

In [1]:

```
from wordcloud import WordCloud
import re
import pandas as pd
import datetime as dt
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem.snowball import SnowballStemmer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.multiclass import OneVsRestClassifier
from sklearn.linear_model import SGDClassifier
from sklearn import metrics
from sklearn.metrics import f1_score, precision_score, recall_score
from sklearn import svm
from sklearn.linear_model import LogisticRegression
from skmultilearn.adapt import mlknn
from skmultilearn.problem_transform import ClassifierChain
from skmultilearn.problem_transform import BinaryRelevance
from skmultilearn.problem_transform import LabelPowerset
from sklearn.naive_bayes import GaussianNB
from datetime import datetime
```

In [2]:

```
import numpy as np
import matplotlib.pyplot as plt
import os as os
import seaborn as sns
import warnings
import csv
from PIL import Image
```

In [3]:

```
plt.style.use('fivethirtyeight')
plt.rcParams['figure.figsize'] = [10, 5]
warnings.filterwarnings("ignore", category=FutureWarning)
%config InlineBackend.figure_format = 'retina'
```

In [4]:

```
data = pd.read_csv('mpst_full_data.csv')
```

In [5]:

```
data.head()
```

Out[5]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source
0	tt0057603	I tre volti della paura	Note: this synopsis is for the original Italian...	cult, horror, gothic, murder, atmospheric	train	imdb
1	tt1733125	Dungeons & Dragons: The Book of Vile Darkness	Two thousand years ago, Nhagruul the Foul, a s...	violence	train	imdb
2	tt0033045	The Shop Around the Corner	Matuschek's, a gift store in Budapest, is the ...	romantic	test	imdb
3	tt0113862	Mr. Holland's Opus	Glenn Holland, not a morning person by anyone'...	inspiring, romantic, stupid, feel-good	train	imdb
4	tt0086250	Scarface	In May 1980, a Cuban man named Tony Montana (A...	cruelty, murder, dramatic, cult, violence, atm...	val	imdb

In [6]:

```
data.shape
```

```
Out[6]:  
(14828, 6)
```

```
In [7]:
```

```
data.drop_duplicates(subset={'imdb_id', 'title', 'plot_synopsis', 'synopsis_source'},  
                    keep='first', inplace=True)
```

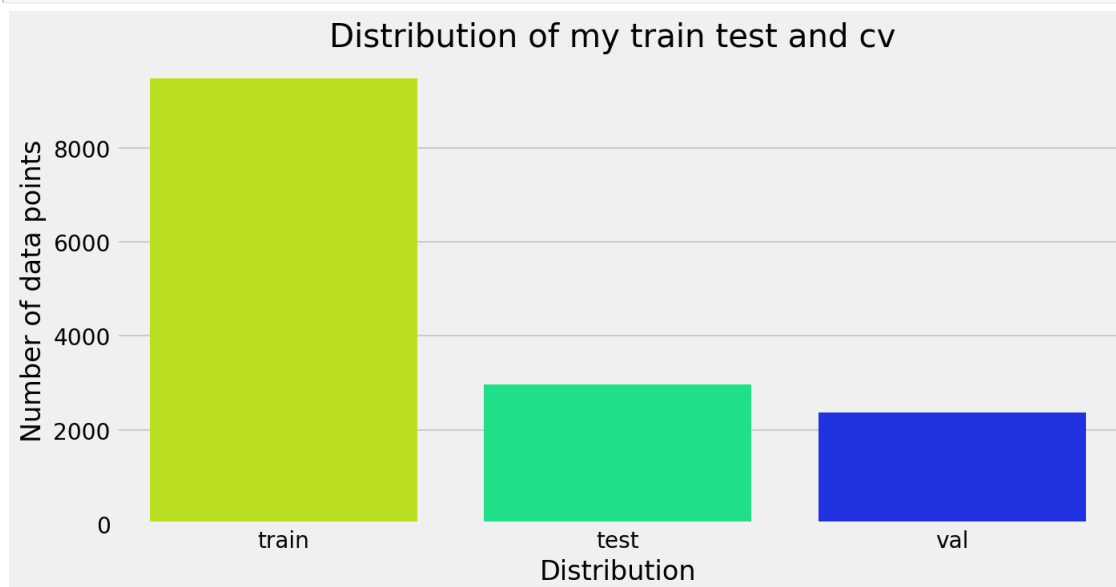
```
In [8]:
```

```
data.shape
```

```
Out[8]:  
(14828, 6)
```

```
In [9]:
```

```
sns.countplot(data['split'], palette='gist_rainbow')  
plt.title("Distribution of my train test and cv ")  
plt.xlabel("Distribution")  
plt.ylabel("Number of data points")  
plt.show()
```



```
In [10]:
```

```
data['split'].value_counts()
```

```
Out[10]:
```

```
train    9489  
test     2966  
val      2373  
Name: split, dtype: int64
```

```
In [11]:
```

```
data = data.sort_values(by=['split'], ascending=False)
```

```
In [ ]:
```

```
In [12]:
```

```
# from sklearn import preprocessing

# le = preprocessing.LabelEncoder()

# a = le.fit(data['split'])

# a.classes_

# a = le.transform(data['split'])
```

In [13]:

```
data["count"] = data["tags"].apply(lambda text: len(text.split(" ")))
data.head()
```

Out[13]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
7414	tt1710626	Zenobureido	=== Setting and characters ===\nThe setting of...	revenge	val	wikipedia	1
11142	tt0119642	Mean Guns	Vincent Moon (Ice-T) is the leader of a crime ...	revenge, action, murder, violence, flashback	val	wikipedia	5
11066	tt0183659	Pollock	The film begins showing the abstract expressio...	tragedy, romantic	val	wikipedia	2
11062	tt0250494	Legally Blonde	Fashion merchandising student Elle Woods is ta...	entertaining, cute, comedy, murder, stupid	val	wikipedia	5
11060	tt0085176	Antony and Cleopatra	Mark Antony – one of the triumvirs of the Roma...	tragedy	val	wikipedia	1

In []:

In [14]:

```
sorted_data = data.sort_values(['count'], ascending=False)# a new df whwre the biggest value of th
e counts coulun is first.
sorted_counts = sorted_data['count'].values #array of the counts in a descending order
```

In [15]:

```
sorted_data.head()
```

Out[15]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
3930	tt0457430	El laberinto del fauno	An unseen narrator (Pablo Adán's voiceover) te...	cruelty, murder, paranormal, allegory, cult, p...	val	imdb	27
1007	tt0110912	Pulp Fiction	Late one morning in the Hawthorne Grill, a res...	comedy, murder, stupid, cult, action, revenge,...	train	imdb	27
3610	tt0246578	Donnie Darko	Donnie Darko (Jake Gyllenhall) is a troubled t...	entertaining, dark, avant garde, gothic, murde...	train	imdb	26
3335	tt0401792	Sin City	The Customer is Always Right\n Basin City, The...	comedy, cruelty, murder, bleak, cult, action, ...	train	imdb	26
3711	tt0434409	V for Vendetta	"Remember, remember the fifth of November, the...	dark, suspenseful, neo noir, murder, avant gar...	train	imdb	24

In [16]:

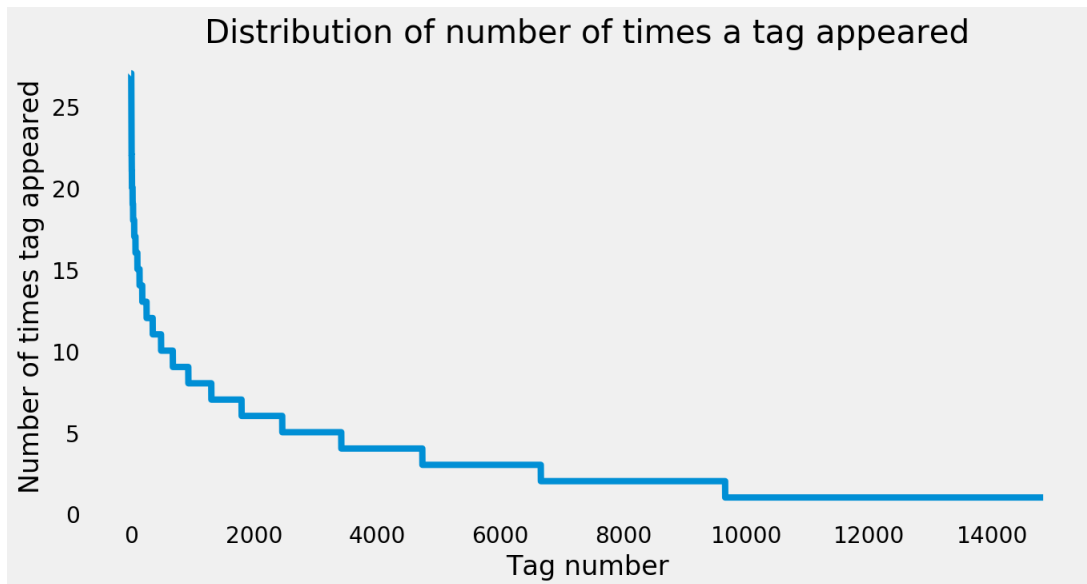
```
sorted_counts
```

Out[16]:

```
array([27, 27, 26, ..., 1, 1, 1], dtype=int64)
```

In [17]:

```
plt.plot(sorted_counts)
plt.title("Distribution of number of times a tag appeared")
plt.grid()
plt.xlabel("Tag number")
plt.ylabel("Number of times tag appeared")
plt.show()
```



