [ 539    1]]
Confusion matrix for label Fantasy:
[[8140    0]
 [ 464    1]]
Confusion matrix for label Foreign:
[[8277    0]
 [ 328    0]]
Confusion matrix for label History:
[[8314    0]
 [ 291    0]]
Confusion matrix for label Horror:
[[7697    0]
 [ 908    0]]
Confusion matrix for label Music:
[[8296    0]
 [ 309    0]]
Confusion matrix for label Mystery:
[[8127    0]
 [ 478    0]]
Confusion matrix for label Romance:
[[7266    0]
 [1338    1]]
Confusion matrix for label Science Fiction:
[[8023    0]
 [ 579    3]]
Confusion matrix for label TV Movie:
[[8453    0]
 [ 152    0]]
Confusion matrix for label Thriller:
[[7111    0]
 [1494    0]]
Confusion matrix for label War:
[[8334    0]
 [ 271    0]]
Confusion matrix for label Western:
[[8398    0]
 [ 207    0]]
```

[74]:
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,knnPredictions)
```

[74]:
```
0.08692620569436374
```

[75]:
```
labelPlots ={}
from sklearn import metrics
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):

    #knnClf = KNeighborsClassifier()

    #knnClf.fit(X_train, ytrain[:,i])
    knnClf = KNeighborsClassifier()
    knnClf = OneVsRestClassifier(knnClf)
    knnClf.fit(X_train, ytrain[:,i])
    ax = plt.gca()
 # labelPlots[i]= plot_roc_curve(rfClassifier, X_test, ytest[:,i], name=␣
 ↪('label_'+str(i)), ax=ax, alpha=0.8)
  #labelPlots[i]= plot_roc_curve(rfClassifier, X_test, ytest[:,i], name=␣
 ↪('label_'+str(i)), ax=ax, alpha=0.8)
    labelPlots[i]= plot_roc_curve(knnClf, X_test, ytest[:,i], name=␣
 ↪(label_names[i]), ax=ax, alpha=0.8)


plt.show()
```



[77]:
```
from sklearn.metrics import classification_report

label_names = multilabel_binarizer.classes_

print(classification_report(ytest, knnPredictions,target_names=label_names))
```

```
              precision    recall  f1-score   support

      Action       0.33      0.00      0.00      1305
   Adventure       0.50      0.00      0.00       706
   Animation       1.00      0.01      0.01       378
      Comedy       0.32      0.29      0.31      2660
       Crime       1.00      0.00      0.00       842
 Documentary       1.00      0.00      0.01       812
       Drama       0.47      0.49      0.48      4017
      Family       0.50      0.00      0.00       540
     Fantasy       1.00      0.00      0.00       465
     Foreign       0.00      0.00      0.00       328
```

```
          History       0.00      0.00      0.00       291
           Horror       0.00      0.00      0.00       908
            Music       0.00      0.00      0.00       309
          Mystery       0.00      0.00      0.00       478
          Romance       1.00      0.00      0.00      1339
  Science Fiction       1.00      0.01      0.01       582
        TV Movie        0.00      0.00      0.00       152
         Thriller       0.00      0.00      0.00      1494
              War       0.00      0.00      0.00       271
          Western       0.00      0.00      0.00       207

        micro avg       0.42      0.15      0.22     18084
        macro avg       0.41      0.04      0.04     18084
     weighted avg       0.45      0.15      0.15     18084
      samples avg       0.32      0.18      0.22     18084
```

[ ]:

[78]:
```
import sklearn.metrics as metrics

metrics.hamming_loss(ytest, knnPredictions) #hyperparameter tuning
```

[78]: 0.11143521208599651

[79]:
```
#5 Decision Tress

from sklearn.tree import DecisionTreeClassifier
dtClassifier = DecisionTreeClassifier()
nbClassifier = OneVsRestClassifier(dtClassifier)
nbClassifier.fit(X_train, ytrain)
nbPreds = nbClassifier.predict(X_test)
f1_score(ytest, nbPreds, average="micro")
```

