

```
In [492]: import pandas as pd
import numpy as np
import json
import matplotlib.pyplot as plt
import seaborn as sns
import collections
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression, LinearRegression
from sklearn import metrics
from sklearn.linear_model import Ridge
from sklearn.linear_model import RidgeCV
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import roc_auc_score
```

```
In [230]: tmdb_movies = pd.read_csv(r'C:\Users\preet\Desktop\Movies\tmdb_movies_data.csv')
```

Replacing 0's in budget with corresponding revenue and vice-versa

```
In [231]: tmdb_movies[(tmdb_movies['budget']==0)].shape[0]
```

```
Out[231]: 5696
```

```
In [232]: tmdb_movies[(tmdb_movies['revenue']==0)].shape[0]
```

```
Out[232]: 6016
```

```
In [233]: tmdb_movies.budget.replace(0,tmdb_movies.revenue,inplace=True)
```

```
In [234]: tmdb_test.revenue.replace(0,tmdb_test.budget,inplace=True)
```

```
In [235]: cast_actress = pd.read_csv(r'C:\Users\preet\Desktop\Movies\The_Best_Actresses_Ever.csv',encoding='latin-1')
cast_actor=pd.read_csv(r'C:\Users\preet\Desktop\Movies\The_Most_Popular_Actors.csv',encoding='latin-1')
```

```
In [236]: cast=pd.concat([cast_actress, cast_actor], axis=0)
```

```
In [237]: cast['Description']=cast['Description'].map(lambda x:x.rstrip('points'))
```

```
In [238]: cast['Description']=cast['Description'].astype(float)
```

```
In [239]: cast_new=cast[['Name','Description']]
```

```
In [240]: director=pd.read_csv(r'C:\Users\preet\Desktop\Movies\My Complete Directors Lis  
t.csv',encoding='latin-1')
```

In [241]: director

Out[241]:

	Position	Const	Created	Modified	Description	Name	Known For	Birth Date
0	1	nm0000229	29-07-12	29-09-16	65417 point	Steven Spielberg	Schindler's List	18-12-46
1	2	nm0634240	11-08-12	24-04-16	54135 point	Christopher Nolan	Inception	30-07-70
2	3	nm0000217	29-07-12	28-02-14	38209 point	Martin Scorsese	Taxi Driver	17-11-42
3	4	nm0000233	05-08-12	21-01-16	36893 point	Quentin Tarantino	Reservoir Dogs	27-03-63
4	5	nm0000399	06-08-12	28-02-15	34425 point	David Fincher	Se7en	28-08-62
5	6	nm0000040	27-07-12	11-08-12	28290 point	Stanley Kubrick	2001: A Space Odyssey	26-07-28
6	7	nm0000709	04-08-12	19-07-16	27236 point	Robert Zemeckis	Back to the Future	14-05-51
7	8	nm0000142	03-08-12	15-07-17	26647 point	Clint Eastwood	Million Dollar Baby	31-05-30
8	9	nm0000631	11-08-12	02-01-16	25778 point	Ridley Scott	The Martian	30-11-37
9	10	nm0000338	29-07-12	15-01-17	24184 point	Francis Ford Coppola	Apocalypse Now	07-04-39
10	11	nm0001054	06-08-12	15-01-17	23710 point	Joel Coen	The Ballad of Buster Scruggs	29-11-54
11	12	nm0000116	29-08-14	08-04-15	21103 point	James Cameron	Avatar	16-08-54
12	13	nm0000033	26-07-12	17-01-16	20849 point	Alfred Hitchcock	Psycho	1899-08-13
13	14	nm0001104	05-08-12	11-08-12	20079 point	Frank Darabont	The Green Mile	28-01-59
14	15	nm0000318	06-08-12	25-10-15	16623 point	Tim Burton	Edward Scissorhands	25-08-58
15	16	nm0005222	11-08-12	24-12-12	15812 point	Sam Mendes	Skyfall	01-08-65
16	17	nm0000095	29-07-12	24-02-17	15604 point	Woody Allen	Manhattan	01-12-35
17	18	nm0001752	11-08-12	06-12-15	15493 point	Steven Soderbergh	Sex, Lies, and Videotape	14-01-63
18	19	nm0001716	11-01-13	10-10-17	14932 point	Tony Scott	Domino	21-06-44
19	20	nm0004716	11-08-12	22-12-12	13943 point	Darren Aronofsky	Pi	12-02-69
20	21	nm0000965	18-08-12	19-03-16	13928 point	Danny Boyle	127 Hours	20-10-56
21	22	nm0000165	06-08-12	10-07-17	13817 point	Ron Howard	Arrested Development	01-03-54

	Position	Const	Created	Modified	Description	Name	Known For	Birth Date
22	23	nm0449984	30-09-16	04-07-17	13643 point	Irvin Kershner	Star Wars: Episode V - The Empire Strikes Back	29-04-23
23	24	nm0000591	28-07-12	04-01-14	12689 point	Roman Polanski	Chinatown	18-08-33
24	25	nm0000154	06-08-12	01-06-17	12663 point	Mel Gibson	Braveheart	03-01-56
25	26	nm0327944	11-08-12	04-02-16	12188 point	Alejandro G. Iñárritu	Biutiful	15-08-63
26	27	nm0005363	30-05-13	18-10-14	11708 point	Guy Ritchie	Sherlock Holmes	10-09-68
27	28	nm0905152	08-12-12	11-09-15	10683 point	Lilly Wachowski	The Matrix	29-12-67
28	29	nm0905154	08-12-12	11-09-15	10683 point	Lana Wachowski	The Matrix	21-06-65
29	30	nm0001232	03-08-12	21-12-13	10597 point	Milos Forman	One Flew Over the Cuckoo's Nest	18-02-32
...
2573	2574	nm0314777	26-07-12	26-07-12	0.5 point	Kurt Gerron	Der blaue Engel	1897-05-11
2574	2575	nm0664846	15-07-15	15-07-15	0.5 point	Jerzy Passendorfer	Powrót	08-04-23
2575	2576	nm0678959	05-08-12	05-08-12	0.5 point	Claus Peymann	Meine Mutter macht Mist mit mir	07-06-37
2576	2577	nm0906056	19-08-12	19-08-12	0.5 point	Petra Katharina Wagner	Staub vor der Sonne	NaN
2577	2578	nm0761400	05-08-12	05-08-12	0.5 point	Helke Sander	Nr. 1 - Aus Berichten der Wach- und Patrouille...	31-01-37
2578	2579	nm2889953	19-08-12	19-08-12	0.5 point	Kuba Czekaj	Baby Bump	18-06-84
2579	2580	nm0124044	05-08-12	05-08-12	0.5 point	Christel Buschmann	Gibbi Westgermany	19-03-42
2580	2581	nm0794638	26-07-12	26-07-12	0.5 point	Vladimir Shnejderov	Dzhulbars	08-07-00
2581	2582	nm0414985	28-07-12	28-07-12	0.5 point	Georg Jacoby	Maske in Blau	1883-07-23
2582	2583	nm0475360	04-08-12	04-08-12	0.5 point	Reiner Kunze	Die wunderbaren Jahre	16-08-33
2583	2584	nm0237897	04-08-12	04-08-12	0.5 point	Burkhard Driest	Annas Mutter	28-04-39

	Position	Const	Created	Modified	Description	Name	Known For	Birth Date
2584	2585	nm1127315	11-08-12	11-08-12	0.5 point	Clarissa Ruge	A Woman and a Half: Hildegard Knef	NaN
2585	2586	nm0880880	03-08-12	03-08-12	0.5 point	Fritz Umgelter	Bratkartoffeln inbegriffen	18-08-22
2586	2587	nm0903088	04-08-12	04-08-12	0.5 point	Rolf von Sydow	Tatort	18-06-24
2587	2588	nm0583311	18-08-12	18-08-12	0.5 point	Paul Meyer	Der Hauptmann von Muffrika	NaN
2588	2589	nm1265135	11-08-12	11-08-12	0.5 point	Javad Ardakani	Choori	NaN
2589	2590	nm0434071	12-08-12	12-08-12	0.5 point	Christine Kabisch	Va Banque	NaN
2590	2591	nm0595436	12-08-12	12-08-12	0.5 point	Felix Moeller	Harlan - Im Schatten von Jud Süß	NaN
2591	2592	nm0297502	18-08-12	18-08-12	0.5 point	Bodo Fürneisen	Der Staatsanwalt hat das Wort	30-06-50
2592	2593	nm0650568	26-08-12	26-08-12	0.5 point	Joseph Orr	Ninas Geschichte	NaN
2593	2594	nm4186195	26-08-12	26-08-12	0.5 point	Marta Minorowicz	Zud	05-06-79
2594	2595	nm0938352	04-08-12	04-08-12	0.5 point	Rainer Wolffhardt	Berliner Antigone	27-08-27
2595	2596	nm0926801	12-08-12	12-08-12	0.5 point	Karsten Wichniarz	Blue Moon	NaN
2596	2597	nm0589865	29-07-12	29-07-12	0.5 point	Willy Millowitsch	National Lampoon's European Vacation	08-01-09
2597	2598	nm1322631	12-08-12	12-08-12	0.5 point	Tanya Detkina	Pakostnik	NaN
2598	2599	nm0160189	12-08-12	12-08-12	0.5 point	Kai Christiansen	Der Traum von der neuen Welt	NaN
2599	2600	nm0903287	06-08-12	12-08-12	0.5 point	Irmgard von zur Mühlen	Die Frauen des 20. Juli	22-04-36
2600	2601	nm0486237	11-08-12	12-08-12	0.5 point	Irene Langemann	Rubljovka - Straße zur Glückseligkeit	NaN
2601	2602	nm0117714	12-08-12	12-08-12	0.5 point	Jürgen Brühns	Heimatfront	NaN
2602	2603	nm0902234	04-08-12	12-08-12	0.5 point	Alexander von Eschwege	Der Kandidat	15-09-48

2603 rows × 8 columns

```
In [242]: director[director['Description']== '729 üpom']
```

```
Out[242]:
```

Position	Const	Created	Modified	Description	Name	Known For	Birth Date
----------	-------	---------	----------	-------------	------	-----------	------------

```
In [243]: director=director.drop(director.index[507])
director['Description']=director['Description'].map(lambda x:x.rstrip('point'))
director['Description']=director['Description'].astype(float)
```

```
In [244]: director_new=director[['Description','Name']]
director_new=director_new.rename(columns={"Name":"director"})
```

```
In [245]: production=pd.read_csv(r'C:\Users\preet\Desktop\Movies\production_companies_data.csv')
```

```
In [246]: production['Total Domestic Box Office']=production['Total Domestic Box Office'].map(lambda x:x.lstrip('$'))
production['Total Domestic Box Office']=production['Total Domestic Box Office'].replace('\,',',')
production['Total Domestic Box Office']=production['Total Domestic Box Office'].apply(lambda x: x.replace(',',''))
production['Total Worldwide Box Office']=production['Total Worldwide Box Office'].map(lambda x:x.lstrip('$'))
production['Total Worldwide Box Office']=production['Total Worldwide Box Office'].apply(lambda x: x.replace(',',''))
```

```
In [247]: from sklearn import preprocessing

min_max_scaler = preprocessing.MinMaxScaler()
scaled_array = min_max_scaler.fit_transform(production[['No. of Movies', 'Total
Domestic Box Office', 'Total Worldwide Box Office']])
df_normalized = pd.DataFrame(scaled_array)
df_normalized
```


Out[247]:

	0	1	2
0	1.000000	1.000000	1.000000
1	0.986957	0.923446	0.960298
2	0.965217	0.941385	0.911028
3	0.534783	0.825408	0.862746
4	0.252174	0.703369	0.781542
5	0.569565	0.639794	0.653083
6	0.447826	0.534664	0.571998
7	0.300000	0.341818	0.376814
8	0.247826	0.341801	0.371694
9	0.191304	0.306640	0.346403
10	0.495652	0.392068	0.345390
11	0.273913	0.321250	0.330546
12	0.100000	0.325886	0.328424
13	0.447826	0.333782	0.321125
14	0.330435	0.305086	0.304401
15	0.352174	0.362181	0.303253
16	0.621739	0.316484	0.289477
17	0.386957	0.271880	0.274646
18	0.100000	0.190480	0.248706
19	0.095652	0.278760	0.238109
20	0.195652	0.214803	0.224526
21	0.095652	0.194118	0.213989
22	0.665217	0.265856	0.211935
23	0.304348	0.182148	0.188404
24	0.265217	0.228753	0.182617
25	0.408696	0.130114	0.170625
26	0.134783	0.148566	0.168874
27	0.086957	0.168023	0.163546
28	0.143478	0.130606	0.157746
29	0.169565	0.132561	0.149439
...
470	0.004348	0.013766	0.000502
471	0.047826	0.011973	0.000494
472	0.000000	0.007548	0.000455
473	0.004348	0.007762	0.000433

	0	1	2
474	0.013043	0.014040	0.000433
475	0.265217	0.002158	0.000428
476	0.013043	0.011452	0.000416
477	0.013043	0.011511	0.000413
478	0.004348	0.000673	0.000364
479	0.017391	0.012604	0.000360
480	0.008696	0.000035	0.000333
481	0.004348	0.017612	0.000256
482	0.000000	0.007675	0.000221
483	0.013043	0.010705	0.000217
484	0.000000	0.008790	0.000193
485	0.013043	0.011744	0.000108
486	0.095652	0.016643	0.000057
487	0.000000	0.004341	0.000048
488	0.000000	0.004341	0.000048
489	0.004348	0.000043	0.000012
490	0.000000	0.000038	0.000000
491	0.000000	0.000038	0.000000
492	0.000000	0.000038	0.000000
493	0.000000	0.000038	0.000000
494	0.000000	0.000038	0.000000
495	0.000000	0.000038	0.000000
496	0.000000	0.000038	0.000000
497	0.000000	0.000038	0.000000
498	0.000000	0.000038	0.000000
499	0.000000	0.000038	0.000000

500 rows × 3 columns

```
In [248]: production['Total Domestic Box Office']=production['Total Domestic Box Office']
          .astype('float')
          production['Total Worldwide Box Office']=production['Total Worldwide Box Office']
          .astype('float')
```

```
In [249]: #production['Score']=production['No. of Movies']*(0.01)+production['Total Domestic Box Office']*(0.005)+production['Total Worldwide Box Office']*(0.008)
          production['Score']=production['Total Worldwide Box Office']/(100000000) + production['No. of Movies']
```

In [250]: production

Out[250]:

	Production Companies	No. of Movies	Total Domestic Box Office	Total Worldwide Box Office	Score
0	Warner Bros.	231	1.845186e+10	4.303795e+10	661.379510
1	Universal Pictures	228	1.703929e+10	4.134544e+10	641.454416
2	Columbia Pictures	223	1.737031e+10	3.924508e+10	615.450787
3	Walt Disney Pictures	124	1.523031e+10	3.718680e+10	495.867979
4	Marvel Studios	59	1.297846e+10	3.372509e+10	396.250863
5	Paramount Pictures	132	1.180539e+10	2.824885e+10	414.488522
6	20th Century Fox	104	9.865534e+09	2.479222e+10	351.922163
7	Dune Entertainment	70	6.307178e+09	1.647153e+10	234.715337
8	Legendary Pictures	58	6.306866e+09	1.625327e+10	220.532699
9	DreamWorks Animation	45	5.658073e+09	1.517509e+10	196.750862
10	Relativity Media	115	7.234386e+09	1.513189e+10	266.318944
11	Amblin Entertainment	64	5.927656e+09	1.449909e+10	208.990882
12	Disney-Pixar	24	6.013201e+09	1.440867e+10	168.086675
13	New Line Cinema	104	6.158901e+09	1.409748e+10	244.974836
14	Village Roadshow Productions	77	5.629397e+09	1.338454e+10	210.845371
15	DreamWorks Pictures	82	6.682910e+09	1.333559e+10	215.355867
16	Regency Enterprises	144	5.839723e+09	1.274833e+10	271.483345
17	Metro-Goldwyn-Mayer Pictures	90	5.016691e+09	1.211610e+10	211.160982
18	Heyday Films	24	3.514718e+09	1.101027e+10	134.102663
19	Lucasfilm	23	5.143633e+09	1.055850e+10	128.584973
20	RatPac Entertainment	46	3.963516e+09	9.979481e+09	145.794809
21	Walt Disney Animation Studios	23	3.581836e+09	9.530293e+09	118.302928
22	Lionsgate	154	4.905540e+09	9.442724e+09	248.427241
23	Summit Entertainment	71	3.360967e+09	8.439587e+09	155.395870
24	Touchstone Pictures	62	4.220914e+09	8.192895e+09	143.928950
25	Working Title Films	95	2.400837e+09	7.681660e+09	171.816604
26	di Bonaventura Pictures	32	2.741315e+09	7.607033e+09	108.070332
27	Jerry Bruckheimer	21	3.100345e+09	7.379908e+09	94.799076
28	Original Film	34	2.409932e+09	7.132645e+09	105.326452
29	TSG Entertainment	40	2.445990e+09	6.778502e+09	107.785020
...
470	John Davis	2	2.540145e+08	4.293145e+08	6.293145
471	Alliance Films	12	2.209178e+08	4.290008e+08	16.290008
472	EON Enterprises	1	1.392709e+08	4.273003e+08	5.273003
473	Huahua Film	2	1.432324e+08	4.263708e+08	6.263708

