Data Analysis Workflow for Microsoft's Movie Industry Venture

Overview

I have been tasked with assisting Microsoft in their venture into the movie industry. The goal is to explore what type of films are currently performing the best at the box office and provide these findings to Microsoft's new movie studio executives.

Importing Data.

We imported the relevant libraries for our analysis.

```
# importing libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import datetime
```

Loaded the first data frame.

```
df1 = pd.read_csv(r"C:\Users\pc\Videos\projects\Phase_One_Proect\data\
MovieData.csv")
df1
                                                     production year \
                                         movie name
0
                            Madea's Family Reunion
                                                                 2006
1
                                                                 2006
                                             Krrish
2
                                  End of the Spear
                                                                 2006
3
                          A Prairie Home Companion
                                                                 2006
4
                                            Saw III
                                                                 2006
               The Nutcracker and the Four Realms
1931
                                                                 2018
1932
                                            Aguaman
                                                                 2018
1933
                         Ralph Breaks The Internet
                                                                 2018
1934
                      Mission: Impossible—Fallout
                                                                 2018
1935
      Fantastic Beasts: The Crimes of Grindelwald
                                                                 2018
      movie odid
                  production budget
                                      domestic box office \
0
         8220100
                            10000000
                                                  63257940
1
        58540100
                            10000000
                                                   1430721
2
        34620100
                            10000000
                                                  11748661
3
        24910100
                            10000000
                                                  20342852
```

4	5840100	1000	0000	80238724	
1931 1932 1933 1934 1935	298170100 213100100 263730100 248680100 222990100	13290 16000 17500 17800 20000	0000 0000 0000	54858851 333804251 200236625 220159104 159555901	
0 1 2 3 4 1931 1932 1933 1934 1935	internation	al_box_office 62581 31000000 175380 6373339 83638091 115435048 805605026 319167373 567297448 492664185	rating PG-13 Not Rated PG-13 PG-13 R PG PG-13 PG-13 PG-13 PG-13	Creative_type Contemporary Fiction Science Fiction Historical Fiction Contemporary Fiction Contemporary Fiction Fantasy Super Hero Kids Fiction Contemporary Fiction Fantasy	
genre	\		source	production_method	
0 Comedy		Bas	ed on Play	Live Action	
1 Action		Original :	Screenplay	Live Action	
2 Drama	I	Original :	Screenplay	Live Action	
3		Original	Screenplay	Live Action	
Comedy 4 Horror		Original :	Screenplay	Live Action	
1931 Advent		lk Tale/Legend,	/Fairytale	Live Action	
1932	Based	d on Comic/Gra	phic Novel	Live Action	
Action 1933		Original	Screenplay	Digital Animation	
Advent 1934		В	ased on TV	Live Action	
Actior 1935 Advent			Spin-Off	Live Action	
0 1 2	sequel runn 1.0 1.0 0.0	ning_time NaN NaN NaN			

```
3
          0.0
                       105.0
4
          1.0
                         NaN
          . . .
1931
          0.0
                        99.0
                       143.0
1932
          0.0
1933
          1.0
                       112.0
1934
          1.0
                       147.0
1935
         1.0
                       134.0
[1936 rows x 13 columns]
```

For consistency. I renamed some columns to names that would it easier to understand and analyse the data.

```
df1 = df1.rename(columns = {'international box office':'
internationalBoxOffice', 'domestic_box_office':'domesticBoxoffice',
'production budget':'productionBudget'})
df1
                                                    production_year \
                                             movie
0
                            Madea's Family Reunion
                                                                2006
1
                                            Krrish
                                                                2006
2
                                  End of the Spear
                                                                2006
3
                         A Prairie Home Companion
                                                                2006
4
                                           Saw III
                                                                2006
                                                                 . . .
               The Nutcracker and the Four Realms
1931
                                                                2018
1932
                                                                2018
                                           Aquaman
1933
                        Ralph Breaks The Internet
                                                                2018
1934
                      Mission: Impossible—Fallout
                                                                2018
      Fantastic Beasts: The Crimes of Grindelwald
1935
                                                                2018
      movie_odid productionBudget
                                     domesticBoxoffice
internationalBoxOffice \
         8220100
                           10000000
0
                                              63257940
62581
        58540100
                           10000000
                                               1430721
31000000
        34620100
                           10000000
                                              11748661
175380
        24910100
                           10000000
                                              20342852
6373339
         5840100
                                              80238724
                           10000000
83638091
                          132900000
1931
       298170100
                                              54858851
115435048
```

1932 213100100 805605026	160000000	333804251	
1933 263730100 319167373	175000000	200236625	
1934 248680100 567297448	178000000	220159104	
1935 222990100 492664185	20000000	159555901	
rating	creative_typ	oe	
source \ 0 PG-13	Contemporary Fiction	on	Based on
Play 1 Not Rated	Science Fiction	on	Original
Screenplay 2 PG-13	Historical Fiction	on	Original
Screenplay 3 PG-13 Screenplay	Contemporary Fiction	on	Original
4 R Screenplay	Contemporary Fiction	on	Original
1931 PG	Fantas	sy Based on Folk	
Tale/Legend/Fair		-	6 ' (6 1 '
1932 PG-13 Novel	Super Her	o Based on	Comic/Graphic
1933 PG	Kids Fictio	n	Original
Screenplay			0. <u>_</u> ga. :
1934 PG-13 on TV	Contemporary Fiction	on	Based
1935 PG-13 Spin-Off	Fantas	Sy	
production	method genre	sequel running_tim	ne
0 Live	Action Comedy	1.0 Na	aN
1 Live 2 Live	Action Action Action Drama	1.0 Na	
2 Live	Action Drama Action Comedy	0.0 Na 0.0 105	
	Action Horror	1.0 Na	
 1931 Live	Action Adventure	0.0 99	
	Action Action	0.0 143	
1933 Digital An		1.0 112	
	Action Action Action Adventure	1.0 147. 1.0 134.	
[1936 rows x 13	columns]		

Renamed the column movie names to movie

```
df1 = df1.rename(columns = {'movie name':' movie'})
df1
                                                     production year \
                                              movie
                            Madea's Family Reunion
                                                                 2006
1
                                             Krrish
                                                                 2006
2
                                  End of the Spear
                                                                 2006
3
                          A Prairie Home Companion
                                                                 2006
4
                                            Saw III
                                                                 2006
               The Nutcracker and the Four Realms
1931
                                                                 2018
1932
                                            Aquaman
                                                                 2018
1933
                         Ralph Breaks The Internet
                                                                 2018
                       Mission: Impossible—Fallout
1934
                                                                 2018
      Fantastic Beasts: The Crimes of Grindelwald
1935
                                                                 2018
      movie odid productionBudget domesticBoxoffice
internationalBoxOffice
         8220100
                           10000000
0
                                               63257940
62581
        58540100
                           10000000
                                                1430721
1
31000000
        34620100
                           10000000
                                               11748661
175380
3
        24910100
                           10000000
                                               20342852
6373339
         5840100
                           10000000
                                               80238724
83638091
1931
       298170100
                          132900000
                                               54858851
115435048
1932
       213100100
                          160000000
                                              333804251
805605026
1933
       263730100
                          175000000
                                              200236625
319167373
1934
       248680100
                          178000000
                                              220159104
567297448
1935
       222990100
                          200000000
                                              159555901
492664185
         rating
                         creative type
source
                 Contemporary Fiction
                                                                Based on
          PG-13
Play
                                                         Original
      Not Rated
                       Science Fiction
Screenplay
```

```
PG-13
                    Historical Fiction
                                                          Original
Screenplay
          PG-13
                  Contemporary Fiction
                                                          Original
Screenplay
              R
                  Contemporary Fiction
                                                          Original
Screenplay
. . .
             PG
                               Fantasy
                                        Based on Folk
1931
Tale/Legend/Fairytale
          PG-13
                            Super Hero
                                                 Based on Comic/Graphic
1932
Novel
             PG
                          Kids Fiction
1933
                                                          Original
Screenplay
1934
          PG-13
                 Contemporary Fiction
                                                                   Based
on TV
1935
          PG-13
                                Fantasy
Spin-Off
      production method
                              genre
                                      sequel
                                               running time
0
            Live Action
                             Comedy
                                         1.0
                                                        NaN
1
            Live Action
                             Action
                                         1.0
                                                        NaN
2
            Live Action
                              Drama
                                         0.0
                                                        NaN
3
            Live Action
                                                      105.0
                             Comedy
                                         0.0
4
            Live Action
                             Horror
                                         1.0
                                                        NaN
1931
                                         0.0
                                                       99.0
            Live Action Adventure
1932
            Live Action
                                         0.0
                                                      143.0
                             Action
1933
                                         1.0
      Digital Animation Adventure
                                                      112.0
1934
            Live Action
                                         1.0
                                                      147.0
                             Action
1935
            Live Action
                          Adventure
                                         1.0
                                                      134.0
[1936 rows x 13 columns]
```

Checked for null values in the running_time column.

```
df1['running_time'].isnull().sum()
114
```

Loaded the second data frame for the analysis.

1 2 3	2 3	12, 16, 10751] 10191 [12, 28, 878] 10138 [16, 35, 10751] 862	en en en
4	4	[28, 878, 12] 27205	en
26512 26513 26514 26515 26516	26512 26513 26514 26515 26516	[27, 18] 488143 [18, 53] 485975 [14, 28, 12] 381231 [10751, 12, 28] 366854 [53, 27] 309885	en en en en en
		original title	nonularity
release date	e \	originat_titte	popularity
_		the Deathly Hallows: Part 1	33.533
1 2010-03-26		How to Train Your Dragon	28.734
2 2010 - 05 - 07		Iron Man 2	28.515
3		Toy Story	28.005
1995 - 11 - 22 4		Inception	27.920
2010-07-16			
20512			0.000
26512 2018-10-13		Laboratory Conditions	
26513 2018-05-01		_EXHIBIT_84xxx_	0.600
26514 2018-10-01		The Last One	0.600
26515		Trailer Made	0.600
2018-06-22		The Church	0.600
26516 2018-10-05		The Church	0.600
		title	vote_average
	y Potter and	the Deathly Hallows: Part 1	7.7
10788 1		How to Train Your Dragon	7.7
7610 2		Iron Man 2	6.8
12368 3		Toy Story	7.9
10174		109 3:019	7.15
4 22186		Inception	8.3
22100			

26512	Laboratory Conditions	0.0
1		
26513	_EXHIBIT_84xxx_	0.0
1		
26514	The Last One	0.0
1		
26515	Trailer Made	0.0
1		
26516	The Church	0.0
1		
[26517 rows x 10 columns]		

