In [85]:
```
import numpy as np
import pandas as pd
import seaborn as sb
import plotly.express as px
import matplotlib.pyplot as mpl
from wordcloud import WordCloud
from sklearn.model_selection import train_test_split
```

In [2]:
```
movie_file = pd.read_csv('prediction.csv', encoding='latin1')
```

In [3]:
```
movie_file
```

Out[3]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1
0		NaN	NaN	Drama	NaN	NaN	J.S. Randhawa	Manmauji
1	#Gadhvi (He thought he was Gandhi)	(2019)	109 min	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal
2	#Homecoming	(2021)	90 min	Drama, Musical	NaN	NaN	Soumyajit Majumdar	Sayani Gupta
3	#Yaaram	(2019)	110 min	Comedy, Romance	4.4	35	Ovais Khan	Prateik
4	...And Once Again	(2010)	105 min	Drama	NaN	NaN	Amol Palekar	Rajat Kapoor
...	...	...	...	...	...	...	...	...
15504	Zulm Ko Jala Doonga	(1988)	NaN	Action	4.6	11	Mahendra Shah	Naseeruddin Shah
15505	Zulmi	(1999)	129 min	Action, Drama	4.5	655	Kuku Kohli	Akshay Kumar
15506	Zulmi Raj	(2005)	NaN	Action	NaN	NaN	Kiran Thej	Sangeeta Tiwari
15507	Zulmi Shikari	(1988)	NaN	Action	NaN	NaN	NaN	NaN
15508	Zulm-O-Sitam	(1998)	130 min	Action, Drama	6.2	20	K.C. Bokadia	Dharmendra

15509 rows × 10 columns

In [4]:
```
movie_file.head(11)
```

Out[4]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	
0		NaN	NaN	Drama	NaN	NaN	J.S. Randhawa	Manmauji	
1	#Gadhvi (He thought he was Gandhi)	(2019)	109 min	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Gh
2	#Homecoming	(2021)	90 min	Drama, Musical	NaN	NaN	Soumyajit Majumdar	Sayani Gupta	B
3	#Yaaram	(2019)	110 min	Comedy, Romance	4.4	35	Ovais Khan	Prateik	I
4	...And Once Again	(2010)	105 min	Drama	NaN	NaN	Amol Palekar	Rajat Kapoor	R S
5	...Aur Pyaar Ho Gaya	(1997)	147 min	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Ai B
6	...Yahaan	(2005)	142 min	Drama, Romance, War	7.4	1,086	Shoojit Sircar	Jimmy Sheirgill	I
7	.in for Motion	(2008)	59 min	Documentary	NaN	NaN	Anirban Datta	NaN	
8	?: A Question Mark	(2012)	82 min	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	
9	@Andheri	(2014)	116 min	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	
10	1:1.6 An Ode to Lost Love	(2004)	96 min	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	

In [5]:
```
movie_file.describe()
```

Out[5]:

	Rating
count	7919.000000
mean	5.841621
std	1.381777
min	1.100000
25%	4.900000
50%	6.000000
75%	6.800000
max	10.000000

In [6]:
```
movie_file.dtypes
```

Out[6]:
```
Name         object
Year         object
Duration     object
Genre        object
Rating      float64
Votes        object
Director     object
Actor 1      object
Actor 2      object
Actor 3      object
dtype: object
```

In [7]:
```
movie_file.isnull().sum()
```

Out[7]:
```
Name           0
Year         528
Duration    8269
Genre       1877
Rating      7590
Votes       7589
Director     525
Actor 1     1617
Actor 2     2384
Actor 3     3144
dtype: int64
```