[79]: 0.403357466324382

[80]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = nbPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize='true')
```

```
for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[0.90287671 0.09712329]
 [0.66590038 0.33409962]]
Confusion matrix for label Adventure:
[[0.94568933 0.05431067]
 [0.82436261 0.17563739]]
Confusion matrix for label Animation:
[[0.97265103 0.02734897]
 [0.71428571 0.28571429]]
Confusion matrix for label Comedy:
[[0.76433978 0.23566022]
 [0.53834586 0.46165414]]
Confusion matrix for label Crime:
[[0.93700889 0.06299111]
 [0.7304038  0.2695962 ]]
Confusion matrix for label Documentary:
[[0.96137559 0.03862441]
 [0.50738916 0.49261084]]
Confusion matrix for label Drama:
[[0.61355711 0.38644289]
 [0.40179238 0.59820762]]
Confusion matrix for label Family:
[[0.95461872 0.04538128]
 [0.77037037 0.22962963]]
Confusion matrix for label Fantasy:
[[0.96584767 0.03415233]
 [0.8172043  0.1827957 ]]
Confusion matrix for label Foreign:
[[0.97704482 0.02295518]
 [0.92682927 0.07317073]]
Confusion matrix for label History:
[[0.97810921 0.02189079]
 [0.87972509 0.12027491]]
Confusion matrix for label Horror:
[[0.94283487 0.05716513]
 [0.5969163  0.4030837 ]]
Confusion matrix for label Music:
[[0.97902604 0.02097396]
 [0.67313916 0.32686084]]
Confusion matrix for label Mystery:
[[0.96087117 0.03912883]
 [0.84309623 0.15690377]]
Confusion matrix for label Romance:
```