Renamed the column original_title to movie.

```
df2 = df2.rename(columns = {'original title':' movie'})
df2
       Unnamed: 0
                               genre_ids
                                               id original_language
0
                         [12, 14, 10751]
                                            12444
1
                 1
                    [14, 12, 16, 10751]
                                            10191
                                                                   en
2
                 2
                           [12, 28, 878]
                                            10138
                                                                   en
3
                 3
                         [16, 35, 10751]
                                              862
                                                                   en
4
                           [28, 878, 12]
                 4
                                            27205
                                                                   en
. . .
                                                                  . . .
                                [27, 18]
26512
             26512
                                           488143
                                                                   en
                                [18, 53]
             26513
26513
                                           485975
                                                                   en
26514
             26514
                            [14, 28, 12]
                                           381231
                                                                   en
26515
             26515
                         [10751, 12, 28]
                                           366854
                                                                   en
                                [53, 27]
26516
             26516
                                           309885
                                                                   en
                                                 movie
                                                         popularity
release date \
       Harry Potter and the Deathly Hallows: Part 1
                                                             33.533
2010-11-19
                                                             28.734
                             How to Train Your Dragon
2010-03-26
                                                             28.515
                                            Iron Man 2
2010-05-07
                                                             28.005
                                             Toy Story
1995-11-22
                                             Inception
                                                             27.920
2010-07-16
. . .
26512
                                Laboratory Conditions
                                                              0.600
2018-10-13
26513
                                       EXHIBIT 84xxx
                                                              0.600
```

```
2018-05-01
                                         The Last One
                                                             0.600
26514
2018-10-01
                                         Trailer Made
26515
                                                             0.600
2018-06-22
                                           The Church
26516
                                                             0.600
2018-10-05
                                                title vote_average
vote_count
       Harry Potter and the Deathly Hallows: Part 1
                                                                 7.7
10788
1
                            How to Train Your Dragon
                                                                 7.7
7610
                                           Iron Man 2
                                                                 6.8
12368
                                            Toy Story
                                                                 7.9
3
10174
                                            Inception
                                                                 8.3
22186
. . .
26512
                               Laboratory Conditions
                                                                 0.0
                                                                 0.0
26513
                                      _EXHIBIT_84xxx_
1
26514
                                         The Last One
                                                                 0.0
26515
                                         Trailer Made
                                                                 0.0
26516
                                           The Church
                                                                 0.0
[26517 rows x 10 columns]
```

Renamed the column release_date to releaseDate.

```
df = df2.rename(columns = {'release date':'releaseDate'}, inplace =
True)
df
df2
       Unnamed: 0
                              genre ids
                                              id original_language \
0
                        [12, 14, 10751]
                 0
                                           12444
                                                                 en
                    [14, 12, 16, 10751]
1
                 1
                                           10191
                                                                 en
2
                          [12, 28, 878]
                 2
                                           10138
                                                                 en
3
                 3
                        [16, 35, 10751]
                                             862
                                                                 en
4
                          [28, 878, 12]
                                           27205
```

26512 26513 26514 26515 26516	26512 26513 26514 26515 26516	[27, 18] 488143 [18, 53] 485975 [14, 28, 12] 381231 [10751, 12, 28] 366854 [53, 27] 309885	en en en en en
_		movie	popularity
releas 0 11-19	•	and the Deathly Hallows: Part 1	33.533 2010-
1		How to Train Your Dragon	28.734 2010-
03-26 2		Iron Man 2	28.515 2010-
05-07 3		Toy Story	28.005 1995-
11-22 4		Inception	27.920 2010-
07 - 16 			
26512		Laboratory Conditions	0.600 2018-
10-13		EVIITETT 04vov	0 600 2010
26513 05-01		_EXHIBIT_84xxx_	0.600 2018-
26514		The Last One	0.600 2018-
10-01 26515		Trailer Made	0.600 2018-
06-22			
26516 10-05		The Church	0.600 2018-
10-03			
vote c	ount	title	vote_average
0 -		and the Deathly Hallows: Part 1	7.7
10788 1		How to Train Your Dragon	7.7
7610		•	
2 12368		Iron Man 2	6.8
3 10174		Toy Story	7.9
4		Inception	8.3
22186			
 26512		Laboratory Conditions	0.0
1		•	
26513		_EXHIBIT_84xxx_	0.0

1		
26514	The Last One	0.0
1		
26515	Trailer Made	0.0
1		_
26516	The Church	0.0
1		
[26517 rows x 10 columns]		

Loaded the third data frame.

```
df3= pd.read csv(r"C:\Users\pc\Videos\projects\Phase One Proect\data\
tn.movie budgets.csv")
df3
      id
         release date
                                                                  movie
       1
          Dec 18, 2009
                                                                 Avatar
          May 20, 2011
       2
1
                         Pirates of the Caribbean: On Stranger Tides
2
       3
           Jun 7, 2019
                                                          Dark Phoenix
3
       4
           May 1, 2015
                                               Avengers: Age of Ultron
4
       5
          Dec 15, 2017
                                    Star Wars Ep. VIII: The Last Jedi
          Dec 31, 2018
5777
      78
                                                                 Red 11
      79
                                                              Following
5778
           Apr 2, 1999
          Jul 13, 2005
5779
      80
                                        Return to the Land of Wonders
                                                  A Plague So Pleasant
5780
      81
          Sep 29, 2015
           Aug 5, 2005
5781
     82
                                                     My Date With Drew
     production_budget domestic_gross worldwide_gross
                                         $2,776,345,279
0
          $425,000,000
                          $760,507,625
1
          $410,600,000
                          $241,063,875
                                         $1,045,663,875
2
          $350,000,000
                            $42,762,350
                                            $149,762,350
3
          $330,600,000
                          $459,005,868
                                         $1,403,013,963
4
          $317,000,000
                          $620,181,382
                                         $1,316,721,747
. . .
5777
                 $7,000
                                                      $0
                                     $0
                                                $240,495
5778
                 $6,000
                                $48,482
5779
                                 $1,338
                                                  $1,338
                 $5,000
                 $1,400
5780
                                     $0
                                                      $0
5781
                 $1,100
                               $181,041
                                                $181,041
[5782 rows x 6 columns]
```

Data Cleaning

- Checked for missing or erroneous data points.
- Standardized and cleaned the dataset to ensure accuracy in subsequent analyses.A

In order for the data cleaning process to be effective. We merged the three datasets together.

```
df3= df3.rename(columns = {'worldwide gross':'
internationalBoxOffice',
'domestic gross':'domesticBoxoffice',
 'production_budget':'productionBudget'},
df3
      id release date
                                                                 movie \
          Dec 18, 2009
       1
                                                                Avatar
       2
          May 20, 2011
1
                         Pirates of the Caribbean: On Stranger Tides
2
       3
          Jun 7, 2019
                                                         Dark Phoenix
3
       4
          May 1, 2015
                                              Avengers: Age of Ultron
4
       5
         Dec 15, 2017
                                   Star Wars Ep. VIII: The Last Jedi
          Dec 31, 2018
5777
                                                                Red 11
      78
          Apr 2, 1999
                                                             Following
5778
     79
          Jul 13, 2005
5779
      80
                                        Return to the Land of Wonders
          Sep 29, 2015
                                                 A Plague So Pleasant
5780
      81
5781
     82
           Aug 5, 2005
                                                    My Date With Drew
     productionBudget domesticBoxoffice
                                           internationalBoxOffice
0
         $425,000,000
                            $760,507,625
                                                   $2,776,345,279
1
         $410,600,000
                            $241,063,875
                                                   $1,045,663,875
2
         $350,000,000
                             $42,762,350
                                                      $149,762,350
3
         $330,600,000
                            $459,005,868
                                                   $1,403,013,963
4
         $317,000,000
                            $620,181,382
                                                   $1,316,721,747
. . .
                                                               . . .
5777
               $7,000
                                       $0
                                                                $0
5778
               $6,000
                                 $48,482
                                                          $240,495
               $5,000
5779
                                  $1,338
                                                            $1,338
5780
               $1,400
                                       $0
                                                                $0
5781
               $1,100
                                $181,041
                                                          $181,041
[5782 rows x 6 columns]
df1.columns = df1.columns.str.strip()
df2.columns = df2.columns.str.strip()
df1['movie']
df2['movie']
df3['movie']
0
                                               Avatar
1
        Pirates of the Caribbean: On Stranger Tides
2
                                         Dark Phoenix
3
                             Avengers: Age of Ultron
```

4	Star Wars Ep. VIII: The Last Jedi
	• • •
5777	Red 11
5778	Following
5779	Return to the Land of Wonders
5780	A Plague So Pleasant
5781	My Date With Drew
Name: movie,	Length: 5782, dtype: object

Merged the datasets together.

```
df = pd.merge(df1, df2, on ='movie', how = 'right')
df = pd.merge(df, df3, on ='movie', how = 'left')
df
                                                 movie
production year \
       Harry Potter and the Deathly Hallows: Part 1
                                                                     NaN
                            How to Train Your Dragon
1
                                                                  2010.0
2
                                           Iron Man 2
                                                                  2010.0
                                            Toy Story
                                                                     NaN
                                            Toy Story
                                                                     NaN
26621
                                Laboratory Conditions
                                                                     NaN
26622
                                      _EXHIBIT_84xxx_
                                                                     NaN
26623
                                         The Last One
                                                                     NaN
26624
                                         Trailer Made
                                                                     NaN
26625
                                           The Church
                                                                     NaN
                                          domesticBoxoffice_x \
        movie odid
                     productionBudget x
0
                NaN
                                                           NaN
1
       116630100.0
                             165000000.0
                                                   217581232.0
2
                             170000000.0
       117940100.0
                                                   312433331.0
3
                NaN
                                     NaN
                                                           NaN
4
                NaN
                                     NaN
                                                           NaN
                . . .
                                                            . . .
26621
                NaN
                                     NaN
                                                           NaN
26622
                NaN
                                     NaN
                                                           NaN
26623
                NaN
                                     NaN
                                                           NaN
```

26624	NaN			aN		NaN		
26625	NaN		Na	aN		NaN		
0	internationa	lBoxOffice NaN	rating NaN	crea	tive_type \ NaN			
1 2 3 4		77289760.0 98723058.0 NaN NaN	PG PG-13 NaN NaN	S	Fantasy uper Hero NaN NaN			
 26621		NaN	 NaN		 NaN			
26622 26623 26624		NaN NaN NaN	NaN NaN NaN		NaN NaN NaN			
26625		NaN	NaN		NaN			
_			SOL	ırce	production_	method		
popula 0	rity \			NaN		NaN		
33.533 1	Based on Fic	tion Book/9	Short Si	torv	Digital Ani	mation		
28.734 2		n Comic/Gra		•	_	Action		
28.515	baseu oi	T COMITC/ GT &	тритс м		LIVE			
3 28.005				NaN		NaN		
4 28.005				NaN		NaN		
26621				NaN		NaN		
0.600 26622				NaN		NaN		
0.600 26623				NaN		NaN		
0.600 26624				NaN		NaN		
0.600 26625				NaN		NaN		
0.600				NUN		IVAIV		
0	releaseDate	Harris Date			Daath3		itle	\
0	2010-11-19 2010-03-26	Harry Poti	er and		Deathly Hall ow to Train	Your Dr	agon	
2 3 4	2010-05-07 1995-11-22					Iron Ma Toy S		
4	1995-11-22					Toy S	_	
26621 26622	2018-10-13 2018-05-01				Laboratory _EXH	Condit IBIT_84	ions	