In [8]:
```
movie_file.isnull().sum().sum()
```

Out[8]:
```
33523
```

In [9]:
```
movie_file.shape
```

Out[9]:
```
(15509, 10)
```

In [10]:
```
movie_file.dropna(inplace = True)
```

In [11]:
```
movie_file.head(11)
```

Out[11]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2
1	#Gadhvi (He thought he was Gandhi)	(2019)	109 min	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande
3	#Yaaram	(2019)	110 min	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj
5	...Aur Pyaar Ho Gaya	(1997)	147 min	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan
6	...Yahaan	(2005)	142 min	Drama, Romance, War	7.4	1,086	Shoojit Sircar	Jimmy Sheirgill	Minissha Lamba
8	?: A Question Mark	(2012)	82 min	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Muntazir Ahmad
9	@Andheri	(2014)	116 min	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fathima Babu
10	1:1.6 An Ode to Lost Love	(2004)	96 min	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulshan Grover
11	1:13:7 Ek Tera Saath	(2016)	120 min	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anubhav Dhir
12	100 Days	(1991)	161 min	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madhuri Dixit
13	100% Love	(2012)	166 min	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Koyel Mallick
15	102 Not Out	(2018)	102 min	Comedy, Drama	7.4	6,619	Umesh Shukla	Amitabh Bachchan	Rishi Kapoor

In [13]:
```
movie_file.isnull().sum()
```

Out[13]:
```
Name        0
Year        0
Duration    0
Genre       0
Rating      0
Votes       0
Director    0
Actor 1     0
Actor 2     0
Actor 3     0
dtype: int64
```

In [14]:
```
movie_file.isnull().sum().sum()
```

Out[14]:
```
0
```

In [15]:
```
movie_file.shape
```

Out[15]:
```
(5659, 10)
```

In [18]:
```
movie_file['Duration'] = movie_file['Duration'].str.extract('(\d+)')
movie_file['Duration'] = pd.to_numeric(movie_file['Duration'], errors='c
```

In [19]:
```
movie_file["Year"].head()
```

Out[19]:
```
1    2019
3    2019
5    1997
6    2005
8    2012
Name: Year, dtype: int64
```

In [20]:
```
genre = movie_file['Genre']
```

In [21]:
```
genre.head(11)
```

Out[21]:
```
1                          Drama
3                 Comedy, Romance
5          Comedy, Drama, Musical
6             Drama, Romance, War
8       Horror, Mystery, Thriller
9         Action, Crime, Thriller
10                         Drama
11                        Horror
12      Horror, Romance, Thriller
13         Comedy, Drama, Romance
15                  Comedy, Drama
Name: Genre, dtype: object
```

In [22]:
```
genres = movie_file['Genre'].str.split(', ', expand=True)
```

In [23]:
```
genres.head(11)
```

Out[23]:

	0	1	2
1	Drama	None	None
3	Comedy	Romance	None
5	Comedy	Drama	Musical
6	Drama	Romance	War
8	Horror	Mystery	Thriller
9	Action	Crime	Thriller
10	Drama	None	None
11	Horror	None	None
12	Horror	Romance	Thriller
13	Comedy	Drama	Romance
15	Comedy	Drama	None

In [24]:
```
genre_counts = {}
for genre in genres.values.flatten():
    if genre is not None:
        if genre in genre_counts:
            genre_counts[genre] += 1
        else:
            genre_counts[genre] = 1

genereCounts = {genre: count for genre, count in sorted(genre_counts.ite
for genre, count in genereCounts.items():
    print(f"{genre}: {count}")
```

```
Action: 1686
Adventure: 277
Animation: 40
Biography: 115
Comedy: 1344
Crime: 875
Documentary: 48
Drama: 3796
Family: 416
Fantasy: 146
History: 99
Horror: 202
Music: 53
Musical: 412
Mystery: 304
News: 1
Romance: 1380
Sci-Fi: 32
Sport: 40
Thriller: 679
War: 33
Western: 1
```

In [25]:
```
genresPie = movie_file['Genre'].value_counts()
```

In [26]:
```
genresPie.head(11)
```

Out[26]:
```
Genre
Drama                       844
Drama, Romance              332
Action, Crime, Drama        329
Action, Drama               206
Comedy, Drama               205
Comedy, Drama, Romance      188
Comedy                      187
Action                      170
Drama, Family               160
Romance                     126
Action, Comedy, Drama       118
Name: count, dtype: int64
```

In [27]:
```
genrePie = pd.DataFrame(list(genresPie.items()))
genrePie = genrePie.rename(columns={0: 'Genre', 1: 'Count'})
```

In [28]:
```
genrePie.head(11)
```

Out[28]:

	Genre	Count
0	Drama	844
1	Drama, Romance	332
2	Action, Crime, Drama	329
3	Action, Drama	206
4	Comedy, Drama	205
5	Comedy, Drama, Romance	188
6	Comedy	187
7	Action	170
8	Drama, Family	160
9	Romance	126
10	Action, Comedy, Drama	118

In [29]:
```
movie_file['Votes'] = movie_file['Votes'].str.replace(',', '').astype(in
```

In [30]:
```
movie_file['Votes'].head(11)
```

Out[30]:
```
1        8
3       35
5      827
6     1086
8      326
9       11
10      17
11      59
12     983
13     512
15    6619
Name: Votes, dtype: int32
```

In [31]:
```
movie_file["Director"].nunique()
```

Out[31]:
```
2431
```

In [32]:
```
directors = movie_file["Director"].value_counts()
```

In [33]:
```
directors.head(11)
```

Out[33]:
```
Director
David Dhawan            41
Mahesh Bhatt            39
Ram Gopal Varma         33
Hrishikesh Mukherjee    33
Shakti Samanta          33
Vikram Bhatt            30
Priyadarshan            29
Basu Chatterjee         23
Rama Rao Tatineni       23
Shyam Benegal           22
Yash Chopra             22
Name: count, dtype: int64
```

In [34]:
```
actors = pd.concat([movie_file['Actor 1'], movie_file['Actor 2'], movie_
```

In [35]:
```
actors.head(11)
```

Out[35]:
```
Mithun Chakraborty    160
Amitabh Bachchan      148
Dharmendra            146
Ashok Kumar           124
Akshay Kumar          120
Shashi Kapoor         117
Jeetendra             116
Sanjay Dutt           101
Rekha                  92
Ajay Devgn             92
Govinda                90
Name: count, dtype: int64
```

In [40]:
```
sb.set(style = "darkgrid", font = "Calibri")
```

In [49]:

```
Q1 = movie_file['Duration'].quantile(0.25)
Q3 = movie_file['Duration'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
df = movie_file[(movie_file['Duration'] >= lower_bound) & (movie_file['D
df.head(11)
```

Out[49]:

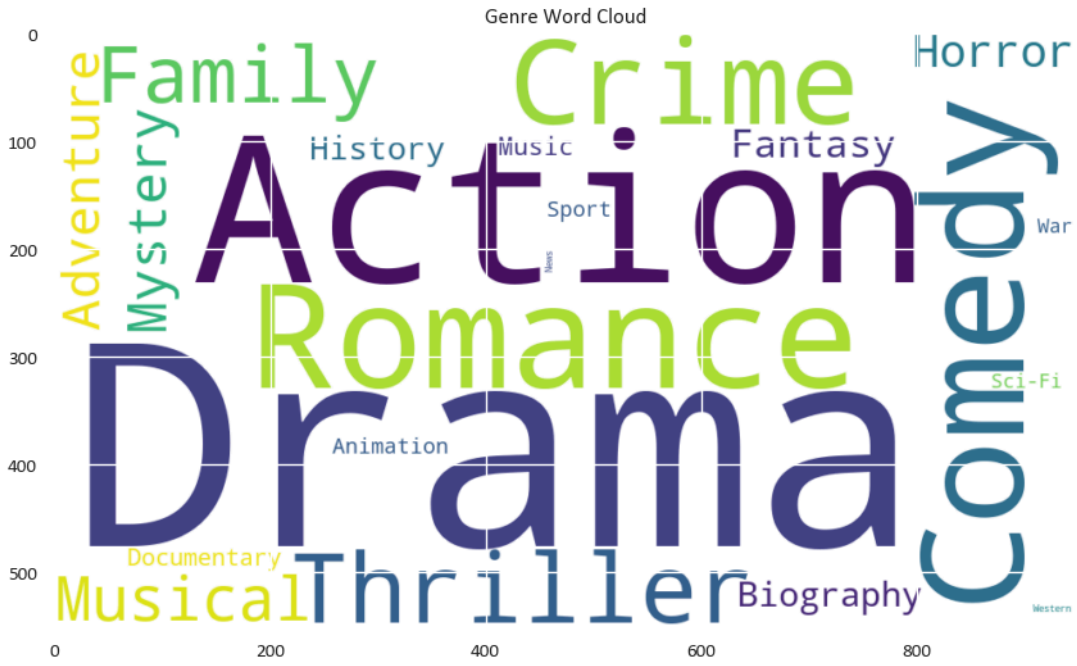
	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj
5	...Aur Pyaar Ho Gaya	1997	147	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan
6	...Yahaan	2005	142	Drama, Romance, War	7.4	1086	Shoojit Sircar	Jimmy Sheirgill	Minissha Lamba
8	?: A Question Mark	2012	82	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Muntazir Ahmad
9	@Andheri	2014	116	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fathima Babu
10	1:1.6 An Ode to Lost Love	2004	96	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulshan Grover
11	1:13:7 Ek Tera Saath	2016	120	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anubhav Dhir
12	100 Days	1991	161	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madhuri Dixit
13	100% Love	2012	166	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Koyel Mallick
15	102 Not Out	2018	102	Comedy, Drama	7.4	6619	Umesh Shukla	Amitabh Bachchan	Rishi Kapoor

In [55]:

```
genre_counts = movie_file['Genre'].str.split(', ', expand=True).stack().

wordcloud = WordCloud(width=950, height=550, background_color='white').g

mpl.figure(figsize=(16, 6))
mpl.imshow(wordcloud, interpolation='bilinear')
mpl.axis('on')
mpl.title('Genre Word Cloud')
mpl.show()
```



In [62]:

```
Q1 = movie_file['Rating'].quantile(0.25)
Q3 = movie_file['Rating'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
movie_file = movie_file[(movie_file['Rating'] >= lower_bound) & (movie_f
movie_file.head(16)
```

Out[62]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Acto
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vi Ghamai
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita
5	...Aur Pyaar Ho Gaya	1997	147	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwa Bachcl
6	...Yahaan	2005	142	Drama, Romance, War	7.4	1086	Shoojit Sircar	Jimmy Sheirgill	Minis Lan
8	?: A Question Mark	2012	82	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Munt Ahn
9	@Andheri	2014	116	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fath Ba
10	1:1.6 An Ode to Lost Love	2004	96	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulsl Grc
11	1:13:7 Ek Tera Saath	2016	120	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anub [
12	100 Days	1991	161	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madl [
13	100% Love	2012	166	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Kc Mal
15	102 Not Out	2018	102	Comedy, Drama	7.4	6619	Umesh Shukla	Amitabh Bachchan	R Kap
18	10ml LOVE	2010	87	Comedy, Drama, Romance	6.3	162	Sharat Katariya	Neil Bhoopalam	Anu B
21	12 O'Clock	1958	132	Crime, Drama, Mystery	7.2	72	Pramod Chakravorty	Guru Dutt	Wahe Rehn
22	12 O'Clock	2021	105	Horror, Thriller	4.4	63	Ram Gopal Varma	Mithun Chakraborty	Makr Deshpa
25	127 B	2017	116	Comedy, Horror	6.6	26	Seshu Kmr	Mast Ali	Aziz Na
28	13B: Fear Has a New Address	2009	146	Drama, Horror, Mystery	7.3	6329	Vikram K. Kumar	Madhavan	Ne Chan

In [70]:
```
movie_file.head(16)
```

Out[70]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Acto
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vi Ghamai
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita
5	...Aur Pyaar Ho Gaya	1997	147	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwa Bachcl
6	...Yahaan	2005	142	Drama, Romance, War	7.4	1086	Shoojit Sircar	Jimmy Sheirgill	Minis Lan
8	?: A Question Mark	2012	82	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Munt Ahn
9	@Andheri	2014	116	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fath Ba
10	1:1.6 An Ode to Lost Love	2004	96	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulsl Grc
11	1:13:7 Ek Tera Saath	2016	120	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anub D
12	100 Days	1991	161	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madl D
13	100% Love	2012	166	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Kc Mal
15	102 Not Out	2018	102	Comedy, Drama	7.4	6619	Umesh Shukla	Amitabh Bachchan	R Kap
18	10ml LOVE	2010	87	Comedy, Drama, Romance	6.3	162	Sharat Katariya	Neil Bhoopalam	Anu Bl

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Acto
21	12 O'Clock	1958	132	Crime, Drama, Mystery	7.2	72	Pramod Chakravorty	Guru Dutt	Wahe Rehn
22	12 O'Clock	2021	105	Horror, Thriller	4.4	63	Ram Gopal Varma	Mithun Chakraborty	Makr Deshpa
25	127 B	2017	116	Comedy, Horror	6.6	26	Seshu Kmr	Mast Ali	Aziz Na
28	13B: Fear Has a New Address	2009	146	Drama, Horror, Mystery	7.3	6329	Vikram K. Kumar	Madhavan	Ne Chan

In [74]:

```
Q1 = movie_file['Genres'].quantile(0.25)
Q3 = movie_file['Genres'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
movie_file = movie_file[(movie_file['Genres'] >= lower_bound) & (movie_f
```

In [76]:
```
movie_file.head(11)
```

Out[76]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj
5	...Aur Pyaar Ho Gaya	1997	147	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan
6	...Yahaan	2005	142	Drama, Romance, War	7.4	1086	Shoojit Sircar	Jimmy Sheirgill	Minissha Lamba
8	?: A Question Mark	2012	82	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Muntazir Ahmad
9	@Andheri	2014	116	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fathima Babu
10	1:1.6 An Ode to Lost Love	2004	96	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulshan Grover
11	1:13:7 Ek Tera Saath	2016	120	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anubhav Dhir
12	100 Days	1991	161	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madhuri Dixit
13	100% Love	2012	166	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Koyel Mallick
15	102 Not Out	2018	102	Comedy, Drama	7.4	6619	Umesh Shukla	Amitabh Bachchan	Rishi Kapoor

In [78]:

```
Q1 = movie_file['Directors'].quantile(0.25)
Q3 = movie_file['Directors'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
movie_file = movie_file[(movie_file['Directors'] >= lower_bound) & (movi
```

In [79]:
```
movie_file.head(11)
```

Out[79]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj
5	...Aur Pyaar Ho Gaya	1997	147	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan
6	...Yahaan	2005	142	Drama, Romance, War	7.4	1086	Shoojit Sircar	Jimmy Sheirgill	Minissha Lamba
8	?: A Question Mark	2012	82	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Muntazir Ahmad
9	@Andheri	2014	116	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fathima Babu
10	1:1.6 An Ode to Lost Love	2004	96	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulshan Grover
11	1:13:7 Ek Tera Saath	2016	120	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anubhav Dhir
12	100 Days	1991	161	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madhuri Dixit
13	100% Love	2012	166	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Koyel Mallick
15	102 Not Out	2018	102	Comedy, Drama	7.4	6619	Umesh Shukla	Amitabh Bachchan	Rishi Kapoor

In [82]:

```
Q1 = movie_file['Actors'].quantile(0.25)
Q3 = movie_file['Actors'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
movie_file = movie_file[(movie_file['Actors'] >= lower_bound) & (movie_f
```

In [84]:
```
movie_file.head(11)
```

Out[84]:

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2
1	#Gadhvi (He thought he was Gandhi)	2019	109	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghamande
3	#Yaaram	2019	110	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj
5	...Aur Pyaar Ho Gaya	1997	147	Comedy, Drama, Musical	4.7	827	Rahul Rawail	Bobby Deol	Aishwarya Rai Bachchan
6	...Yahaan	2005	142	Drama, Romance, War	7.4	1086	Shoojit Sircar	Jimmy Sheirgill	Minissha Lamba
8	?: A Question Mark	2012	82	Horror, Mystery, Thriller	5.6	326	Allyson Patel	Yash Dave	Muntazir Ahmad
9	@Andheri	2014	116	Action, Crime, Thriller	4.0	11	Biju Bhaskar Nair	Augustine	Fathima Babu
10	1:1.6 An Ode to Lost Love	2004	96	Drama	6.2	17	Madhu Ambat	Rati Agnihotri	Gulshan Grover
11	1:13:7 Ek Tera Saath	2016	120	Horror	5.9	59	Arshad Siddiqui	Pankaj Berry	Anubhav Dhir
12	100 Days	1991	161	Horror, Romance, Thriller	6.5	983	Partho Ghosh	Jackie Shroff	Madhuri Dixit
13	100% Love	2012	166	Comedy, Drama, Romance	5.7	512	Rabi Kinagi	Jeet	Koyel Mallick
15	102 Not Out	2018	102	Comedy, Drama	7.4	6619	Umesh Shukla	Amitabh Bachchan	Rishi Kapoor

In [86]:
```
Input = movie_file.drop(['Name', 'Genre', 'Rating', 'Director', 'Actor 1
Output = movie_file['Rating']
```

In [87]:
```
Input.head(16)
```

Out[87]:

	Year	Duration	Votes	Directors	Genres	Actors
1	2019	109	8	629	229	3858
3	2019	110	35	1334	184	3326
5	1997	147	827	1527	157	1112
6	2005	142	1086	2037	289	2075
8	2012	82	326	135	320	5535
9	2014	116	11	401	37	967
10	2004	96	17	1083	229	3863
11	2016	120	59	266	316	3213
12	1991	161	983	1373	323	1890
13	2012	166	512	1503	159	1971
15	2018	102	6619	2270	151	508
18	2010	87	162	1990	159	3113
21	1958	132	72	1425	199	1716
22	2021	105	63	1637	325	2768
25	2017	116	26	1948	171	2637
28	2009	146	6329	2340	268	2463

In [88]:
```
Output.head(16)
```

Out[88]:
```
1     7.0
3     4.4
5     4.7
6     7.4
8     5.6
9     4.0
10    6.2
11    5.9
12    6.5
13    5.7
15    7.4
18    6.3
21    7.2
22    4.4
25    6.6
28    7.3
Name: Rating, dtype: float64
```

In [89]:
```
x_train, x_test, y_train, y_test = train_test_split(Input, Output, test_
```

In [91]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score as score
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegr
from sklearn.tree import DecisionTreeRegressor
from xgboost import XGBRegressor
from lightgbm import LGBMRegressor
from catboost import CatBoostRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.svm import SVR
```

In [92]:
```
def evaluate_model(y_true, y_pred, model_name):
    print("Model: ", model_name)
    print("Accuracy = {:0.2f}%".format(score(y_true, y_pred)*1000))
    print("Mean Squared Error = {:0.2f}\n".format(mean_squared_error(y_t
    return round(score(y_true, y_pred)*1000, 2)
```

In [94]:
```
LRScore = evaluate_model(y_test, lr_preds, "LINEAR REGRESSION")
RFScore = evaluate_model(y_test, rf_preds, "RANDOM FOREST")
DTScore = evaluate_model(y_test, dt_preds, "DECEISION TREE")
XGBScore = evaluate_model(y_test, xgb_preds, "EXTENDED GRADIENT BOOSTING
GBScore = evaluate_model(y_test, gb_preds, "GRADIENT BOOSTING")
LGBScore = evaluate_model(y_test, lgbm_preds, "LIGHT GRADIENT BOOSTING")
CBRScore = evaluate_model(y_test, catboost_preds, "CAT BOOST")
KNNScore = evaluate_model(y_test, knn_preds, "K NEAREST NEIGHBORS")
```

```
Model:  LINEAR REGRESSION
Accuracy = -64.53%
Mean Squared Error = 1.39

Model:  RANDOM FOREST
Accuracy = 354.06%
Mean Squared Error = 1.09

Model:  DECEISION TREE
Accuracy = -277.69%
Mean Squared Error = 1.53

Model:  EXTENDED GRADIENT BOOSTING
Accuracy = 348.12%
Mean Squared Error = 1.09

Model:  GRADIENT BOOSTING
Accuracy = 380.38%
Mean Squared Error = 1.06

Model:  LIGHT GRADIENT BOOSTING
Accuracy = 381.56%
Mean Squared Error = 1.06

Model:  CAT BOOST
Accuracy = 358.62%
Mean Squared Error = 1.08

Model:  K NEAREST NEIGHBORS
Accuracy = 30.25%
Mean Squared Error = 1.33
```

In [95]:

```
models = pd.DataFrame(
    {
        "MODELS": ["Linear Regression", "Random Forest", "Decision Tree"
        "SCORES": [LRScore, RFScore, DTScore, GBScore, XGBScore, LGBScor
    }
)
models.sort_values(by='SCORES', ascending=False)
```

Out[95]:

	MODELS	SCORES
5	Light Gradient Boosting	381.56
3	Gradient Boosting	380.38
6	Cat Boosting	358.62
1	Random Forest	354.06
4	Extended Gradient Boosting	348.12
7	K Nearest Neighbors	30.25
0	Linear Regression	-64.53
2	Decision Tree	-277.69