	Production Companies	No. of Movies	Total Domestic Box Office	Total Worldwide Box Office	Score
474	Runteldat Entertainment	4	2.590636e+08	4.263621e+08	8.263621
475	France Television	62	3.981157e+07	4.261601e+08	66.261601
476	Karz Entertainment	4	2.113045e+08	4.256546e+08	8.256546
477	Conundrum Entertainment	4	2.123942e+08	4.255481e+08	8.255481
478	Tokuma Shoten	2	1.242399e+07	4.234250e+08	6.234250
479	Lawrence Bender	5	2.325627e+08	4.232686e+08	9.232686
480	Dream Sky Pictures	3	6.369090e+05	4.221315e+08	7.221315
481	Jack Giarraputo	2	3.249714e+08	4.188329e+08	6.188329
482	Silverbell Films	1	1.416215e+08	4.173413e+08	5.173413
483	Wayne Rice	4	1.975232e+08	4.171734e+08	8.171734
484	Captivate Entertainment	1	1.621929e+08	4.161683e+08	5.161683
485	Mirage Enterprises	4	2.166996e+08	4.125114e+08	8.125114
486	LD Entertainment	23	3.070895e+08	4.103642e+08	27.103642
487	Conspiracy Factory	1	8.010112e+07	4.099539e+08	5.099539
488	Secret Hideout	1	8.010112e+07	4.099539e+08	5.099539
489	Asia Pacific Films	2	7.862700e+05	4.084315e+08	6.084315
490	Xiu Shanfang Culture and Art Company	1	7.065720e+05	4.079233e+08	5.079233
491	Beijing Huaming Xingkong International Media	1	7.065720e+05	4.079233e+08	5.079233
492	Zhongzhi Enterprise Group	1	7.065720e+05	4.079233e+08	5.079233
493	Sichuan Aviation Culture Media Advertising Com...	1	7.065720e+05	4.079233e+08	5.079233
494	Shanghai Fuyi Media	1	7.065720e+05	4.079233e+08	5.079233
495	Zhejiang Bona	1	7.065720e+05	4.079233e+08	5.079233
496	CAAC News	1	7.065720e+05	4.079233e+08	5.079233
497	Chong Qing Film Group	1	7.065720e+05	4.079233e+08	5.079233
498	Lamb Film	1	7.065720e+05	4.079233e+08	5.079233
499	Yunnan Jincai Shijie Pictures	1	7.065720e+05	4.079233e+08	5.079233

500 rows × 5 columns

```
In [251]: production=production.rename(columns={"Production Companies":"Name"})
```

```
In [252]: mydict_prod = dict(zip(production.Name, production.Score))
mydict_cast = dict(zip(cast_new.Name, cast_new.Description))

tmdb_movies['cast'] = tmdb_movies['cast'].astype('str')
```

```
In [253]: l = []
for index, row in tmdb_movies.iterrows():

    s = row['cast']
    x= s.split('|')
    sum =0
    for i in x:
        if i in mydict_cast:
            sum = sum + mydict_cast.get(i)
        else:
            sum = sum + 0
    l.append(sum)

tmdb_movies['cast_points'] = pd.Series(l)
```

```
In [254]: tmdb_movies['production_points'] = 0

tmdb_movies['production_companies'] = tmdb_movies['production_companies'].astype('str')

l = []
for index, row in tmdb_movies.iterrows():

    s = row['production_companies']
    x=s.split('|')
    sum =0
    for i in x:
        if i in mydict_prod:
            sum = sum + mydict_prod.get(i)
        else:
            sum = sum + 0
    l.append(sum)

tmdb_movies['production_points'] = pd.Series(l)
```

In [255]: `tmdb_movies`

Out[255]:

	id	imdb_id	popularity	budget	revenue	original_title	cas
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vin Diesel
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays Byrne Nicholas Hoult
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theodore James Kate Winslet Ansel Elgort
3	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam Driver
4	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle Rodriguez
5	281957	tt1663202	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Will Poulter Domhnall Gleeson
6	87101	tt1340138	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clarke
7	286217	tt3659388	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff Bridges
8	211672	tt2293640	7.404165	74000000	1156730962	Minions	Sandra Bullock Jordan Peele Michael Keaton Allison Janney
9	150540	tt2096673	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Hader
10	206647	tt2379713	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz Léa Seydoux Ralph Fiennes
11	76757	tt1617661	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sebastian Bean Eddie Redmayne
12	264660	tt0470752	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac Sofia Boutella
13	257344	tt2120120	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Pete Dinklage Tim Allen
14	99861	tt2395427	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo

	id	imdb_id	popularity	budget	revenue	original_title	cas
15	273248	tt3460252	5.898400	44000000	155760117	The Hateful Eight	Samuel L Jackson Kurt Russell Jennifer Jason ..
16	260346	tt2446042	5.749758	48000000	325771424	Taken 3	Liam Neeson Forest Whitaker Maggie Grace Famke ..
17	102899	tt0478970	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michael Douglas Evangeline Lilly Cor..
18	150689	tt1661199	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen..
19	131634	tt1951266	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennifer Lawrence Josh Hutcherson Liam Hemsworth..
20	158852	tt1964418	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ..
21	307081	tt1798684	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rachel McAdams Forest Whitaker..
22	254128	tt2126355	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carlos Guejano..
23	216015	tt2322441	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennifer Ehle Eloi..
24	318846	tt1596363	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Gosling Brad ..
25	177677	tt2381249	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Fe..
26	214756	tt2637276	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ..
27	207703	tt2802144	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L Jackson Mi..
28	314365	tt1895587	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michael Keaton Rachel McAdams Lie..
29	294254	tt4046784	3.968891	61000000	311256926	Maze Runner: The Scorch Trials	Dylan O'Brien Kaya Scodelario Thomas Brodie-Sa..
...
10836	38720	tt0061170	0.239435	0	0	Walk Don't Run	Cary Grant Samantha Eggar Jim Hutton John Stan..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10837	19728	tt0060177	0.291704	0	0	The Blue Max	George Peppard James Mason Ursula Andress Jere...
10838	22383	tt0060862	0.151845	0	0	The Professionals	Burt Lancaster Lee Marvin Rober Ryan Woody St...
10839	13353	tt0060550	0.276133	0	0	It's the Great Pumpkin, Charlie Brown	Christopher Shea Sally Dryer Kathy Steinberg A...
10840	34388	tt0060437	0.102530	0	0	Funeral in Berlin	Michael Caine Paul Hubschmid Oskar Homolka Eva...
10841	42701	tt0062262	0.264925	75000	0	The Shooting	Will Hutchins Millie Perkins Jack Nicholson Wa...
10842	36540	tt0061199	0.253437	0	0	Winnie the Pooh and the Honey Tree	Sterling Holloway Junius Matthews Sebastian Ca...
10843	29710	tt0060588	0.252399	0	0	Khartoum	Charlton Heston Laurence Olivier Richard Johns...
10844	23728	tt0059557	0.236098	0	0	Our Man Flint	James Coburn Lee J Cobb Gil Golan Edward Mul...
10845	5065	tt0059014	0.230873	0	0	Carry On Cowboy	Sid James Jim Dale Angela Douglas Kenneth Will...
10846	17102	tt0059127	0.212716	0	0	Dracula: Prince of Darkness	Christopher Lee Barbara Shelley Andrew Keir Fr...
10847	28763	tt0060548	0.034555	0	0	Island of Terror	Peter Cushing Edward Judd Carole Gray Eddie By...
10848	2161	tt0060397	0.207257	5115000	12000000	Fantastic Voyage	Stephen Boyd Raque Welch Edmond O'Brien Donal...
10849	28270	tt0060445	0.206537	0	0	Gambit	Michael Caine Shirley MacLaine Herbert Lom Joh...
10850	26268	tt0060490	0.202473	0	0	Harper	Paul Newman Laurie Bacall Julie Harris Arthur ..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10851	15347	tt0060182	0.342791	0	0	Born Free	Virginia McKenna Bill Travers Geoffrey Keen Pe..
10852	37301	tt0060165	0.227220	0	0	A Big Hand for the Little Lady	Henry Fonda Joanne Woodward Jasor Robards Paul..
10853	15598	tt0060086	0.163592	0	0	Alfie	Michael Caine Shelley Winters Millicen Martin..
10854	31602	tt0060232	0.146402	0	0	The Chase	Marlon Brando Jane Fonda Rober Redford E.G. M..
10855	13343	tt0059221	0.141026	700000	0	The Ghost & Mr. Chicken	Don Knotts Joar Staley Liam Redmond Dick Sarge..
10856	20277	tt0061135	0.140934	0	0	The Ugly Dachshund	Dean Jones Suzanne Pleshette Charles Ruggles K..
10857	5921	tt0060748	0.131378	0	0	Nevada Smith	Steve McQueen Kar Malden Briar Keith Arthur K..
10858	31918	tt0060921	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner Eva Marie Saint Alan Arkin Briar K..
10859	20620	tt0060955	0.089072	0	0	Seconds	Rock Hudson Salome Jens Johr Randolph Will Gee..
10860	5060	tt0060214	0.087034	0	0	Carry On Screaming!	Kenneth Williams Jirr Dale Harry H Corbett Joa..
10861	21	tt0060371	0.080598	0	0	The Endless Summer	Michael Hynson Rober August Lord 'Tally Ho B..
10862	20379	tt0060472	0.065543	0	0	Grand Prix	James Garner Eva Marie Saint Yves Montand Tosh..
10863	39768	tt0060161	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy Oleg Efremov Georgi Z..
10864	21449	tt0061177	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi Akiko Wakabayashi Mie Hama Joh..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10865	22293	tt0060666	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren Tor Neyman Johr Reynolds Dian..

10866 rows × 23 columns

```
In [256]: new_tmdb=pd.merge(tmdb_movies,director_new,on='director',how='left')
```

```
In [257]: new_tmdb = new_tmdb.rename(columns = {"Description":"Director_points"})
```

Genres

```
In [258]: def get_decade_by_year(year):
           decade = year - (year%10)
           decade_plus_10 = decade +10
           decade_as_str = str(decade)+'-'+ str(decade_plus_10)
           return decade_as_str
```

```
In [259]: df_copy = new_tmdb.copy()
           year_of_release = []
           decade_of_release = []
           month_of_release = []
           for x in range(len(df_copy)):
               year = int(df_copy.iloc[x]['release_date'][-4:])
               month = int(df_copy.iloc[x]['release_date'].split('/')[0])
               year_of_release.append(year)
               decade_of_release.append(get_decade_by_year(year))
               month_of_release.append(month)
           df_copy['year_of_release'] = year_of_release
           df_copy['decade_of_release'] = decade_of_release
           df_copy['month_of_release'] = month_of_release
```

```
In [260]: df_copy_for_genres = df_copy.copy()
```

```
In [261]: import collections
           genre_by_decade = collections.OrderedDict()
           for i in range(len(df_copy_for_genres)):
               decade = df_copy_for_genres.iloc[i]['decade_of_release']
               if decade not in genre_by_decade:
                   genre_by_decade[decade] = {}
               if type(df_copy_for_genres.iloc[i]['genres'])==str:
                   genres_as_list = df_copy_for_genres.iloc[i]['genres'].split('|')
                   for each_genre_in_movie in genres_as_list:
                       if each_genre_in_movie not in genre_by_decade[decade]:
                           genre_by_decade[decade][each_genre_in_movie] =1
                       else:
                           genre_by_decade[decade][each_genre_in_movie]+=1
```

```
In [488]: df = pd.DataFrame(genre_by_decade)
cols= sorted(list(df.columns))
df[cols]
```

Out[488]:

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010	2010-2020
Action	78	123	275	457	782	681
Adventure	64	79	177	278	506	376
Animation	16	17	38	79	297	254
Comedy	113	122	429	738	1357	1044
Crime	43	83	153	270	468	341
Documentary	2	13	13	23	144	325
Drama	168	240	424	865	1621	1458
Family	39	39	105	274	472	305
Fantasy	23	30	124	189	320	233
Foreign	9	3	8	32	103	33
History	31	24	34	53	115	79
Horror	47	104	222	189	486	594
Music	21	25	52	49	122	141
Mystery	32	56	72	149	284	220
Romance	67	51	175	341	636	445
Science Fiction	34	78	174	213	373	366
TV Movie	2	8	9	19	53	76
Thriller	64	149	262	497	980	969
War	31	25	33	29	88	65
Western	36	36	14	25	26	28

```
In [499]: df_with_scores = df.copy()
columns = list(df.columns)
for column in columns:
    df_with_scores[column] = 100*df_with_scores[column] / df_with_scores[column].sum()
cols= sorted(list(df.columns))
df_with_scores[cols]
```

Out[499]:

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010	2010-2020
Action	8.478261	9.425287	9.846044	9.582722	8.469620	8.477530
Adventure	6.956522	6.053640	6.337272	5.829314	5.480342	4.680692
Animation	1.739130	1.302682	1.360544	1.656532	3.216723	3.161957
Comedy	12.282609	9.348659	15.359828	15.474942	14.697281	12.996390
Crime	4.673913	6.360153	5.477981	5.661564	5.068775	4.244989
Documentary	0.217391	0.996169	0.465449	0.482281	1.559623	4.045811
Drama	18.260870	18.390805	15.180809	18.137974	17.556590	18.150131
Family	4.239130	2.988506	3.759398	5.745439	5.112098	3.796838
Fantasy	2.500000	2.298851	4.439671	3.963095	3.465829	2.900535
Foreign	0.978261	0.229885	0.286430	0.671000	1.115564	0.410805
History	3.369565	1.839080	1.217329	1.111344	1.245532	0.983443
Horror	5.108696	7.969349	7.948443	3.963095	5.263728	7.394498
Music	2.282609	1.915709	1.861797	1.027469	1.321347	1.755260
Mystery	3.478261	4.291188	2.577873	3.124345	3.075923	2.738703
Romance	7.282609	3.908046	6.265664	7.150346	6.888335	5.539649
Science Fiction	3.695652	5.977011	6.229860	4.466345	4.039857	4.556206
TV Movie	0.217391	0.613027	0.322234	0.398406	0.574028	0.946097
Thriller	6.956522	11.417625	9.380594	10.421472	10.614102	12.062741
War	3.369565	1.915709	1.181525	0.608094	0.953103	0.809162
Western	3.913043	2.758621	0.501253	0.524219	0.281599	0.348562

```
In [264]: df_with_scores_add_new_column = df_with_scores.copy()
```

```

In [265]: tmdb_with_genre_score = df_copy.copy()
genre_score = []
current_genre_score = []

for i in range(len(tmdb_with_genre_score)):
    if type(tmdb_with_genre_score.iloc[i]['genres'])==str:
        each_movie_genre_score =0
        each_movie_current_genre_score = 0

        decade_of_release = tmdb_with_genre_score.iloc[i]['decade_of_release']
        genres_as_list = df_copy_for_genres.iloc[i]['genres'].split('|')
        for each_genre_in_movie in genres_as_list:
            each_movie_genre_score+=df_with_scores_add_new_column[decade_of_re
lease][each_genre_in_movie]
            each_movie_current_genre_score+=df_with_scores_add_new_column['201
0-2020'][each_genre_in_movie]
        else:
            each_movie_genre_score = 0
            each_movie_current_genre_score = 0

        genre_score.append(each_movie_genre_score)
        current_genre_score.append(each_movie_current_genre_score)

tmdb_with_genre_score['genre_score_atRelease'] = genre_score
tmdb_with_genre_score['genre_score_atPresent'] = current_genre_score

```

```

In [266]: tmdb_with_genre_score.describe()

```

Out[266]:

	id	popularity	budget	revenue	runtime	vote_count	vc
count	10900.000000	10900.000000	1.090000e+04	1.090000e+04	10900.000000	10900.000000	10
mean	65980.661284	0.649498	1.631633e+07	3.995642e+07	102.101743	218.236789	
std	92054.450204	1.034034	3.192169e+07	1.170185e+08	31.358108	577.849608	
min	5.000000	0.000065	0.000000e+00	0.000000e+00	0.000000	10.000000	
25%	10587.750000	0.207759	0.000000e+00	0.000000e+00	90.000000	17.000000	
50%	20617.000000	0.384145	1.175280e+06	0.000000e+00	99.000000	38.000000	
75%	75480.250000	0.716003	2.000000e+07	2.420708e+07	111.000000	146.000000	
max	417859.000000	32.985763	4.250000e+08	2.781506e+09	900.000000	9767.000000	

```

In [267]: oscars = pd.read_csv(r'C:\Users\preet\Desktop\Movies\oscars.csv')

```

In [268]:

oscars

Out[268]:

	Year	Ceremony	Award	Winner	Name	original_title
0	1930/1931	4	Actor	NaN	Adolphe Menjou	The Front Page
1	1975	48	Actor	NaN	Al Pacino	Dog Day Afternoon
2	1973	46	Actor	NaN	Al Pacino	Serpico
3	1974	47	Actor	NaN	Al Pacino	The Godfather Part II
4	1968	41	Actor	NaN	Alan Arkin	The Heart Is a Lonely Hunter
5	1966	39	Actor	NaN	Alan Arkin	The Russians Are Coming The Russians Are Coming
6	1968	41	Actor	NaN	Alan Bates	The Fixer
7	1974	47	Actor	NaN	Albert Finney	Murder on the Orient Express
8	1963	36	Actor	NaN	Albert Finney	Tom Jones
9	1957	30	Actor	1.0	Alec Guinness	The Bridge on the River Kwai
10	1952	25	Actor	NaN	Alec Guinness	The Lavender Hill Mob
11	1944	17	Actor	NaN	Alexander Knox	Wilson
12	1931/1932	5	Actor	NaN	Alfred Lunt	The Guardsman
13	1957	30	Actor	NaN	Anthony Franciosa	A Hatful of Rain
14	1957	30	Actor	NaN	Anthony Quinn	Wild Is the Wind
15	1964	37	Actor	NaN	Anthony Quinn	Zorba the Greek
16	1974	47	Actor	1.0	Art Carney	Harry and Tonto
17	1951	24	Actor	NaN	Arthur Kennedy	Bright Victory
18	1944	17	Actor	NaN	Barry Fitzgerald	Going My Way
19	1944	17	Actor	1.0	Bing Crosby	Going My Way
20	1945	18	Actor	NaN	Bing Crosby	The Bells of St. Mary's
21	1954	27	Actor	NaN	Bing Crosby	The Country Girl
22	1949	22	Actor	1.0	Broderick Crawford	All the King's Men
23	1962	35	Actor	NaN	Burt Lancaster	Birdman of Alcatraz
24	1960	33	Actor	1.0	Burt Lancaster	Elmer Gantry
25	1953	26	Actor	NaN	Burt Lancaster	From Here to Eternity
26	1944	17	Actor	NaN	Cary Grant	None but the Lonely Heart
27	1941	14	Actor	NaN	Cary Grant	Penny Serenade
28	1938	11	Actor	NaN	Charles Boyer	Algiers
29	1937	10	Actor	NaN	Charles Boyer	Conquest
...