In [18]:

```
# Importing & Initializing the "CountVectorizer" object, which
# is scikit-learn's bag of words tool.

# by default 'split()' will tokenize each tag using space.
vectorizer = CountVectorizer(tokenizer = lambda x: x.split())
# fit_transform() does two functions: First, it fits the model
# and learns the vocabulary; second, it transforms our training data
# into feature vectors. The input to fit_transform should be a list of strings.
tags = vectorizer.fit_transform(data['tags'])
```

In [19]:

```
# 'get_feature_name()' gives us the vocabulary.
tag = vectorizer.get_feature_names()
# Lets look at the tags we have.
print("Some of the tags we have :", tag[:10])
```

Some of the tags we have : ['absurd', 'absurd,', 'action', 'action,', 'adult', 'allegory', 'allegory,', 'alternate', 'anti', 'atmospheric']

In [20]:

```
# https://stackoverflow.com/questions/15115765/how-to-access-sparse-matrix-elements
# Lets now store the document term matrix in a dictionary.
freqs = tags.sum(axis=0).A1
result = dict(zip(tag, freqs))
```

In [21]:

```
result.keys()
```

Out[21]:

```
dict_keys(['absurd', 'absurd,', 'action', 'action,', 'adult', 'allegory', 'allegory,', 'alternate', 'anti', 'atmospheric', 'atmospheric,', 'autobiographical', 'autobiographical,', 'avant', 'blaxploitation', 'blaxploitation,', 'bleak', 'bleak,', 'boring', 'boring,', 'brainwashing', 'brainwashing,', 'christian', 'claustrophobic', 'claustrophobic,', 'clever',
```

```

brainwashing, 'brainwashing,', 'christian', 'claustrophobic', 'claustrophobic,', 'clever',
'clever,', 'comedy', 'comedy,', 'comic', 'comic,', 'cruelty', 'cruelty,', 'cult', 'cult,', 'cute',
'cute,', 'dark', 'dark,', 'depressing', 'depressing,', 'dramatic', 'dramatic,', 'entertaining', 'e
ntertaining,', 'evil', 'evil,', 'fantasy', 'fantasy,', 'feel-good', 'feel-good,', 'fiction',
'fiction,', 'film', 'film,', 'flashback', 'flashback,', 'garde', 'garde,', 'good', 'gothic',
'gothic,', 'grindhouse', 'haunting', 'haunting,', 'historical', 'historical,', 'history',
'history,', 'home', 'horror', 'horror,', 'humor', 'humor,', 'insanity', 'insanity,', 'inspiring',
'inspiring,', 'intrigue', 'intrigue,', 'magical', 'melodrama', 'melodrama,', 'movie', 'movie,', 'm
urder', 'murder,', 'mystery', 'mystery,', 'neo', 'noir', 'noir,', 'non', 'paranormal',
'paranormal,', 'philosophical', 'philosophical,', 'plot', 'pornographic', 'pornographic,',
'prank', 'prank,', 'psychedelic', 'psychedelic,', 'psychological', 'psychological,', 'queer', 'que
er,', 'realism', 'realism,', 'reality', 'reality,', 'revenge', 'revenge,', 'romantic',
'romantic,', 'sadist', 'sadist,', 'satire', 'satire,', 'sci-fi', 'sci-fi,', 'sentimental',
'sentimental,', 'storytelling', 'storytelling,', 'stupid', 'stupid,', 'suicidal', 'suicidal,', 'su
spenseful', 'suspenseful,', 'thought-provoking', 'thought-provoking,', 'tragedy', 'tragedy,',
'twist', 'twist,', 'versus', 'violence', 'violence,', 'war', 'war,', 'western', 'western,',
'whimsical', 'whimsical,']))

```

In [22]:

```
result.values()
```

Out[22]:

```

dict_values([11, 259, 111, 553, 128, 18, 121, 307, 118, 80, 316, 22, 22, 220, 57, 17, 25, 186, 84,
441, 22, 85, 42, 23, 61, 19, 68, 341, 1646, 74, 40, 30, 412, 284, 2363, 24, 173, 37, 368, 54, 151,
30, 382, 417, 332, 105, 770, 217, 347, 36, 40, 146, 34, 42, 66, 1938, 999, 22, 198, 875, 88, 353,
66, 77, 74, 270, 145, 57, 45, 153, 71, 415, 183, 639, 50, 586, 28, 90, 36, 132, 54, 310, 147, 150,
3, 2471, 3311, 47, 472, 745, 71, 674, 37, 107, 439, 67, 161, 205, 40, 123, 220, 35, 582, 1315, 31,
258, 74, 24, 56, 202, 36, 169, 657, 1811, 1416, 1490, 518, 134, 326, 490, 292, 17, 187, 46, 232, 1
32, 68, 122, 15, 39, 235, 851, 2, 118, 102, 495, 66, 139, 875, 1424, 3002, 26, 92, 12, 61, 2, 77])

```

In [23]:

```
result
```

Out[23]:

```

{'absurd': 11,
'absurd,': 259,
'action': 111,
'action,': 553,
'adult': 128,
'allegory': 18,
'allegory,': 121,
'alternate': 307,
'anti': 118,
'atmospheric': 80,
'atmospheric,': 316,
'autobiographical': 22,
'autobiographical,': 22,
'avant': 220,
'blaxploitation': 57,
'blaxploitation,': 17,
'bleak': 25,
'bleak,': 186,
'boring': 84,
'boring,': 441,
'brainwashing': 22,
'brainwashing,': 85,
'christian': 42,
'claustrophobic': 23,
'claustrophobic,': 61,
'clever': 19,
'clever,': 68,
'comedy': 341,
'comedy,': 1646,
'comic': 74,
'comic,': 40,
'cruelty': 30,
'cruelty,': 412,
'cult': 284,
'cult,': 2363,
'cute': 24,
'cute,': 173,

```

'dark': 37,
'dark,': 368,
'depressing': 54,
'depressing,': 151,
'dramatic': 30,
'dramatic,': 382,
'entertaining': 417,
'entertaining,': 332,
'evil': 105,
'evil,': 770,
'fantasy': 217,
'fantasy,': 347,
'feel-good': 36,
'feel-good,': 40,
'fiction': 146,
'fiction,': 34,
'film': 42,
'film,': 66,
'flashback': 1938,
'flashback,': 999,
'garde': 22,
'garde,': 198,
'good': 875,
'gothic': 88,
'gothic,': 353,
'grindhouse': 66,
'haunting': 77,
'haunting,': 74,
'historical': 270,
'historical,': 145,
'history': 57,
'history,': 45,
'home': 153,
'horror': 71,
'horror,': 415,
'humor': 183,
'humor,': 639,
'insanity': 50,
'insanity,': 586,
'inspiring': 28,
'inspiring,': 90,
'intrigue': 36,
'intrigue,': 132,
'magical': 54,
'melodrama': 310,
'melodrama,': 147,
'movie': 150,
'movie,': 3,
'murder': 2471,
'murder,': 3311,
'mystery': 47,
'mystery,': 472,
'neo': 745,
'noir': 71,
'noir,': 674,
'non': 37,
'paranormal': 107,
'paranormal,': 439,
'philosophical': 67,
'philosophical,': 161,
'plot': 205,
'pornographic': 40,
'pornographic,': 123,
'prank': 220,
'prank,': 35,
'psychedelic': 582,
'psychedelic,': 1315,
'psychological': 31,
'psychological,': 258,
'queer': 74,
'queer,': 24,
'realism': 56,
'realism,': 202,
'reality': 36,
'reality,': 169,
'revenge': 657,
'revenge,': 1811,

```

'romantic': 1416,
'romantic,': 1490,
'sadist': 518,
'sadist,': 134,
'satire': 326,
'satire,': 490,
'sci-fi': 292,
'sci-fi,': 17,
'sentimental': 187,
'sentimental,': 46,
'storytelling': 232,
'storytelling,': 132,
'stupid': 68,
'stupid,': 122,
'suicidal': 15,
'suicidal,': 39,
'suspenseful': 235,
'suspenseful,': 851,
'thought-provoking': 2,
'thought-provoking,': 118,
'tragedy': 102,
'tragedy,': 495,
'twist': 66,
'twist,': 139,
'versus': 875,
'violence': 1424,
'violence,': 3002,
'war': 26,
'war,': 92,
'western': 12,
'western,': 61,
'whimsical': 2,
'whimsical,': 77}

```

In [24]:

```

#Saving this dictionary to csv files.
if not os.path.isfile('tag_counts.csv'):
    with open('tag_counts', 'w') as csv_file:
        writer = csv.writer(csv_file)
        for key, value in result.items():
            writer.writerow([key, value])
tag_df = pd.read_csv("tag_counts", names=['Tags', 'Counts'])
tag_df.head()

```

Out[24]:

	Tags	Counts
0	absurd	11
1	absurd,	259
2	action	111
3	action,	553
4	adult	128

In [25]:

```

tag_df_sorted = tag_df.sort_values(['Counts'], ascending=False)
tag_counts = tag_df_sorted['Counts'].values

```

In [26]:

```

tag_df_sorted['Counts'].head()