```
[[0.89622901 0.10377099]
 [0.69305452 0.30694548]]
Confusion matrix for label Science Fiction:
[[0.96510034 0.03489966]
 [0.59793814 0.40206186]]
Confusion matrix for label TV Movie:
[[0.98450254 0.01549746]
 [0.96052632 0.03947368]]
Confusion matrix for label Thriller:
[[0.87512305 0.12487695]
 [0.69879518 0.30120482]]
Confusion matrix for label War:
[[0.97924166 0.02075834]
 [0.75645756 0.24354244]]
Confusion matrix for label Western:
[[0.98904501 0.01095499]
 [0.57004831 0.42995169]]
```

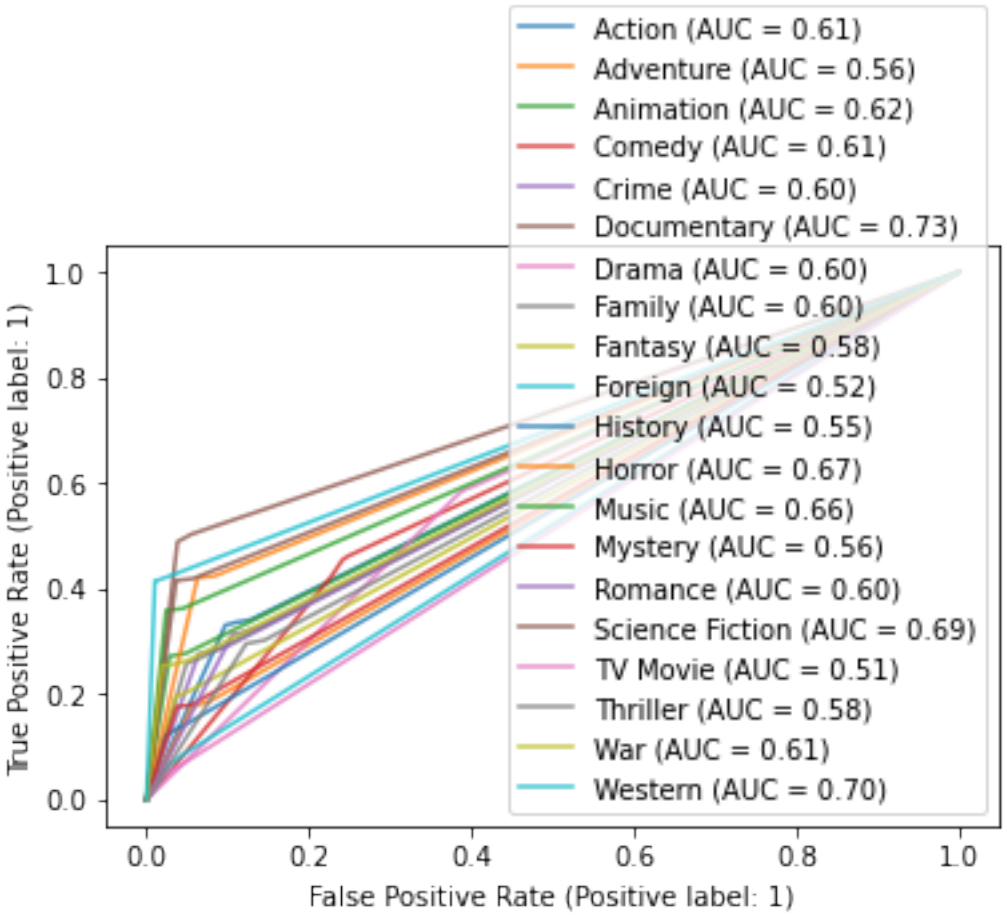
[81]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = nbPreds[:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label)

for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[6591  709]
 [ 869  436]]
Confusion matrix for label Adventure:
[[7470  429]
 [ 582  124]]
Confusion matrix for label Animation:
[[8002  225]
 [ 270  108]]
Confusion matrix for label Comedy:
[[4544 1401]
 [1432 1228]]
Confusion matrix for label Crime:
[[7274  489]
 [ 615  227]]
Confusion matrix for label Documentary:
```

```
[[7492  301]
 [ 412  400]]
Confusion matrix for label Drama:
[[2815 1773]
 [1614 2403]]
Confusion matrix for label Family:
[[7699  366]
 [ 416  124]]
Confusion matrix for label Fantasy:
[[7862  278]
 [ 380   85]]
Confusion matrix for label Foreign:
[[8087  190]
 [ 304   24]]
Confusion matrix for label History:
[[8132  182]
 [ 256   35]]
Confusion matrix for label Horror:
[[7257  440]
 [ 542  366]]
Confusion matrix for label Music:
[[8122  174]
 [ 208  101]]
Confusion matrix for label Mystery:
[[7809  318]
 [ 403   75]]
Confusion matrix for label Romance:
[[6512  754]
 [ 928  411]]
Confusion matrix for label Science Fiction:
[[7743  280]
 [ 348  234]]
Confusion matrix for label TV Movie:
[[8322  131]
 [ 146    6]]
Confusion matrix for label Thriller:
[[6223  888]
 [1044  450]]
Confusion matrix for label War:
[[8161  173]
 [ 205   66]]
Confusion matrix for label Western:
[[8306   92]
 [ 118   89]]
```

[82]:
```
labelPlots ={}
from sklearn import metrics
```

```
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):
    dtClassifier = DecisionTreeClassifier()
    nbClassifier = OneVsRestClassifier(dtClassifier)
    nbClassifier.fit(X_train,  ytrain[:,i])

    ax = plt.gca()

    labelPlots[i]= plot_roc_curve(nbClassifier, X_test, ytest[:,i], name=␣
 ↪(label_names[i]), ax=ax, alpha=0.8)


plt.show()
```