	Year	Ceremony	Award	Winner	Name	original_title
9934	1959	32	Writing (Story and Screenplay, Written Directl...	NaN	Operation Petticoat	Story by Paul King, Joseph Stone; Screenplay b...
9935	1959	32	Writing (Story and Screenplay, Written Directl...	1.0	Pillow Talk	Story by Russell Rouse, Clarence Greene; Scree...
9936	1961	34	Writing (Story and Screenplay, Written Directl...	1.0	Splendor in the Grass	William Inge
9937	1958	31	Writing (Story and Screenplay, Written Directl...	NaN	Teacher's Pet	Fay Kanin, Michael Kanin
9938	1964	37	Writing (Story and Screenplay, Written Directl...	NaN	That Man from Rio	Jean-Paul Rappeneau, Ariane Mnouchkine, Daniel...
9939	1962	35	Writing (Story and Screenplay, Written Directl...	NaN	That Touch of Mink	Stanley Shapiro, Nate Monaster
9940	1959	32	Writing (Story and Screenplay, Written Directl...	NaN	The 400 Blows	Francois Truffaut, Marcel Moussy
9941	1960	33	Writing (Story and Screenplay, Written Directl...	NaN	The Angry Silence	Story by Richard Gregson, Michael Craig; Scree...
9942	1960	33	Writing (Story and Screenplay, Written Directl...	1.0	The Apartment	Billy Wilder, I. A. L. Diamond
9943	1968	41	Writing (Story and Screenplay, Written Directl...	NaN	The Battle of Algiers	Franco Solinas, Gillo Pontecorvo
9944	1958	31	Writing (Story and Screenplay, Written Directl...	1.0	The Defiant Ones	Nedrick Young, Harold Jacob Smith
9945	1960	33	Writing (Story and Screenplay, Written Directl...	NaN	The Facts of Life	Norman Panama, Melvin Frank
9946	1966	39	Writing (Story and Screenplay, Written Directl...	NaN	The Fortune Cookie	Billy Wilder, I. A. L. Diamond
9947	1963	36	Writing (Story and Screenplay, Written Directl...	NaN	The Four Days of Naples	Story by Pasquale Festa Campanile, Massimo Fra...
9948	1958	31	Writing (Story and Screenplay, Written Directl...	NaN	The Goddess	Paddy Chayefsky
9949	1966	39	Writing (Story and Screenplay, Written Directl...	NaN	The Naked Prey	Clint Johnston, Don Peters
9950	1964	37	Writing (Story and Screenplay, Written Directl...	NaN	The Organizer	Age, Scarpelli, Mario Monicelli

	Year	Ceremony	Award	Winner	Name	original_title
9951	1968	41	Writing (Story and Screenplay, Written Directl...	1.0	The Producers	Mel Brooks
9952	1958	31	Writing (Story and Screenplay, Written Directl...	NaN	The Sheepman	Story by James Edward Grant; Screenplay by Wil...
9953	1957	30	Writing (Story and Screenplay, Written Directl...	NaN	The Tin Star	Story by Barney Slater, Joel Kane; Screenplay ...
9954	1965	38	Writing (Story and Screenplay, Written Directl...	NaN	The Train	Franklin Coen, Frank Davis
9955	1965	38	Writing (Story and Screenplay, Written Directl...	NaN	The Umbrellas of Cherbourg	Jacques Demy
9956	1965	38	Writing (Story and Screenplay, Written Directl...	NaN	Those Magnificent Men in Their Flying Machines	Jack Davies, Ken Annakin
9957	1962	35	Writing (Story and Screenplay, Written Directl...	NaN	Through a Glass Darkly	Ingmar Bergman
9958	1967	40	Writing (Story and Screenplay, Written Directl...	NaN	Two for the Road	Frederic Raphael
9959	1957	30	Writing (Story and Screenplay, Written Directl...	NaN	Vitelloni	Story by Federico Fellini, Ennio Flaiano, Tull...
9960	1959	32	Writing (Story and Screenplay, Written Directl...	NaN	Wild Strawberries	Ingmar Bergman
9961	1927/1928	1	Writing (Title Writing)	NaN	George Marion, Jr.	NaN
9962	1927/1928	1	Writing (Title Writing)	NaN	Gerald Duffy	The Private Life of Helen of Troy
9963	1927/1928	1	Writing (Title Writing)	1.0	Joseph Farnham	NaN

9964 rows × 6 columns

```
In [269]: oscars_new = oscars[['original_title', 'Winner']]
#director_new=director[['Description', 'Name']]
```

In [270]: `oscars_new.fillna(0)`

Out[270]:

	original_title	Winner
0	The Front Page	0.0
1	Dog Day Afternoon	0.0
2	Serpico	0.0
3	The Godfather Part II	0.0
4	The Heart Is a Lonely Hunter	0.0
5	The Russians Are Coming The Russians Are Coming	0.0
6	The Fixer	0.0
7	Murder on the Orient Express	0.0
8	Tom Jones	0.0
9	The Bridge on the River Kwai	1.0
10	The Lavender Hill Mob	0.0
11	Wilson	0.0
12	The Guardsman	0.0
13	A Hatful of Rain	0.0
14	Wild Is the Wind	0.0
15	Zorba the Greek	0.0
16	Harry and Tonto	1.0
17	Bright Victory	0.0
18	Going My Way	0.0
19	Going My Way	1.0
20	The Bells of St. Mary's	0.0
21	The Country Girl	0.0
22	All the King's Men	1.0
23	Birdman of Alcatraz	0.0
24	Elmer Gantry	1.0
25	From Here to Eternity	0.0
26	None but the Lonely Heart	0.0
27	Penny Serenade	0.0
28	Algiers	0.0
29	Conquest	0.0
...
9934	Story by Paul King, Joseph Stone; Screenplay b...	0.0
9935	Story by Russell Rouse, Clarence Greene; Scree...	1.0
9936	William Inge	1.0
9937	Fay Kanin, Michael Kanin	0.0

	original_title	Winner
9938	Jean-Paul Rappeneau, Ariane Mnouchkine, Daniel...	0.0
9939	Stanley Shapiro, Nate Monaster	0.0
9940	Francois Truffaut, Marcel Moussy	0.0
9941	Story by Richard Gregson, Michael Craig; Scree...	0.0
9942	Billy Wilder, I. A. L. Diamond	1.0
9943	Franco Solinas, Gillo Pontecorvo	0.0
9944	Nedrick Young, Harold Jacob Smith	1.0
9945	Norman Panama, Melvin Frank	0.0
9946	Billy Wilder, I. A. L. Diamond	0.0
9947	Story by Pasquale Festa Campanile, Massimo Fra...	0.0
9948	Paddy Chayefsky	0.0
9949	Clint Johnston, Don Peters	0.0
9950	Age, Scarpelli, Mario Monicelli	0.0
9951	Mel Brooks	1.0
9952	Story by James Edward Grant; Screenplay by Wil...	0.0
9953	Story by Barney Slater, Joel Kane; Screenplay ...	0.0
9954	Franklin Coen, Frank Davis	0.0
9955	Jacques Demy	0.0
9956	Jack Davies, Ken Annakin	0.0
9957	Ingmar Bergman	0.0
9958	Frederic Raphael	0.0
9959	Story by Federico Fellini, Ennio Flaiano, Tull...	0.0
9960	Ingmar Bergman	0.0
9961	0	0.0
9962	The Private Life of Helen of Troy	0.0
9963	0	1.0

9964 rows × 2 columns

```
In [271]: oscar_sum = oscars_new.groupby(['original_title']).agg({'Winner': 'sum'})
```

In [272]: oscar_sum

Out[272]:

	original_title	Winner
	"Charlie McCarthy"	1.0
'Made by Fred Zinnemann with the cooperation of Paramount Pictures Corporation for the Los Angeles Orthopaedic Hospital'		1.0
	'Round Midnight	0.0
	(Arthur Edeson)	0.0
(First National Studio Sound Department, George Groves, Sound Director)		0.0
	(Gaetano Gaudio), (Harry Perry)	0.0
	(George Abbott), (Maxwell Anderson), (Del Andrews)	0.0
	(Hans Dreier)	0.0
	(Howard Estabrook)	0.0
	(Jack Okey)	0.0
	(John Meehan)	0.0
	(Julian Josephson)	0.0
(Paramount Famous Lasky Studio Sound Department, Franklin Hansen, Sound Director)		0.0
	(RKO Radio Studio Sound Department, John Tribby, Sound Director)	0.0
(United Artists Studio Sound Department, Oscar Lagerstrom, Sound Director)		0.0
	(Victor Milner)	0.0
	(William Cameron Menzies)	0.0
	(William Daniels)	0.0
	...And Justice for All	0.0
	1000 Love Songs number from Cain and Mabel	0.0
	12 Monkeys	0.0
	12 Years a Slave	1.0
	127 Hours	0.0
	20th Century	0.0
	20th Century [Third]	0.0
	20th Century-Fox	3.0
	20th Century-Fox Studio	0.0
20th Century-Fox Studio Music Department, Louis Silvers, head of department (no composer credit)		0.0
	20th Century-Fox Studio Sound Department	0.0
20th Century-Fox Studio Sound Department, Carl Faulkner, Sound Director		1.0

	Written for the Screen by Paul Thomas Anderson	0.0
	Written for the screen by Joel Coen & Ethan Coen	1.0

	Winner
original_title	
Written for the screen by Paul Thomas Anderson	0.0
Written on the Wind	1.0
Wuthering Heights	0.0
Wylie Stateman	0.0
Wylie Stateman, Gregg Baxter	0.0
Xavier Legrand and Alexandre Gavras	0.0
Yan England	0.0
Yankee Doodle Dandy	1.0
Yee Chung Man	0.0
Yentl	0.0
Yoshiro Muraki	0.0
You Can Count on Me	0.0
You Can't Take It with You	0.0
You're a Big Boy Now	0.0
Yugoslavia	0.0
Yuri Faier, G. Rozhdestvensky	0.0
Yvon Mallette and Robert Verrall, Producers	0.0
Yvonne Blake, Antonio Castillo	1.0
Yvonne Blake, Ron Talsky	0.0
Zach Staenberg	1.0
Zachary Heinzerling and Lydia Dean Pilcher	0.0
Zagreb Film	1.0
Zbigniew Rybczynski, Producer	1.0
Zero Dark Thirty	0.0
Zhao Xiaoding	0.0
Zorba the Greek	1.0
dell Istituto Nazionale Luce, Comitato Organizzatore Del Giochi Della XVII Olimpiade	0.0
"Mickey Mouse"	1.0

6363 rows × 1 columns

```
In [273]: tmdb_with_oscar=pd.merge(tmdb_with_genre_score,oscar_sum,on='original_title',how='left')
```

In [274]: `tmdb_with_oscar`

Out[274]:

	id	imdb_id	popularity	budget	revenue	original_title	cas
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays Byrne Nic...
2	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays Byrne Nic...
3	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...
4	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...
5	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle ...
6	281957	tt1663202	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Wil Poulter Domhnall...
7	87101	tt1340138	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clar...
8	286217	tt3659388	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff ...
9	211672	tt2293640	7.404165	74000000	1156730962	Minions	Sandra Bullock Jordan Peele Michael Keaton Allison...
10	150540	tt2096673	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Hader...
11	206647	tt2379713	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz Léa Seydoux Ralph Fi...
12	76757	tt1617661	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sebastian Bean Eddie Redm...
13	264660	tt0470752	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac S...
14	257344	tt2120120	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Pete Dinklage ...
15	99861	tt2395427	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...

	id	imdb_id	popularity	budget	revenue	original_title	cas
16	273248	tt3460252	5.898400	44000000	155760117	The Hateful Eight	Samuel L Jackson Kurt Russell Jennifer Jason ..
17	260346	tt2446042	5.749758	48000000	325771424	Taken 3	Liam Neeson Forest Whitaker Maggie Grace Famke ..
18	102899	tt0478970	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michael Douglas Evangeline Lilly Corey ..
19	150689	tt1661199	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen ..
20	131634	tt1951266	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennifer Lawrence Josh Hutcherson Liam Hemsworth ..
21	158852	tt1964418	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ..
22	307081	tt1798684	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rache McAdams Forest Whitaker ..
23	254128	tt2126355	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carlos Guejuna ..
24	216015	tt2322441	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennifer Ehle Eloise ..
25	318846	tt1596363	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Reynolds Greg Kinnear ..
26	177677	tt2381249	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Ferguson ..
27	214756	tt2637276	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ..
28	207703	tt2802144	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L Jackson Michael ..
29	314365	tt1895587	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michael Keaton Rachel Watson Liev Schreiber ..
...
10870	38720	tt0061170	0.239435	0	0	Walk Don't Run	Cary Grant Samantha Eggar Jim Hutton John Stan ..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10871	19728	tt0060177	0.291704	0	0	The Blue Max	George Peppard James Mason Ursula Andress Jere...
10872	22383	tt0060862	0.151845	0	0	The Professionals	Burt Lancaster Lee Marvin Robert Ryan Woody St...
10873	13353	tt0060550	0.276133	0	0	It's the Great Pumpkin, Charlie Brown	Christopher Shea Sally Dryer Kathy Steinberg A...
10874	34388	tt0060437	0.102530	0	0	Funeral in Berlin	Michael Caine Paul Hubschmid Oskar Homolka Eva...
10875	42701	tt0062262	0.264925	75000	0	The Shooting	Will Hutchins Millie Perkins Jack Nicholson Wa...
10876	36540	tt0061199	0.253437	0	0	Winnie the Pooh and the Honey Tree	Sterling Holloway Junius Matthews Sebastian Ca...
10877	29710	tt0060588	0.252399	0	0	Khartoum	Charlton Heston Laurence Olivier Richard Johns...
10878	23728	tt0059557	0.236098	0	0	Our Man Flint	James Coburn Lee J. Cobb Gil Golan Edward Mul...
10879	5065	tt0059014	0.230873	0	0	Carry On Cowboy	Sid James Jim Dale Angela Douglas Kenneth Will...
10880	17102	tt0059127	0.212716	0	0	Dracula: Prince of Darkness	Christopher Lee Barbara Shelley Andrew Keir Fr...
10881	28763	tt0060548	0.034555	0	0	Island of Terror	Peter Cushing Edward Judd Carole Gray Eddie By...
10882	2161	tt0060397	0.207257	5115000	12000000	Fantastic Voyage	Stephen Boyd Raque Welch Edmond O'Brien Donal...
10883	28270	tt0060445	0.206537	0	0	Gambit	Michael Caine Shirley MacLaine Herbert Lom Joh...
10884	26268	tt0060490	0.202473	0	0	Harper	Paul Newman Laurer Bacall Julie Harris Arthur ..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10885	15347	tt0060182	0.342791	0	0	Born Free	Virginia McKenna Bill Travers Geoffrey Keen Pe..
10886	37301	tt0060165	0.227220	0	0	A Big Hand for the Little Lady	Henry Fonda Joanne Woodward Jasor Robards Paul..
10887	15598	tt0060086	0.163592	0	0	Alfie	Michael Caine Shelley Winters Millicen Martin..
10888	31602	tt0060232	0.146402	0	0	The Chase	Marlon Brando Jane Fonda Rober Redford E.G. M..
10889	13343	tt0059221	0.141026	700000	0	The Ghost & Mr. Chicken	Don Knotts Joar Staley Lian Redmond Dick Sarge..
10890	20277	tt0061135	0.140934	0	0	The Ugly Dachshund	Dean Jones Suzanne Pleshette Charles Ruggles K..
10891	5921	tt0060748	0.131378	0	0	Nevada Smith	Steve McQueen Kar Malden Briar Keith Arthur K..
10892	31918	tt0060921	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner Eva Marie Saint Alan Arkin Briar K..
10893	20620	tt0060955	0.089072	0	0	Seconds	Rock Hudson Salome Jens Johr Randolph Will Gee..
10894	5060	tt0060214	0.087034	0	0	Carry On Screaming!	Kenneth Williams Jirr Dale Harry H Corbett Joa..
10895	21	tt0060371	0.080598	0	0	The Endless Summer	Michael Hynson Rober August Lord 'Tally Ho B..
10896	20379	tt0060472	0.065543	0	0	Grand Prix	James Garner Eva Marie Saint Yves Montand Tosh..
10897	39768	tt0060161	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy Oleg Efremov Georgi Z..
10898	21449	tt0061177	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi Akiko Wakabayashi Mie Hama Joh..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10899	22293	tt0060666	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren Tom Neyman John Reynolds Dian..