```

Out[26]:

```

86      3311
140     3002
85      2471
34      2363
55      1938

```

```
__ _1000
Name: Counts, dtype: int64
```

In [27]:

```
print( "Maximum number of tags per question: %d"%max(sorted_counts))
print( "Minimum number of tags per question: %d"%min(sorted_counts))
print( "Avg. number of tags per question: %f"% ((sum(sorted_counts)*1.0)/len(sorted_counts)))
```

```
Maximum number of tags per question: 27
Minimum number of tags per question: 1
Avg. number of tags per question: 3.248854
```

In [28]:

```
unique_genere = data['tags'].value_counts()
print('Number of Unique Genere :', unique_genere.shape[0])
print(unique_genere.head(27))
```

```
Number of Unique Genere : 5604
murder                1004
romantic              731
violence              584
psychedelic          437
flashback            332
violence, murder     262
revenge              232
cult                 226
comedy               157
revenge, murder     150
fantasy              131
murder, flashback    124
satire               113
melodrama            106
paranormal           106
romantic, flashback  99
good versus evil     95
romantic, murder     95
tragedy              77
revenge, murder, violence 76
cult, murder         76
cult, violence       73
sci-fi               60
humor                56
revenge, violence   56
prank                50
suspenseful         50
Name: tags, dtype: int64
```

In [29]:

```
# Plotting word cloud
start = datetime.now()

# Lets first convert the 'result' dictionary to 'list of tuples'
tup = dict(result.items())
#Initializing WordCloud using frequencies of tags.
wordcloud = WordCloud(    background_color='white',
                        width=1600,
                        height=800,
                        ).generate_from_frequencies(tup)

fig = plt.figure(figsize=(30,20))
plt.imshow(wordcloud)
plt.axis('off')
plt.tight_layout(pad=0)
fig.savefig("tag.png")
plt.show()
print("Time taken to run this cell :", datetime.now() - start)
```



Observations:

1. Majority of the most frequent Genres are related with murders, violence, revenge etc.
2. Murder, sadly is the top most genre.
3. Violence, Cult, Comedy and Romantic are among the top most frequent Genres.

In []:

In []:

In [31]:

```
data['plot_synopsis'][4]
```

Out[31]:

'In May 1980, a Cuban man named Tony Montana (Al Pacino) claims asylum, in Florida, USA, and is in search of the "American Dream" after departing Cuba in the Mariel boatlift of 1980. When questioned by three tough-talking INS officials, they notice a tattoo on Tony's left arm of a black heart with a pitchfork through it, which identifies him as a hitman, and detain him in a camp called 'Freedomtown' with other Cubans, including Tony's best friend and former Cuban Army buddy Manolo "Manny Ray" Ribiera (Steven Bauer), under the local I-95 expressway while the government evaluates their visa petitions. After 30 days of governmental dithering and camp rumors, Manny receives an offer from the Cuban Mafia which he quickly relays to Tony. If they kill Emilio Rebenga (Roberto Contreras) a former aide to Fidel Castro who is now detained in Freedomtown, they will receive green cards. Tony agrees, and kills Rebenga during a riot at Freedomtown. The murder of Rebenga was requested by Frank López, a wealthy, politically astute man who deals cars and trades in cocaine, as Rebenga had tortured López's brother to death while still in Cuba many years earlier. After getting their Green Cards, Tony Montana and Manny Ray find work as dishwashers in a corner sandwich/taco shop. Some weeks later, a López henchman and underboss, Omar Suárez (F. Murray Abraham), the man who contacted Manny for the Rebenga hit job, offers Tony and Manny a low-risk job of unloading marijuana from a boat from Mexico to arrive in Miami the following night for \$500 each. Tony insults Suárez by turning down the job over the little money they will receive, and demands at least \$1,000 for the work. After an altercation, Suárez sets Tony up on another job to purchase two kilograms of cocaine worth around \$25,000 a piece from a Colombian dealer, named Hector The Toad, a medium to high-risk job for which Tony and Manny will receive \$5,000 for their work. That weekend, Tony, Manny, and two other Marielitos in his crew whom they met in Freedomtown, Angel Fernández (Pepe Serna), and Chi Chi (Ángel Salazar) then set out to meet "Hector the Toad" (Al Israel) at a seedy motel on the boulevard in Miami Beach. While Manny and Chi Chi wait in the car on the street, Tony and Angel go up to the hotel room to meet with Hector. The meeting does not go smoothly, as Tony grows irritated with Hector, who is slow to give him the cocaine in exchange for money. Suddenly, Tony and Angel are double-crossed by the Colombian. It becomes apparent that Hector does not intend to sell Tony the cocaine he has; he only wants to steal the money Tony has been given to purchase the product. To convince Tony to give over the cash, Hector dismembers Angel in a shower stall with a chainsaw. After Angel is dead, Tony, about to suffer the same fate, is saved by Chi Chi and Manny who arrive in the nick of time to gun down Hector's henchmen. Manny receives a minor bullet wound in his shoulder when his Uzi sub-machine gun jams. Hector escapes but Tony vengefully confronts him in the street and shoots him dead in the middle of the crowded Ocean Drive (the now famous Miami South Beach boulevard). Tony and his crew then get away with both the cocaine and the money before the police arrive. The following night, Tony and Manny meet López (Robert Loggia) at his house for the first time where Tony impresses Lopez with not only the return of his cash but with a gift of the cocaine, a prize from the botched rip off. Frank immediately hires Tony and his crew into his criminal hierarchy, a representative of a Cuban mafia. But during this initial get together Tony also meets Lopez's lady, the blond and beautiful Elvira Hancock (Michelle Pfeiffer), who will eventually become the source of tension between the two men. Taking Tony and Manny out to a local nightclub, called The Babylon Club where Frank frequently attends, Tony and Manny see first-hand the high standard of living they have come to acquire. Though Frank actually warns against these excesses, Tony is seduced by them regardless. Thus, Tony Montana begins his rise through the ranks of the Miami cocaine underworld. Three months later, Tony Montana attempts to make amends by meeting with his estranged family. It is implied that Tony's father, a former U.S. Navy sailor, abandoned the family when Tony was little. Since then, his mother (Miriam Colon) and younger 19-year-old sister Gina (Mary Elizabeth Mastrantonio) have been living in Miami. Tony shows up at his mother's and Gina's house one evening, fashionably dressed, and offers them \$1,000 in cash for financial support. Gina is overjoyed to see her older brother whom they have not seen for five years. But Tony's mother has only scorn for him since he turned his back on them many years ago for the quick and easy life of crime back in Cuba, and wants nothing to do with Tony, and she is too full of pride to accept his money despite being financially stricken. But Gina, who idolizes her brother, follows him outside where he slips her the money secretly. Gina tells Tony that she wants in on the flashy life that he has going for him. Tony's love for Gina is clearly genuine for she's the only person that he trusts, and is also very protective of her. Afterwards, Manny makes a comment to Tony about how attractive Gina is, but Tony angrily warns him to avoid co