[84]:
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,nbPreds)
```

[84]: 0.08588030214991284

[85]:
```
import sklearn.metrics as metrics

metrics.hamming_loss(ytest, nbPreds) #hyperparameter tuning
```

[85]: 0.12019174898314933

[86]:
```
from sklearn.metrics import classification_report

label_names = multilabel_binarizer.classes_

print(classification_report(ytest, nbPreds,target_names=label_names))
```

```
                precision    recall  f1-score   support

        Action       0.38      0.33      0.36      1305
     Adventure       0.22      0.18      0.20       706
     Animation       0.32      0.29      0.30       378
        Comedy       0.47      0.46      0.46      2660
         Crime       0.32      0.27      0.29       842
   Documentary       0.57      0.49      0.53       812
         Drama       0.58      0.60      0.59      4017
        Family       0.25      0.23      0.24       540
       Fantasy       0.23      0.18      0.21       465
       Foreign       0.11      0.07      0.09       328
       History       0.16      0.12      0.14       291
        Horror       0.45      0.40      0.43       908
         Music       0.37      0.33      0.35       309
       Mystery       0.19      0.16      0.17       478
       Romance       0.35      0.31      0.33      1339
Science Fiction       0.46      0.40      0.43       582
       TV Movie       0.04      0.04      0.04       152
       Thriller       0.34      0.30      0.32      1494
            War       0.28      0.24      0.26       271
       Western       0.49      0.43      0.46       207

     micro avg       0.42      0.39      0.40     18084
     macro avg       0.33      0.29      0.31     18084
  weighted avg       0.41      0.39      0.40     18084
   samples avg       0.42      0.43      0.38     18084
```

[87]:
```
#6 Randomforest

classifier = OneVsRestClassifier(RandomForestClassifier())
classifier.fit(X_train, ytrain)
```

```
rdPreds = classifier.predict(X_test)
f1_score(ytest, rdPreds, average="micro")
```

[87]: 0.39137292011671004

[88]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rdPreds [:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label)

for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
[[7256   44]
 [1150  155]]
Confusion matrix for label Adventure:
[[7893    6]
 [ 690   16]]
Confusion matrix for label Animation:
[[8193   34]
 [ 315   63]]
Confusion matrix for label Comedy:
[[5630  315]
 [1886  774]]
Confusion matrix for label Crime:
[[7727   36]
 [ 806   36]]
Confusion matrix for label Documentary:
[[7761   32]
 [ 494  318]]
Confusion matrix for label Drama:
[[3083 1505]
 [1152 2865]]
Confusion matrix for label Family:
[[8051   14]
 [ 521   19]]
Confusion matrix for label Fantasy:
[[8138    2]
 [ 457    8]]
Confusion matrix for label Foreign:
```

```
[[8263   14]
 [ 325    3]]
Confusion matrix for label History:
[[8310    4]
 [ 289    2]]
Confusion matrix for label Horror:
[[7632   65]
 [ 709  199]]
Confusion matrix for label Music:
[[8288    8]
 [ 288   21]]
Confusion matrix for label Mystery:
[[8119    8]
 [ 469    9]]
Confusion matrix for label Romance:
[[7167   99]
 [1191  148]]
Confusion matrix for label Science Fiction:
[[7970   53]
 [ 403  179]]
Confusion matrix for label TV Movie:
[[8447    6]
 [ 151    1]]
Confusion matrix for label Thriller:
[[7047   64]
 [1385  109]]
Confusion matrix for label War:
[[8332    2]
 [ 267    4]]
Confusion matrix for label Western:
[[8394    4]
 [ 173   34]]
```