10900 rows × 30 columns

```
In [275]: tmdb_with_oscar=tmdb_with_oscar.rename(columns={"Winner":"oscars_won"})
```

```
In [276]: tmdb_with_oscar.isna().sum()
```

```
Out[276]: id                                0
imdb_id                                10
popularity                             0
budget                                 0
revenue                                0
original_title                         0
cast                                   0
homepage                             7956
director                              44
tagline                              2829
keywords                             1497
overview                              4
runtime                                0
genres                                23
production_companies                  0
release_date                          0
vote_count                            0
vote_average                          0
release_year                          0
budget_adj                            0
revenue_adj                           0
cast_points                           0
production_points                     0
Director_points                       6738
year_of_release                       0
decade_of_release                     0
month_of_release                      0
genre_score_atRelease                 0
genre_score_atPresent                 0
oscars_won                           10268
dtype: int64
```

Inflation

```
In [277]: inflation = pd.read_csv(r'C:\Users\preet\Desktop\Movies\infla.csv')
inflation['DATE'] = pd.to_datetime(inflation['DATE'])
inflation["year"] = inflation["DATE"].dt.year
inflation["month"] = inflation["DATE"].dt.month
```

```
In [278]: inflated_budget = []
inflated_revenue = []
inflated_profit = []

for i in range(len(tmdb_with_oscar)):
    cpi = inflation.loc[inflation.year == tmdb_with_oscar.iloc[i].year_of_release][inflation.month == tmdb_with_oscar.iloc[i].month_of_release]['CPIAUCNS_NB19130101'].item()
    inflated_budget.append(tmdb_with_oscar.iloc[i].budget/cpi)
    inflated_revenue.append(tmdb_with_oscar.iloc[i].revenue/cpi)

tmdb_with_oscar['inflated_budget'] = inflated_budget
tmdb_with_oscar['inflated_revenue'] = inflated_revenue
tmdb_with_oscar['inflated_profit'] = tmdb_with_oscar['inflated_revenue'] - tmdb_with_oscar['inflated_budget']
```

C:\Users\preet\Anaconda3\lib\site-packages\ipykernel_launcher.py:6: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

In [280]: `tmdb_with_oscar`

Out[280]:

	id	imdb_id	popularity	budget	revenue	original_title	cas
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays Byrne Nic...
2	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays Byrne Nic...
3	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...
4	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...
5	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle ...
6	281957	tt1663202	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Wil Poulter Domhnall...
7	87101	tt1340138	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clar...
8	286217	tt3659388	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff ...
9	211672	tt2293640	7.404165	74000000	1156730962	Minions	Sandra Bullock Jordan Peele Michael Keaton Allison...
10	150540	tt2096673	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Hader...
11	206647	tt2379713	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz Léa Seydoux Ralph...
12	76757	tt1617661	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sebastian Bean Eddie Redm...
13	264660	tt0470752	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac S...
14	257344	tt2120120	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Pete Dinklage ...
15	99861	tt2395427	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...

	id	imdb_id	popularity	budget	revenue	original_title	cas
16	273248	tt3460252	5.898400	44000000	155760117	The Hateful Eight	Samuel L Jackson Kur Russell Jennifer Jasor ..
17	260346	tt2446042	5.749758	48000000	325771424	Taken 3	Liam Neeson Fores Whitaker Maggie Grace Famke..
18	102899	tt0478970	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michae Douglas Evangeline Lilly Cor..
19	150689	tt1661199	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen..
20	131634	tt1951266	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennife Lawrence Josh Hutcherson Liam Hemswor..
21	158852	tt1964418	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ..
22	307081	tt1798684	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rache McAdams Fores Whitaker..
23	254128	tt2126355	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carla Gugino..
24	216015	tt2322441	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennife Ehle Eloi..
25	318846	tt1596363	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Gosling Brad ..
26	177677	tt2381249	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Fe..
27	214756	tt2637276	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ..
28	207703	tt2802144	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L Jackson Mi..
29	314365	tt1895587	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michae Keaton Rache McAdams Lie..
...
10870	38720	tt0061170	0.239435	0	0	Walk Don't Run	Cary Grant Samantha Eggar Jim Hutton John Stan..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10871	19728	tt0060177	0.291704	0	0	The Blue Max	George Peppard James Mason Ursula Andress Jere...
10872	22383	tt0060862	0.151845	0	0	The Professionals	Burt Lancaster Lee Marvin Rober Ryan Woody St...
10873	13353	tt0060550	0.276133	0	0	It's the Great Pumpkin, Charlie Brown	Christopher Shea Sally Dryer Kathy Steinberg A...
10874	34388	tt0060437	0.102530	0	0	Funeral in Berlin	Michael Caine Paul Hubschmid Oskar Homolka Eva...
10875	42701	tt0062262	0.264925	75000	0	The Shooting	Will Hutchins Millie Perkins Jack Nicholson Wa...
10876	36540	tt0061199	0.253437	0	0	Winnie the Pooh and the Honey Tree	Sterling Holloway Junius Matthews Sebastian Ca...
10877	29710	tt0060588	0.252399	0	0	Khartoum	Charlton Heston Laurence Olivier Richard Johns...
10878	23728	tt0059557	0.236098	0	0	Our Man Flint	James Coburn Lee J Cobb Gil Golan Edward Mul...
10879	5065	tt0059014	0.230873	0	0	Carry On Cowboy	Sid James Jim Dale Angela Douglas Kenneth Will...
10880	17102	tt0059127	0.212716	0	0	Dracula: Prince of Darkness	Christopher Lee Barbara Shelley Andrew Keir Fr...
10881	28763	tt0060548	0.034555	0	0	Island of Terror	Peter Cushing Edward Judd Carole Gray Eddie By...
10882	2161	tt0060397	0.207257	5115000	12000000	Fantastic Voyage	Stephen Boyd Raque Welch Edmond O'Brien Donal...
10883	28270	tt0060445	0.206537	0	0	Gambit	Michael Caine Shirley MacLaine Herbert Lom Joh...
10884	26268	tt0060490	0.202473	0	0	Harper	Paul Newman Laurie Bacall Julie Harris Arthur ...

	id	imdb_id	popularity	budget	revenue	original_title	cas
10885	15347	tt0060182	0.342791	0	0	Born Free	Virginia McKenna Bill Travers Geoffrey Keen Pe..
10886	37301	tt0060165	0.227220	0	0	A Big Hand for the Little Lady	Henry Fonda Joanne Woodward Jason Robards Paul..
10887	15598	tt0060086	0.163592	0	0	Alfie	Michael Caine Shelley Winters Millicen Martin..
10888	31602	tt0060232	0.146402	0	0	The Chase	Marlon Brando Jane Fonda Rober Redford E.G. M..
10889	13343	tt0059221	0.141026	700000	0	The Ghost & Mr. Chicken	Don Knotts Joar Staley Liam Redmond Dick Sarge..
10890	20277	tt0061135	0.140934	0	0	The Ugly Dachshund	Dean Jones Suzanne Pleshette Charles Ruggles K..
10891	5921	tt0060748	0.131378	0	0	Nevada Smith	Steve McQueen Kar Malden Briar Keith Arthur K..
10892	31918	tt0060921	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner Eva Marie Saint Alan Arkin Briar K..
10893	20620	tt0060955	0.089072	0	0	Seconds	Rock Hudson Salome Jens John Randolph Will Gee..
10894	5060	tt0060214	0.087034	0	0	Carry On Screaming!	Kenneth Williams Jirr Dale Harry H Corbett Joa..
10895	21	tt0060371	0.080598	0	0	The Endless Summer	Michael Hynson Rober August Lord 'Tally Ho B..
10896	20379	tt0060472	0.065543	0	0	Grand Prix	James Garner Eva Marie Saint Yves Montand Tosh..
10897	39768	tt0060161	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy Oleg Efremov Georgi Z..
10898	21449	tt0061177	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi Akiko Wakabayashi Mie Hama Joh..

	id	imdb_id	popularity	budget	revenue	original_title	cas
10899	22293	tt0060666	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren Tor Neyman Johr Reynolds Dian..

10900 rows × 33 columns

Duration vs Decade

```
In [281]: def get_category_by_length(duration):
            if duration >= 0 and duration < 90:
                return "Less (dur)"
            elif duration >= 90 and duration < 120:
                return "Med (dur)"
            else:
                return "Long (dur)"
```

```
In [282]: duration_by_decade = collections.OrderedDict()
            for i in range(len(tmdb_with_oscar)):
                decade = tmdb_with_oscar.iloc[i]['decade_of_release']
                if decade not in duration_by_decade:
                    duration_by_decade[decade] = {}
                movie_duration_category = get_category_by_length(tmdb_with_oscar.iloc[i][
'runtime'])
                if movie_duration_category not in duration_by_decade[decade]:
                    duration_by_decade[decade][movie_duration_category] = 1
                else:
                    duration_by_decade[decade][movie_duration_category] += 1
```

```
In [283]: df = pd.DataFrame(duration_by_decade)
            cols = sorted(list(df.columns))
            df[cols]
```

Out[283]:

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010	2010-2020
Less (dur)	69	73	181	277	810	1008
Long (dur)	126	118	158	357	560	410
Med (dur)	167	319	729	1137	2203	2198

```
In [285]: df_with_dur_scores = df.copy()
columns = list(df.columns)
for column in columns:
    df_with_dur_scores[column] = 100*df_with_dur_scores[column] / df_with_dur_scores[column].sum()
cols= sorted(list(df.columns))
df_with_dur_scores[cols]
```

Out[285]:

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010	2010-2020
Less (dur)	19.060773	14.313725	16.947566	15.640881	22.670025	27.876106
Long (dur)	34.806630	23.137255	14.794007	20.158103	15.673104	11.338496
Med (dur)	46.132597	62.549020	68.258427	64.201016	61.656871	60.785398

```
In [286]: df_with_dur_scores
```

Out[286]:

	2010-2020	1970-1980	2000-2010	1990-2000	1980-1990	1960-1970
Less (dur)	27.876106	14.313725	22.670025	15.640881	16.947566	19.060773
Long (dur)	11.338496	23.137255	15.673104	20.158103	14.794007	34.806630
Med (dur)	60.785398	62.549020	61.656871	64.201016	68.258427	46.132597

```
In [287]: dur_score = []
for i in range(len(tmdb_with_oscar)):
    decade_of_release = tmdb_with_oscar.iloc[i]['decade_of_release']
    d_score = df_with_dur_scores[decade_of_release][get_category_by_length(tmd
b_with_oscar.iloc[i]['runtime'])]
    dur_score.append(d_score)
tmdb_with_oscar['dur_score'] = dur_score
```

In [288]: `tmdb_with_oscar.dtypes`

```
Out[288]: id                int64
imdb_id              object
popularity           float64
budget              int64
revenue             int64
original_title       object
cast                object
homepage            object
director            object
tagline             object
keywords            object
overview            object
runtime             int64
genres              object
production_companies object
release_date        object
vote_count          int64
vote_average        float64
release_year        int64
budget_adj          float64
revenue_adj         float64
cast_points         float64
production_points   float64
Director_points     float64
year_of_release     int64
decade_of_release   object
month_of_release    int64
genre_score_atRelease float64
genre_score_atPresent float64
oscars_won         float64
inflated_budget     float64
inflated_revenue    float64
inflated_profit     float64
dur_score           float64
dtype: object
```



```
In [289]: tmdb_with_oscar.head(5)
```

Out[289]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	
2	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	
3	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...	http://www.
4	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...	http

5 rows × 34 columns

```
In [290]: tmdb_with_oscar.dtypes
```

```
Out[290]: id                int64
imdb_id                object
popularity             float64
budget                 int64
revenue                int64
original_title         object
cast                   object
homepage               object
director               object
tagline                object
keywords               object
overview               object
runtime                int64
genres                 object
production_companies   object
release_date           object
vote_count             int64
vote_average           float64
release_year           int64
budget_adj             float64
revenue_adj            float64
cast_points            float64
production_points      float64
Director_points        float64
year_of_release        int64
decade_of_release      object
month_of_release        int64
genre_score_atRelease  float64
genre_score_atPresent  float64
oscars_won             float64
inflated_budget        float64
inflated_revenue       float64
inflated_profit        float64
dur_score              float64
dtype: object
```

```
In [291]: final_tmdb = tmdb_with_oscar.copy()
```

```
In [292]: final_tmdb = final_tmdb.drop(['imdb_id', 'homepage', 'tagline', 'keywords', 'overview'], axis=1)
```

In [293]: final_tmdb

Out[293]:

	id	popularity	budget	revenue	original_title	cast	di
0	135397	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	Colin Tre
1	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	George
2	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	George
3	262500	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...	I Schv
4	140607	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...	J.J. A
5	168259	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle ...	Jame
6	281957	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Will Poulter Domhn...	Ale Gon IÃ±
7	87101	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clar...	Alan
8	286217	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff ...	Ridley
9	211672	7.404165	74000000	1156730962	Minions	Sandra Bullock Jon Hamm Michael Keaton Allison...	Balda
10	150540	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Ha...	Pete
11	206647	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz LÃ©a Seydoux Ralp...	Sam M
12	76757	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sean Bean Eddie Redm...	Wachows Wacl
13	264660	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac S...	Alex G
14	257344	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Peter Dinklage ...	Coli
15	99861	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...	Joss W

	id	popularity	budget	revenue	original_title	cast	di
16	273248	5.898400	44000000	155760117	The Hateful Eight	Samuel L. Jackson Kurt Russell Jennifer Jason ...	Q Tar
17	260346	5.749758	48000000	325771424	Taken 3	Liam Neeson Forest Whitaker Maggie Grace Famke...	Olivier M
18	102899	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michael Douglas Evangeline Lilly Cor...	Peytor
19	150689	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen...	K Br
20	131634	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennifer Lawrence Josh Hutcherson Liam Hemswor...	F Lav
21	158852	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ...	Br
22	307081	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rachel McAdams Forest Whitaker...	Antoine
23	254128	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carla Gugino...	Brad F
24	216015	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennifer Ehle Eloi...	Sam Jc
25	318846	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Gosling Brad ...	Adam I
26	177677	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Fe...	Chris McC
27	214756	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ...	MacF
28	207703	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L. Jackson Mi...	M V
29	314365	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michael Keaton Rachel McAdams Lie...	Tom Mc
...
10870	38720	0.239435	0	0	Walk Don't Run	Cary Grant Samantha Eggar Jim Hutton John Stan...	Charles V
10871	19728	0.291704	0	0	The Blue Max	George Peppard James Mason Ursula Andress Jere...	John Gui

	id	popularity	budget	revenue	original_title	cast	di
10872	22383	0.151845	0	0	The Professionals	Burt Lancaster Lee Marvin Robert Ryan Woody St...	Richard E
10873	13353	0.276133	0	0	It's the Great Pumpkin, Charlie Brown	Christopher Shea Sally Dryer Kathy Steinberg A...	Bill Me
10874	34388	0.102530	0	0	Funeral in Berlin	Michael Caine Paul Hubschmid Oskar Homolka Eva...	Guy He
10875	42701	0.264925	75000	0	The Shooting	Will Hutchins Millie Perkins Jack Nicholson Wa...	Monte H
10876	36540	0.253437	0	0	Winnie the Pooh and the Honey Tree	Sterling Holloway Junius Matthews Sebastian Ca...	Wc Reith
10877	29710	0.252399	0	0	Khartoum	Charlton Heston Laurence Olivier Richard Johns...	Dearde E
10878	23728	0.236098	0	0	Our Man Flint	James Coburn Lee J. Cobb Gila Golan Edward Mul...	Daniel
10879	5065	0.230873	0	0	Carry On Cowboy	Sid James Jim Dale Angela Douglas Kenneth Will...	Gerald TI
10880	17102	0.212716	0	0	Dracula: Prince of Darkness	Christopher Lee Barbara Shelley Andrew Keir Fr...	Terence
10881	28763	0.034555	0	0	Island of Terror	Peter Cushing Edward Judd Carole Gray Eddie By...	Terence
10882	2161	0.207257	5115000	12000000	Fantastic Voyage	Stephen Boyd Raquele Welch Edmond O'Brien Donal...	R Fle
10883	28270	0.206537	0	0	Gambit	Michael Caine Shirley MacLaine Herbert Lom Joh...	Ronald M
10884	26268	0.202473	0	0	Harper	Paul Newman Lauren Bacall Julie Harris Arthur ...	Jack S
10885	15347	0.342791	0	0	Born Free	Virginia McKenna Bill Travers Geoffrey Keen Pe...	Jam
10886	37301	0.227220	0	0	A Big Hand for the Little Lady	Henry Fonda Joanne Woodward Jason Robards Paul...	Fielde
10887	15598	0.163592	0	0	Alfie	Michael Caine Shelley Winters Millicent Martin...	Lewis
10888	31602	0.146402	0	0	The Chase	Marlon Brando Jane Fonda Robert Redford E.G. M...	Arthu

	id	popularity	budget	revenue	original_title	cast	di
10889	13343	0.141026	700000	0	The Ghost & Mr. Chicken	Don Knotts Joan Staley Liam Redmond Dick Sarge...	Alan
10890	20277	0.140934	0	0	The Ugly Dachshund	Dean Jones Suzanne Pleshette Charles Ruggles K...	Norman
10891	5921	0.131378	0	0	Nevada Smith	Steve McQueen Karl Malden Brian Keith Arthur K...	Hat
10892	31918	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner Eva Marie Saint Alan Arkin Brian K...	N Jé
10893	20620	0.089072	0	0	Seconds	Rock Hudson Salome Jens John Randolph Will Gee...	Frankenl
10894	5060	0.087034	0	0	Carry On Screaming!	Kenneth Williams Jim Dale Harry H. Corbett Joa...	Gerald TI
10895	21	0.080598	0	0	The Endless Summer	Michael Hynson Robert August Lord 'Tally Ho' B...	Bruce
10896	20379	0.065543	0	0	Grand Prix	James Garner Eva Marie Saint Yves Montand Tosh...	Frankenl
10897	39768	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy Oleg Efremov Georgi Z...	Eldar Rye
10898	21449	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi Akiko Wakabayashi Mie Hama Joh...	Wood
10899	22293	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren Tom Neyman John Reynolds Dian...	Ha v

10900 rows × 29 columns

In [207]: `final_tmdb['oscars_won'].fillna(-1, inplace=True)`

In [208]: final_tmdb

Out[208]:

	id	popularity	budget	revenue	original_title	cast	di
0	135397	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	Colin Tre
1	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	George
2	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	George
3	262500	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...	I Schv
4	140607	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...	J.J. A
5	168259	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle ...	Jame
6	281957	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Will Poulter Domhn...	Ale Gon IÃ±
7	87101	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clar...	Alan
8	286217	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff ...	Ridley
9	211672	7.404165	74000000	1156730962	Minions	Sandra Bullock Jon Hamm Michael Keaton Allison...	Balda
10	150540	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Ha...	Pete
11	206647	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz LÃ©a Seydoux Ralp...	Sam M
12	76757	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sean Bean Eddie Redm...	Wachows Wacl
13	264660	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac S...	Alex G
14	257344	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Peter Dinklage ...	Coli
15	99861	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...	Joss W

	id	popularity	budget	revenue	original_title	cast	di
16	273248	5.898400	44000000	155760117	The Hatef Eight	Samuel L. Jackson Kurt Russell Jennifer Jason ...	Q Tar
17	260346	5.749758	48000000	325771424	Taken 3	Liam Neeson Forest Whitaker Maggie Grace Famke...	Olivier M
18	102899	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michael Douglas Evangeline Lilly Cor...	Peytor
19	150689	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen...	K Br
20	131634	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennifer Lawrence Josh Hutcherson Liam Hemswor...	F Lav
21	158852	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ...	Br
22	307081	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rachel McAdams Forest Whitaker...	Antoine
23	254128	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carla Gugino...	Brad F
24	216015	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennifer Ehler Eloi...	Sam Jc
25	318846	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Gosling Brad ...	Adam I
26	177677	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Fe...	Chris McC
27	214756	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ...	MacF
28	207703	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L. Jackson Mi...	M V
29	314365	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michael Keaton Rachel McAdams Lie...	Tom Mc
...
10870	38720	0.239435	0	0	Walk Don't Run	Cary Grant Samantha Eggar Jim Hutton John Stan...	Charles V
10871	19728	0.291704	0	0	The Blue Max	George Peppard James Mason Ursula Andress Jere...	John Gui

	id	popularity	budget	revenue	original_title	cast	di
10872	22383	0.151845	0	0	The Professionals	Burt Lancaster Lee Marvin Robert Ryan Woody St...	Richard E
10873	13353	0.276133	0	0	It's the Great Pumpkin, Charlie Brown	Christopher Shea Sally Dryer Kathy Steinberg A...	Bill Me
10874	34388	0.102530	0	0	Funeral in Berlin	Michael Caine Paul Hubschmid Oskar Homolka Eva...	Guy He
10875	42701	0.264925	75000	0	The Shooting	Will Hutchins Millie Perkins Jack Nicholson Wa...	Monte H
10876	36540	0.253437	0	0	Winnie the Pooh and the Honey Tree	Sterling Holloway Junius Matthews Sebastian Ca...	Wc Reith
10877	29710	0.252399	0	0	Khartoum	Charlton Heston Laurence Olivier Richard Johns...	Dearde E
10878	23728	0.236098	0	0	Our Man Flint	James Coburn Lee J. Cobb Gila Golan Edward Mul...	Daniel
10879	5065	0.230873	0	0	Carry On Cowboy	Sid James Jim Dale Angela Douglas Kenneth Will...	Gerald TI
10880	17102	0.212716	0	0	Dracula: Prince of Darkness	Christopher Lee Barbara Shelley Andrew Keir Fr...	Terence
10881	28763	0.034555	0	0	Island of Terror	Peter Cushing Edward Judd Carole Gray Eddie By...	Terence
10882	2161	0.207257	5115000	12000000	Fantastic Voyage	Stephen Boyd Raquele Welch Edmond O'Brien Donal...	R Fle
10883	28270	0.206537	0	0	Gambit	Michael Caine Shirley MacLaine Herbert Lom Joh...	Ronald M
10884	26268	0.202473	0	0	Harper	Paul Newman Lauren Bacall Julie Harris Arthur ...	Jack S
10885	15347	0.342791	0	0	Born Free	Virginia McKenna Bill Travers Geoffrey Keen Pe...	Jam
10886	37301	0.227220	0	0	A Big Hand for the Little Lady	Henry Fonda Joanne Woodward Jason Robards Paul...	Fielde
10887	15598	0.163592	0	0	Alfie	Michael Caine Shelley Winters Millicent Martin...	Lewis
10888	31602	0.146402	0	0	The Chase	Marlon Brando Jane Fonda Robert Redford E.G. M...	Arthu

	id	popularity	budget	revenue	original_title	cast	di
10889	13343	0.141026	700000	0	The Ghost & Mr. Chicken	Don Knotts Joan Staley Liam Redmond Dick Sarge...	Alan
10890	20277	0.140934	0	0	The Ugly Dachshund	Dean Jones Suzanne Pleshette Charles Ruggles K...	Norman
10891	5921	0.131378	0	0	Nevada Smith	Steve McQueen Karl Malden Brian Keith Arthur K...	Hat
10892	31918	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner Eva Marie Saint Alan Arkin Brian K...	N Jé
10893	20620	0.089072	0	0	Seconds	Rock Hudson Salome Jens John Randolph Will Gee...	Frankenl
10894	5060	0.087034	0	0	Carry On Screaming!	Kenneth Williams Jim Dale Harry H. Corbett Joa...	Gerald TI
10895	21	0.080598	0	0	The Endless Summer	Michael Hynson Robert August Lord 'Tally Ho' B...	Bruce
10896	20379	0.065543	0	0	Grand Prix	James Garner Eva Marie Saint Yves Montand Tosh...	Frankenl
10897	39768	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy Oleg Efremov Georgi Z...	Eldar Rye
10898	21449	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi Akiko Wakabayashi Mie Hama Joh...	Wood
10899	22293	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren Tom Neyman John Reynolds Dian...	Ha v

10900 rows × 29 columns

```
In [209]: final_tmdb.isna().sum()
```

```
Out[209]: id                0
popularity                0
budget                   0
revenue                  0
original_title            0
cast                     0
director                 44
runtime                  0
genres                   23
production_companies      0
release_date              0
vote_count                0
vote_average              0
release_year              0
budget_adj                0
revenue_adj               0
cast_points               0
production_points         0
Director_points           6738
year_of_release           0
decade_of_release         0
month_of_release          0
genre_score_atRelease     0
genre_score_atPresent     0
oscars_won                0
inflated_budget           0
inflated_revenue          0
inflated_profit           0
dur_score                 0
dtype: int64
```

filling nan with -1 in oscars_won

```
In [296]: final_tmdb.oscars_won.fillna(-1, inplace=True)
```

```
In [299]: final_tmdb.isna().sum()
```

```
Out[299]: id                                0
popularity                                0
budget                                  0
revenue                                0
original_title                          0
cast                                    0
director                               44
runtime                                0
genres                                 23
production_companies                   0
release_date                           0
vote_count                             0
vote_average                           0
release_year                           0
budget_adj                             0
revenue_adj                             0
cast_points                             0
production_points                       0
Director_points                         6738
year_of_release                         0
decade_of_release                       0
month_of_release                        0
genre_score_atRelease                   0
genre_score_atPresent                   0
oscars_won                             0
inflated_budget                         0
inflated_revenue                        0
inflated_profit                         0
dur_score                              0
dtype: int64
```

replacing NAN in director's points

```
In [300]: final_tmdb.describe()
```

```
Out[300]:
```

	id	popularity	budget	revenue	runtime	vote_count	vc
count	10900.000000	10900.000000	1.090000e+04	1.090000e+04	10900.000000	10900.000000	10
mean	65980.661284	0.649498	1.631633e+07	3.995642e+07	102.101743	218.236789	
std	92054.450204	1.034034	3.192169e+07	1.170185e+08	31.358108	577.849608	
min	5.000000	0.000065	0.000000e+00	0.000000e+00	0.000000	10.000000	
25%	10587.750000	0.207759	0.000000e+00	0.000000e+00	90.000000	17.000000	
50%	20617.000000	0.384145	1.175280e+06	0.000000e+00	99.000000	38.000000	
75%	75480.250000	0.716003	2.000000e+07	2.420708e+07	111.000000	146.000000	
max	417859.000000	32.985763	4.250000e+08	2.781506e+09	900.000000	9767.000000	

8 rows × 22 columns

replacing NAN with 25th percentile score

```
In [301]: final_tmdb.Director_points.fillna(284, inplace=True)
```

```
In [304]: final_tmdb.isna().sum()
```

```
Out[304]: id                0
popularity                0
budget                   0
revenue                  0
original_title            0
cast                     0
director                 44
runtime                  0
genres                   23
production_companies      0
release_date              0
vote_count                0
vote_average              0
release_year              0
budget_adj                0
revenue_adj               0
cast_points               0
production_points         0
Director_points           0
year_of_release           0
decade_of_release         0
month_of_release          0
genre_score_atRelease     0
genre_score_atPresent     0
oscars_won                0
inflated_budget           0
inflated_revenue          0
inflated_profit           0
dur_score                 0
dtype: int64
```

current popularity


```
In [454]: min_max_scaler = preprocessing.MinMaxScaler()
scaled_array = min_max_scaler.fit_transform(final_tmdb[['vote_average', 'cast_p
oints', 'production_points', 'Director_points', 'inflated_profit', 'genre_score_at
Present']])
df_normalized_Pc = pd.DataFrame(scaled_array)
df_normalized_Pc
```

Out[454]:

	0	1	2	3	4	5
0	0.649351	0.065287	0.302339	0.004334	0.515915	0.520339
1	0.727273	0.151838	0.000000	0.020507	0.193348	0.520339
2	0.727273	0.151838	0.000000	0.000818	0.193348	0.520339
3	0.623377	0.036320	0.150001	0.015646	0.181397	0.372199
4	0.779221	0.268112	0.136104	0.077190	0.664172	0.360235
5	0.753247	0.082903	0.540338	0.055238	0.505696	0.433109
6	0.740260	0.540998	0.232813	0.004334	0.242337	0.615836
7	0.558442	0.109930	0.325594	0.004334	0.209377	0.520339
8	0.792208	0.327656	0.070103	0.394052	0.267162	0.478573
9	0.649351	0.080652	0.468700	0.004334	0.436062	0.430498
10	0.844156	0.000000	0.322509	0.004334	0.321167	0.348706
11	0.610390	0.353574	0.400284	0.241705	0.309538	0.304111
12	0.480519	0.053907	0.582813	0.004334	0.130467	0.360235
13	0.792208	0.089971	0.052909	0.046877	0.134508	0.396781
14	0.558442	0.008542	0.400284	0.153379	0.172414	0.454862
15	0.766234	0.253233	0.257719	0.051432	0.450850	0.309550
16	0.766234	0.207058	0.010534	0.563963	0.160224	0.445290
17	0.597403	0.198281	0.052219	0.004334	0.208816	0.433109
18	0.714286	0.040935	0.257719	0.026454	0.238660	0.309550
19	0.688312	0.011681	0.322509	0.011213	0.256732	0.530998
20	0.649351	0.215128	0.297550	0.044736	0.268417	0.175549
21	0.610390	0.138306	0.322509	0.078474	0.133590	0.423755
22	0.753247	0.196711	0.027021	0.061429	0.145707	0.465303
23	0.597403	0.000000	0.603484	0.004334	0.231032	0.676093
24	0.493506	0.002769	0.069363	0.002836	0.281292	0.413966
25	0.753247	0.767766	0.484854	0.025292	0.158384	0.544268
26	0.727273	0.231523	0.378067	0.004334	0.279530	0.148140
27	0.623377	0.146682	0.433828	0.004334	0.170207	0.227105
28	0.792208	0.520747	0.070103	0.075875	0.221892	0.531216
29	0.818182	0.238956	0.125921	0.055361	0.147701	0.545138
...
10870	0.558442	0.026362	0.000000	0.003172	0.128158	0.323907
10871	0.519481	0.019406	0.000000	0.007850	0.128158	0.561236
10872	0.584416	0.025568	0.400284	0.013292	0.128158	0.236023
10873	0.740260	0.000000	0.000000	0.004334	0.128158	0.121601

	0	1	2	3	4	5
10874	0.545455	0.315349	0.000000	0.033913	0.128158	0.210790
10875	0.519481	0.157867	0.000000	0.000252	0.128003	0.006091
10876	0.831169	0.000000	0.000000	0.019789	0.128158	0.121601
10877	0.558442	0.034407	0.000000	0.004334	0.128158	0.578421
10878	0.532468	0.050623	0.228887	0.001842	0.128158	0.439199
10879	0.571429	0.000000	0.000000	0.004334	0.128158	0.233196
10880	0.545455	0.046604	0.000000	0.002056	0.128158	0.129215
10881	0.493506	0.000000	0.000000	0.002056	0.128158	0.208832
10882	0.675325	0.016455	0.000000	0.016227	0.142446	0.161410
10883	0.597403	0.325546	0.417197	0.008645	0.128158	0.449424
10884	0.584416	0.051576	0.430156	0.001017	0.128158	0.798129
10885	0.662338	0.000000	0.000000	0.000589	0.128158	0.620622
10886	0.584416	0.074002	0.000000	0.000955	0.128158	0.006091
10887	0.610390	0.326784	0.000000	0.023198	0.128158	0.641070
10888	0.584416	0.123338	0.000000	0.020354	0.128158	0.602132
10889	0.597403	0.000000	0.417197	0.004334	0.126664	0.438112
10890	0.545455	0.000000	0.322509	0.004334	0.128158	0.610616
10891	0.571429	0.025499	0.269580	0.015203	0.128158	0.154231
10892	0.519481	0.063469	0.000000	0.047519	0.128158	0.241244
10893	0.662338	0.000000	0.000000	0.043666	0.128158	0.655427
10894	0.714286	0.000000	0.000000	0.004334	0.128158	0.227105
10895	0.766234	0.000000	0.000000	0.004334	0.128158	0.070698
10896	0.545455	0.033627	0.000000	0.043666	0.128158	0.547096
10897	0.649351	0.000000	0.000000	0.002255	0.128158	0.274962
10898	0.506494	0.000000	0.000000	0.238525	0.128158	0.375245
10899	0.000000	0.000000	0.000000	0.004334	0.128119	0.129215

10900 rows × 6 columns

```
In [464]: final_tmdb['Current_popularity'] = (0.5)*df_normalized_Pc[0] + (0.1)*df_normalized_Pc[1] + (0.1)*df_normalized_Pc[2] + (0.1)*df_normalized_Pc[3] + (0.1)*df_normalized_Pc[4] + (0.1)*df_normalized_Pc[5]
```

In [465]: final_tmdb

Out[465]:

	id	popularity	budget	revenue	original_title	cast	di
0	135397	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	Colin Tre
1	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	George
2	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	George
3	262500	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...	I Schv
4	140607	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...	J.J. A
5	168259	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle ...	Jame
6	281957	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Will Poulter Domhn...	Ale Gon IÃ±
7	87101	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clar...	Alan
8	286217	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff ...	Ridley
9	211672	7.404165	74000000	1156730962	Minions	Sandra Bullock Jon Hamm Michael Keaton Allison...	Balda
10	150540	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Ha...	Pete
11	206647	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz LÃ©a Seydoux Ralp...	Sam M
12	76757	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sean Bean Eddie Redm...	Wachows Wacl
13	264660	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac S...	Alex G
14	257344	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Peter Dinklage ...	Coli
15	99861	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...	Joss W

	id	popularity	budget	revenue	original_title	cast	di
16	273248	5.898400	44000000	155760117	The Hateful Eight	Samuel L. Jackson Kurt Russell Jennifer Jason ...	Q Tar
17	260346	5.749758	48000000	325771424	Taken 3	Liam Neeson Forest Whitaker Maggie Grace Famke...	Olivier Me
18	102899	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michael Douglas Evangeline Lilly Cor...	Peytor
19	150689	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen...	Ké Br
20	131634	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennifer Lawrence Josh Hutcherson Liam Hemswor...	F Lav
21	158852	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ...	Bré
22	307081	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rachel McAdams Forest Whitaker...	Antoine
23	254128	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carla Gugino...	Brad F
24	216015	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennifer Ehle Eloi...	Sam · Jc
25	318846	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Gosling Brad ...	Adam I
26	177677	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Fe...	Chris McC
27	214756	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ...	MacF
28	207703	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L. Jackson Mi...	M: V
29	314365	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michael Keaton Rachel McAdams Lie...	Tom Mc
...
10870	38720	0.239435	0	0	Walk Don't Run	Cary Grant Samantha Eggar Jim Hutton John Stan...	Charles V

	id	popularity	budget	revenue	original_title	cast	di
10871	19728	0.291704	0	0	The Blue Max	George Peppard James Mason Ursula Andress Jere...	John Gui
10872	22383	0.151845	0	0	The Professionals	Burt Lancaster Lee Marvin Robert Ryan Woody St...	Richard E
10873	13353	0.276133	0	0	It's the Great Pumpkin, Charlie Brown	Christopher Shea Sally Dryer Kathy Steinberg A...	Bill Me
10874	34388	0.102530	0	0	Funeral in Berlin	Michael Caine Paul Hubschmid Oskar Homolka Eva...	Guy Ha
10875	42701	0.264925	75000	0	The Shooting	Will Hutchins Millie Perkins Jack Nicholson Wa...	Monte H
10876	36540	0.253437	0	0	Winnie the Pooh and the Honey Tree	Sterling Holloway Junius Matthews Sebastian Ca...	Wc Reith
10877	29710	0.252399	0	0	Khartoum	Charlton Heston Laurence Olivier Richard Johns...	Dearde E
10878	23728	0.236098	0	0	Our Man Flint	James Coburn Lee J. Cobb Gila Golan Edward Mul...	Daniel
10879	5065	0.230873	0	0	Carry On Cowboy	Sid James Jim Dale Angela Douglas Kenneth Will...	Gerald TI
10880	17102	0.212716	0	0	Dracula: Prince of Darkness	Christopher Lee Barbara Shelley Andrew Keir Fr...	Terence
10881	28763	0.034555	0	0	Island of Terror	Peter Cushing Edward Judd Carole Gray Eddie By...	Terence
10882	2161	0.207257	5115000	12000000	Fantastic Voyage	Stephen Boyd Raquel Welch Edmond O'Brien Donal...	R Fle
10883	28270	0.206537	0	0	Gambit	Michael Caine Shirley MacLaine Herbert Lom Joh...	Ronald M
10884	26268	0.202473	0	0	Harper	Paul Newman Lauren Bacall Julie Harris Arthur ...	Jack S
10885	15347	0.342791	0	0	Born Free	Virginia McKenna Bill Travers Geoffrey Keen Pe...	Jam
10886	37301	0.227220	0	0	A Big Hand for the Little Lady	Henry Fonda Joanne Woodward Jason Robards Paul...	Fielde

	id	popularity	budget	revenue	original_title	cast	di
10887	15598	0.163592	0	0	Alfie	Michael Caine Shelley Winters Millicent Martin...	Lewis
10888	31602	0.146402	0	0	The Chase	Marlon Brando Jane Fonda Robert Redford E.G. M...	Arthu
10889	13343	0.141026	700000	0	The Ghost & Mr. Chicken	Don Knotts Joan Staley Liam Redmond Dick Sarge...	Alan
10890	20277	0.140934	0	0	The Ugly Dachshund	Dean Jones Suzanne Pleshette Charles Ruggles K...	Norman
10891	5921	0.131378	0	0	Nevada Smith	Steve McQueen Karl Malden Brian Keith Arthur K...	Hat
10892	31918	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner Eva Marie Saint Alan Arkin Brian K...	N Je
10893	20620	0.089072	0	0	Seconds	Rock Hudson Salome Jens John Randolph Will Gee...	Frankenl
10894	5060	0.087034	0	0	Carry On Screaming!	Kenneth Williams Jim Dale Harry H. Corbett Joa...	Gerald TI
10895	21	0.080598	0	0	The Endless Summer	Michael Hynson Robert August Lord 'Tally Ho' B...	Bruce
10896	20379	0.065543	0	0	Grand Prix	James Garner Eva Marie Saint Yves Montand Tosh...	Frankenl
10897	39768	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy Oleg Efremov Georgi Z...	Eldar Ryë
10898	21449	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi Akiko Wakabayashi Mie Hama Joh...	Wood
10899	22293	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren Tom Neyman John Reynolds Dian...	Ha v

10900 rows × 32 columns

Release popularity


```
In [466]: min_max_scaler = preprocessing.MinMaxScaler()
scaled_array = min_max_scaler.fit_transform(final_tmdb[['genre_score_atRelease', 'dur_score', 'inflated_budget']])
df_normalized_Rp = pd.DataFrame(scaled_array)
df_normalized_Rp
```