26623 26624 26625	2018-10-01 2018-06-22 2018-10-05					The Last One Trailer Made The Church
	vote_average vote	_count	id_y	release_	date	productionBudget_y
0	7.7	10788	NaN		NaN	NaN
1	7.7	7610	30.0	Mar 26,	2010	\$165,000,000
2	6.8	12368	15.0	May 7,	2010	\$170,000,000
3	7.9	10174	37.0	Nov 22,	1995	\$30,000,000
4	7.9	10174	37.0	Nov 22,	1995	\$30,000,000
26621	0.0	1	NaN		NaN	NaN
26622	0.0	1	NaN		NaN	NaN
26623	0.0	1	NaN		NaN	NaN
26624	0.0	1	NaN		NaN	NaN
26625	0.0	1	NaN		NaN	NaN
0 1 2 3 4 26621 26622 26623	domesticBoxoffice_ Na \$217,581,23 \$312,433,33 \$191,796,23 \$191,796,23 Na Na	N 2 1 3 3 N N	rnatio	\$494,870, \$621,156, \$364,545, \$364,545,	NaN 992 389 516 516 NaN NaN	
26624 26625	Na Na				NaN NaN	
[26626	rows x 27 columns]				

Dropped the duplicates from column movie of the new dataset.

```
df = df.drop_duplicates(subset = ['movie'], keep = 'first')
df = df.drop(columns = ['source'], axis = 1)
df
```

produc	tion_year \				movie		
0	Harry Potte	r and the De	eathly H	allows: F	Part 1		NaN
1		Hov	v to Tra	in Your [ragon	2	2010.0
2				Iron	Man 2	2	2010.0
3				Toy	Story		NaN
5				Ince	eption	2	2010.0
26621			Laborat	ory Condi	itions		NaN
26622			_	EXHIBIT_8	34xxx_		NaN
26623				The Las	st One		NaN
26624				Traile	^ Made		NaN
26625				The (Church		NaN
0 1 2 3 5 26621 26622 26623 26624 26625	movie_odid NaN 116630100.0 117940100.0 NaN 105240100.0 NaN NaN NaN NaN NaN NaN tion method	170 160	Na 50000000 . 00000000 . Na 00000000 . Na Na Na Na	N 0 0 N 0 N N N N	217581 312433 292576	NaN 232.0 331.0 NaN	
0	tion_method	\ NaN	NaN		NaN		
NaN 1 Animat		277289760.0	PG		Fantasy	Digi	ital
2 Action		308723058.0	PG-13	Sup	er Hero		Live
3		NaN	NaN		NaN		
NaN 5	!	539825887.0	PG-13	Science	Fiction	Animatio	on/Live
Action							

```
NaN
                                    NaN
                                                      NaN
26621
NaN
                            NaN
                                                      NaN
26622
                                    NaN
NaN
26623
                            NaN
                                    NaN
                                                      NaN
NaN
26624
                            NaN
                                    NaN
                                                      NaN
NaN
26625
                            NaN
                                    NaN
                                                      NaN
NaN
                    genre
                                  popularity
                                               releaseDate
0
                                      33.533
                                                2010-11-19
                      NaN
1
                Adventure
                                      28.734
                            . . .
                                                2010-03-26
2
                   Action
                                      28.515
                                                2010-05-07
3
                                      28,005
                                                1995-11-22
                      NaN
                            . . .
       Thriller/Suspense
5
                                      27.920
                                                2010-07-16
                            . . .
26621
                                       0.600
                                                2018-10-13
                       NaN
26622
                       NaN
                                       0.600
                                                2018-05-01
                                       0.600
                                                2018-10-01
26623
                       NaN
26624
                                       0.600
                                                2018-06-22
                       NaN
                                       0.600
                                                2018-10-05
26625
                      NaN
                                                  title vote average
vote count \
       Harry Potter and the Deathly Hallows: Part 1
                                                                  7.7
10788
                             How to Train Your Dragon
                                                                  7.7
7610
                                            Iron Man 2
                                                                  6.8
12368
                                                                  7.9
                                             Toy Story
10174
5
                                              Inception
                                                                  8.3
22186
. . .
26621
                                Laboratory Conditions
                                                                  0.0
26622
                                       EXHIBIT 84xxx
                                                                  0.0
1
26623
                                          The Last One
                                                                  0.0
26624
                                          Trailer Made
                                                                  0.0
26625
                                            The Church
                                                                  0.0
1
```

```
id y
              release_date productionBudget_y domesticBoxoffice_y
        NaN
0
                       NaN
                                                                  NaN
                                            NaN
1
       30.0
             Mar 26, 2010
                                  $165,000,000
                                                        $217,581,232
2
       15.0
              May 7, 2010
                                  $170,000,000
                                                        $312,433,331
3
       37.0
             Nov 22, 1995
                                   $30,000,000
                                                        $191,796,233
5
       38.0
             Jul 16, 2010
                                  $160,000,000
                                                        $292,576,195
        . . .
26621
        NaN
                       NaN
                                                                 NaN
                                            NaN
26622
        NaN
                       NaN
                                            NaN
                                                                  NaN
26623
        NaN
                       NaN
                                            NaN
                                                                  NaN
26624
        NaN
                       NaN
                                            NaN
                                                                  NaN
26625
        NaN
                       NaN
                                            NaN
                                                                  NaN
        internationalBoxOffice
0
                             NaN
1
                   $494,870,992
2
                   $621,156,389
3
                   $364,545,516
5
                   $835,524,642
26621
                             NaN
26622
                             NaN
26623
                             NaN
26624
                             NaN
26625
                             NaN
[24835 rows x 26 columns]
df = df.drop duplicates(subset = ['movie'], keep = 'first')
df.isnull().sum()
df = df.dropna()
df
                                                       movie
production year \
                                  How to Train Your Dragon
2010.0
                                                 Iron Man 2
2010.0
                                                  Inception
2010.0
       Percy Jackson & the Olympians: The Lightning T...
2010.0
                                                      Avatar
2009.0
. . .
. . .
24538
                                                       Gotti
```

2016.0				
24575			Proud	Mary
2017.0 24597			Rene	gades
2016.0				
25388 2016.0		Bilal:	A New Breed of	Hero
26207			The	e Box
2009.0				
movie_ 1 1166301	- ·	nBudget_ 5000000.		office_x \ 581232.0
2 1179401		0000000.		433331.0
5 1052401		0000000.		576195.0
6 1075501 7 1220401		5000000. 5000000.		768303.0 507625.0
/ 1220401	42			
24538 2468201		0000000.		286367.0
24575 2816301		0000000.		368638.0
24597 2276101 25388 2651901		7500000. 00000000.		0.0 490973.0
26207 1154701		5000000.		951977.0
intorna	tionalBoxOffice	rating	croativ	ve type \
1	2.772898e+08	_		Fantasy
2	3.087231e+08		Supe	er Hero
5	5.398259e+08 1.342826e+08		Science I	Fiction Fantasy
6 7	2.015838e+09		Science I	
24538 24575	1.802733e+06 8.409010e+05			ization
24597	1.521672e+06		Contemporary Contemporary	
25388	1.576260e+05			ization
26207	1.128992e+07	PG-13		Fantasy
pro releaseDate \	duction_method		genre	. popularity
	ital Animation		Adventure	. 28.734
2	Live Action		Action	. 28.515
2010-05-07 5 Animati 2010-07-16	on/Live Action	Thrille	r/Suspense	. 27.920
6	Live Action		Adventure	. 26.691
2010-02-11	on/livo Astis		1 at - a -	26 526
7 Animati 2009-12-18	on/Live Action		Action	. 26.526

24538 Li 2018-06-15	ve Action		Drama		10.034	
24575 Li	ve Action		Action		9.371	
	ve Action		Action		9.022	
2018-12-21 25388 Digital	Animation	А	dventure		2.707	
2018-02-02 26207 Li	ve Action	Thriller/	Suspense		0.840	
2018-03-04		,				
V				title	vote_ave	rage
1		How to	Train You	r Dragon		7.7
2			Ir	on Man 2		6.8
5			I	nception		8.3
6 Percy Jackso	n & the Ol	ympians: T	he Lightn	ing T		6.1
7				Avatar		7.4
24538				Gotti		5.2
24575			Pr	oud Mary		5.5
24597				enegades		5.8
25388		Bilal: A		_		6.8
		bitat. A	New breed	The Box		
26207				THE DOX		8.0
vote_count		.ease_date	productio	nBudget_y		
<pre>domesticBoxoffice_y 1 7610</pre>		26, 2010	\$16	5,000,000		
\$217,581,232 2 12368	15.0 Ma	y 7, 2010	\$17	0,000,000		
\$312,433,331 5 22186		16, 2010		0,000,000		
\$292,576,195		12, 2010	·	5,000,000		
6 4229 \$88,768,303		-	·			
7 18676 \$760,507,625	1.0 Dec	18, 2009	\$42	5,000,000		

```
24538
               231
                     64.0 Jun 15, 2018
                                                 $10,000,000
$4,286,367
24575
               259
                     50.0
                           Jan 12, 2018
                                                 $30,000,000
$20,868,638
                           Jan 22, 2019
24597
               156
                     20.0
                                                 $77,500,000
$0
                             Feb 2, 2018
25388
                54
                    100.0
                                                 $30,000,000
$490,973
                            Nov 6, 2009
26207
                 1
                     66.0
                                                 $25,000,000
$15,051,977
        internationalBoxOffice
1
                   $494,870,992
2
                   $621,156,389
5
                   $835,524,642
6
                   $223,050,874
7
                 $2,776,345,279
. . .
                     $6,089,100
24538
24575
                    $21,709,539
24597
                     $1,521,672
25388
                       $648,599
26207
                    $34,356,760
[1110 rows \times 26 columns]
```