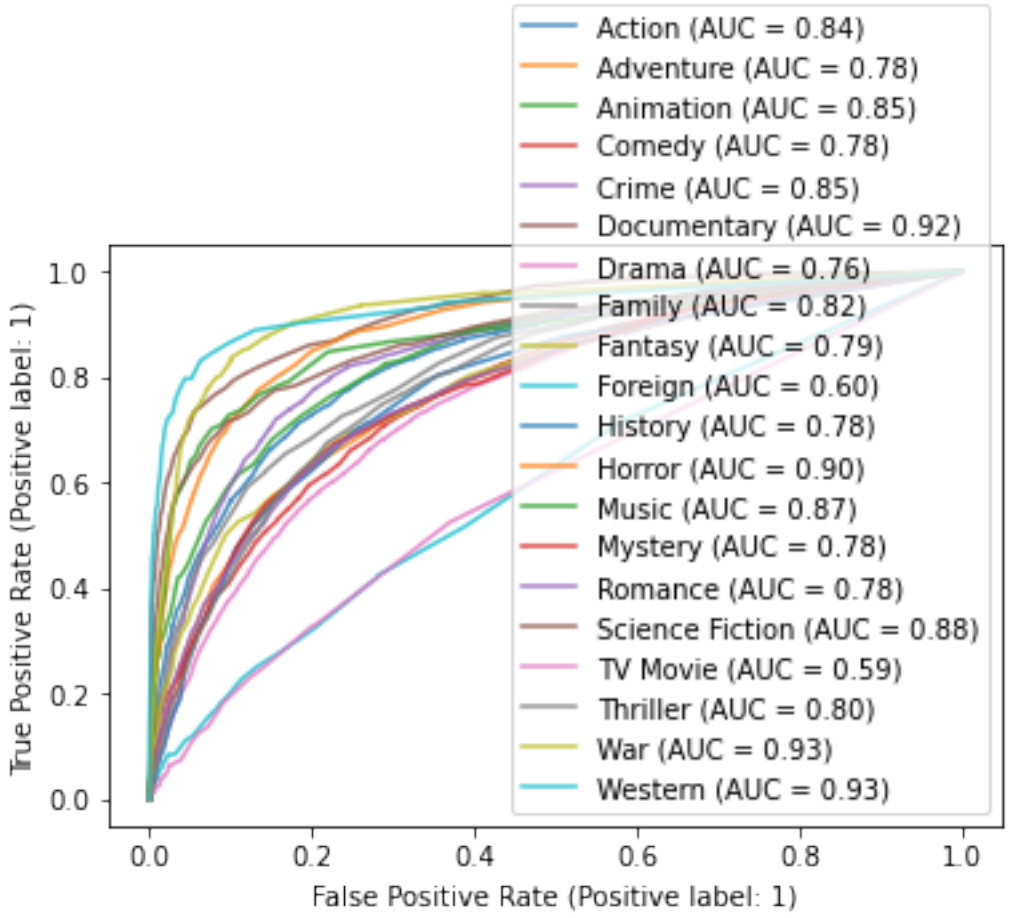
[89]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rdPreds [:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize='true')

for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

```
Confusion matrix for label Action:
```

```
[[0.9939726  0.0060274 ]
 [0.88122605 0.11877395]]
Confusion matrix for label Adventure:
[[9.99240410e-01 7.59589821e-04]
 [9.77337110e-01 2.26628895e-02]]
Confusion matrix for label Animation:
[[0.99586727 0.00413273]
 [0.83333333 0.16666667]]
Confusion matrix for label Comedy:
[[0.9470143  0.0529857 ]
 [0.70902256 0.29097744]]
Confusion matrix for label Crime:
[[0.99536262 0.00463738]
 [0.95724466 0.04275534]]
Confusion matrix for label Documentary:
[[0.99589375 0.00410625]
 [0.60837438 0.39162562]]
Confusion matrix for label Drama:
[[0.67197036 0.32802964]
 [0.28678118 0.71321882]]
Confusion matrix for label Family:
[[0.9982641  0.0017359 ]
 [0.96481481 0.03518519]]
Confusion matrix for label Fantasy:
[[9.99754300e-01 2.45700246e-04]
 [9.82795699e-01 1.72043011e-02]]
Confusion matrix for label Foreign:
[[0.99830857 0.00169143]
 [0.99085366 0.00914634]]
Confusion matrix for label History:
[[9.99518884e-01 4.81116190e-04]
 [9.93127148e-01 6.87285223e-03]]
Confusion matrix for label Horror:
[[0.99155515 0.00844485]
 [0.780837   0.219163  ]]
Confusion matrix for label Music:
[[9.99035680e-01 9.64320154e-04]
 [9.32038835e-01 6.79611650e-02]]
Confusion matrix for label Mystery:
[[9.99015627e-01 9.84373077e-04]
 [9.81171548e-01 1.88284519e-02]]
Confusion matrix for label Romance:
[[0.9863749  0.0136251 ]
 [0.88946975 0.11053025]]
Confusion matrix for label Science Fiction:
[[0.99339399 0.00660601]
 [0.69243986 0.30756014]]
Confusion matrix for label TV Movie:
```

```
[[9.99290193e-01 7.09807169e-04]
 [9.93421053e-01 6.57894737e-03]]
Confusion matrix for label Thriller:
[[0.99099986 0.00900014]
 [0.9270415  0.0729585 ]]
Confusion matrix for label War:
[[9.99760019e-01 2.39980802e-04]
 [9.85239852e-01 1.47601476e-02]]
Confusion matrix for label Western:
[[9.99523696e-01 4.76303882e-04]
 [8.35748792e-01 1.64251208e-01]]
```