urting her. Several months later, Tony is sent to Bolivia to help Omar set up a new distribution deal with Bolivian kingpin Alejandro Sosa (Paul Shenar), since Frank is having legal troubles that preclude him from leaving the country. Though Tony was supposed to let Omar do all the talking, Omar proves to be a poor negotiator, prompting Tony to step in and save the deal. They seem to negotiate a deal that, on the surface seems favorable to both sides, but Omar insists that Frank would not approve. Sosa then sides with Omar on this and suggests that Omar use his phone to call Frank. A few minutes later, Sosa hands Tony binoculars, and he sees two menacing assassins, Alberto the Shadow (Mark Margolis) and the Skull (Geno Silva), execute Omar by hanging him by the neck from an airborne helicopter. Sosa reveals that Alberto recognized Omar as once being an informant for the police several years ago, and he has a zero tolerance for disloyalty. Tony insists that he never goes back on his word, and that he never trusted Omar. Sosa agrees to bring Tony on board with him as his North American distributor of cocaine and other drugs. But upon their agreement, Sosa sternly warns Tony never to betray or double-cross him in any way. Upon his return to Florida, Tony is chewed out by Frank for overstepping his authority as well as hearing about Omar's death. Tony explains to Frank that for a price of \$18 million to pay Sosa for the manufacturing and transportation costs, they will receive 2,000 kilograms of cocaine from Bolivia for nationwide sale and distribution which will earn them \$75 million over a period of one year. Frank is worried because he does not have the many millions to pay Sosa for the cocaine, but Tony says that he is in tight with Sosa and he has established a "credit line" with him as well as work out a payment plan where they will pay Sosa \$5 million up front and the rest in monthly installments. Plus, in case Frank comes up short a few million, Tony will earn the money needed through his own street contacts. Frank angrily tells Tony that he did not negotiate a good deal, for Sosa merely tricked him into thinking he did. Tony replies that it's time for them to "think big," and to expand the cartel for nationwide distribution of cocaine. With them as the main North American distributors and wholesalers of the Sosa cartel, they will make millions and become the biggest cartel in the continent. Frank warns Tony that Sosa cannot be trusted and that he will sooner or later turn against them for any slight deviation or compromise of his business. Frank orders Tony to stall his deal with Sosa for the time being. Frank then promptly tells Tony that ambitious drug dealers, such as himself, who want too much and crave money, power and attention, do not last long in the business. Tony leaves shrugging with indifference and strikes out on his own. (Note: Frank López's two warnings about Tony's greed and Sosa's violence will be later proven true.) After this incident, Tony then seeks out Elvira to whom he makes an unexpected marriage proposal. She is shaken by this, but agrees to think about it. Frank López is none too happy when he hears about this and decides to take out Tony. At the Babylon nightclub that evening, Tony is approached and shaken down by a Miami police detective, named Mel Bernstein (Harris Yulin). He proposes to "tax" Tony on his transactions in return for police protection and information. Tony is distracted by the sight of Gina dancing with a local drug dealer. He follows the two to a restroom stall where he berates Gina for her promiscuous conduct. He asks Manny to take her home. On the way Gina admits she is attracted to Manny. Manny wards her off, mindful of Tony's extreme protectiveness. Back at the nightclub, Tony is attacked by two gunmen but manages to escape, killing them both despite being wounded by a gunshot to his left shoulder. Suspecting Frank sent Bernstein and the hitmen, Tony asks one of his bodyguards, Nick The Pig, to call Frank after Tony arrives at Frank's office at 3:00 a.m. that very night and inform him the hit failed. Tony, Manny and Chi Chi visit Frank at his car dealership back office, who is with Det. Bernstein. Nick calls Frank, who confirms his involvement by playing the call off as Elvira telling him she'll be late home. When it becomes apparent that Bernstein (who is armed) will not help him, Frank begs for Tony's forgiveness, saying that he can have Elvira and ten million dollars in exchange for sparing his life. Tony will have none of it, and Manny coldly executes Frank. Bernstein insists that he could be a valuable ally for Tony, but Tony disagrees, and kills him too. His problems apparently solved, Tony begins a profitable relationship with Sosa, marries Elvira, buys a new mansion complete with surveillance cameras and numerous luxury items, and Tony even sets Gina up in business with her own beauty salon. Manny and Gina soon begin a romantic relationship, but they keep it secret from Tony who had firmly stated to Gina that he does not want her dating anybody. As Tony's business grows, so does his cocaine addiction and paranoia, and he begins to spiral out of control... the beginning of the end. His wife, who becomes further addicted to cocaine, becomes bored and emotionally distant. Tony's banker informs him that laundering the increasing flow of drug money has become increasingly difficult, so he will be charging higher fees, up to 10%. A Jewish mob boss, Mel Seidelbaum (Ted Beniades), contacts Manny, offering his assistance. However, as they are cleaning out the money, Seidelbaum reveals himself to be an undercover cop and arrests Tony. After posting a \$5 million in bail, Tony's corrupt lawyer, Sheffield, tells him that although he may get him cleared of the corruption and money laundering charges, Tony will probably have to serve at least three years in prison for tax evasion. Manny suggests that he take it, as the American prison system is nowhere near as harsh as its Cuban counterpart, and the right legal loopholes could trim the sentence down to six months. However, the strung-out Tony yells that he would rather die than spend a single day in jail. After hearing about Tony's arrest, Sosa, not wanting to lose his main distributor, steps in to intervene by offering Tony a way out of going to prison. He calls Tony back to Bolivia where he introduces him to his cocaine "board of directors" a group that includes a sugar baron, Bolivia's military chief, and a mysterious American named Charles Goodson (Gregg Henry). We assume he is a corrupt CIA officer because Sosa guarantees that the IRS will not be able to send Tony to jail. But this help comes at a price. A Bolivian journalist is attempting to expose the ongoing corruption in the Bolivian government involvement in drug trafficking, and his crusade is beginning to hurt Sosa and his partners. Sosa will be sending Alberto to New York assassinate the journalist, but he needs Tony and his crew to provide some extra muscle. Tony is clearly disturbed by the assassination since it is against his custom to kill a man whom he sees as a civilian, plus Tony has never killed anybody who didn't wrong him personally. But seeing no other options, Tony reluctantly agrees to help Sosa with the hit. In the meantime, Tony's marriage with Elvira finally ends when after a bitter altercation at a local restaurant, she finally expresses her contempt for him and the lives he had

led her on, and walks out of the restaurant, and out of his life. Tony, punch-drunk on cocaine, tells the restaurant's other patrons that his existence is necessary since society needs a man like him to call a criminal. Tony also informs Manny to look after things while he travels to New York on business (but he doesn't tell him about the Sosa assassination deal.) Tony, with his henchmen, Chi Chi and Reuben, and Alberto travel to New York City and Alberto places a bomb under the journalist's car in with the intention of detonating it outside the UN building before the journalist addresses the General Assembly and exposes Sosa's cartel. But Tony has second thoughts when the journalist unexpectedly picks up his wife and children. Tony, saying that the team was only supposed to kill only the journalist, shoots Alberto to prevent the journalist's family from being killed. When authorities later discover the unexploded bomb underneath the journalist's car, they realize that an execution had been planned and increase the amount of security protecting the journalist. Sosa is now the primary suspect and Sosa vows to get even with Tony. Returning to Miami, Tony discovers that Gina and Manny (who opposed the trip to New York) have disappeared. Tony has long harbored an apparent unnatural obsession for his sister and is overly protective of her for reasons that he may not understand himself. Tony visits his mother again where she angrily tells him about Gina's descent into an immoral life and accuses him of corrupting her with his flashy lifestyle. After getting Gina's home address from Mrs. Montana, who doesn't know who else lives there, Tony goes to the house in nearby Palm Grove. Much to Tony's surprise, Manny unexpectedly opens the door in his bathrobe. Tony then sees Gina in a night gown at the top of the stairs. Enraged that another man has obviously slept with his sister, Tony shoots Manny dead. Hysterical, Gina reveals to Tony that they had just been married and were going to surprise him. Tony, riddled with guilt, has Gina taken back to his mansion. In revenge for Tony's failure to kill the journalist, who has now exposed Sosa and his partners to the world as drug lords, Sosa sends a Latino mercenary hit squad (the size of a large platoon), to Tony's mansion to kill him that very evening. Sitting at his desk snorting from an enormous pile of cocaine, Tony realizes and regrets what he has done to his best friend. When Tony is contemplating his actions, Sosa's mercenaries breach the main gate at Tony's estate and quietly begin to kill all the guards around the mansion. At the same time, a distraught Gina, wearing only an unbuttoned sleep shirt and armed with a revolver, enters Tony's office to confront him with the truth about his feelings for her. She now realizes that Tony loves her in an unnatural way and demands, at gunpoint, that he make love to her. She begins to shoot at him while demanding he take her. A Sosa assassin hiding on the balcony, thinking Gina is shooting at him, leaps in and riddles her with bullets. An enraged Tony throws the man off the balcony and kills him with his sub-machine gun creating a storm of chaos at the mansion. At this point, the mercenaries, robbed of the element of surprise by the gunshots, swarm in to attack Tony's mansion from all directions. As all his men are being killed, Tony, still delirious from the cocaine, leans over Gina's dead body begging for forgiveness, at the same time the mercs break into the mansion, Chi Chi opens fire with an Uzi as he falls back and ends up banging on the door to Tony's office (it has been locked from the inside by Gina who was planning to kill Tony). Unfortunately, Tony does not seem to hear him. Chi Chi is shot in the back and Tony sees it on the security cameras. As the hit men prepare to storm his office, Tony finally snaps out of his drug-induced state, arms himself with an AR-15 assault rifle with an under-mounted M203 grenade launcher and blows down the door. A huge climatic gun battle begins as Tony takes position atop the grand staircase and guns down dozens of Sosa's men who try to storm the balcony. Tony is hit a number of times by return fire, but he keeps shooting. With most of Sosa's men dead, Tony, strung-out on drugs, defiantly yells out at the assassins, not realizing that the Skull has sneaked into the room behind him. The Skull shoots Tony in the back with a 12-Gauge shotgun. Tony falls off the balcony and into a reflecting pool at the base of the grand staircase. In the final shot, as the Skull and the few surviving assassins look on, Tony Montana lies dead... face down in the reflecting pool which is located below a large brass globe that says: THE WORLD IS YOURS.'