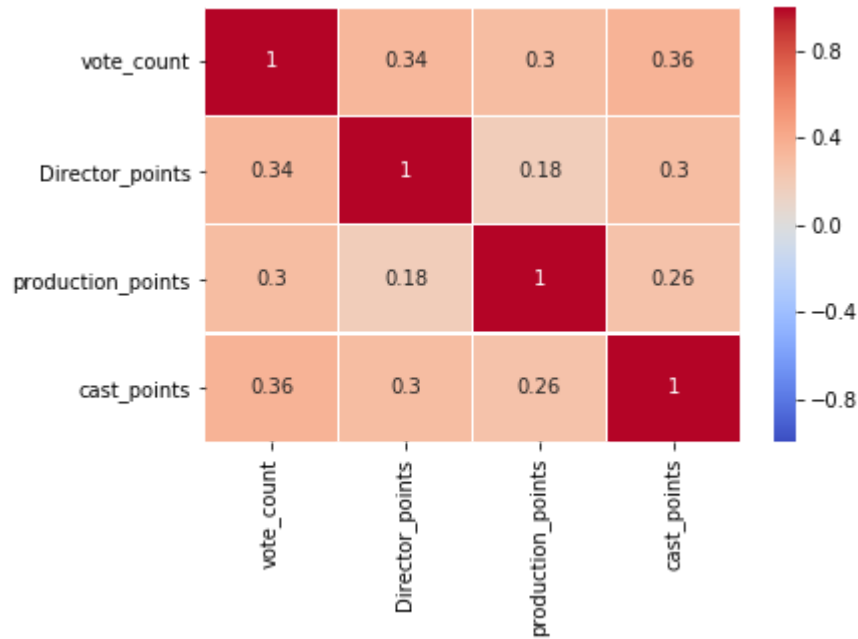
Out[466]:

	0	1	2
0	0.500908	0.000000	0.299064
1	0.500908	0.000000	0.300112
2	0.500908	0.000000	0.300112
3	0.358300	0.868710	0.221653
4	0.346782	0.000000	0.402314
5	0.416934	0.000000	0.382079
6	0.592838	0.000000	0.271562
7	0.500908	0.000000	0.309033
8	0.460701	0.000000	0.215953
9	0.414422	0.868710	0.147538
10	0.335684	0.868710	0.348908
11	0.292755	0.000000	0.490114
12	0.346782	0.000000	0.356756
13	0.381963	0.868710	0.030537
14	0.437875	0.868710	0.175439
15	0.297990	0.000000	0.563064
16	0.428661	0.000000	0.088509
17	0.416934	0.868710	0.097720
18	0.297990	0.868710	0.259171
19	0.511169	0.868710	0.191428
20	0.168994	0.000000	0.320752
21	0.407930	0.000000	0.380141
22	0.447927	0.000000	0.059813
23	0.650845	0.868710	0.220082
24	0.398506	0.000000	0.081081
25	0.523943	0.000000	0.056324
26	0.142608	0.000000	0.299044
27	0.218624	0.868710	0.135576
28	0.511378	0.000000	0.164902
29	0.524780	0.000000	0.040094
...
10870	0.329124	0.611282	0.000000
10871	0.623506	0.412301	0.000000
10872	0.325467	0.611282	0.000000
10873	0.100566	0.135669	0.000000

	0	1	2
10874	0.117022	0.611282	0.000000
10875	0.065825	0.135669	0.001085
10876	0.100566	0.135669	0.000000
10877	0.680189	0.412301	0.000000
10878	0.427861	0.611282	0.000000
10879	0.272441	0.611282	0.000000
10880	0.085938	0.611282	0.000000
10881	0.148106	0.135669	0.000000
10882	0.179189	0.611282	0.074424
10883	0.427861	0.611282	0.000000
10884	0.703959	0.412301	0.000000
10885	0.654590	0.611282	0.000000
10886	0.065825	0.611282	0.000000
10887	0.636305	0.611282	0.000000
10888	0.502828	0.412301	0.000000
10889	0.458944	0.611282	0.010473
10890	0.585109	0.611282	0.000000
10891	0.208445	0.412301	0.000000
10892	0.263299	0.412301	0.000000
10893	0.544882	0.611282	0.000000
10894	0.206616	0.135669	0.000000
10895	0.003657	0.611282	0.000000
10896	0.566824	0.412301	0.000000
10897	0.265127	0.611282	0.000000
10898	0.349237	0.135669	0.000000
10899	0.085938	0.135669	0.000275

10900 rows × 3 columns

In [367]:



```
In [467]: final_tmdb['Release popularity'] = (0.4)*df_normalized_Rp[0] + (0.3)*df_normalized_Rp[1] + (0.3)*df_normalized_Rp[2]
```

```
In [468]: final_tmdb.describe()
```

Out[468]:

	id	popularity	budget	revenue	runtime	vote_count	vc
count	10900.000000	10900.000000	1.090000e+04	1.090000e+04	10900.000000	10900.000000	10900.000000
mean	65980.661284	0.649498	1.631633e+07	3.995642e+07	102.101743	218.236789	218.236789
std	92054.450204	1.034034	3.192169e+07	1.170185e+08	31.358108	577.849608	577.849608
min	5.000000	0.000065	0.000000e+00	0.000000e+00	0.000000	10.000000	10.000000
25%	10587.750000	0.207759	0.000000e+00	0.000000e+00	90.000000	17.000000	17.000000
50%	20617.000000	0.384145	1.175280e+06	0.000000e+00	99.000000	38.000000	38.000000
75%	75480.250000	0.716003	2.000000e+07	2.420708e+07	111.000000	146.000000	146.000000
max	417859.000000	32.985763	4.250000e+08	2.781506e+09	900.000000	9767.000000	9767.000000

8 rows × 25 columns

```
In [469]: final_tmdb.columns
```

```
Out[469]: Index(['id', 'popularity', 'budget', 'revenue', 'original_title', 'cast',  
                'director', 'runtime', 'genres', 'production_companies', 'release_date',  
                'vote_count', 'vote_average', 'release_year', 'budget_adj',  
                'revenue_adj', 'cast_points', 'production_points', 'Director_points',  
                'year_of_release', 'decade_of_release', 'month_of_release',  
                'genre_score_atRelease', 'genre_score_atPresent', 'oscars_won',  
                'inflated_budget', 'inflated_revenue', 'inflated_profit', 'dur_score',  
                'Current_popularity', 'Release popularity', 'endurance'],  
              dtype='object')
```

```
In [470]: final_tmdb['endurance'] = (0.6 * (final_tmdb['Current_popularity'] - final_tmdb['Release popularity'])  
                                     + 0.4 * (final_tmdb['Current_popularity'] + final_tmdb['Release popularity'])) / 2
```

In [471]: `final_tmdb.nlargest(10, 'endurance')`

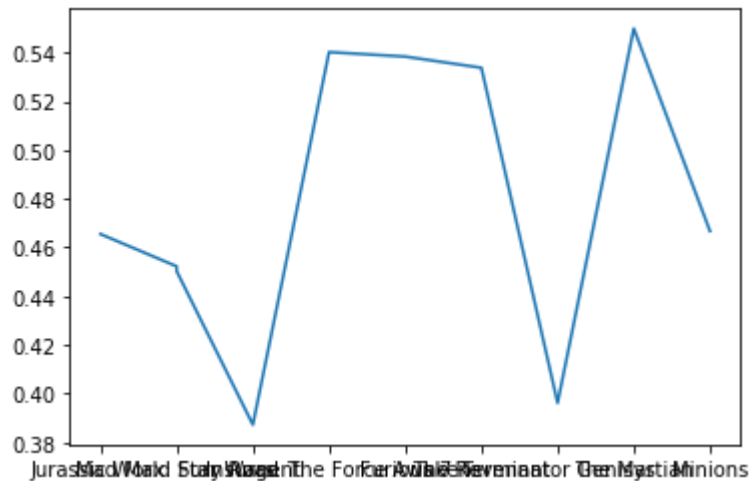
Out[471]:

	id	popularity	budget	revenue	original_title	cast	director
2881	155	8.466668	185000000	1001921825	The Dark Knight	Christian Bale Michael Caine Heath Ledger Aaro...	Christophe Nolar
632	157336	24.949134	165000000	621752480	Interstellar	Matthew McConaughey Jessica Chastain Anne Hath...	Christophe Nolar
1925	27205	9.363643	160000000	825500000	Inception	Leonardo DiCaprio Joseph Gordon-Levitt Ellen P...	Christophe Nolar
4375	49026	6.591277	250000000	1081041287	The Dark Knight Rises	Christian Bale Michael Caine Gary Oldman Anne ...	Christophe Nolar
10256	424	2.377288	22000000	321265768	Schindler's List	Liam Neeson Ben Kingsley Ralph Fiennes Carolín...	Stever Spielberg
6581	1422	2.819750	90000000	289847354	The Departed	Leonardo DiCaprio Matt Damon Jack Nicholson Ma...	Martin Scorsese
6208	272	5.400826	150000000	374218673	Batman Begins	Christian Bale Michael Caine Liam Neeson Katie...	Christophe Nolar
7290	238	5.738034	6000000	245066411	The Godfather	Marlon Brando Al Pacino James Caan Richard S. ...	Francis Ford Coppola
9004	857	2.170136	70000000	481840909	Saving Private Ryan	Tom Hanks Tom Sizemore Dennis Farina Max Marti...	Stever Spielberg
3930	640	2.973115	52000000	352114312	Catch Me If You Can	Leonardo DiCaprio Tom Hanks Christopher Walken...	Stever Spielberg

10 rows × 32 columns

```
In [522]: plt.plot(final_tmdb['original_title'].head(n=10),final_tmdb['Current_popularity'].head(n=10))
```

```
Out[522]: [<matplotlib.lines.Line2D at 0x1ebdd1b5e80>]
```



```
In [472]: Linear_RegModel = final_tmdb.filter(['popularity', 'cast_points', 'production_
points', 'Director_points', 'year_of_release',
        'genre_score_atRelease', 'genre_score_atPresent', 'oscars_won', 'inflated_
budget', 'inflated_revenue', 'inflated_profit', 'dur_score',
        'Current_popularity', 'Release popularity'])
Linear_RegModel.isna().sum()
Y = final_tmdb.filter(['endurance'])
Y
le = LabelEncoder()
Linear_RegModel = Linear_RegModel.apply(le.fit_transform)
#feature,result = linear_reg.loc[:,linear_reg.columns != 'popularity'], linear_
reg['popularity']
X_train, X_test, y_train, y_test = train_test_split(Linear_RegModel,Y, test_si
ze=0.2)
lm = LinearRegression().fit(X_train,y_train)
y_pred = lm.predict(X_test)
rmse = np.sqrt(metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:'+str(rmse))
```

Root Mean Squared Error:0.00929721142570161

```
In [478]: Ridge_RegModel = final_tmdb.filter(['popularity', 'cast_points', 'production_p
oints', 'Director_points', 'year_of_release',
        'genre_score_atRelease', 'genre_score_atPresent', 'oscars_won', 'inflated
d_budget', 'inflated_revenue', 'inflated_profit', 'dur_score',
        'Current_popularity', 'Release popularity'])
Ridge_RegModel.isna().sum()
Y = final_tmdb.filter(['endurance'])
Y
le = LabelEncoder()
Ridge_RegModel = Ridge_RegModel.apply(le.fit_transform)
#feature,result = linear_reg.loc[:,linear_reg.columns != 'popularity'], linear_
reg['popularity']
X_train, X_test, y_train, y_test = train_test_split(Ridge_RegModel,Y, test_siz
e=0.2)
lm = Ridge(alpha = 10).fit(X_train,y_train)
y_pred = lm.predict(X_test)
rmse = np.sqrt(metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:'+str(rmse))
```

Root Mean Squared Error:0.010735150095225006

```
In [481]: Lasso_RegModel = final_tmdb.filter(['popularity', 'cast_points', 'production_p
oints', 'Director_points', 'year_of_release',
        'genre_score_atRelease', 'genre_score_atPresent', 'oscars_won', 'inflated
d_budget', 'inflated_revenue', 'inflated_profit', 'dur_score',
        'Current_popularity', 'Release popularity'])
Lasso_RegModel.isna().sum()
Y = final_tmdb.filter(['endurance'])
Y
le = LabelEncoder()
Lasso_RegModel = Lasso_RegModel.apply(le.fit_transform)
#feature,result = linear_reg.loc[:,linear_reg.columns != 'popularity'], linear_
reg['popularity']
X_train, X_test, y_train, y_test = train_test_split(Lasso_RegModel,Y, test_siz
e=0.2)
lm = Lasso(alpha = 0.1).fit(X_train,y_train)
y_pred = lm.predict(X_test)
rmse = np.sqrt(metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:'+str(rmse))
```

Root Mean Squared Error:0.011120279629658275


```
In [475]: forest_model = RandomForestRegressor(
            n_estimators=400, max_features=0.4,
            min_samples_leaf=30, n_jobs=-1)

X = final_tmdb.filter(['popularity', 'cast_points', 'production_points', 'Director_points', 'year_of_release',
                      'genre_score_atRelease', 'genre_score_atPresent', 'oscars_won', 'inflated_budget', 'inflated_revenue', 'inflated_profit', 'dur_score',
                      'Current_popularity', 'Release popularity', 'endurance'])
X.isna().sum()
Y = final_tmdb.filter(['endurance'])
X = X.apply(le.fit_transform)

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
lm = Lasso(alpha = 100).fit(X_train, y_train)
y_pred = lm.predict(X_test)
rmse = np.sqrt(metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:' + str(rmse))
```

Root Mean Squared Error:0.033863158241783355

```
In [390]: import seaborn as sns
```

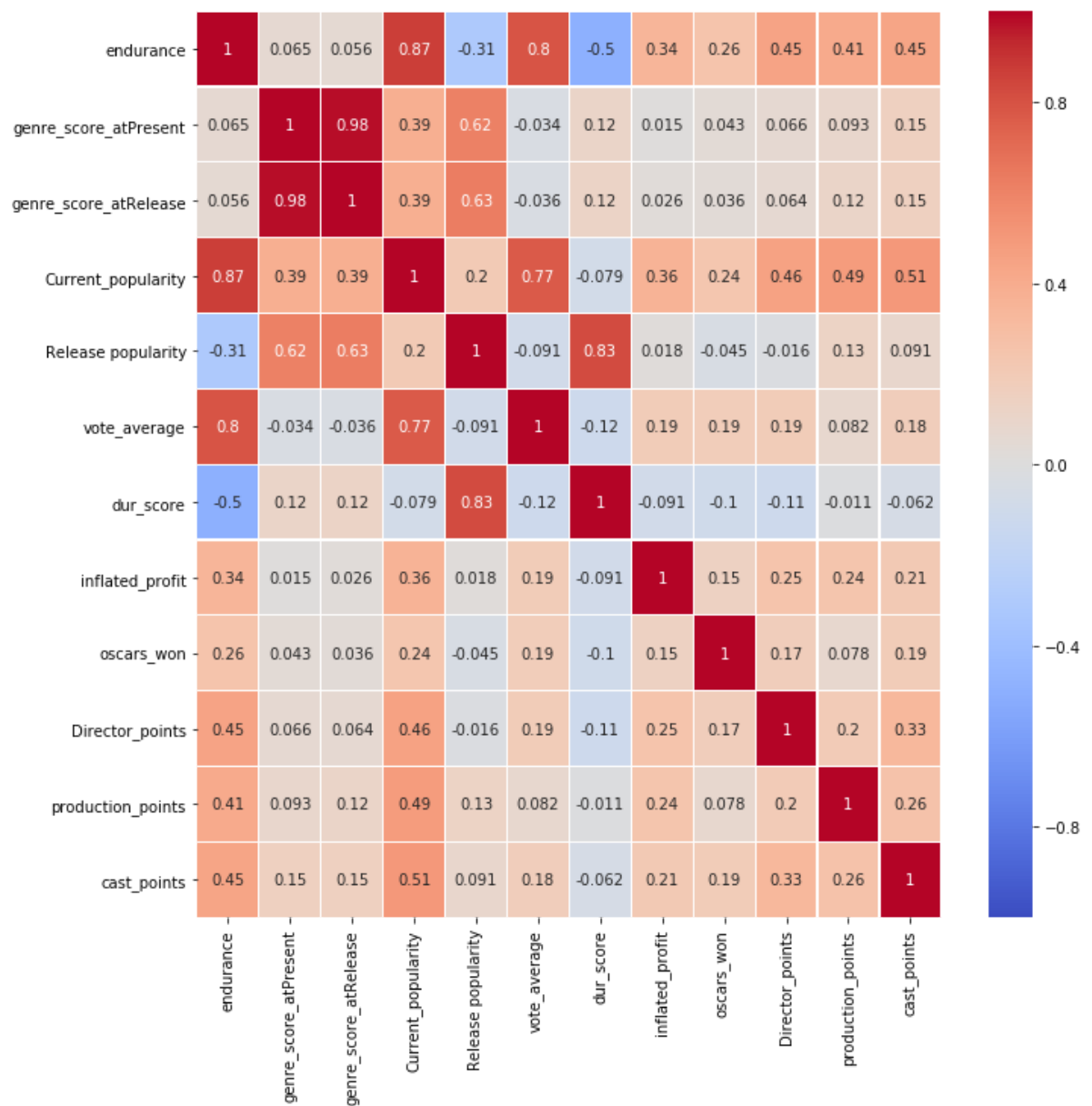
```
In [391]: final_tmdb.dtypes
```

```
Out[391]: id                int64
popularity                float64
budget                   int64
revenue                  int64
original_title            object
cast                     object
director                 object
runtime                  int64
genres                   object
production_companies      object
release_date              object
vote_count                int64
vote_average              float64
release_year              int64
budget_adj                float64
revenue_adj               float64
cast_points               float64
production_points         float64
Director_points           float64
year_of_release           int64
decade_of_release         object
month_of_release           int64
genre_score_atRelease     float64
genre_score_atPresent     float64
oscars_won                float64
inflated_budget           float64
inflated_revenue          float64
inflated_profit           float64
dur_score                 float64
Current_popularity        float64
Release popularity        float64
endurance                 float64
dtype: object
```

```
In [392]: train_1 = final_tmdb.filter(['endurance', 'genre_score_atPresent', 'genre_score_atRelease', 'Current_popularity', 'Release popularity', 'vote_average', 'dur_score', 'inflated_profit', 'oscars_won', 'Director_points', 'production_points', 'cast_points'])
```

```
In [393]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [394]: plt.figure(figsize=(11,11))
sns.heatmap(train_1.corr(),vmin=-1,vmax=1,cmap='coolwarm',linewidths=0.2,annot
=True);
```



```
In [403]: final_tmdb.dtypes
```

```
Out[403]: id                int64
popularity                float64
budget                   int64
revenue                  int64
original_title            object
cast                     object
director                 object
runtime                  int64
genres                   object
production_companies      object
release_date              object
vote_count                int64
vote_average              float64
release_year              int64
budget_adj                float64
revenue_adj               float64
cast_points               float64
production_points         float64
Director_points           float64
year_of_release           int64
decade_of_release         object
month_of_release           int64
genre_score_atRelease     float64
genre_score_atPresent     float64
oscars_won                float64
inflated_budget           float64
inflated_revenue          float64
inflated_profit           float64
dur_score                 float64
Current_popularity        float64
Release_popularity        float64
endurance                 float64
dtype: object
```

Oscars effect on endurance

```
In [408]: df_winning = final_tmdb.loc[(final_tmdb['oscars_won'] > 0)]
```

```
In [413]: df_nominated = final_tmdb.loc[(final_tmdb['oscars_won'] == 0)]
```

```
In [426]: final_tmdb['endurance'].max()
```

```
Out[426]: 0.33499061356446613
```

```
In [427]: final_tmdb['endurance'].mean()
```

```
Out[427]: 0.09551030803635098
```

```
In [420]: df_winning['endurance'].mean()
```

```
Out[420]: 0.13283207114937368
```

```
In [421]: df_nominated['endurance'].mean()
```

```
Out[421]: 0.13154501914278316
```

Genres effect on movie's popularity over time

```
In [430]: western_past = final_tmdb.loc[(final_tmdb['decade_of_release'] == '1960-1970')]
```

```
In [513]: end = 0
count = 0
for index, row in western_past.iterrows():
    s = row['genres']
    l = s.split('|')
    #print(l)
    if 'Western' in l or 'War' in l:
        count+=1
    end = end + row['Current_popularity']
print(end)
print(count)
```

```
17.279343103183525
```

```
66
```

```
In [450]: western_current = final_tmdb.loc[(final_tmdb['decade_of_release'] == '2010-2020')]
```

In [453]: western_current

Out[453]:

	id	popularity	budget	revenue	original_title	cast
0	135397	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...
1	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...
2	76341	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...
3	262500	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel...
4	140607	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam D...
5	168259	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle ...
6	281957	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Will Poulter Domhn...
7	87101	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clar...
8	286217	7.667400	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff ...
9	211672	7.404165	74000000	1156730962	Minions	Sandra Bullock Jon Hamm Michael Keaton Allison...
10	150540	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Ha...
11	206647	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz L��a Seydoux Ralp...
12	76757	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sean Bean Eddie Redm...
13	264660	6.118847	15000000	36869414	Ex Machina	Domhnall Gleeson Alicia Vikander Oscar Isaac S...
14	257344	5.984995	88000000	243637091	Pixels	Adam Sandler Michelle Monaghan Peter Dinklage ...
15	99861	5.944927	280000000	1405035767	Avengers: Age of Ultron	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...

	id	popularity	budget	revenue	original_title	cast
16	273248	5.898400	44000000	155760117	The Hateful Eight	Samuel L. Jackson Kurt Russell Jennifer Jason ...
17	260346	5.749758	48000000	325771424	Taken 3	Liam Neeson Forest Whitaker Maggie Grace Famke...
18	102899	5.573184	130000000	518602163	Ant-Man	Paul Rudd Michael Douglas Evangeline Lilly Cor...
19	150689	5.556818	95000000	542351353	Cinderella	Lily James Cate Blanchett Richard Madden Helen...
20	131634	5.476958	160000000	650523427	The Hunger Games: Mockingjay - Part 2	Jennifer Lawrence Josh Hutcherson Liam Hemswor...
21	158852	5.462138	190000000	209035668	Tomorrowland	Britt Robertson George Clooney Raffey Cassidy ...
22	307081	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rachel McAdams Forest Whitaker...
23	254128	4.907832	110000000	470490832	San Andreas	Dwayne Johnson Alexandra Daddario Carla Gugino...
24	216015	4.710402	40000000	569651467	Fifty Shades of Grey	Dakota Johnson Jamie Dornan Jennifer Ehle Eloi...
25	318846	4.648046	28000000	133346506	The Big Short	Christian Bale Steve Carell Ryan Gosling Brad ...
26	177677	4.566713	150000000	682330139	Mission: Impossible - Rogue Nation	Tom Cruise Jeremy Renner Simon Pegg Rebecca Fe...
27	214756	4.564549	68000000	215863606	Ted 2	Mark Wahlberg Seth MacFarlane Amanda Seyfried ...
28	207703	4.503789	81000000	403802136	Kingsman: The Secret Service	Taron Egerton Colin Firth Samuel L. Jackson Mi...
29	314365	4.062293	20000000	88346473	Spotlight	Mark Ruffalo Michael Keaton Rachel McAdams Lie...
...
6067	230158	0.112041	0	0	Aziz Ansari: Buried Alive	Aziz Ansari
6068	187628	0.021150	0	0	K�nn ingen sorg	Adam Lundgren Josefin Neld�n Disa �strand Jo...

	id	popularity	budget	revenue	original_title	cast
6069	159770	0.146962	0	0	The Dirties	Matt Johnson Owen Williams Krista Madison Bran...
6070	216790	0.019930	0	0	Sacro GRA	Roberto Giuliani Francesco De Santis
6071	184149	0.019755	0	0	Somm	nan
6072	182127	0.080410	37884	37884	Yip Man: The Final Fight	Anthony Wong Anita Yuen Gillian Chung Jordan C...
6073	225564	0.016326	1400000	0	100% Cachemire	ValÃ©rie Lemercier Gilles Lellouche Marina FoÃ...
6074	158752	0.015951	650	0	Escape from Tomorrow	Roy Abramsohn Elena Schuber Katelynn Rodriguez...
6075	193524	0.015722	729043	729043	The Stone Roses: Made of Stone	Ian Brown Gary 'Mani' Mounfield John Squire Al...
6076	173491	0.007821	0	0	TINY: A Story About Living Small	Daryl Gibson Christopher Smith Paul H. Smith W...
6077	225703	0.015247	4500000	0	The Starving Games	Brant Daugherty Maiara Walsh Lauren Bowles Cod...
6078	124623	0.013510	0	0	I Know That Voice	Kevin Conroy Jim Cummings Rachael MacFarlane J...
6079	213417	0.011298	0	0	How to Be a Man	Gavin McInnes Liam Aiken Paulo Costanzo Brian ...
6080	190250	0.017565	0	0	ABE	Sam Hoare Claire Huskisson Emily Baxter
6081	203793	0.010335	30000000	68129518	ĐjÑ,Đ°Đ»Đ,Đ½Đ³Ñ€Đ°Đ,	Thomas Kretschmann Yanina Studilina Philippe R...
6082	141058	0.010213	0	0	A Viking Saga: The Darkest Day	Elen Rhys Mark Lewis Jones Gary Mavers Marc Pi...
6083	178850	0.008540	0	0	Beyonce: Life Is But A Dream	BeyoncÃ© Knowles Jay-Z Tina Knowles
6084	238749	0.035367	0	0	A Very Merry Mix-Up	Alicia Witt Mark Wiebe Scott Gibson Susan Hoga...
6085	194121	0.008078	0	0	Dark Circles	Johnathon Schaech Andrea Frankle Philippe Bren...

	id	popularity	budget	revenue	original_title	cast
6086	246308	0.005853	0	0	Master of the Universe	Rainer Voss Angela Merkel
6087	279954	0.004323	500	0	Portal: Survive!	Monique Blanchard Bradley Mixon
6088	203217	0.040099	0	0	Minha Mãe é uma Peça	Paulo Gustavo Mariana Xavier Rodrigo Pandolfo ...
6089	160220	0.003611	348164	0	Bronies: The Extremely Unexpected Adult Fans o...	Tara Strong John de Lancie Cathy Weseluck Pete...
6090	228009	0.002838	0	0	Russell Peters: Notorious	Russell Peters
6091	224908	0.002719	0	0	Embrace of the Vampire	Sharon Hinnendael Kaniehtio Horn Chelsey Reis...
6092	173455	0.006115	137460	137460	Mistaken for Strangers	Matt Berninger Tom Berninger Aaron Dessner Bry...
6093	203351	0.002648	0	0	Atlantic Rim	Graham Greene David Chokachi Jackie Moore Nico...
6094	376823	0.002647	0	0	Trophy Kids	nan
6095	231401	0.002165	0	0	The Sound of Music Live!	Laura Benanti Stephen Moyer Carrie Underwood C...
6096	174323	0.000620	0	0	G.B.F.	Michael J. Willett Paul Iacono Sasha Pieterse ...

3616 rows × 32 columns

```
In [516]: end =0
count =0
for index, row in western_current.iterrows():
    s = str(row['genres'])
    l = s.split('|')
    #print(l)
    if 'Western' in l or 'War' in l:
        count+=1
    end = end + row['Current_popularity']
print(end)
print(count)
```

25.55045373379339

93

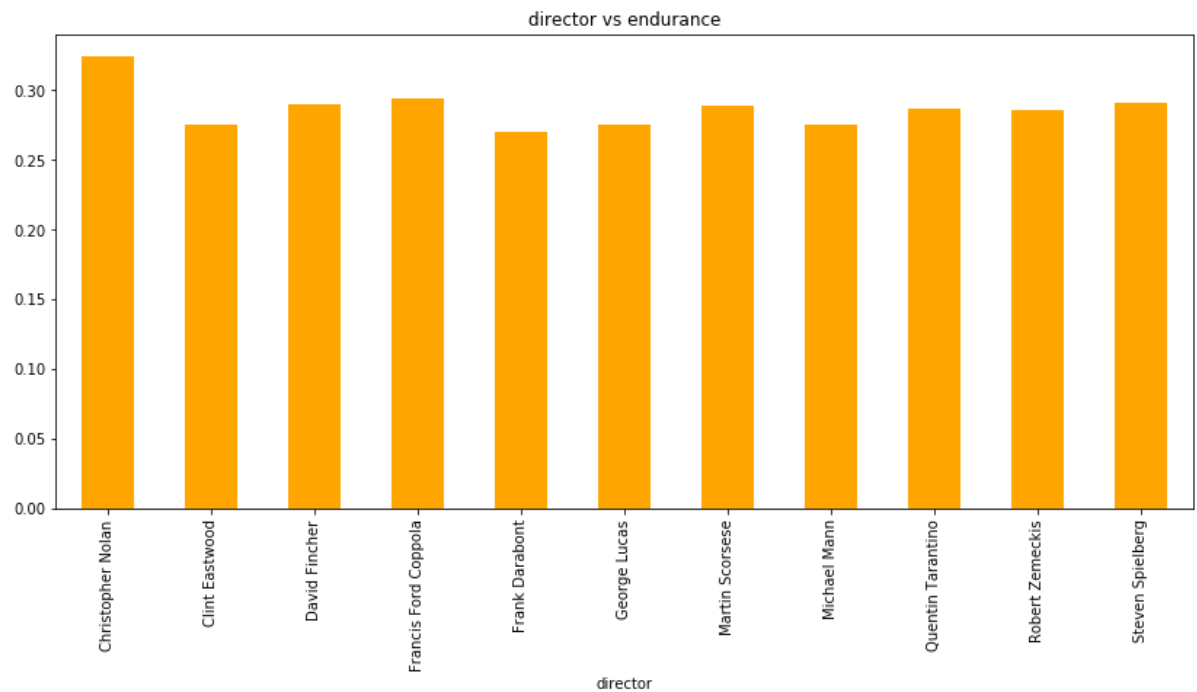
```
In [524]: final_tmdb.isna().sum()
```

```
Out[524]: id                                0
popularity                                0
budget                                    0
revenue                                  0
original_title                           0
cast                                      0
director                                44
runtime                                  0
genres                                   23
production_companies                     0
release_date                             0
vote_count                               0
vote_average                             0
release_year                             0
budget_adj                               0
revenue_adj                              0
cast_points                              0
production_points                         0
Director_points                          0
year_of_release                          0
decade_of_release                        0
month_of_release                         0
genre_score_atRelease                    0
genre_score_atPresent                    0
oscars_won                              0
inflated_budget                          0
inflated_revenue                         0
inflated_profit                          0
dur_score                                0
Current_popularity                       0
Release_popularity                       0
endurance                                0
dtype: int64
```

```
In [540]: tmdb_high_endurance = final_tmdb.loc[(final_tmdb['endurance'] > 0.27)]
```

```
In [542]: plt.figure(figsize=(14, 6))  
tmdb_high_endurance.groupby(['director']).mean()['endurance'].plot(kind="bar",  
title='director vs endurance',color='orange')
```

Out[542]: <matplotlib.axes._subplots.AxesSubplot at 0x1ebdd5ae240>



```
In [569]: import math
sorted_final = final_tmdb.copy()
sorted_final = sorted_final.sort_values('Current_popularity',ascending=False)
result = sorted_final.head(math.ceil(0.023*10900))

result
```

Out[569]:

	id	popularity	budget	revenue	original_title	cast	director
2881	155	8.466668	185000000	1001921825	The Dark Knight	Christian Bale Michael Caine Heath Ledger Aaro...	Christoph N...
4375	49026	6.591277	250000000	1081041287	The Dark Knight Rises	Christian Bale Michael Caine Gary Oldman Anne ...	Christoph N...
632	157336	24.949134	165000000	621752480	Interstellar	Matthew McConaughey Jessica Chastain Anne Hath...	Christoph N...
1925	27205	9.363643	160000000	825500000	Inception	Leonardo DiCaprio Joseph Gordon-Levitt Ellen P...	Christoph N...
6208	272	5.400826	150000000	374218673	Batman Begins	Christian Bale Michael Caine Liam Neeson Katie...	Christoph N...
10256	424	2.377288	22000000	321265768	Schindler's List	Liam Neeson Ben Kingsley Ralph Fiennes Carolyn...	Steven Spielt...
6581	1422	2.819750	90000000	289847354	The Departed	Leonardo DiCaprio Matt Damon Jack Nicholson Ma...	McScors
1393	16869	6.098027	70000000	319131050	Inglourious Basterds	Brad Pitt Mélanie Laurent Christoph Waltz Eli...	Quentin Taran...
7290	238	5.738034	6000000	245066411	The Godfather	Marlon Brando Al Pacino James Caan Richard S. ...	Francis Ford Cop...
4191	13	6.715966	55000000	677945399	Forrest Gump	Tom Hanks Robin Wright Gary Sinise Mykelti Wil...	Rob Zeme...
9004	857	2.170136	70000000	481840909	Saving Private Ryan	Tom Hanks Tom Sizemore Dennis Farina Max Marti...	Steven Spielt...
6583	1124	2.582095	40000000	109676311	The Prestige	Hugh Jackman Christian Bale Michael Caine Scar...	Christoph N...
9839	578	2.563191	7000000	470654000	Jaws	Roy Scheider Robert Shaw Richard Dreyfuss Lorr...	Steven Spielt...
5443	106646	4.877927	100000000	392000694	The Wolf of Wall Street	Leonardo DiCaprio Jonah Hill Margot Robbie Kyl...	McScors
3930	640	2.973115	52000000	352114312	Catch Me If You Can	Leonardo DiCaprio Tom Hanks Christopher Walken...	Steven Spielt...

	id	popularity	budget	revenue	original_title	cast	direc
8089	807	4.765359	33000000	327311859	Se7en	Brad Pitt Morgan Freeman Gwyneth Paltrow John ...	David Finc
8101	949	1.847583	60000000	187436818	Heat	Al Pacino Robert De Niro Val Kilmer Jon Voight...	Michael M
9791	240	3.264571	13000000	47542841	The Godfather: Part II	Al Pacino Robert Duvall Diane Keaton Robert De...	Francis F Cop
1933	11324	3.176169	80000000	294804195	Shutter Island	Leonardo DiCaprio Mark Ruffalo Ben Kingsley Em...	Ma Scors
4376	68718	5.944518	100000000	425368238	Django Unchained	Jamie Foxx Christoph Waltz Leonardo DiCaprio K...	Que Taran
8919	601	2.900556	10500000	792910554	E.T. the Extra-Terrestrial	Henry Thomas Drew Barrymore Robert MacNaughton...	Ste Spielb
10013	769	2.659926	25000000	46836394	Goodfellas	Robert De Niro Ray Liotta Joe Pesci Lorraine B...	Ma Scors
5245	597	4.355219	200000000	1845034188	Titanic	Kate Winslet Leonardo DiCaprio Frances Fisher ...	Jar Came
2891	4922	2.537342	150000000	333932083	The Curious Case of Benjamin Button	Cate Blanchett Brad Pitt Tilda Swinton Julia O...	David Finc
10257	329	2.204926	63000000	920100000	Jurassic Park	Sam Neill Laura Dern Jeff Goldblum Richard Att...	Ste Spielb
2415	550	8.947905	63000000	100853753	Fight Club	Edward Norton Brad Pitt Meat Loaf Jared Leto H...	David Finc
8400	85	4.578300	18000000	389925971	Raiders of the Lost Ark	Harrison Ford Karen Allen Paul Freeman Ronald ...	Ste Spielb
9211	89	3.536655	48000000	474171806	Indiana Jones and the Last Crusade	Harrison Ford Sean Connery Denholm Elliott Ali...	Ste Spielb
6097	105	6.095293	19000000	381109762	Back to the Future	Michael J. Fox Christopher Lloyd Lea Thompson ...	Rol Zeme
7444	1949	1.464555	65000000	84785914	Zodiac	Jake Gyllenhaal Robert Downey Jr. Mark Ruffalo...	David Finc
...

	id	popularity	budget	revenue	original_title	cast	direc
1452	11322	1.138277	80000000	214104620	Public Enemies	Christian Bale Johnny Depp Giovanni Ribisi Bil...	Michael M
4401	68734	2.003232	44500000	232324128	Argo	Ben Affleck Bryan Cranston Alan Arkin John Goo...	Ben Affl
9842	3175	1.019569	11000000	20000000	Barry Lyndon	Ryan O'Neal Marisa Berenson Hardy KrÃ¼ger Dian...	Star Kub
5442	101299	4.921275	130000000	847423452	The Hunger Games: Catching Fire	Jennifer Lawrence Josh Hutcherson Liam Hemswor...	Frai Lawre
9017	73	1.333006	20000000	23875127	American History X	Edward Norton Edward Furlong Beverly D'Angelo ...	Tony K
10475	4808	0.920094	4000000	13474588	Charade	Cary Grant Audrey Hepburn Walter Matthau James...	Star Do
10260	5503	1.708867	44000000	368875760	The Fugitive	Harrison Ford Tommy Lee Jones Sela Ward Julian...	Andrew Di
9215	207	2.470407	16400000	235860116	Dead Poets Society	Robin Williams Ethan Hawke Robert Sean Leonard...	Peter V
9250	11352	0.494235	31000000	74134790	Always	Richard Dreyfuss Holly Hunter John Goodman Aud...	Ste Spielt
5395	8197	0.245162	35000000	25105255	Midnight in the Garden of Good and Evil	John Cusack Kevin Spacey Jude Law Alison Eastw...	(Eastw
10254	137	2.571339	14600000	70906973	Groundhog Day	Bill Murray Andie MacDowell Chris Elliott Step...	Harold Ra
7409	675	4.878162	150000000	938212738	Harry Potter and the Order of the Phoenix	Daniel Radcliffe Rupert Grint Emma Watson Mich...	David Ye
9023	9631	1.242498	50000000	44547681	The Negotiator	Samuel L. Jackson Kevin Spacey David Morse Ron...	F. Gary C
4989	616	1.622536	140000000	456758981	The Last Samurai	Tom Cruise Ken Watanabe William Atherton Chad ...	Edward Zv
5939	241287	0.107333	0	0	Tropico	Lana Del Rey Shaun Ross	Anth Man

	id	popularity	budget	revenue	original_title	cast	direc
1403	18785	3.153060	35000000	459270619	The Hangover	Bradley Cooper Ed Helms Zach Galifianakis Justin Theroux	Todd Phillips
6098	2108	2.834739	1000000	51525171	The Breakfast Club	Emilio Estevez Anthony Michael Hall Judd Nelson Molly Ringwald Andie MacDowell	John Hughes
7427	4347	2.066397	30000000	129266061	Atonement	Keira Knightley James McAvoy Romola Garai Saoirse Ronan	Joe Wright
4002	8051	0.741980	25000000	17000000	Punch-Drunk Love	Adam Sandler Emily Watson Philip Seymour Hoffman Catherine Keener	Paul Thomas Anderson
6989	652	3.177517	175000000	497409852	Troy	Brad Pitt Orlando Bloom Eric Bana Brian Cox Seaneen Williams	Wolfgang Peter
7525	4512	0.653746	30000000	14711793	The Assassination of Jesse James by the Coward Robert Ford	Brad Pitt Casey Affleck Sam Shepard Mary-Louise Parker Timothy Spall	Andrew Dominik
3476	39538	1.045866	60000000	135458097	Contagion	Matt Damon Gwyneth Paltrow Kate Winslet Jude Law	Steinberg
9546	11051	0.353630	7000000	8373585	The Last Temptation of Christ	Willem Dafoe Harvey Keitel Barbara Hershey David Byrne	McScors
10506	679	2.485419	18500000	131060248	Aliens	Sigourney Weaver Carrie Henn Michael Biehn Lance Rothenberg	Jarvis
5248	782	3.077454	36000000	12532777	Gattaca	Ethan Hawke Jude Law Gore Vidal Uma Thurman Alfred Molina	And Nic
8529	1592	0.759509	30000000	102616183	Primal Fear	Richard Gere Edward Norton Laura Linney John M. Larson	Gregg
5091	11615	0.431188	50000000	38955598	The Life of David Gale	Kevin Spacey Kate Winslet Laura Linney Gabriel Byrne	Alan
1396	14160	4.908902	175000000	731342744	Up	Ed Asner Christopher Plummer Jordan Nagai Bob Odenkirk	Peterson F Do
9803	5121	0.415866	3000000	12800000	The Sugarland Express	Goldie Hawn William Atherton Ben Johnson Louis L. Armato	Ste Spielt
10216	10747	0.597297	3700000	31800000	The Outlaw Josey Wales	Clint Eastwood Chief Dan George Sondra Locke Burt Reynolds	C Eastw

251 rows × 32 columns

```
In [570]: imdb_top_250 = pd.read_csv(r'C:\Users\preet\Desktop\Movies\imdb-top-250.csv')
```

```
In [571]: imdb_top_250=imdb_top_250.dropna()  
imdb_top_250
```

Out[571]:

	rank	original_title
1	1.0	The Shawshank Redemption
2	2.0	The Godfather
3	3.0	The Godfather: Part II
4	4.0	The Dark Knight
5	5.0	12 Angry Men
6	6.0	Schindler's List
7	7.0	The Lord of the Rings: The Return of the King
8	8.0	Pulp Fiction
9	9.0	Il buono, il brutto, il cattivo
10	10.0	Fight Club
11	11.0	The Lord of the Rings: The Fellowship of the Ring
12	12.0	Forrest Gump
13	13.0	Star Wars: Episode V - The Empire Strikes Back
14	14.0	Inception
15	15.0	The Lord of the Rings: The Two Towers
16	16.0	One Flew Over the Cuckoo's Nest
17	17.0	Goodfellas
18	18.0	The Matrix
19	19.0	Shichinin no samurai
20	20.0	Cidade de Deus
21	21.0	Se7en
22	22.0	Star Wars
23	23.0	The Silence of the Lambs
24	24.0	It's a Wonderful Life
25	25.0	La vita è bella
26	26.0	The Usual Suspects
27	27.0	Sen to Chihiro no kamikakushi
28	28.0	Saving Private Ryan
29	29.0	Léon
30	30.0	The Green Mile
...
221	221.0	The Maltese Falcon
222	222.0	Les diaboliques
223	223.0	Donnie Darko
224	224.0	Sanjuro

	rank	original_title
225	225.0	The Terminator
226	226.0	Gandhi
227	227.0	Dead Poets Society
228	228.0	La haine
229	229.0	Groundhog Day
230	230.0	Blade Runner 2049
231	231.0	La La Land
232	232.0	Le notti di Cabiria
233	233.0	The Wizard of Oz
234	234.0	Jaws
235	235.0	The Help
236	236.0	Faa yeung nin wa
237	237.0	Paper Moon
238	238.0	Gangs of Wasseypur
239	239.0	Mandariinid
240	240.0	Before Sunset
241	241.0	A Wednesday
242	242.0	Paris, Texas
243	243.0	Ah-ga-ssi
244	244.0	The Bourne Ultimatum
245	245.0	Guardians of the Galaxy
246	246.0	PK
247	247.0	Pirates of the Caribbean: The Curse of the Bla...
248	248.0	Tenkû no shiro Rapyuta
249	249.0	The Best Years of Our Lives
250	250.0	Swades: We, the People

250 rows × 2 columns

```
In [572]: imdb_top_250=pd.merge(imdb_top_250,result,on='original_title')
```

In [573]: `imdb_top_250`

Out[573]:

	rank	original_title	id	popularity	budget	revenue	cast	
0	1.0	The Shawshank Redemption	278	7.192039	25000000	28341469	Tim Robbins Morgan Freeman Bob Gunton William ...	Frank
1	2.0	The Godfather	238	5.738034	6000000	245066411	Marlon Brando Al Pacino James Caan Richard S. ...	Fr
2	3.0	The Godfather: Part II	240	3.264571	13000000	47542841	Al Pacino Robert Duvall Diane Keaton Robert De...	Fr
3	4.0	The Dark Knight	155	8.466668	185000000	1001921825	Christian Bale Michael Caine Heath Ledger Aaro...	C
4	6.0	Schindler's List	424	2.377288	22000000	321265768	Liam Neeson Ben Kingsley Ralph Fiennes Carol...	
5	7.0	The Lord of the Rings: The Return of the King	122	7.122455	94000000	1118888979	Elijah Wood Ian McKellen Viggo Mortensen Liv T...	Pete
6	8.0	Pulp Fiction	680	8.093754	8000000	213928762	John Travolta Samuel L. Jackson Uma Thurman Br...	
7	10.0	Fight Club	550	8.947905	63000000	100853753	Edward Norton Brad Pitt Meat Loaf Jared Leto H...	Dav
8	11.0	The Lord of the Rings: The Fellowship of the Ring	120	8.575419	93000000	871368364	Elijah Wood Ian McKellen Viggo Mortensen Liv T...	Pete
9	12.0	Forrest Gump	13	6.715966	55000000	677945399	Tom Hanks Robin Wright Gary Sinise Mykelti Wil...	Robert
10	14.0	Inception	27205	9.363643	160000000	825500000	Leonardo DiCaprio Joseph Gordon-Levitt Ellen P...	C
11	15.0	The Lord of the Rings: The Two Towers	121	8.095275	79000000	926287400	Elijah Wood Ian McKellen Viggo Mortensen Liv T...	Pete
12	16.0	One Flew Over the Cuckoo's Nest	510	3.258151	3000000	108981275	Jack Nicholson Louise Fletcher Danny DeVito Wi...	Milo
13	17.0	Goodfellas	769	2.659926	25000000	46836394	Robert De Niro Ray Liotta Joe Pesci Lorraine B...	Martin
14	18.0	The Matrix	603	7.753899	63000000	463517383	Keanu Reeves Laurence Fishburne Carrie-Anne Mo...	Wacho V

	rank	original_title	id	popularity	budget	revenue	cast	
15	21.0	Se7en	807	4.765359	33000000	327311859	Brad Pitt Morgan Freeman Gwyneth Paltrow John ...	Dav
16	22.0	Star Wars	11	12.037933	11000000	775398007	Mark Hamill Harrison Ford Carrie Fisher Peter ...	Gec
17	23.0	The Silence of the Lambs	274	0.695385	19000000	272742922	Jodie Foster Anthony Hopkins Scott Glenn Ted L...	
18	26.0	The Usual Suspects	629	3.440238	6000000	23341568	Stephen Baldwin Gabriel Byrne Chazz Palminteri...	Bry
19	28.0	Saving Private Ryan	857	2.170136	70000000	481840909	Tom Hanks Tom Sizemore Dennis Farina Max Marti...	
20	30.0	The Green Mile	497	2.719461	60000000	284600000	Tom Hanks Michael Clarke Duncan David Morse Bo...	Frank
21	31.0	Interstellar	157336	24.949134	165000000	621752480	Matthew McConaughey Jessica Chastain Anne Hath...	C
22	32.0	Psycho	539	2.610362	806948	32000000	Anthony Perkins Vera Miles John Gavin Janet Le...	Alfred
23	33.0	American History X	73	1.333006	20000000	23875127	Edward Norton Edward Furlong Beverly D'Angelo
24	40.0	The Departed	1422	2.819750	90000000	289847354	Leonardo DiCaprio Matt Damon Jack Nicholson Ma...	Martin
25	42.0	Back to the Future	105	6.095293	19000000	381109762	Michael J. Fox Christopher Lloyd Lea Thompson ...	Robert
26	45.0	Raiders of the Lost Ark	85	4.578300	18000000	389925971	Harrison Ford Karen Allen Paul Freeman Ronald ...	
27	46.0	Gladiator	98	4.271452	103000000	457640427	Russell Crowe Joaquin Phoenix Connie Nielsen O...	R
28	47.0	The Lion King	8587	4.782688	45000000	788241776	Jonathan Taylor Thomas Matthew Broderick James...	.
29	48.0	The Prestige	1124	2.582095	40000000	109676311	Hugh Jackman Christian Bale Michael Caine Scar...	C
...	

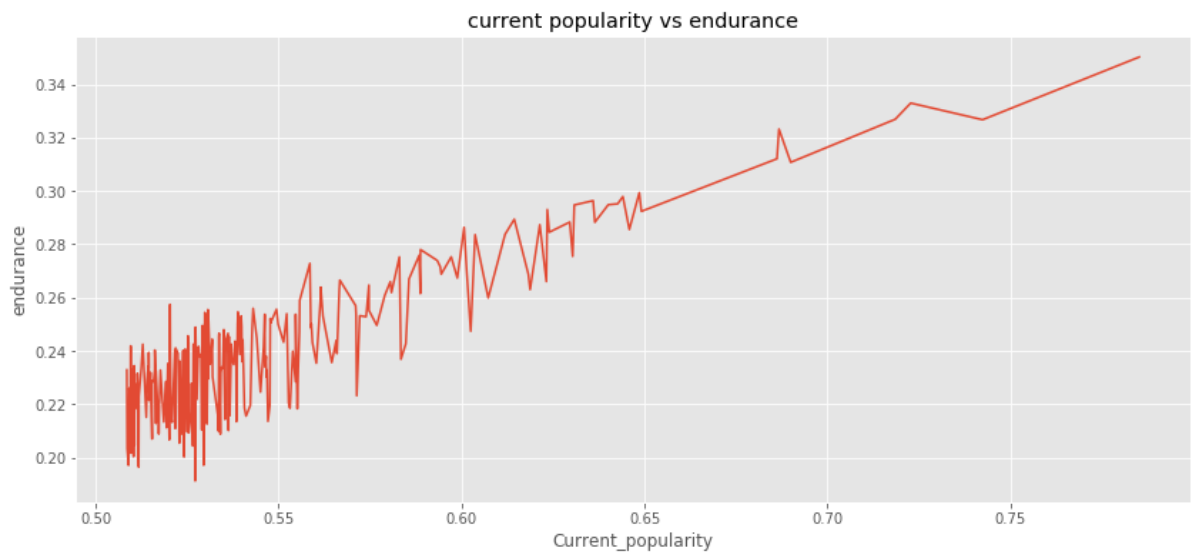
	rank	original_title	id	popularity	budget	revenue	cast	
58	131.0	Chinatown	829	1.650848	6000000	30000000	Jack Nicholson Faye Dunaway John Huston Perry ...	Roma
59	140.0	A Beautiful Mind	453	2.169826	60000000	313542341	Russell Crowe Ed Harris Jennifer Connelly Chri...	Rc
60	143.0	Casino	524	1.552666	52000000	116112375	Robert De Niro Sharon Stone Joe Pesci James Wo...	Martin
61	145.0	Inside Out	150540	6.326804	175000000	853708609	Amy Poehler Phyllis Smith Richard Kind Bill Ha...	Pi
62	149.0	The Wolf of Wall Street	106646	4.877927	100000000	392000694	Leonardo DiCaprio Jonah Hill Margot Robbie Kyl...	Martin
63	152.0	Blade Runner	78	4.215642	28000000	32768670	Harrison Ford Rutger Hauer Sean Young Edward J...	R
64	157.0	No Country for Old Men	6977	2.730114	25000000	171600000	Tommy Lee Jones Javier Bardem Josh Brolin Wood...	Joel C
65	159.0	The Sixth Sense	745	2.578731	40000000	672806292	Bruce Willis Haley Joel Osment Toni Collette O...	S
66	162.0	Fargo	275	2.775756	7000000	60611975	William H. Macy Frances McDormand Steve Buscem...	
67	163.0	Gran Torino	13223	2.245306	33000000	269958228	Clint Eastwood Christopher Carley Bee Vang Ahn...	Clint
68	164.0	The Deer Hunter	11778	0.959754	15000000	50000000	Robert De Niro John Cazale John Savage Christo...	Micha
69	168.0	The Big Lebowski	115	2.556082	15000000	27739163	Jeff Bridges John Goodman Steve Buscemi Julian...	
70	170.0	Shutter Island	11324	3.176169	80000000	294804195	Leonardo DiCaprio Mark Ruffalo Ben Kingsley Em...	Martin
71	178.0	Gone Girl	210577	6.438727	61000000	369330363	Ben Affleck Rosamund Pike Carrie Coon Neil Pat...	Dav
72	181.0	Jurassic Park	329	2.204926	63000000	920100000	Sam Neill Laura Dern Jeff Goldblum Richard Att...	

	rank	original_title	id	popularity	budget	revenue	cast	
73	188.0	The Grand Budapest Hotel	120467	4.930820	30000000	174600318	Ralph Fiennes Tony Revolori F. Murray Abraham ...	Wes
74	190.0	The Truman Show	37165	4.180540	60000000	264118201	Jim Carrey Laura Linney Noah Emmerich Natascha...	I
75	197.0	12 Years a Slave	76203	3.625529	20000000	187000000	Chiwetel Ejiofor Michael Fassbender Lupita Nyong'o ...	Steve
76	200.0	Million Dollar Baby	70	1.648554	30000000	216763646	Clint Eastwood Hilary Swank Morgan Freeman Jay...	Clint
77	207.0	Spotlight	314365	4.062293	20000000	88346473	Mark Ruffalo Michael Keaton Rachel McAdams Lieke...	Tom
78	210.0	Prisoners	146233	2.032167	46000000	122126687	Hugh Jackman Jake Gyllenhaal Paul Dano Maria B...	Denis
79	214.0	Catch Me If You Can	640	2.973115	52000000	352114312	Leonardo DiCaprio Tom Hanks Christopher Walken ...	
80	216.0	Harry Potter and the Deathly Hallows: Part 2	12445	5.711315	125000000	1327817822	Daniel Radcliffe Rupert Grint Emma Watson Alan...	D.
81	218.0	Barry Lyndon	3175	1.019569	11000000	20000000	Ryan O'Neal Marisa Berenson Hardy Krüger Dian...	Stanl
82	227.0	Dead Poets Society	207	2.470407	16400000	235860116	Robin Williams Ethan Hawke Robert Sean Leonard ...	I
83	229.0	Groundhog Day	137	2.571339	14600000	70906973	Bill Murray Andie MacDowell Chris Elliott Step...	Har
84	234.0	Jaws	578	2.563191	7000000	470654000	Roy Scheider Robert Shaw Richard Dreyfuss Lorr...	
85	244.0	The Bourne Ultimatum	2503	2.755051	70000000	442824138	Matt Damon Julia Stiles David Strathairn Scott...	Paul G
86	245.0	Guardians of the Galaxy	118340	14.311205	170000000	773312399	Chris Pratt Zoe Saldana Dave Bautista Vin Dies...	Ja
87	247.0	Pirates of the Caribbean: The Curse of the Bla...	22	6.864067	140000000	655011224	Johnny Depp Geoffrey Rush Orlando Bloom Keira ...	Gore

88 rows × 33 columns

```
In [580]: plt.figure(figsize=(14, 6))
sns.lineplot(result['Current_popularity'],result['endurance']).set_title('current popularity vs endurance')
```

Out[580]: Text(0.5, 1.0, 'current popularity vs endurance')



```
In [582]: plt.figure(figsize=(14, 6))
plt.style.use('ggplot')
sns.lineplot(result['Release_popularity'],result['endurance']).set_title('Release popularity vs endurance')
```

Out[582]: Text(0.5, 1.0, 'Release popularity vs endurance')