Dropped the domesticBoxOffice columns.

```
df['domesticBoxoffice y'] == df['domesticBoxoffice x']
df = df.drop(columns = ['domesticBoxoffice y'], axis = 1)
df
                                                     movie
production year \
                                 How to Train Your Dragon
2010.0
                                                Iron Man 2
2010.0
                                                 Inception
2010.0
       Percy Jackson & the Olympians: The Lightning T...
2010.0
                                                    Avatar
2009.0
. . .
24538
                                                     Gotti
2016.0
                                                Proud Mary
24575
```

```
2017.0
24597
                                                  Renegades
2016.0
                                Bilal: A New Breed of Hero
25388
2016.0
26207
                                                    The Box
2009.0
        movie odid
                     productionBudget x
                                          domesticBoxoffice x
1
       116630100.0
                             165000000.0
                                                   217581232.0
2
       117940100.0
                             170000000.0
                                                   312433331.0
5
       105240100.0
                             160000000.0
                                                   292576195.0
6
       107550100.0
                             95000000.0
                                                    88768303.0
7
                             425000000.0
       122040100.0
                                                   760507625.0
24538
       246820100.0
                              10000000.0
                                                     4286367.0
24575
       281630100.0
                                                    20868638.0
                              30000000.0
24597
       227610100.0
                              77500000.0
                                                           0.0
25388
       265190100.0
                              3000000.0
                                                      490973.0
26207
       115470100.0
                              25000000.0
                                                    15051977.0
       internationalBoxOffice rating
                                                creative type \
                  2.772898e+08
1
                                    PG
                                                      Fantasy
2
                  3.087231e+08
                                 PG-13
                                                   Super Hero
5
                                 PG-13
                  5.398259e+08
                                              Science Fiction
6
                  1.342826e+08
                                    PG
                                                      Fantasy
7
                  2.015838e+09
                                 PG-13
                                              Science Fiction
                  1.802733e+06
                                                Dramatization
                                     R
24538
                  8.409010e+05
                                        Contemporary Fiction
24575
                                     R
24597
                  1.521672e+06
                                 PG-13
                                        Contemporary Fiction
                  1.576260e+05
                                 PG-13
                                                Dramatization
25388
26207
                  1.128992e+07
                                 PG-13
                                                      Fantasy
           production method
                                             genre ...
original language \
           Digital Animation
                                        Adventure
1
en
2
                  Live Action
                                           Action
en
5
       Animation/Live Action Thriller/Suspense
en
                  Live Action
6
                                        Adventure
en
       Animation/Live Action
7
                                           Action
en
24538
                  Live Action
                                             Drama
en
```

```
24575
                  Live Action
                                            Action ...
en
24597
                  Live Action
                                            Action ...
fr
25388
           Digital Animation
                                        Adventure ...
en
26207
                  Live Action Thriller/Suspense
en
       popularity
                    releaseDate \
1
           28.734
                     2010-03-26
2
           28.515
                     2010-05-07
5
           27.920
                     2010-07-16
6
           26.691
                     2010-02-11
7
           26.526
                     2009 - 12 - 18
. . .
               . . .
           10.034
                     2018-06-15
24538
24575
            9.371
                     2018-01-12
24597
            9.022
                     2018-12-21
25388
            2.707
                     2018-02-02
            0.840
                     2018-03-04
26207
                                                              vote average
                                                      title
/
1
                                  How to Train Your Dragon
                                                                        7.7
2
                                                                        6.8
                                                 Iron Man 2
5
                                                                        8.3
                                                  Inception
       Percy Jackson & the Olympians: The Lightning T...
                                                                        6.1
                                                                        7.4
                                                     Avatar
24538
                                                                        5.2
                                                      Gotti
24575
                                                 Proud Mary
                                                                        5.5
24597
                                                                        5.8
                                                  Renegades
25388
                                Bilal: A New Breed of Hero
                                                                        6.8
26207
                                                    The Box
                                                                        8.0
                           release_date productionBudget_y \
      vote_count
                    id_y
1
            7610
                    30.0
                          Mar 26, 2010
                                               $165,000,000
2
           12368
                    15.0
                           May 7, 2010
                                               $170,000,000
5
                          Jul 16, 2010
           22186
                    38.0
                                               $160,000,000
```

```
6
                    17.0
                          Feb 12, 2010
            4229
                                                $95,000,000
7
            18676
                     1.0
                          Dec 18, 2009
                                               $425,000,000
                     . . .
              . . .
                    64.0
                          Jun 15, 2018
              231
                                                $10,000,000
24538
24575
              259
                    50.0
                          Jan 12, 2018
                                                $30,000,000
                          Jan 22, 2019
24597
              156
                    20.0
                                                $77,500,000
                            Feb 2, 2018
25388
                   100.0
               54
                                                $30,000,000
26207
                1
                    66.0
                            Nov 6, 2009
                                                $25,000,000
        internationalBoxOffice
1
                   $494,870,992
2
                   $621,156,389
5
                   $835,524,642
6
                   $223,050,874
7
                 $2,776,345,279
. . .
                     $6,089,100
24538
24575
                    $21,709,539
24597
                     $1,521,672
                       $648,599
25388
26207
                    $34,356,760
[1110 rows \times 25 columns]
df = df.drop(columns = ['productionBudget y'], axis = 1)
df
                                                       movie
production_year \
                                  How to Train Your Dragon
2010.0
                                                 Iron Man 2
2010.0
                                                  Inception
2010.0
       Percy Jackson & the Olympians: The Lightning T...
2010.0
7
                                                      Avatar
2009.0
                                                         . . .
24538
                                                       Gotti
2016.0
24575
                                                 Proud Mary
2017.0
24597
                                                  Renegades
2016.0
                                Bilal: A New Breed of Hero
25388
2016.0
```

```
26207
                                                     The Box
2009.0
                                           domesticBoxoffice_x
        movie odid
                     productionBudget x
1
       116630100.0
                             165000000.0
                                                    217581232.0
2
       117940100.0
                             170000000.0
                                                    312433331.0
5
                             160000000.0
                                                    292576195.0
       105240100.0
6
       107550100.0
                              95000000.0
                                                     88768303.0
7
       122040100.0
                             425000000.0
                                                    760507625.0
24538
       246820100.0
                              10000000.0
                                                      4286367.0
24575
       281630100.0
                                                     20868638.0
                              30000000.0
24597
       227610100.0
                              77500000.0
                                                            0.0
                              30000000.0
25388
       265190100.0
                                                       490973.0
26207
       115470100.0
                              25000000.0
                                                     15051977.0
       internationalBoxOffice rating
                                                creative type
1
                  2.772898e+08
                                     PG
                                                       Fantasy
2
                  3.087231e+08
                                 PG-13
                                                    Super Hero
5
                                 PG-13
                  5.398259e+08
                                              Science Fiction
6
                  1.342826e+08
                                    PG
                                                       Fantasv
7
                                 PG-13
                  2.015838e+09
                                              Science Fiction
                  1.802733e+06
                                                Dramatization
24538
                                      R
24575
                  8.409010e+05
                                      R
                                         Contemporary Fiction
24597
                  1.521672e+06
                                 PG-13
                                         Contemporary Fiction
                                 PG-13
25388
                  1.576260e+05
                                                Dramatization
                  1.128992e+07
                                 PG-13
26207
                                                       Fantasy
            production method
                                             genre
                                                            id x
1
            Digital Animation
                                                           10191
                                         Adventure
2
                  Live Action
                                            Action
                                                           10138
5
                                Thriller/Suspense
       Animation/Live Action
                                                           27205
6
                  Live Action
                                         Adventure
                                                           32657
7
       Animation/Live Action
                                                           19995
                                            Action
                  Live Action
                                                          339103
24538
                                             Drama
                  Live Action
24575
                                            Action
                                                          442064
24597
                  Live Action
                                            Action
                                                          335788
25388
            Digital Animation
                                                          332718
                                         Adventure
                                Thriller/Suspense
26207
                  Live Action
                                                          509314
       original language
                            popularity releaseDate
1
                                28.734
                                         2010-03-26
                        en
2
                       en
                                28.515
                                         2010-05-07
5
                                27.920
                                         2010-07-16
                       en
6
                                26.691
                                         2010-02-11
                       en
7
                                26.526
                                         2009 - 12 - 18
                        en
                                10.034
                                         2018-06-15
24538
                        en
```

```
24575
                                        2018-01-12
                                 9.371
                       en
24597
                                 9.022
                       fr
                                        2018-12-21
25388
                                 2.707
                                        2018-02-02
                       en
26207
                                 0.840
                                        2018-03-04
                       en
                                                      title vote average
/
1
                                  How to Train Your Dragon
                                                                       7.7
2
                                                 Iron Man 2
                                                                      6.8
5
                                                  Inception
                                                                      8.3
       Percy Jackson & the Olympians: The Lightning T...
                                                                       6.1
                                                                       7.4
                                                     Avatar
                                                                       . . .
24538
                                                      Gotti
                                                                       5.2
24575
                                                 Proud Mary
                                                                       5.5
24597
                                                  Renegades
                                                                       5.8
                                Bilal: A New Breed of Hero
25388
                                                                       6.8
26207
                                                    The Box
                                                                       8.0
                            release date
                                            internationalBoxOffice
       vote count
                     id y
1
             7610
                     30.0
                           Mar 26, 2010
                                                       $494,870,992
2
            12368
                     15.0
                            May 7, 2010
                                                      $621,156,389
                           Jul 16, 2010
5
                     38.0
            22186
                                                      $835,524,642
6
                           Feb 12, 2010
              4229
                     17.0
                                                      $223,050,874
7
             18676
                      1.0
                           Dec 18, 2009
                                                    $2,776,345,279
                                                         $6,089,100
               231
                     64.0
                           Jun 15, 2018
24538
24575
               259
                     50.0
                           Jan 12, 2018
                                                        $21,709,539
                           Jan 22, 2019
24597
               156
                     20.0
                                                         $1,521,672
                             Feb 2, 2018
25388
                54
                    100.0
                                                           $648,599
26207
                 1
                     66.0
                            Nov 6, 2009
                                                        $34,356,760
[1110 rows x 24 columns]
df = df.drop(columns = ['production year'], axis = 1)
                                                               movie odid
                                                      movie
/
```

1 How to Train Your Dragon 116630100.0 2 Iron Man 2 117940100.0 5 Inception 105240100.0 6 Percy Jackson & the Olympians: The Lightning T 107550100.0 7 Avatar 122040100.0	sequel	creative_type	e production_method	genre
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Proud Mary 281630100.0 24575 Proud Mary 281630100.0 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 ProductionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG 1 165000000.0 312433331.0 3.087231e+08 PG-13 5 160000000.0 292576195.0 5.398259e+08 PG-13 6 95000000.0 88768303.0 1.342826e+08 PG 7 425000000.0 88768303.0 1.342826e+08 PG 7 425000000.0 760507625.0 2.015838e+09 PG-13	26207	25000000.0	15051977.0	1.128992e+07
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Second Mary 281630100.0 Proud Mary 281630100.0 Renegades 227610100.0 Estating 1 165000000.0 Bilal: A New Breed of Hero 265190100.0 ProductionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG 2 170000000.0 312433331.0 3.087231e+08 PG 3 16000000.0 292576195.0 5.398259e+08 PG 425000000.0 88768303.0 1.342826e+08 PG 425000000.0 760507625.0 2.015838e+09 PG-13	25388	30000000.0	490973.0	1.576260e+05
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Proud Mary 281630100.0 24575 Proud Mary 281630100.0 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 Earling 1 165000000.0 217581232.0 2.772898e+08 PG 2 170000000.0 312433331.0 3.087231e+08 PG-13 16000000.0 292576195.0 5.398259e+08 PG-13 5 16000000.0 88768303.0 1.342826e+08 PG 7 425000000.0 760507625.0 2.015838e+09 PG-13 24538 10000000.0 4286367.0 1.802733e+06 R 24575 30000000.0 20868638.0 8.409010e+05 R		77500000.0	0.0	1.521672e+06
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Proud Mary 281630100.0 24575 Proud Mary 281630100.0 24597 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 25388 Bilal: A New Breed of Hero 265190100.0 The Box 115470100.0 productionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG 1 170000000.0 312433331.0 3.087231e+08 PG-13 5 160000000.0 292576195.0 5.398259e+08 PG-13 6 95000000.0 88768303.0 1.342826e+08 PG 7 425000000.0 88768303.0 1.342826e+08 PG 96 425000000.0 760507625.0 2.015838e+09 PG-13	R			
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Proud Mary 281630100.0 Renegades 227610100.0 Establish A New Breed of Hero 265190100.0 Bilal: A New Breed of Hero 265190100.0 The Box 115470100.0 productionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG I 170000000.0 312433331.0 3.087231e+08 PG-13 I 16000000.0 292576195.0 5.398259e+08 PG-13 I 10000000.0 88768303.0 1.342826e+08 PG-13 I 10000000.0 760507625.0 2.015838e+09 PG-13 I 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	R			
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Proud Mary 281630100.0 Renegades 227610100.0 Bilal: A New Breed of Hero 265190100.0 Bilal: A New Breed of Hero 265190100.0 The Box 115470100.0 Proud Mary 281630100.0 24597 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 The Box 115470100.0 ProductionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG 2 170000000.0 312433331.0 3.087231e+08 PG-13 5 160000000.0 292576195.0 5.398259e+08 PG-13 6 95000000.0 88768303.0 1.342826e+08 PG 7 425000000.0 760507625.0 2.015838e+09 PG-13				
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Compared to the State of Hero 246820100.0 Renegades 227610100.0 24597 Renegades 227610100.0 24597 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 26207 The Box 115470100.0 ProductionBudget_x domesticBoxoffice_x internationalBoxOffice internationalBoxOffice internationalBoxOffice 2 772898e+08 PG 170000000.0 312433331.0 3.087231e+08 PG-13 5 160000000.0 292576195.0 5.398259e+08 PG-13 6 95000000.0 88768303.0 1.342826e+08 PG 7 425000000.0 760507625.0 2.015838e+09				
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Control Many 2 246820100.0 Proud Mary 281630100.0 Renegades 227610100.0 Enable A New Breed of Hero 265190100.0 Proud Mary 281630100.0 The Box 115470100.0 Proud Mary 281630100.0 24597 Renegades 227610100.0 Proud Mary 281630100.0 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 The Box 115470100.0 ProductionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG The Box 115470100.0 PG-13 The Box 115470100.0 Renegades 227610100.0 Renegades 227610100.0 The Box 115470100.0 Renegades 227610100.0 The Box 115470100.0 Renegades 227610100.0 Renegades 227610100.	7	425000000.0	760507625.0	2.015838e+09
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Proud Mary 281630100.0 24575 Proud Mary 281630100.0 24597 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 25388 Bilal: A New Breed of Hero 265190100.0 productionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG 2 170000000.0 312433331.0 3.087231e+08 PG-13 5 160000000.0 292576195.0 5.398259e+08		95000000.0	88768303.0	1.342826e+08
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 122040100.0 Proud Mary 281630100.0 24575 Proud Mary 281630100.0 24597 Renegades 227610100.0 25388 Bilal: A New Breed of Hero 265190100.0 26207 The Box 115470100.0 productionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08 PG 2 170000000.0 312433331.0 3.087231e+08 PG-13		160000000.0	2925/6195.0	5.398259e+08
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Gotti 246820100.0 Proud Mary 281630100.0 Renegades 227610100.0 Signature of the control of the con	PG-13			
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Compared to the State of Hero 265190100.0 Bilal: A New Breed of Hero 265190100.0 The Box 115470100.0 productionBudget_x domesticBoxoffice_x internationalBoxOffice rating 1 165000000.0 217581232.0 2.772898e+08		17000000 0	312433331 0	3 087231e+08
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Control 246820100.0 Proud Mary 281630100.0 Renegades 227610100.0 Ending T 107550100.0 The Box 115470100.0 ProductionBudget_x domesticBoxoffice_x internationalBoxOffice	1	,	217581232.0	2.772898e+08
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Gotti 246820100.0 Proud Mary 281630100.0 Renegades 227610100.0 Bilal: A New Breed of Hero 265190100.0	rating	· .	domesticBoxoffice_x internati	onalBoxOffice
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Gotti 246820100.0 Proud Mary 281630100.0 Renegades 227610100.0 Bilal: A New Breed of Hero 265190100.0	26207		The Box	115470100.0
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Gotti 246820100.0 Proud Mary 281630100.0 Renegades 227610100.0				
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Avatar 5246820100.0 24575 Proud Mary 281630100.0			·	
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0 Gotti 246820100.0			·	227610100 0
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0 Avatar 122040100.0	24575		Proud Mary	281630100.0
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0	24538		Gotti	246820100.0
Iron Man 2 117940100.0 Inception 105240100.0 Percy Jackson & the Olympians: The Lightning T 107550100.0				
2 Iron Man 2 117940100.0 5 Inception 105240100.0	7		Avatar	122040100.0
2 Iron Man 2 117940100.0	6	Percy Jackson & the	Olympians: The Lightning T	107550100.0
	5		Inception	105240100.0
	2		Iron Man 2	117940100.0
			<u> </u>	116630100.0

1	F	antasy	Diait	al An	imation		Adventure
0.0		-	Digit				
2 1.0	Supe	er Hero		Live	Action		Action
5	Science F	iction	Animation	/Live	Action	Thrille	r/Suspense
0.0 6	F	antasy		Live	Action		Adventure
0.0 7	Science F	iction	Animation	/Live	Action		Action
0.0							
24538 0.0	Dramati	ization		Live	Action		Drama
24575 0.0	Contemporary F	iction		Live	Action		Action
24597 0.0	Contemporary F	iction		Live	Action		Action
25388 0.0	Dramati	ization	Digit	al An:	imation		Adventure
26207 0.0	F	antasy		Live	Action	Thrille	r/Suspense
1 2 5 6 7	id_x c 10191 10138 27205 32657 19995	original_	_language en en en en en	28 28 27 20	arity 3.734 3.515 7.920 5.691 5.526	releaseDat 2010-03-2 2010-05-0 2010-07-2 2010-02-2	26 97 16 11
24538 24575 24597 25388 26207	339103 442064 335788 332718 509314		en en fr en en	(0.034 0.371 0.022 2.707 0.840	2018-06-1 2018-01-1 2018-12-2 2018-02-0 2018-03-0	12 21 92
						title vo	ote_average
1			How to	Trai	n Your [ragon	7.7
2					Iron	Man 2	6.8
5					Ince	eption	8.3
6	Percy Jackson	& the 01	ympians:	The L	ightning	ј Т	6.1
7					I	Avatar	7.4

24538	Gotti	5.2
24575	Proud Mary	5.5
24597	Renegades	5.8
25388	Bilal: A New Breed of Hero	6.8
26207	The Box	8.0
	release_date internationalBoxOffice	
1 7610 30.0 2 12368 15.0	Mar 26, 2010 \$494,870,992 May 7, 2010 \$621,156,389	
	Jul 16, 2010\$835,524,642Feb 12, 2010\$223,050,874Dec 18, 2009\$2,776,345,279	
24538 231 64.0 24575 259 50.0 24597 156 20.0 25388 54 100.0 26207 1 66.0	Jun 15, 2018 \$6,089,100 Jan 12, 2018 \$21,709,539 Jan 22, 2019 \$1,521,672 Feb 2, 2018 \$648,599 Nov 6, 2009 \$34,356,760	
[1110 rows x 23 columns]		

Dropped unnecessary columns for the analysis.

```
df = df.drop(columns = ['movie_odid', 'creative_type', 'id_x',
'original_language', 'title', 'id_y', 'release_date'], axis = 1)
                                                                movie
productionBudget_x \
                                        How to Train Your Dragon
165000000.0
                                                         Iron Man 2
170000000.0
                                                           Inception
160000000.0
        Percy Jackson & the Olympians: The Lightning T...
95000000.0
                                                              Avatar
425000000.0
24538
                                                                Gotti
10000000.0
24575
                                                         Proud Mary
```

200000	00.0						
30000000.0 24597 Renegades							
775000	00.0						
25388 Bilal: A New Breed of Hero 30000000.0							
26207		Th	ne Box				
250000	00.0						
	domesticBoxoffice_x	internationalBoxOffice	rating \				
1	217581232.0	2.772898e+08	PG DC 13				
2 5	312433331.0 292576195.0	3.087231e+08 5.398259e+08	PG-13 PG-13				
6	88768303.0	1.342826e+08	PG				
7	760507625.0	2.015838e+09	PG-13				
24538	4286367.0	1.802733e+06	R				
24575	20868638.0	8.409010e+05	R				
24597 25388	0.0 490973.0	1.521672e+06 1.576260e+05	PG-13 PG-13				
26207	15051977.0	1.128992e+07	PG-13				
	production method	genre se	equel running	n time			
\	· –	_	·	_			
1	Digital Animation	Adventure	0.0	91.0			
2	Live Action	Action	1.0	125.0			
5	Animation/Live Action	Thriller/Suspense	0.0	147.0			
6	Live Action	Adventure	0.0	119.0			
7	Animation/Live Action	Action	0.0	162.0			
24538	Live Action	Drama	0.0	110.0			
24575	Live Action	Action	0.0	88.0			
24597	Live Action	Action	0.0	105.0			
25388	Digital Animation	Adventure	0.0	103.0			
26207	Live Action	Thriller/Suspense	0.0	115.0			
			_				
vote a	Unnamed: 0 verage \	<pre>genre_ids popularity</pre>	releaseDate				
1		16, 10751] 28.734	2010-03-26				
7.7 2	2 [12	2, 28, 878] 28.515	2010-05-07				
2	2 [12	., 20, 0/0] 20.313	2010-03-07				

```
6.8
                           [28, 878, 12]
                                               27.920 2010-07-16
5
8.3
                        [12, 14, 10751]
                                               26.691
                                                       2010-02-11
6
6.1
7
                      [28, 12, 14, 878]
                                               26.526
                                                       2009 - 12 - 18
7.4
. . .
                                               10.034
24538
            24168
                       [80, 18, 36, 53]
                                                       2018-06-15
5.2
24575
            24212
                            [53, 28, 80]
                                                9.371
                                                       2018-01-12
5.5
                                                9.022
24597
            24239
                                [53, 28]
                                                       2018-12-21
5.8
25388
                            [28, 12, 16]
                                                2.707 2018-02-02
            25148
6.8
26207
            26040
                                       []
                                                0.840 2018-03-04
8.0
       vote_count
                    internationalBoxOffice
1
             7610
                               $494,870,992
2
             12368
                               $621,156,389
5
            22186
                               $835,524,642
6
             4229
                               $223,050,874
7
                             $2,776,345,279
             18676
               . . .
24538
               231
                                 $6,089,100
               259
                                $21,709,539
24575
                                 $1,521,672
24597
               156
                54
                                   $648,599
25388
26207
                 1
                                $34,356,760
[1110 rows \times 16 columns]
df = df.drop(columns = ['Unnamed: 0'], axis = 1)
df
df.isnull().sum()
df = df.dropna()
df
                                                       movie
productionBudget_x \
                                  How to Train Your Dragon
165000000.0
                                                 Iron Man 2
170000000.0
                                                  Inception
160000000.0
```

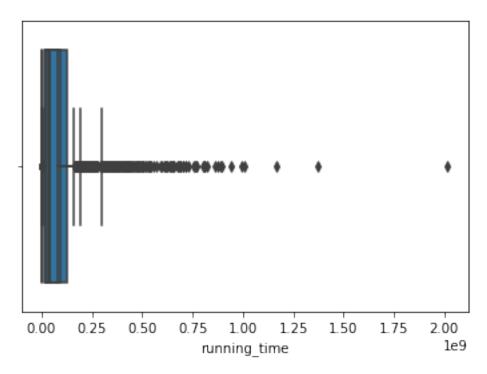
6 950000		the Ol	ympians: The Lightni	ng T	
7 425000				Avatar	
24538	000			Gotti	
100000 24575			Pro	oud Mary	
300000 24597			Re	enegades	
775000 25388			Bilal: A New Breed	of Hero	
300000 26207	000.0			The Box	
250000	000.0				
1 2	domesticBoxoffs 2175812 3124333	232.0	nternationalBoxOffic 2.772898e+6 3.087231e+6)8 PG	\
5 6 7	2925763 887683 7605076	303.0	5.398259e+6 1.342826e+6 2.015838e+6)8 PG-13)8 PG	
24538 24575 24597 25388	208686	367.0 38.0 0.0 973.0	1.802733e+6 8.409010e+6 1.521672e+6 1.576260e+6)5 R)6 PG-13	
26207	150519	977.0	1.128992e+6	7 PG-13	
\	production_	_method	genre	sequel	running_time
ì	Digital Ani	imation	Adventure	0.0	91.0
2	Live	Action	Action	1.0	125.0
5	Animation/Live	Action	Thriller/Suspense	0.0	147.0
6	Live	Action	Adventure	0.0	119.0
7	Animation/Live	Action	Action	0.0	162.0
24538	Live	Action	Drama	0.0	110.0
24575	Live	Action	Action	0.0	88.0
24597	Live	Action	Action	0.0	105.0

25388	Digital Animation	Д	dventure	0.0	103.0
26207	Live Action	Thriller/	'Suspense	0.0	115.0
vote c		popularity	releaseDate	vote_averag	е
1 7610	[14, 12, 16, 10751]	28.734	2010-03-26	7.	7
2	[12, 28, 878]	28.515	2010-05-07	6.	8
12368 5	[28, 878, 12]	27.920	2010-07-16	8.	3
22186 6	[12, 14, 10751]	26.691	2010-02-11	6.	1
4229 7	[28, 12, 14, 878]	26.526	2009-12-18	7.	4
18676					
	[80, 18, 36, 53]	10.034	2018-06-15	5.	
24538 231					
24575 259	[53, 28, 80]	9.371	2018-01-12	5.	
24597 156	[53, 28]	9.022	2018-12-21	5.	8
25388 54	[28, 12, 16]	2.707	2018-02-02	6.	8
26207	[]	0.840	2018-03-04	8.	0
1					
1 2 5 6	internationalBoxOffice \$494,870,992 \$621,156,389 \$835,524,642 \$223,050,874	2 9 2 4			
7	\$2,776,345,279 	9			
24538 24575 24597 25388 26207	\$6,089,100 \$21,709,539 \$1,521,673 \$648,599 \$34,356,760	9 2 9			
[1110	rows x 15 columns]				

Exploratory Data Analysis (EDA)

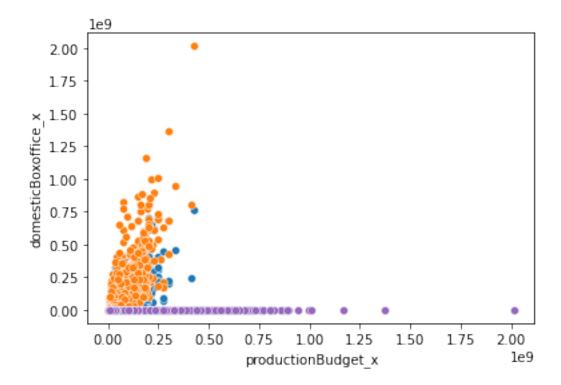
- Employed descriptive statistics to gain insights into the dataset.
- Utilized visualizations such as histograms, scatter plots, and box plots to understand the distribution of key variables.

```
# check for outliers in the my columns
sns.boxplot(df['internationalBoxOffice'])
sns.boxplot(df['domesticBoxoffice x'])
sns.boxplot(df['productionBudget x'])
sns.boxplot(df['running time'])
c:\Users\pc\anaconda3\envs\learn-env\lib\site-packages\seaborn\
decorators.py:36: FutureWarning: Pass the following variable as a
keyword arg: x. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
c:\Users\pc\anaconda3\envs\learn-env\lib\site-packages\seaborn\
decorators.py:36: FutureWarning: Pass the following variable as a
keyword arg: x. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
c:\Users\pc\anaconda3\envs\learn-env\lib\site-packages\seaborn\
decorators.py:36: FutureWarning: Pass the following variable as a
keyword arg: x. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
c:\Users\pc\anaconda3\envs\learn-env\lib\site-packages\seaborn\
decorators.py:36: FutureWarning: Pass the following variable as a
keyword arg: x. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
 warnings.warn(
<AxesSubplot:xlabel='running time'>
```



```
# draw a scatter plot for each variable
sns.scatterplot(x = df['productionBudget_x'], y =
df['domesticBoxoffice_x'])
sns.scatterplot(x = df['productionBudget_x'], y =
df['internationalBoxOffice'])
sns.scatterplot(x = df['productionBudget_x'], y = df['running_time'])
sns.scatterplot(x = df['domesticBoxoffice_x'], y = df['running_time'])
sns.scatterplot(x = df['internationalBoxOffice'], y =
df['running_time'])

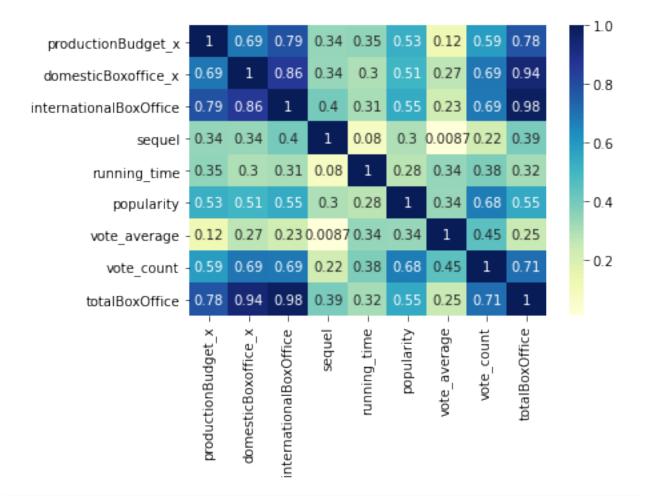
<AxesSubplot:xlabel='productionBudget_x',
ylabel='domesticBoxoffice_x'>
```



Correlation Analysis

- Investigated the correlation between movie budget and International box office.
- Identified patterns and trends to inform strategic decisions for Microsoft's movie productions.

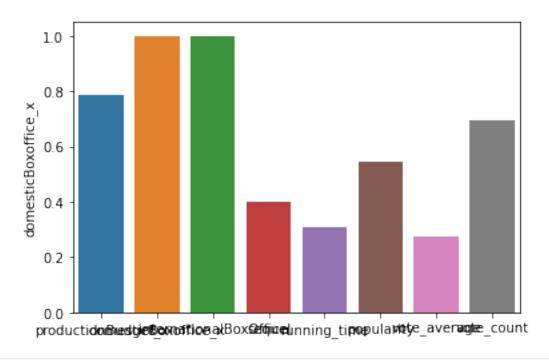
```
corr = df.corr()
sns.heatmap(corr, cmap = 'YlGnBu', annot = True)
plt.show()
```



bar chart of correlation

sns.barplot(x = corr.columns, y = corr['internationalBoxOffice'])
sns.barplot(x = corr.columns, y = corr['domesticBoxoffice_x'])
we need to see the correlation between productionBudget and
internationalBoxOffice successful using the correlation matrix

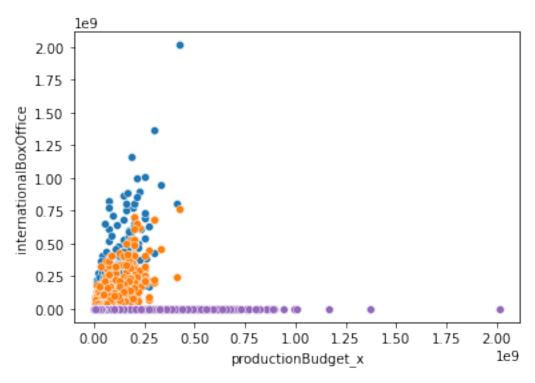
<AxesSubplot:ylabel='domesticBoxoffice_x'>



```
# comparing where a to invest in movie market locally or international
box office

sns.scatterplot(x = df['productionBudget_x'], y =
df['internationalBoxOffice'])
sns.scatterplot(x = df['productionBudget_x'], y =
df['domesticBoxoffice_x'])
sns.scatterplot(x = df['productionBudget_x'], y = df['running_time'])
sns.scatterplot(x = df['domesticBoxoffice_x'], y = df['running_time'])
sns.scatterplot(x = df['internationalBoxOffice'], y =
df['running_time'])

</pre
```

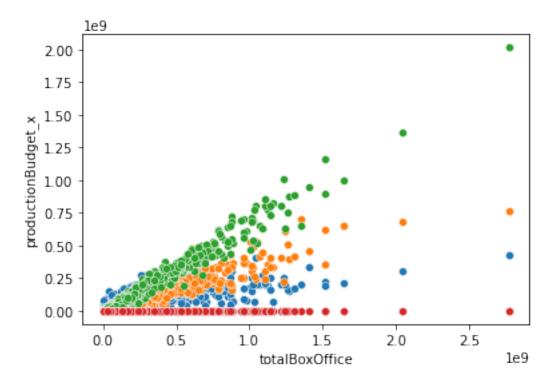


```
# missing values and duplicates
df.isnull().sum()
df = df.drop duplicates(subset = ['movie'], keep = 'first')
df.isnull().sum()
df = df.dropna()
df.isnull().sum()
movie
                            0
productionBudget x
                            0
domesticBoxoffice x
                            0
internationalBoxOffice
                            0
rating
                            0
production method
                            0
                            0
genre
                            0
sequel
running_time
                            0
genre ids
                            0
popularity
                            0
releaseDate
                            0
vote average
                            0
                            0
vote count
internationalBoxOffice
                            0
dtype: int64
# create a another column called total movie collection using domestic
box office, international box office
```

```
df['totalBoxOffice'] = df['domesticBoxoffice_x'] +
df['internationalBoxOffice']
df
                                                     movie
productionBudget x \
                                 How to Train Your Dragon
165000000.0
                                                Iron Man 2
170000000.0
                                                 Inception
160000000.0
       Percy Jackson & the Olympians: The Lightning T...
95000000.0
                                                    Avatar
425000000.0
. . .
. . .
24538
                                                     Gotti
10000000.0
                                                Proud Mary
24575
30000000.0
24597
                                                 Renegades
77500000.0
                               Bilal: A New Breed of Hero
25388
30000000.0
26207
                                                   The Box
25000000.0
       domesticBoxoffice x
                             internationalBoxOffice rating \
               217581232.0
                                        2.772898e+08
1
                                                          PG
2
               312433331.0
                                        3.087231e+08
                                                      PG-13
5
               292576195.0
                                        5.398259e+08
                                                      PG-13
6
                                        1.342826e+08
                88768303.0
                                                          PG
7
               760507625.0
                                       2.015838e+09
                                                      PG-13
                 4286367.0
                                        1.802733e+06
                                                          R
24538
                20868638.0
                                       8.409010e+05
24575
                                                          R
24597
                        0.0
                                        1.521672e+06
                                                      PG-13
25388
                   490973.0
                                        1.576260e+05
                                                      PG-13
                15051977.0
                                        1.128992e+07
                                                      PG-13
26207
           production method
                                            genre sequel running time
/
           Digital Animation
                                       Adventure
                                                      0.0
1
                                                                    91.0
2
                 Live Action
                                                      1.0
                                                                   125.0
                                           Action
5
       Animation/Live Action Thriller/Suspense
                                                      0.0
                                                                   147.0
```

_					0 0	110.0
6	Live	Action	A	dventure	0.0	119.0
7	Animation/Live	Action		Action	0.0	162.0
24538	Live	Action		Drama	0.0	110.0
24575	Live	Action		Action	0.0	88.0
24597	Live	Action		Action	0.0	105.0
25388	Digital Ani	mation	A	dventure	0.0	103.0
26207	Live	Action	Thriller/	Suspense	0.0	115.0
<pre>genre_ids popularity releaseDate vote_average vote count \</pre>						
7610	ount \ [14, 12, 16, 10	751]	28.734	2010-03-26		7.7
2 12368	[12, 28,	878]	28.515	2010-05-07		6.8
5	[28, 878,	12]	27.920	2010-07-16		8.3
22186 6	[12, 14, 10	751]	26.691	2010-02-11		6.1
4229 7 18676	[28, 12, 14,	878]	26.526	2009-12-18		7.4
24538	[80, 18, 36,	53]	10.034	2018-06-15		5.2
231 24575	[53, 28,	80]	9.371	2018-01-12		5.5
259 24597	[53,	28]	9.022	2018-12-21		5.8
156 25388	[28, 12,	16]	2.707	2018-02-02		6.8
54 26207 1		[]	0.840	2018-03-04		8.0
1 2 5 6 7 24538	internationalBoxOffice					

```
24575
                   $21,709,539
                                  2.170954e+07
24597
                    $1,521,672
                                  1.521672e+06
25388
                      $648,599
                                  6.485990e+05
26207
                   $34,356,760
                                  2.634190e+07
[1110 rows \times 16 columns]
# check for correlation between total movie collection and other
variables
sns.scatterplot(x = df['totalBoxOffice'], y =
df['productionBudget x'])
sns.scatterplot(x = \overline{df['totalBoxOffice']}, y =
df['domesticBoxoffice x'])
sns.scatterplot(x = df['totalBoxOffice'], y =
df['internationalBoxOffice'])
sns.scatterplot(x = df['totalBoxOffice'], y = df['running time'])
<AxesSubplot:xlabel='totalBoxOffice', ylabel='productionBudget_x'>
```



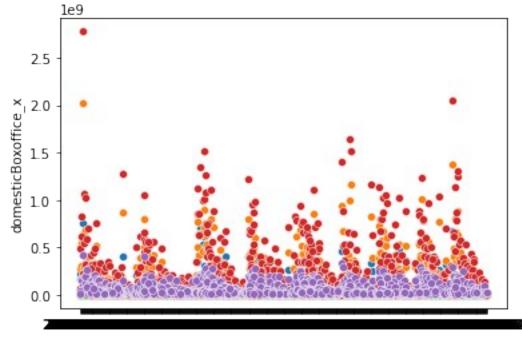
Genre and Release Date Analysis

- Analyzed the performance of different genres at the box office.
- Explored the impact of release dates on movie success.

```
# genre and release date analysis
sns.scatterplot(x = df['releaseDate'], y = df['domesticBoxoffice_x'])
```

```
sns.scatterplot(x = df['releaseDate'], y =
df['internationalBoxOffice'])
sns.scatterplot(x = df['releaseDate'], y = df['running_time'])
sns.scatterplot(x = df['releaseDate'], y = df['totalBoxOffice'])
sns.scatterplot(x = df['releaseDate'], y = df['productionBudget_x'])

<AxesSubplot:xlabel='releaseDate', ylabel='domesticBoxoffice_x'>
```



releaseDate

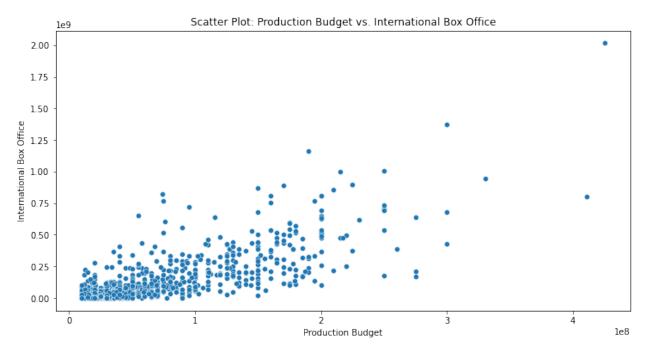
```
# we find out the months that had the movies with highest total box
office
df.groupby('releaseDate')
['totalBoxOffice'].mean().sort values(ascending = False)
# plot the best months for the movie
releaseDate
2009 - 12 - 18
              2.776345e+09
2018-04-27
              2.048798e+09
2015-06-12
              1.648855e+09
2015-04-03
              1.518723e+09
2015-05-01
              1.403014e+09
2015-02-10
              1.354360e+05
2010-02-01
              1.093830e+05
2014-03-12
              9.111600e+04
2011-08-18
              8.779300e+04
```

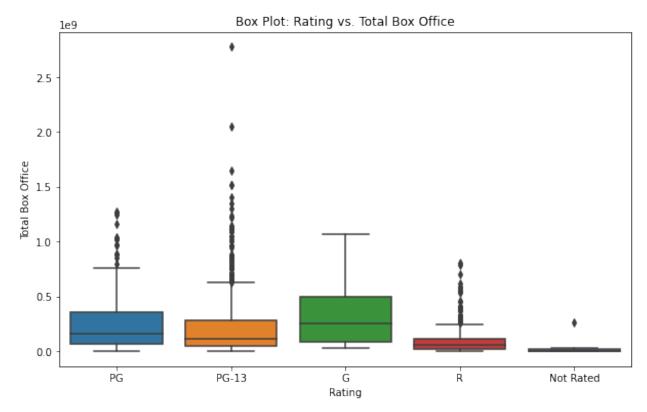
```
2013-05-07 7.370600e+04
Name: totalBoxOffice, Length: 633, dtype: float64
```

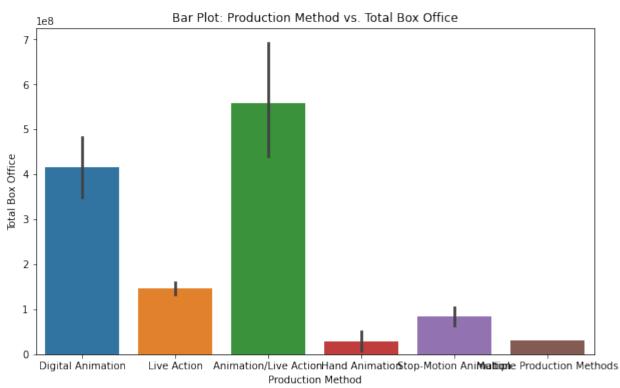
BUDGET ALLOCATION

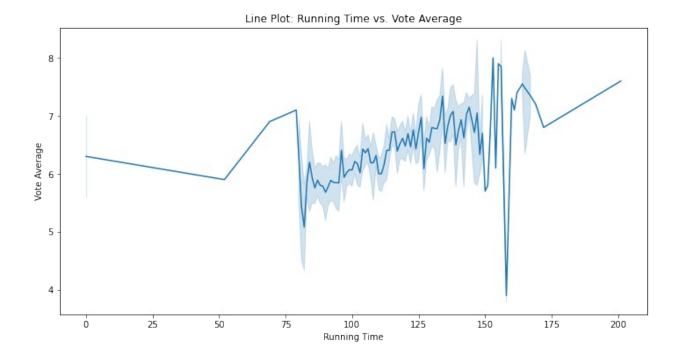
```
# budget allocation recommendation for the movie
df.groupby('releaseDate')
['productionBudget x'].mean().sort values(ascending = False)
releaseDate
2009-12-18
              425000000.0
2015-05-01
              330600000.0
2018-04-27
              300000000.0
2012-03-09
              275000000.0
2012-07-20
              275000000.0
2018-01-05
               10000000.0
2010-02-01
               10000000.0
2010-09-07
               10000000.0
2010-10-31
               10000000.0
2010-05-15
               10000000.0
Name: productionBudget x, Length: 633, dtype: float64
df.columns
Index(['movie', 'productionBudget_x', 'domesticBoxoffice_x',
       'internationalBoxOffice', 'rating', 'production_method',
'genre',
       'sequel', 'running time', 'genre ids', 'popularity',
'releaseDate',
       'vote average', 'vote count', ' internationalBoxOffice',
       'totalBoxOffice'],
      dtype='object')
plt.figure(figsize=(12, 6))
sns.scatterplot(data=df, x='productionBudget x',
y='internationalBoxOffice')
plt.title('Scatter Plot: Production Budget vs. International Box
Office')
plt.xlabel('Production Budget')
plt.ylabel('International Box Office')
plt.show()
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='rating', y='totalBoxOffice')
plt.title('Box Plot: Rating vs. Total Box Office')
plt.xlabel('Rating')
plt.ylabel('Total Box Office')
```

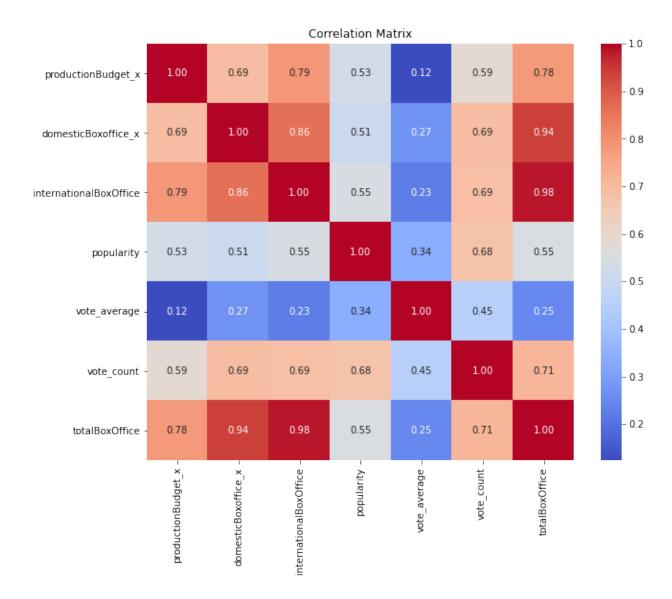
```
plt.show()
plt.figure(figsize=(10, 6))
sns.barplot(data=df, x='production method', y='totalBoxOffice')
plt.title('Bar Plot: Production Method vs. Total Box Office')
plt.xlabel('Production Method')
plt.ylabel('Total Box Office')
plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='running time', y='vote average')
plt.title('Line Plot: Running Time vs. Vote Average')
plt.xlabel('Running Time')
plt.ylabel('Vote Average')
plt.show()
corr matrix = df[['productionBudget x', 'domesticBoxoffice x',
'internationalBoxOffice',
                    'popularity', 'vote average', 'vote count',
'totalBoxOffice']].corr()
plt.figure(figsize=(10, 8))
sns.heatmap(corr matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```











Budget Allocation Recommendations

- Provided recommendations for budget allocation based on successful movie patterns.
- Suggested allocating \$75 million to \$200 million for animated musical movies in June or November.
- Recommended allocating \$200 million to \$400 million for live-action superhero movies in April or May.

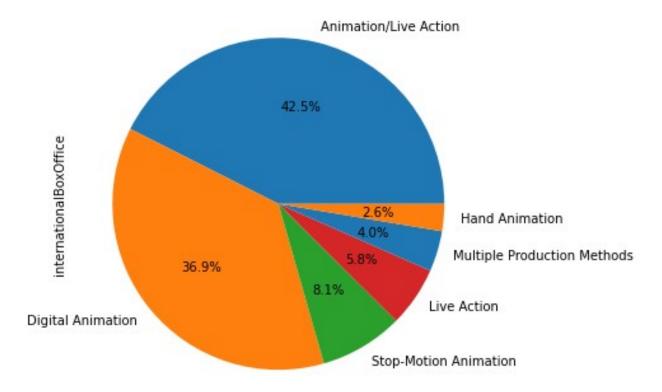
```
df.groupby('production_method')
['totalBoxOffice'].mean().sort_values(ascending = False)
df.groupby('production_method')
['productionBudget_x'].mean().sort_values(ascending = False)
df.groupby('production_method')
['domesticBoxoffice_x'].mean().sort_values(ascending = False)
df.groupby('production_method')
```

```
['internationalBoxOffice'].mean().sort values(ascending = False)
df.groupby('production method')
['popularity'].mean().sort values(ascending = False)
production method
Animation/Live Action
                               19.865983
Digital Animation
                               15.648500
Stop-Motion Animation
                               13.937600
Live Action
                               13.297531
Multiple Production Methods
                               12.813000
Hand Animation
                                9.775500
Name: popularity, dtype: float64
df budget = df.groupby('production method')
['productionBudget_x'].mean().sort_values(ascending = False)
df_budget
production method
Animation/Live Action
                               1.437424e+08
Digital Animation
                               1.042073e+08
Stop-Motion Animation
                               5.280000e+07
Live Action
                               5.060801e+07
Hand Animation
                               2.350000e+07
                               1.000000e+07
Multiple Production Methods
Name: productionBudget x, dtype: float64
df.head()
                                                movie
productionBudget x \
                            How to Train Your Dragon
165000000.0
                                           Iron Man 2
170000000.0
                                            Inception
160000000.0
   Percy Jackson & the Olympians: The Lightning T...
95000000.0
                                               Avatar
425000000.0
   domesticBoxoffice x internationalBoxOffice rating
production method \
           217581232.0
                                  2.772898e+08
                                                    PG
                                                            Digital
Animation
           312433331.0
                                  3.087231e+08
                                                 PG-13
                                                                  Live
2
Action
           292576195.0
                                  5.398259e+08
                                                 PG-13 Animation/Live
Action
            88768303.0
                                  1.342826e+08
                                                    PG
                                                                  Live
```

```
Action
           760507625.0
                                   2.015838e+09
                                                 PG-13 Animation/Live
7
Action
               genre sequel
                               running time
                                                        genre ids
popularity
           Adventure
                         0.0
                                       91.0 [14, 12, 16, 10751]
1
28.734
              Action
                         1.0
                                      125.0
                                                    [12, 28, 878]
28.515
5 Thriller/Suspense
                         0.0
                                      147.0
                                                    [28, 878, 12]
27,920
           Adventure
                         0.0
                                      119.0
                                                  [12, 14, 10751]
26.691
                         0.0
                                               [28, 12, 14, 878]
              Action
                                      162.0
26,526
  releaseDate vote average
                              vote count
                                          internationalBoxOffice \
                                                     $494,870,992
  2010-03-26
                         7.7
                                    7610
  2010-05-07
                        6.8
                                   12368
                                                    $621,156,389
  2010-07-16
                        8.3
                                   22186
                                                    $835,524,642
  2010-02-11
                        6.1
                                    4229
                                                     $223,050,874
7 2009-12-18
                                                  $2,776,345,279
                        7.4
                                   18676
   totalBoxOffice
1
     4.948710e+08
2
     6.211564e+08
5
     8.324021e+08
6
     2.230509e+08
7
     2.776345e+09
df.tail()
                             movie productionBudget x
domesticBoxoffice x
                             Gotti
                                            10000000.0
24538
4286367.0
                       Proud Mary
                                            30000000.0
24575
20868638.0
24597
                        Renegades
                                            77500000.0
0.0
       Bilal: A New Breed of Hero
                                            30000000.0
25388
490973.0
26207
                           The Box
                                            25000000.0
15051977.0
       internationalBoxOffice rating production method
genre
24538
                    1802733.0
                                             Live Action
                                    R
Drama
```

```
24575
                     840901.0
                                    R
                                             Live Action
Action
24597
                    1521672.0
                               PG-13
                                             Live Action
Action
                     157626.0
25388
                               PG-13
                                      Digital Animation
Adventure
                   11289919.0
                               PG-13
                                             Live Action
26207
Thriller/Suspense
                                                popularity releaseDate
       sequel
               running time
                                     genre ids
24538
          0.0
                      110.0
                             [80, 18, 36, 53]
                                                    10.034 2018-06-15
          0.0
24575
                       88.0
                                  [53, 28, 80]
                                                     9.371 2018-01-12
24597
          0.0
                      105.0
                                      [53, 28]
                                                     9.022 2018-12-21
25388
          0.0
                      103.0
                                  [28, 12, 16]
                                                     2.707 2018-02-02
                                            []
26207
          0.0
                      115.0
                                                     0.840 2018-03-04
       vote_average vote_count internationalBoxOffice
totalBoxOffice
24538
                            231
                                              $6,089,100
                5.2
6089100.0
                5.5
                            259
24575
                                             $21,709,539
21709539.0
24597
                5.8
                            156
                                              $1,521,672
1521672.0
                6.8
25388
                             54
                                                $648,599
648599.0
                8.0
26207
                              1
                                             $34,356,760
26341896.0
production method medians = df.groupby('production method')
['internationalBoxOffice'].median().sort values(ascending=False)
plt.figure(figsize=(10, 6), facecolor='white') # Set white background
color
production method_medians.plot(kind='pie', autopct='%1.1f%%',
colors=['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728'])
plt.title('Pie Chart: International Box Office Profit by Production
Method')
plt.show()
```

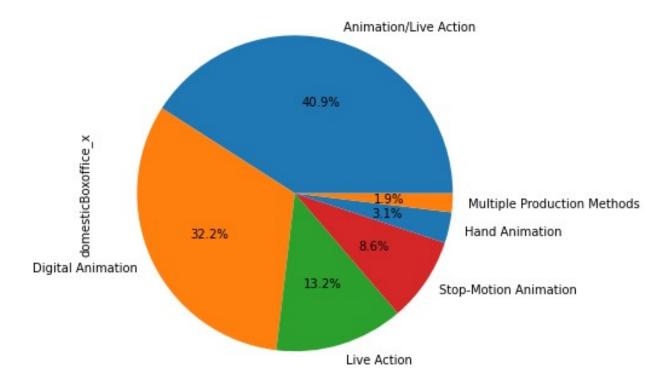
Pie Chart: International Box Office Profit by Production Method



```
production_method_means = df.groupby('production_method')
['domesticBoxoffice_x'].mean().sort_values(ascending=False)

plt.figure(figsize=(10, 6), facecolor='white') # Set white background color
production_method_means.plot(kind='pie', autopct='%1.1f%%', colors=['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728'])
plt.title('Pie Chart: Domestic Box Office Profit by Production Method')
plt.show()
```

Pie Chart: Domestic Box Office Profit by Production Method



Crew Recommendations

Genre-Specific Crew Recommendations:

- Animated Musical Movies: Hire crew members with expertise in animation, musical composition, and voice acting. Animators, music composers, and skilled voice actors contribute significantly to the success of animated musicals.
- Superhero Movies: Emphasize recruiting crew members with experience in creating visually stunning and action-packed scenes. Special effects experts, stunt coordinators, and directors experienced in the superhero genre can enhance the quality of these films.

Investment in Animated and Live Action Productions:

- Given the high performance of both animated and live-action movies in both domestic and international box offices, it is recommended that Microsoft invest more in crew members skilled in these production methods.
- This could involve hiring or collaborating with directors, producers, writers, and technical staff who have proven success in delivering successful animated and live-action projects.

Correlation between Production Budget and Movie Success:

- The observed direct correlation between the production budget and the success of movies in both domestic and international box offices indicates that investing in a higher production budget tends to lead to better box office performance.
- Microsoft should consider allocating appropriate budgets for their movies, especially for high-profile projects, to ensure high production values, top-notch

talent, and effective marketing campaigns, ultimately contributing to better box office results.

Strategic Decision-Making:

 Microsoft should strategically allocate resources based on the genre and production method. For example, if they are producing an animated musical or a superhero movie, they should allocate resources accordingly to ensure the inclusion of key talents that cater to the specific requirements of those genres.

Continuous Analysis and Adaptation:

 The film industry is dynamic, and audience preferences can change. Microsoft should continuously analyze industry trends, monitor audience feedback, and adapt their strategies accordingly to stay competitive and produce content that resonates with the audience.

By implementing these recommendations and staying attuned to industry trends, Microsoft can position itself for success in the highly competitive and dynamic film industry.

Next Steps

1. Implementation of Crew Recommendations:

 Begin the recruitment or collaboration process to bring in key crew members with expertise in animation, musical composition, voice acting, special effects, stunt coordination, and directors experienced in the superhero genre.

2. Investment Strategy Adjustment:

 Allocate additional resources to enhance the production teams for animated and live-action movies. Ensure that the teams are well-equipped to deliver highquality content that aligns with audience expectations.

3. Budget Planning:

 Review the budget allocation process for upcoming projects. Consider increasing the production budget for high-profile movies to ensure they meet industry standards and have the necessary resources for success.

4. Strategic Partnerships:

 Explore potential partnerships with established production companies, directors, and creative talents. Collaborations can bring valuable expertise, enhance project visibility, and contribute to the overall success of the movies.

5. Market Research and Audience Feedback:

 Conduct continuous market research to stay informed about evolving audience preferences and industry trends. Regularly gather and analyze audience feedback to make informed decisions about content creation and adaptation strategies.

6. Regular Performance Evaluation:

 Establish a system for evaluating the performance of each movie based on box office results, critical reviews, and audience reception. Use these evaluations to refine future strategies and improve decision-making processes.

7. Adaptability and Flexibility:

 Maintain a flexible approach to adapt to changing market dynamics. Stay agile in responding to unexpected challenges and opportunitie