[90]:
```
labelPlots ={}
from sklearn import metrics
from sklearn.metrics import plot_roc_curve
import matplotlib.pyplot as plt
labelPlots ={}
for i in range (len(label_names)):
    classifier = OneVsRestClassifier(RandomForestClassifier())
    classifier.fit(X_train,  ytrain[:,i])

    ax = plt.gca()

    labelPlots[i]= plot_roc_curve(classifier, X_test, ytest[:,i], name=␣
 ↪(label_names[i]), ax=ax, alpha=0.8)


plt.show()
```



[ ]:
```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import multilabel_confusion_matrix
conf_mat_dict={}
labels=multilabel_binarizer.classes_
for label_col in range(len(labels)):
    y_true_label = ytest[:, label_col]
    y_pred_label = rdPreds [:, label_col]
    conf_mat_dict[labels[label_col]] = confusion_matrix(y_pred=y_pred_label,␣
 ↪y_true=y_true_label,normalize='all')

for label, matrix in conf_mat_dict.items():
    print("Confusion matrix for label {}:".format(label))
    print(matrix)
```

[92]:
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,rdPreds)
```

[92]:
```
0.16257989540964554
```

[93]:
```
import sklearn.metrics as metrics

metrics.hamming_loss(ytest,rdPreds) #hyperparameter tuning
```

[93]: 0.08969203951191168

[94]:
```
from sklearn.metrics import classification_report

label_names = multilabel_binarizer.classes_

print(classification_report(ytest,rdPreds,target_names=label_names))
```

```
                precision    recall  f1-score   support

        Action       0.78      0.12      0.21      1305
     Adventure       0.73      0.02      0.04       706
     Animation       0.65      0.17      0.27       378
        Comedy       0.71      0.29      0.41      2660
         Crime       0.50      0.04      0.08       842
   Documentary       0.91      0.39      0.55       812
         Drama       0.66      0.71      0.68      4017
        Family       0.58      0.04      0.07       540
       Fantasy       0.80      0.02      0.03       465
       Foreign       0.18      0.01      0.02       328
       History       0.33      0.01      0.01       291
        Horror       0.75      0.22      0.34       908
         Music       0.72      0.07      0.12       309
       Mystery       0.53      0.02      0.04       478
       Romance       0.60      0.11      0.19      1339
Science Fiction       0.77      0.31      0.44       582
      TV Movie       0.14      0.01      0.01       152
      Thriller       0.63      0.07      0.13      1494
           War       0.67      0.01      0.03       271
       Western       0.89      0.16      0.28       207

     micro avg       0.68      0.27      0.39     18084
     macro avg       0.63      0.14      0.20     18084
  weighted avg       0.67      0.27      0.33     18084
   samples avg       0.50      0.34      0.38     18084
```