In []:

In []:

In []:

In [32]:

```
data['title'][4]
```

Out[32]:

```
'Scarface'
```

In [33]:

```
import nltk
from nltk import PorterStemmer
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\ankan\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

Out[33]:

True

In [34]:

```
data['New'] = data['title'] + data['plot_synopsis']
```

In []:

In []:

In []:

In []:

In []:

In []:

In [34]:

```
i=0;
for sent in data['plot_synopsis'].values:
    if (len(re.findall('<.*?>', sent))):
        print(i)
        print(sent)
        break;
    i += 1;
```

86

It's Complicated:

Jake (Alec Baldwin), a divorced lawyer, married to a younger wife (Lake Bell), has a son, Pedro (E mjay Anthony) still feels affectionate to ex-wife, Jane (Meryl Streep), who is owner of a bakery s hop.The films starts off with a nice view of a neighborhood and the view of the ocean. We are then bought to a house party. Jake (Alec) proposes a toast to his friends' anniversary (Ted & Ally). Th e four talk momentarily about their son's graduation, and then Alec's young wife (he divorced Jane , played by Meryl Streep, 10 years ago) walks towards them. Jane seemed shock the way Agness approaches the group and the lack of clothing she had on. Jane leaves as Agness walk towards Jake; they exchange a few words of meeting up in New York at a hotel. Jane and Ally, to much dismay talk ed about Jane's divorce as they walk towards the door and how Ally didn't seem to feel comfortable around the sight of Agness. As Jane closes the gate, Jake turns around just in time to take a glimpse of Jane, letting us know that he still has feelings for her, although they've been divorced ten years.We then follow up to Jane's middle child, Gabby (Zoe Kazan) who is moving out o f the house. The oldest daughter, Lauren (Caitlin Fitzgerald), and her fiancée Harley (John Krasin ski) assist Gabby as Jane arrives home. Although reluctant to let Gabby move out of the house, she agrees. Lauren teases her mother, saying she must be lonely and afraid to sleep alone. Jane denies

that, saying "One of you is always moving out". Harley: "They just grow up so freaking fast." Lauren: "Mom, you should get a dog." The couple leaves and Jane walks into her kitchen, and reminisces about the many memories with all her children at home with her, now all grown up. The following day, at her bakery, Jane's friend Peter brings along another architect (Steve Martin) to show blueprints to remodel parts of her house. They talk briefly before she remembers she has an appointment with a doctor. As she leaves the office, the elevator stops on the Fertility Center floor and Jake walks in. That night, Jane and her friends enjoy laughs over dinner. They recommend Jane should find a date, rather be semi-miserable at home. The next day, at New York, Jane's family all meet up with Luke, the youngest son, one last night before graduation. They book a hotel (mentioned earlier at the house party). The kids leave to a party, leaving Jane in the hotel. Again, Jane gets in an elevator, this time with a random couple hugging and kissing. She decides to go to the bar and grab a drink. Coincidentally, Jake is there and they talk. He tells her that he is "flying it low" because Pedro has a stomach flu, so he is free for the night. Jane asks when was the last time they were together, he says 1999. Next thing they're having a good time on the dance floor and they go back to the hotel room. They do their thing and Jane felt horrible; drunk and half realizing she had just slept with her ex-husband. Jake says "This is amazing, we just had a great time" as Jane is leaning on the toilet, throwing up and regretting it. During the graduation, the whole family is together and the kids were excited to see their father Jake, despite the divorce. At lunch, the family is having dinner and Jake flirts with Jane, and proposes a toast to Jane, however the family thought it was supposed to be to Luke's academic success. They take it and drink. Abby properly finishes the toast to Luke. Adam (Steve Martin) drives his Audi to Jane's house to talk about the project of modeling her home. In his car, he is listening to a CD, lecturing on divorces. He sees Jane jogging, calls her and she apologizes for almost forgetting the appointment. She gets in the car and hears the audio and Adam, embarrassingly tries to turn it off. As he finishes talking about the blueprint, Jake arrives in his Porsche. They introduce each other and Adam leaves. Jake goes inside for "coffee", which actually turns out to be more flirting and another sex affair. Once again, Jake is happy and Jane is dismayed for being a "mistress" (all in a fun, comedic way). He claims fate brought them together again and that he is a walking cliché, father to a kid in kindergarten, but they're all grown up. That night, Jane's three closest friends talk over wine. They all notice Jane is acting differently, more energetic. She tells them she's been having sex and having an affair with Agness's husband, Jake. The four ladies all get excited and accepted the fact that Jane is having an affair with Jake. They all hate Agness, and one observes "Karma is a bitch"). Her energetic passion moves along to the bakery. She tells her bakers to make the breakfast menu more lively. She then receives a phone call from Adam, whose waiting at her house for another appointment. He set up a chunk of land with different color tapes and a ladder acting as stairs to the second floor. All this is just a preview of how Adam is going to fix the house. She receives a call from Jake, but doesn't answer it. That night at Jake's house, we see him leaving a message for Jane to pick up her phone. He is in the bathroom with the water running, to prevent Agness from hearing. Pedro, being a nosy brat, asks Jake why he was "showering" with his clothes on. Agness comes in and Pedro snitches, but Jake leaves without getting caught. He drives over to Jane's house. Meanwhile, Adam is having dinner at her house. They talk and Jake spies on them through the kitchen window, then accidentally falls. As Adam leaves, Jake knocks on the door and says, "Oh Em Gee <omg>, I thought he'd never leave." He grabs her and flirts. Jane says "it's Nine o'clock, shouldn't your wife be concerned?" He reaches for her delicious dish and goes to the bathroom. He is sitting next to the bathtub eating her lavender ice cream, while she bathes. He takes out marijuana and asks if she remembers the old days, when they would smoke and get high. But that night, Jane tells Jake she has changed, but he insists and puts the marijuana in the sink's cabinet, saying one day they will smoke it together for old times sake. The next morning, Jane meets her psychiatrist, Dr. Alan. She asks if an affair is okay. He sees her condition of affair as a good way to reconnecting with Jake, so suggests she continues. That night she arranges a candle-lit dinner in hopes of being back with Jake as a family, which Jake agrees happily over the phone. He was at the fertility center when she called. They meet up in the same hotel the same afternoon, to Harley's surprise while having lunch with Lauren. He kisses her, trying to distract her from noticing Jane and Jake getting into the elevator and kissing as the doors close. In the room, Jake faints and a doctor goes up as Harley sees this as well. Jake tells the doctor he is taking "Flomax" to reduce constant urination. In return, it decreases his semen. (He tells the doctor due to his decrease in semen, it isn't good for his wife, which Jane thinks is the reason for the fertility center, but actually Jake was saying it to her. Both the doctor and Jane were surprised, but Jane agrees to follow the story and says she prefers lots of semen, all in which the doctor said "wow" as if it was a competitive matter to please a woman.) That night, Jane prepares the special dinner for Jake, making his favorite dishes and picking a nice dress. Jake at home, anxious to leave the house, grabs his phone and walks to the bathroom. Agness notices the phone and grows suspicious, preventing him from calling Jane. She tells him to put Pedro back in his bed, and takes off her shirt, saying she is ovulating, compensating the dinner arrangement which Jake had thought was able to attend, since Agness was going to a party, but canceled due to ovulation. Jane drinks a glass wine walking around with the food getting cold and she waits for Jake, unknowingly of the situation. Jake seemed distressed as he sets Pedro in his bed and puts his hand on the boy's chest. At this moment, Jake feels torn apart from two different families. He goes over to Jane's house the next morning, hoping for forgiveness, but Jane ignores him as if he wasn't there. The kids are happy to see him and ask if Jane will be coming to the party with dad. She says "No, I have another date with Adam." to much of Jake's surprise, so he leaves the house since he couldn't stay over because of Agness. Adam and Jane's relationship grows stronger. They have dinner together and one night, she takes him to her bakery to make croissants. They flirt and eventually Adam was going to ask if he could kiss her, but she already kissed him. Afterwards, Adam drives her to Luke's party and before they arrive, Jane takes the marijuana that was supposedly meant for Jake and her to smoke, decided to smoke it with Adam. He was hesitant at first, but let in. Luke's friend walks by and asks, "Who brought the reefer?", unknowingly Jane and Adam were smoking behind the shrubs.

bs. They go inside and dance/laugh. Adam sees Jake at a table, glaring at him with Jane, enviously. He walks over and asks if he can talk with Jane for a second. Jake takes her to the bathroom and confess his love for her. He realizes she was high and she admittedly said she smoked the reefer with Adam instead. Jake gets annoyed because it was meant for him and Jane, but he takes it from her just as Harley opens the door and sees them together, again this time, but smoking the reefer. Harley was going to leave them two alone, but Jane intervened and asks him to take a pull, which at first he didn't want to, but took a big pull. The three walk out and Adam is dancing with Agness, and Jake dances with Jane, which Agness sees and gets emotional and a bit upset, now that the secrets revealed. She sees Jake happily dance and smile affectionately at Jane. Jake was just about to kiss Jane, but she turns away and tells him that it's over and stop this affair immediately. She walks towards Adam. Jake sees Jane smile happily as she approaches Steve and feels sad, as Agness is across the room staring at him, which he doesn't notice. The following morning, everybody is drunk. Harley is wearing female pajamas, Luke forgot what happened last night. Gabby yawns, just as she asks, "What is dad doing outside?" Jane walks out and Jake approaches her. He told her that he left Agness that night, for good. He really loves Jane and wants to live here again, since he has no where to go. The kids come out to see what's going on and see the truth that although the parents are divorced, Jake wants to be back again. Luke wants Dad to stay for the night, which Jane didn't allow but the kids really want dad back home so she allows him to stay. That night, Jake is at the dinner table devouring the delicious food Jane made. The kids stare at him with smiles. Jane however, isn't at the dinner table, but upstairs video-chatting with Adam over the computer. Jake decides to run upstairs, as she's in the bathroom. He takes his clothes off and lays on the bed. The computer is on the bed, which Jake doesn't realize Adam is watching. But both Adam and Jane went to the bathroom, so Jake adjusts the computer as Jane walks out and screams, Adam comes back and sees Jake's private area and yells all different kinds of things. Jake surprised gets off the bed and turns around to pick up his clothes, only to moon at the computer (Adam). The kids run to the room and ask what's going on. Jake confesses that he left Agness for their mother. They have been meeting secretly since New York, and having an affair. The kids, shocked asks Jane if this is correct, which she agrees to. Gabby asks if they're going to be back together, which Jake says yes they will, but Jane strictly says, "No, we're not going to be together again." Adam hears this and immediately turns off this computer, which Jane sees. The kids, all upset and emotional deprived, leaves the house. Jake, astonished and rejected packs his bags and drives off. That night Jane sleeps outside the pool area, staring at the full moon, realizing the mess she has gotten into- an affair with Jake and a growing relationship with Adam, it's complicated. The following morning, Jane arrives at Adam's office and apologizes. Adam tells her that her divorce isn't officially over. He sees the love that Jake is expressing towards her is real. He continues to say that, it's best if he discontinues to meet with Jane, in case of getting into a serious relationship. He wants to resolve this situation simply by leaving her and allowing her divorce with Jake to heal. Jane asks if he will continue with remodeling her kitchen, but he says he might let Peter fill in the job instead. She becomes speechless and walks out. She drives home and sees Jake sitting on a swing outside her house. She sits down with him and they talk. Jake says he loves her and doesn't regret a second chance. Jane smiles and says the same. It starts to rain, but we just see the rain hitting the swing they were sitting on. The camera then focuses on a group of men outside, huddle under a tent discussing the project. Jane walks outside in hopes of seeing Adam, but doesn't. However, he pops out from the crowd and tells her, "It's a lucky day to construct in the rain." She smiles and asks the group if they want coffee. They all walk inside the house as Adam walks with her and asks if she has any of those double fudge chocolate cakes in her house. She laughs, saying "You love that stuff, don't you?" The credits roll.

In [35]:

```
def striphtml(data):
    cleanr = re.compile('<.*?>')
    cleantext = re.sub(cleanr, ' ', str(data))
    return cleantext

stop = set(stopwords.words('english')) #set of stopwords
sno = nltk.stem.SnowballStemmer('english') #initialising the snowball stemmer

def cleanpunc(sentence): #function to clean the word of any punctuation or special characters
    cleaned = re.sub(r'[!|\\"|\'|\"|#|=]', ' ', sentence)
    cleaned = re.sub(r'[\.,| |(|\|/|)]', ' ', cleaned)
    return cleaned

def clean_body(text):
    regex = r"href=\s*'\s*http\s*.*?'"
    return re.sub(regex, ' ', text)

def stemIt(text):
    stemmer = SnowballStemmer("english")
    ct = []
    for each in word_tokenize(text):
        ct.append(stemmer.stem(each))
    return ' '.join(ct)
```


In [36]:

```
stemmer = SnowballStemmer("english")
```

In [37]:

```
data['plot_synopsis'] = data['plot_synopsis'].apply(striphtml)
data['plot_synopsis'] = data['plot_synopsis'].apply(cleanpunc).str.lower()
data['plot_synopsis'] = data['plot_synopsis'].apply(clean_body)
data['plot_synopsis'] = data['plot_synopsis'].apply(stemIt)
```

In [38]:

```
data['plot_synopsis'][7414]
```

Out[38]:

'set and charact the set of xenoblad chronicl origin in a world that was noth but endless ocean un
til two great titan god the bioni and the mechoni came into exist the two god fought a timeless ba
ttl until with one final strike onli their corps remain forev lock in combat in the eon follow the
ir battl the god becam the home of multipl form of life the bioni is home to organ lifeform most p
romin the humanoid hom who are virtual ident to human both in appear and biolog ; the diminut and
furri nopon whose lifespan can doubl or even tripl those of hom ; and the avian humanoid high enti
a whose lifespan span centuri convers the mechoni is home to the mechan humanoid machina whose lif
espan span sever millennia a key weapon use by the hom in the battl against the attack machina is
the monado a mystic sword histor link to the bioni which grant vision of the futur to it wielder w
hen fulli control though the term monado is synonym with the red sword seen on the game cover it i
s also a general term use to describ ani supernatur sword that is awaken by a god key to the exist
of life and technolog on bioni and mechoni is ether a substanc that not onli serv as the primari b
uild block of all live be but also a fuel sourc and a weapon the game main charact is shulk a youn
g hom scientist who live in the settlement of coloni 9 on bioni dure an attack by machina warrior
call mechon shulk come into possess of the monado dure his journey shulk is join by reyn shulk chi
ldhood friend and headstrong member of the defenc forc ; dunban a former wielder of the monado and
the brother of shulk other childhood friend fiara ; sharla a medic and sniper from coloni 6 ; meli
a antiqua crown princess of the high entia and a high entia-hom hybrid ; and riki a member of the
nopon who is chosen as the hero of his villag other import charact includ dickson shulk mentor ; m
umkhar a coward soldier who fought alongsid dunban and wish to wield the monado for himself ; egil
self-proclaim leader of the machina ; and alvi a mysteri man who aid shulk on his journey plot the
game open detail event one year past when dickson dunban and mumkhar were fight a mechon armi mumk
har desert and in the process of defeat the mechon the monado paralyz dunban right arm in the pres
ent shulk studi the monado in coloni 9 where dunban and fiara live coloni 9 is soon attack by a gr
oup of mechon led by a special face mechon call metal face dunban is almost kill when he attempt t
o use the monado again prompt shulk to use it : he wield it with eas and receiv vision of the futu
r from it while the mechon are driven back metal face prove immun to the monado and kill fiara bef
or flee swear reveng against metal face shulk set out togeth with reyn they meet sharla from the d
ecim coloni 6 and work to free it from mechon occup they reunite with dunban on their journey short
thereaft guid by a vision from shulk the group travel to the high entia capit to gain entri to
prison island join by melia and riki shulk also meet alvi who is reveal to share shulk abil to
wield the monado gain entri they encount zanza a be who creat the monado and who offer shulk the a
bil to destroy the face mechon reveal to be human insid mechon mech though shulk accept and the mo
nado is grant the abil to destroy human zanza is kill by metal face and anoth mechon call face
nemesi dure an attack on the capit dure the ensu battl face nemesi is damag to reveal a recreat an
d amnesiac fiara control it while initi dishearten by this shulk is ralli by his comrad and set
out in pursuit of metal face and fiara dure a peac encount with fiara metal face attack them again
reveal itself to be mumkhar egil then interven spirit fiara away on the way to the mechoni the par
ti face off against and defeat mumkhar who is impal as a result of his own attack and then plummet
to his death the parti then face off against egil and fiara in the result fight shulk and fiara ar
e seapar from the group dure their time togeth shulk success awaken fiara memori and learn that
anoth be was control her bodi reunite with the group they meet up with a friend machina name vanea
and her peopl the machina she reveal that the bioni and mechoni were initi at peac befor the bioni
god zanza launch an unprovok attack sinc the battl a year befor egil has been work on his own to c
onvert the life of bioni into mechon to render the monado useless go to face egil fiara is taken o
ver by the other presenc the machina goddess meyneth they reach egil as he reactiv the mechoni and
begin an attack on the bioni intent on destroy the bioni as reveng for the decim it had caus to th
e machina thousand of year ago despit fight him shulk manag to make him see that they both wish fo
r a return to peac and declin the open opportun to kill egil upon the latter defeat at this point
dickson appear and shoot shulk reveal himself as a discipl of zanza who emerg from shulk bodi and
reclaim the monado egil and meyneth both attempt to fight zanza but are overpow and meyneth monado
is stolen by zanza as the parti escap egil attempt to fight onc more but zanza destroy both him an
d the mechoni in the aftermath of the mechoni destruct pure-blood high entia begin transform into
telethia be whose one purpos is to purg bioni of life the parti is helpless befor the telethia but
the telethia are attack by a squadron of high entia howev the high entia minist of research lorith
ia is reveal to be a discipl of zanza release other to transform the high entia squadron into

id is reveal to be a discipl of zanza releas ether to transform the high entia squadron into telethia while the parti griev for shulk shulk awaken and manag to defeat a telethia raid on coloni 6 but alvi is reveal to be the third discipl of zanza make their way to prison island they defeat lorithea and then dickson the parti then travel to face zanza who declar the life of bioni as simpli his food and vessel and with both monado he intend to recreat the world as he has befor destroy the current world and everyth in it in the process zanza then offer shulk the chanc to becom his new discipl shulk reject the offer and dure the ensu battl produc a third monado : promp t by alvi the spirit of the monado who betray zanza shulk use his monado to destroy zanza alvi the n show shulk zanza origin ; both zanza then name klaus and meyneth were origin human scientist from earth work on an experi to creat a new univers aboard a space station klauss obsess desir cau s the experi to end in disast obliter the univers and caus zanza and meyneth to be reborn as god i n the new univers alvi was origin the administr comput of the station after the new univers creati on zanza and meyneth creat life in their imag and zanza creat the cycl of bioni out of fear that h e would eventu die as his creation forgot of his exist and seek life beyond bioni with the current univers threaten with death from zanza action alvi ask shulk to decid how he will remak the univers shulk now a god wish for a world without god where everyon can decid their own fate in the new univers the survivor of bioni and mechoni build a new settlement and live peac togeth fiora is restor to her hom form and both she and shulk optimist look forward to alvi promis of endless worl d and race of peopl beyond their own'

In [39]:

```
data['title'] = data['title'].apply(striphtml)
data['title'] = data['title'].apply(cleanpunc).str.lower()
data['title'] = data['title'].apply(clean_body)
data['title'] = data['title'].apply(stemIt)
```

In [40]:

```
data['title'][7414]
```

Out[40]:

'zenobureido'

In []:

In [41]:

```
data['tags'] = data['tags'].apply(cleanpunc).str.lower()
data['tags'] = data['tags'].apply(striphtml)
```

In [42]:

```
data['tags'][57]
```

Out[42]:

'violence'

In [43]:

```
data.head()
```

Out[43]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
7414	tt1710626	zenobureido	set and charact the set of xenoblad chronicl o...	revenge	val	wikipedia	1
11142	tt0119642	mean gun	vincent moon ice-t is the leader of a crime sy...	revenge action murder violence flashback	val	wikipedia	5
11066	tt0183659	pollock	the film begin show the abstract expressionist...	tragedy romantic	val	wikipedia	2
11062	tt0250494	legal blond	fashion merchandis student ell wood is taken t...	entertaining cute comedy murder stupid	val	wikipedia	5
11000	tt0005170	antoni and	mark antoni – one of the triumvir of the				1

In [44]:

```
data['plot_synopsis'][12999]
```

Out[44]:

'open chapter 1 to 3 in 1801 lockwood a wealthi young man from the south of england who is seek pe ac and recuper rent thrushcross grang in yorkshir he visit his landlord heathcliff who live in a r emot moorland farmhous wuther height there lockwood find an odd assemblag : heathcliff seem to be a gentleman but his manner are uncouth ; the reserv mistress of the hous is in her mid-teen ; and a young man who seem to be a member of the famili yet dress and speak as if he is a servant snow i n lockwood is grudg allow to stay and is shown to a bedchamb where he notic book and graffiti left by a former inhabit name catherin he fall asleep and has a nightmar in which he see the ghost cath erin tri to enter through the window he cri out in fear rous heathcliff who rush into the room loc kwood is convinc that what he saw was real heathcliff believ lockwood to be right examin the windo w and open it hope to allow catherin spirit to enter when noth happen heathcliff show lockwood to his own bedroom and return to keep watch at the window at sunris heathcliff escort lockwood back t o thrushcross grang lockwood ask the housekeep nelli dean about the famili at wuther height and sh e tell him the tale heathcliff childhood chapter 4 to 17 thirti year earlier the owner of wuther h eight is mr earnshaw who live with his son hindley and younger daughter catherin on a trip to live rpool earnshaw encount a homeless boy describ as a dark-skin gypsi in aspect he adopt the boy and name him heathcliff hindley feel that heathcliff has supplant him in his father affect and becom b itter jealous catherin and heathcliff becom friend and spend hour each day play on the moor they g row close hindley is sent to colleg three year later earnshaw die and hindley becom the landown ; he is now master of wuther height he return to live there with his new wife franc he allow heathcliff to stay but onli as a servant and regular mistreat him a few month after hindley return heathcliff and catherin walk to thrushcross grang to spi on edgar and isabella linton who live the re after be discov they tri to run away but are caught catherin is injur by the linton dog and tak en into the hous to recuper while heathcliff is sent home catherin stay with the linton the linton are land gentri and catherin is influenc by their fine appear and genteel manner when she return t o wuther height her appear and manner are more ladylik and she laugh at heathcliff unkempt appear the next day know that the linton are to visit heathcliff upon nelli advic tri to dress up in an e ffort to impress catherin but he and edgar linton get into an argument and hindley humili heathcli ff by lock him in the attic catherin tri to comfort heathcliff but he vow reveng on hindley the fo llow year franc earnshaw give birth to a son name hareton but she die a few month later hindley de scend into drunken two more year pass and catherin and edgar linton becom friend while she becom m ore distant from heathcliff edgar visit catherin while hindley is away and they declar themselv lo ver soon afterward catherin confess to nelli that edgar has propos marriag and she has accept alth ough her love for edgar is not compar to her love for heathcliff whom she can not marri becaus of his low social status and lack of educ she hope to use her posit as edgar wife to rais heathcliff stand heathcliff overhear her say that it would degrad her to marri him but not how much she love him and he run away and disappear without a trace distraught over heathcliff departur catherin mak e herself ill nelli and edgar begin to pander to her everi whim to prevent her from becom ill again three year pass edgar and catherin marri and go to live togeth at thrushcross grang where ca therin enjoy be ladi of the manor six month later heathcliff return now a wealthi gentleman catherin is delight but edgar is not edgar sister isabella soon fall in love with heathcliff who d espis her but encourag the infatu as a mean of reveng one day he embrac isabella lead to an argue me nt with edgar upset catherin lock herself in her room and begin to make herself ill again heathcliff take up resid at wuther height and spend his time gambl with hindley and teach hareton bad habit hindley dissip his wealth and mortgag the farmhous to heathcliff to pay his debt heathcl iff elop with isabella linton two month after their elop they return to wuther height where heathcliff hear that catherin is die with nelli help he visit cathi secret howev catherin is pregnant the follow day she give birth to a daughter cathi short befor die after catherin funer is abella confid in nelli laugh as she tell nelli how glad she is that cathi die and that she has lef t heathcliff reject shelter by edgar isabella eventu find refug in the south of england and give b irth to a son linton hindley die six month after catherin and heathcliff thus find himself master of wuther height heathcliff matur chapter 18 to 31 twelv year pass catherin daughter cathi has bec om a beauti high-spirit girl edgar learn that his sister isabella is die so he leav to retriev her son linton in order to adopt and educ him cathi who has rare left home take advantag of her father absenc to ventur further afield she ride over the moor to wuther height and discov that she has no t one but two cousin : hareton in addit to linton she also let it be known that her father has gon e to fetch linton when edgar return with linton a weak and sick boy heathcliff insist that he live at wuther height three year pass walk on the moor nelli and cathi encount heathcliff who take them to wuther height to see linton and hareton heathcliff hope that linton and cathi will marri so tha t linton will becom the heir to thrushcross grang linton and cathi begin a secret friendship echo the childhood friendship between their respect parent heathcliff and catherin the follow year edga r becom veri ill and take a turn for the wors while nelli and cathi are out on the moor where heat hcliff and linton trick them into enter wuther height heathcliff keep them captiv to enabl the mar riag of cathi and linton to take place after five day nelli is releas and later with linton help c athi escap she return to the grang to see her father short befor he die now master of both wuther height and thrushcross grang cathi father-in-law heathcliff insist on her return to live at wuther height soon after she arriv linton die hareton tri to be kind to cathi but she withdraw from the w orld at this point nelli tale catch up to the present day 1801 time pass and after be ill for a pe riod lockwood grow tire of the moor and inform heathcliff that he will be leav thrushcross grang a

lockwood grow life of the moor and inform heathcliff that he will be leav thrushcross grang e
nd chapter 32 to 34 eight month later lockwood return to the area by chanc given that his tenanc a
t thrushcross grang is still valid he decid to stay there again he find nelli live at wuther height
and enquir what has happen sinc he left she explain that she move to wuther height to replac the h
ousekeep zillah who had left hareton has an accid and is confin to the farmhous dure his convalesc
he and cathi overcom their mutual antipathi and becom close while their friendship develop
heathcliff begin to act strang and has vision of catherin he stop eat and after four day is found
dead in catherin old room he is buri next to catherin lockwood learn that hareton and cathi plan t
o marri on new year day as he get readi to leav he pass the grave of catherin edgar and heathcliff
and paus to contempl the quiet of the moor'

In [45]:

```
data['synopsis_source'] = data['synopsis_source'].apply(striphtml)
data['synopsis_source'] = data['synopsis_source'].apply(cleanpunc).str.lower()
data['synopsis_source'] = data['synopsis_source'].apply(clean_body)
data['synopsis_source'] = data['synopsis_source'].apply(stemIt)
```

In [46]:

```
data.head()
```

Out[46]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
7414	tt1710626	zenobureido	set and charact the set of xenoblad chronicl o...	revenge	val	wikipedia	1
11142	tt0119642	mean gun	vincent moon ice-t is the leader of a crime sy...	revenge action murder violence flashback	val	wikipedia	5
11066	tt0183659	pollock	the film begin show the abstract expressionist...	tragedy romantic	val	wikipedia	2
11062	tt0250494	legal blond	fashion merchandis student ell wood is taken t...	entertaining cute comedy murder stupid	val	wikipedia	5
11060	tt0085176	antoni and cleopatra	mark antoni – one of the triumvir of the roman...	tragedy	val	wikipedia	1

In [47]:

```
train_data = data['split']=='train'
train = data[train_data]
train = pd.DataFrame(train)

test_data = data['split']=='test'
test = data[test_data]
test = pd.DataFrame(test)

validation_data = data['split']=='val'
val = data[validation_data]
val = pd.DataFrame(val)
```

In [48]:

```
train.head()
```

Out[48]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
14173	tt0068688	heat	fbi agent sarah ashburn sandra bullock is an e...	cult satire	train	wikipedia	2
9219	tt0013520	puss in boot	the tale open with the third and youngest son ...	fantasy	train	wikipedia	1
9205	tt0093286	it aliv iii : island of the aliv	this film is set year after the first two the ...	cult satire violence	train	wikipedia	3
9202	tt0022286	the public enemi	as youngster in 1900s chicago tom power jame c...	revenge murder	train	wikipedia	2

9228	imdb_id	title	a singl human narrat from england is	plot_synopsis	insanity romantic avant garde	tags	split	synopsis_source	count
		luomo dell'oceano	transport...		murder				

In [49]:

```
test.head()
```

Out[49]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
13401	tt0112453	balto	in new york citi an elder woman and her grandd...	cult psychedelic	test	wikipedia	2
13940	tt0481610	the retreat	london architect martin kennedi cillian murphi...	violence murder	test	wikipedia	2
13702	tt0339840	undead	after lose her famili farm to the bank local b...	comedy murder bleak cult violence flashback	test	wikipedia	6
1326	tt0934584	supercroc	a small group of soldier is sent north of l a ...	violence	test	imdb	1
1014	tt0075995	equus	martin dysart richard burton is a psychiatrist...	insanity flashback	test	imdb	2

In [50]:

```
val.head()
```

Out[50]:

	imdb_id	title	plot_synopsis	tags	split	synopsis_source	count
7414	tt1710626	zenobureido	set and charact the set of xenoblad chronicl o...	revenge	val	wikipedia	1
11142	tt0119642	mean gun	vincent moon ice-t is the leader of a crime sy...	revenge action murder violence flashback	val	wikipedia	5
11066	tt0183659	pollock	the film begin show the abstract expressionist...	tragedy romantic	val	wikipedia	2
11062	tt0250494	legal blond	fashion merchandis student ell wood is taken t...	entertaining cute comedy murder stupid	val	wikipedia	5
11060	tt0085176	antoni and cleopatra	mark antoni – one of the triumvir of the roman...	tragedy	val	wikipedia	1

In []:

In [51]:

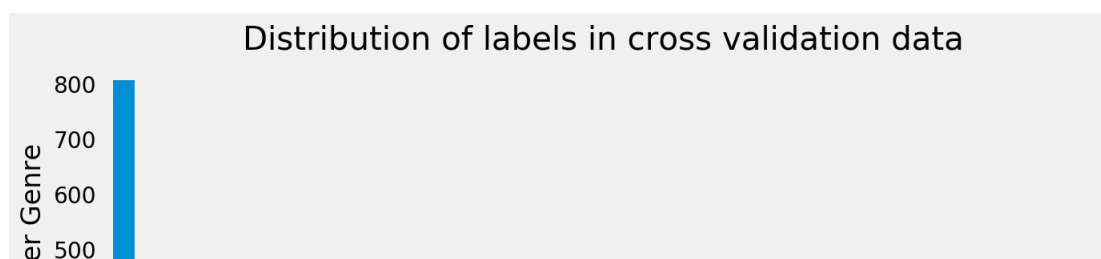
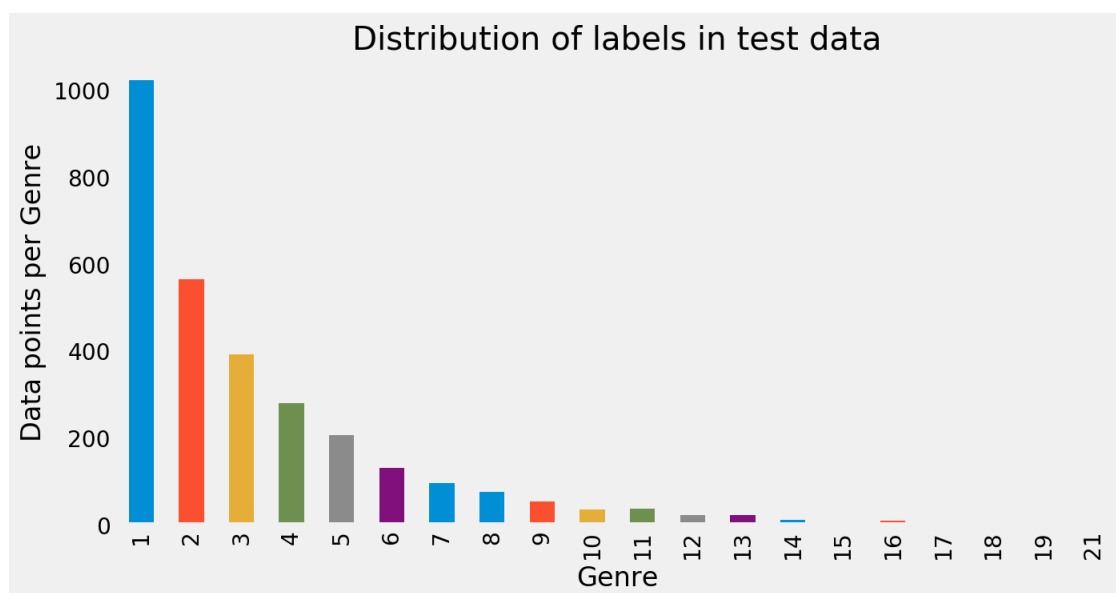
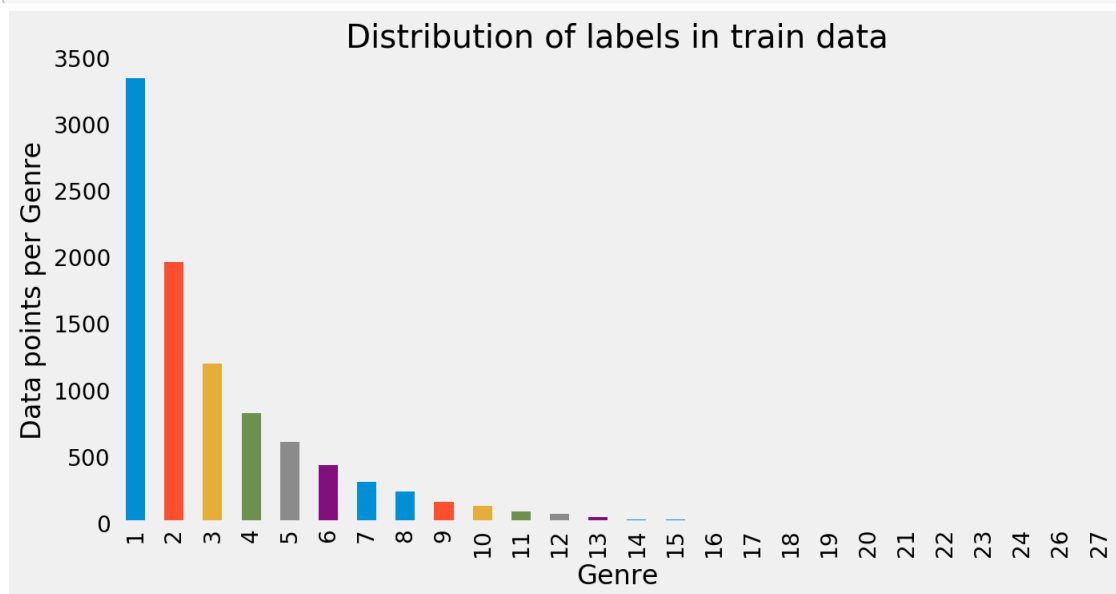
```
# it returns a dict, keys as Genre labels and values as the number of data points in that Genre
train_Genre_distribution = train['count'].value_counts().sortlevel()
test_Genre_distribution = test['count'].value_counts().sortlevel()
cv_Genre_distribution = val['count'].value_counts().sortlevel()

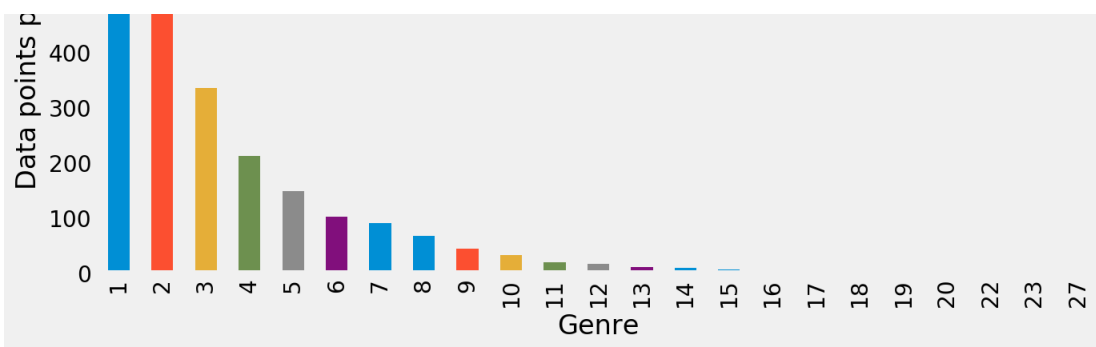
my_colors = 'rgbkymc'
train_Genre_distribution.plot(kind='bar')
plt.xlabel('Genre')
plt.ylabel('Data points per Genre')
plt.title('Distribution of labels in train data')
plt.grid()
plt.show()

print('-'*80)
my_colors = 'rgbkymc'
test_Genre_distribution.plot(kind='bar')
plt.xlabel('Genre')
plt.ylabel('Data points per Genre')
plt.title('Distribution of labels in test data')
plt.grid()
```

```
plt.show()

print('-'*80)
print('-'*80)
my_colors = 'rgbkymc'
cv_Genre_distribution.plot(kind='bar')
plt.xlabel('Genre')
plt.ylabel('Data points per Genre')
plt.title('Distribution of labels in cross validation data')
plt.grid()
plt.show()
```





In [52]:

```
data.to_csv('data_cleaned.csv', index=True)
```

In [53]:

```
train.to_csv('train.csv', index=True)
```

In [54]:

```
test.to_csv('test.csv', index=True)
```

In [55]:

```
val.to_csv('val.csv', index=True)
```

In []: