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

                                       movie_name  production_year  \
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   
...                                           ...              ...   
1931           The Nutcracker and the Four Realms             2018   
1932                                      Aquaman             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           10000000             80238724
...          ...                ...                  ...
1931   298170100          132900000             54858851
1932   213100100          160000000            333804251
1933   263730100          175000000            200236625
1934   248680100          178000000            220159104
1935   222990100          200000000            159555901

      international_box_office     rating         creative_type  \
0                        62581      PG-13  Contemporary Fiction   
1                     31000000  Not Rated       Science Fiction   
2                       175380      PG-13    Historical Fiction   
3                      6373339      PG-13  Contemporary Fiction   
4                     83638091          R  Contemporary Fiction   
...                        ...        ...                   ...   
1931                 115435048         PG               Fantasy   
1932                 805605026      PG-13            Super Hero   
1933                 319167373         PG          Kids Fiction   
1934                 567297448      PG-13  Contemporary Fiction   
1935                 492664185      PG-13               Fantasy   

                                   source  production_method
genre  \
0                           Based on Play        Live Action
Comedy   
1                     Original Screenplay        Live Action
Action   
2                     Original Screenplay        Live Action
Drama   
3                     Original Screenplay        Live Action
Comedy   
4                     Original Screenplay        Live Action
Horror   
...                                   ...                ...         ..
.   
1931  Based on Folk Tale/Legend/Fairytale        Live Action
Adventure   
1932         Based on Comic/Graphic Novel        Live Action
Action   
1933                  Original Screenplay  Digital Animation
Adventure   
1934                          Based on TV        Live Action
Action   
1935                             Spin-Off        Live Action
Adventure   

      sequel  running_time
0        1.0           NaN
1        1.0           NaN
2        0.0           NaN
```

```
3        0.0         105.0
4        1.0           NaN
...      ...           ...
1931     0.0          99.0
1932     0.0         143.0
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

                                            movie  production_year  \
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
...                                           ...              ...
1931           The Nutcracker and the Four Realms             2018
1932                                      Aquaman             2018
1933                    Ralph Breaks The Internet             2018
1934                  Mission: Impossible—Fallout             2018
1935  Fantastic Beasts: The Crimes of Grindelwald             2018

      movie_odid  productionBudget  domesticBoxoffice
internationalBoxOffice  \
0        8220100          10000000           63257940
62581
1       58540100          10000000            1430721
31000000
2       34620100          10000000           11748661
175380
3       24910100          10000000           20342852
6373339
4        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  \
0         PG-13  Contemporary Fiction                        Based on
Play
1     Not Rated       Science Fiction                  Original
Screenplay
2         PG-13    Historical Fiction                  Original
Screenplay
3         PG-13  Contemporary Fiction                  Original
Screenplay
4             R  Contemporary Fiction                  Original
Screenplay
...         ...                   ...
...
1931         PG               Fantasy  Based on Folk
Tale/Legend/Fairytale
1932      PG-13            Super Hero         Based on Comic/Graphic
Novel
1933         PG          Kids Fiction                  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     Comedy     0.0         105.0
4           Live Action     Horror     1.0           NaN
...                 ...        ...     ...           ...
1931        Live Action  Adventure     0.0          99.0
1932        Live Action     Action     0.0         143.0
1933  Digital Animation  Adventure     1.0         112.0
1934        Live Action     Action     1.0         147.0
1935        Live Action  Adventure     1.0         134.0

[1936 rows x 13 columns]
```

Renamed the column movie names to *movie*

```
df1 = df1.rename(columns = {'movie_name':' movie'})
df1

                                         movie  production_year  \
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
...                                        ...              ...
1931           The Nutcracker and the Four Realms             2018
1932                                   Aquaman             2018
1933                    Ralph Breaks The Internet             2018
1934                  Mission: Impossible—Fallout             2018
1935  Fantastic Beasts: The Crimes of Grindelwald             2018

      movie_odid  productionBudget  domesticBoxoffice
internationalBoxOffice  \
0        8220100          10000000           63257940
62581
1       58540100          10000000            1430721
31000000
2       34620100          10000000           11748661
175380
3       24910100          10000000           20342852
6373339
4        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  \
0         PG-13  Contemporary Fiction                          Based on
Play
1     Not Rated       Science Fiction                   Original
Screenplay
```

```
2         PG-13    Historical Fiction                  Original
Screenplay
3         PG-13  Contemporary Fiction                  Original
Screenplay
4             R  Contemporary Fiction                  Original
Screenplay
...         ...                   ...
...
1931         PG               Fantasy  Based on Folk
Tale/Legend/Fairytale
1932      PG-13            Super Hero         Based on Comic/Graphic
Novel
1933         PG          Kids Fiction                  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     Comedy     0.0         105.0
4           Live Action     Horror     1.0           NaN
...                 ...        ...     ...           ...
1931        Live Action  Adventure     0.0          99.0
1932        Live Action     Action     0.0         143.0
1933  Digital Animation  Adventure     1.0         112.0
1934        Live Action     Action     1.0         147.0
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.

```
df2= pd.read_csv(r"C:\Users\pc\Videos\projects\Phase_One_Proect\data\
tmdb.movies.csv")
df2

       Unnamed: 0            genre_ids      id original_language  \
0               0      [12, 14, 10751]   12444                en
```

```
1               1  [14, 12, 16, 10751]   10191                en
2               2        [12, 28, 878]   10138                en
3               3      [16, 35, 10751]     862                en
4               4        [28, 878, 12]   27205                en
...           ...                  ...     ...               ...
26512       26512             [27, 18]  488143                en
26513       26513             [18, 53]  485975                en
26514       26514         [14, 28, 12]  381231                en
26515       26515      [10751, 12, 28]  366854                en
26516       26516             [53, 27]  309885                en

                                     original_title  popularity
release_date  \
0      Harry Potter and the Deathly Hallows: Part 1      33.533
2010-11-19
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
26513                               _EXHIBIT_84xxx_       0.600
2018-05-01
26514                                  The Last One       0.600
2018-10-01
26515                                  Trailer Made       0.600
2018-06-22
26516                                    The Church       0.600
2018-10-05

                                              title  vote_average
vote_count
0      Harry 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
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]
```

Renamed the column original_title to *movie*.

```

df2 = df2.rename(columns = {'original_title':' movie'})
df2

       Unnamed: 0            genre_ids      id original_language  \
0               0      [12, 14, 10751]   12444                en
1               1  [14, 12, 16, 10751]   10191                en
2               2        [12, 28, 878]   10138                en
3               3      [16, 35, 10751]     862                en
4               4        [28, 878, 12]   27205                en
...           ...                  ...     ...               ...
26512       26512             [27, 18]  488143                en
26513       26513             [18, 53]  485975                en
26514       26514         [14, 28, 12]  381231                en
26515       26515      [10751, 12, 28]  366854                en
26516       26516             [53, 27]  309885                en

                                              movie  popularity
release_date  \
0      Harry Potter and the Deathly Hallows: Part 1      33.533
2010-11-19
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
26513                               _EXHIBIT_84xxx_       0.600
```

```
2018-05-01
26514                                   The Last One       0.600
2018-10-01
26515                                   Trailer Made       0.600
2018-06-22
26516                                     The Church       0.600
2018-10-05

                                               title  vote_average
vote_count
0      Harry 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
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]
```

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               0      [12, 14, 10751]    12444                en
1               1  [14, 12, 16, 10751]    10191                en
2               2        [12, 28, 878]    10138                en
3               3      [16, 35, 10751]      862                en
4               4        [28, 878, 12]    27205                en
```

```
...           ...                  ...     ...                 ...
26512       26512             [27, 18]  488143                en
26513       26513             [18, 53]  485975                en
26514       26514         [14, 28, 12]  381231                en
26515       26515      [10751, 12, 28]  366854                en
26516       26516             [53, 27]  309885                en

                                              movie  popularity 
releaseDate  \
0      Harry Potter and the Deathly Hallows: Part 1      33.533  2010-
11-19
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
26513                               _EXHIBIT_84xxx_       0.600  2018-
05-01
26514                                  The Last One       0.600  2018-
10-01
26515                                  Trailer Made       0.600  2018-
06-22
26516                                    The Church       0.600  2018-
10-05

                                              title  vote_average 
vote_count
0      Harry 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
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  \
0      1  Dec 18, 2009                                       Avatar
1      2  May 20, 2011  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
...   ..           ...                                          ...
5777  78  Dec 31, 2018                                       Red 11
5778  79   Apr 2, 1999                                    Following
5779  80  Jul 13, 2005                Return to the Land of Wonders
5780  81  Sep 29, 2015                         A Plague So Pleasant
5781  82   Aug 5, 2005                            My Date With Drew

     production_budget domestic_gross worldwide_gross
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
5779            $5,000         $1,338          $1,338
5780            $1,400             $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  \
0      1  Dec 18, 2009                                       Avatar
1      2  May 20, 2011  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
...   ..           ...                                          ...
5777  78  Dec 31, 2018                                       Red 11
5778  79   Apr 2, 1999                                    Following
5779  80  Jul 13, 2005                Return to the Land of Wonders
5780  81  Sep 29, 2015                         A Plague So Pleasant
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
5779           $5,000            $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  \
0      Harry Potter and the Deathly Hallows: Part 1              NaN

1                          How to Train Your Dragon           2010.0

2                                        Iron Man 2           2010.0

3                                         Toy Story              NaN

4                                         Toy Story              NaN

...                                             ...              ...

26621                         Laboratory Conditions              NaN

26622                               _EXHIBIT_84xxx_              NaN

26623                                  The Last One              NaN

26624                                  Trailer Made              NaN

26625                                    The Church              NaN


        movie_odid  productionBudget_x  domesticBoxoffice_x  \
0              NaN                 NaN                  NaN
1      116630100.0         165000000.0          217581232.0
2      117940100.0         170000000.0          312433331.0
3              NaN                 NaN                  NaN
4              NaN                 NaN                  NaN
...            ...                 ...                  ...
26621          NaN                 NaN                  NaN
26622          NaN                 NaN                  NaN
26623          NaN                 NaN                  NaN
```

```
26624          NaN                 NaN                  NaN
26625          NaN                 NaN                  NaN

       internationalBoxOffice rating creative_type  \
0                         NaN    NaN           NaN
1                 277289760.0     PG       Fantasy
2                 308723058.0  PG-13    Super Hero
3                         NaN    NaN           NaN
4                         NaN    NaN           NaN
...                       ...    ...           ...
26621                     NaN    NaN           NaN
26622                     NaN    NaN           NaN
26623                     NaN    NaN           NaN
26624                     NaN    NaN           NaN
26625                     NaN    NaN           NaN

                                  source  production_method  ...
popularity  \
0                                    NaN                NaN  ...
33.533
1      Based on Fiction Book/Short Story  Digital Animation  ...
28.734
2           Based on Comic/Graphic Novel        Live Action  ...
28.515
3                                    NaN                NaN  ...
28.005
4                                    NaN                NaN  ...
28.005
...                                  ...                ...  ...
...
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

       releaseDate                                         title  \
0       2010-11-19  Harry Potter and the Deathly Hallows: Part 1
1       2010-03-26                      How to Train Your Dragon
2       2010-05-07                                    Iron Man 2
3       1995-11-22                                     Toy Story
4       1995-11-22                                     Toy Story
...            ...                                           ...
26621   2018-10-13                         Laboratory Conditions
26622   2018-05-01                               _EXHIBIT_84xxx_
```

```
26623   2018-10-01                                  The Last One
26624   2018-06-22                                  Trailer Made
26625   2018-10-05                                    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   


      domesticBoxoffice_y  internationalBoxOffice  
0                     NaN                     NaN  
1            $217,581,232            $494,870,992  
2            $312,433,331            $621,156,389  
3            $191,796,233            $364,545,516  
4            $191,796,233            $364,545,516  
...                   ...                     ...  
26621                 NaN                     NaN  
26622                 NaN                     NaN  
26623                 NaN                     NaN  
26624                 NaN                     NaN  
26625                 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
```

```
                                              movie
production_year  \
0      Harry Potter and the Deathly Hallows: Part 1              NaN

1                          How to Train Your Dragon           2010.0

2                                        Iron Man 2           2010.0

3                                         Toy Story              NaN

5                                         Inception           2010.0

...                                             ...              ...

26621                         Laboratory Conditions              NaN

26622                               _EXHIBIT_84xxx_              NaN

26623                                  The Last One              NaN

26624                                  Trailer Made              NaN

26625                                    The Church              NaN


        movie_odid  productionBudget_x  domesticBoxoffice_x  \
0              NaN                 NaN                  NaN
1      116630100.0         165000000.0          217581232.0
2      117940100.0         170000000.0          312433331.0
3              NaN                 NaN                  NaN
5      105240100.0         160000000.0          292576195.0
...            ...                 ...                  ...
26621          NaN                 NaN                  NaN
26622          NaN                 NaN                  NaN
26623          NaN                 NaN                  NaN
26624          NaN                 NaN                  NaN
26625          NaN                 NaN                  NaN

       internationalBoxOffice rating    creative_type
production_method  \
0                         NaN    NaN              NaN
NaN
1                 277289760.0     PG          Fantasy      Digital
Animation
2                 308723058.0  PG-13       Super Hero            Live
Action
3                         NaN    NaN              NaN
NaN
5                 539825887.0  PG-13  Science Fiction  Animation/Live
Action
...                       ...    ...              ...
```

```
...
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

                   genre  ...  popularity  releaseDate  \
0                    NaN  ...      33.533   2010-11-19
1              Adventure  ...      28.734   2010-03-26
2                 Action  ...      28.515   2010-05-07
3                    NaN  ...      28.005   1995-11-22
5      Thriller/Suspense  ...      27.920   2010-07-16
...                  ...  ...         ...          ...
26621                NaN  ...       0.600   2018-10-13
26622                NaN  ...       0.600   2018-05-01
26623                NaN  ...       0.600   2018-10-01
26624                NaN  ...       0.600   2018-06-22
26625                NaN  ...       0.600   2018-10-05

                                              title vote_average
vote_count  \
0      Harry 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
5                                         Inception          8.3
22186
...                                             ...          ...
...
26621                         Laboratory Conditions          0.0
1
26622                               _EXHIBIT_84xxx_          0.0
1
26623                                  The Last One          0.0
1
26624                                  Trailer Made          0.0
1
26625                                    The Church          0.0
1
```

```
       id_y  release_date productionBudget_y domesticBoxoffice_y  \
0       NaN           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
df.isnull().sum()
df = df.dropna()
df
```

```
                                                   movie
production_year  \
1                               How to Train Your Dragon
2010.0
2                                             Iron Man 2
2010.0
5                                              Inception
2010.0
6      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
25388                         Bilal: A New Breed of Hero
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      122040100.0         425000000.0          760507625.0
...            ...                 ...                  ...
24538  246820100.0          10000000.0            4286367.0
24575  281630100.0          30000000.0           20868638.0
24597  227610100.0          77500000.0                  0.0
25388  265190100.0          30000000.0             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                5.398259e+08  PG-13       Science Fiction
6                1.342826e+08     PG               Fantasy
7                2.015838e+09  PG-13       Science Fiction
...                       ...    ...                   ...
24538            1.802733e+06      R         Dramatization
24575            8.409010e+05      R  Contemporary Fiction
24597            1.521672e+06  PG-13  Contemporary Fiction
25388            1.576260e+05  PG-13         Dramatization
26207            1.128992e+07  PG-13               Fantasy

           production_method              genre  ...  popularity
releaseDate  \
1          Digital Animation          Adventure  ...      28.734
2010-03-26
2                Live Action             Action  ...      28.515
2010-05-07
5      Animation/Live Action  Thriller/Suspense  ...      27.920
2010-07-16
6                Live Action          Adventure  ...      26.691
2010-02-11
7      Animation/Live Action             Action  ...      26.526
2009-12-18
...                      ...                ...  ...         ...
...
```

```
24538            Live Action               Drama  ...      10.034
2018-06-15
24575            Live Action              Action  ...       9.371
2018-01-12
24597            Live Action              Action  ...       9.022
2018-12-21
25388      Digital Animation           Adventure  ...       2.707
2018-02-02
26207            Live Action  Thriller/Suspense  ...       0.840
2018-03-04

                                                       title vote_average
\
1                                How to Train Your Dragon            7.7

2                                              Iron Man 2            6.8

5                                               Inception            8.3

6      Percy Jackson & the Olympians: The Lightning T...            6.1

7                                                  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


       vote_count   id_y  release_date productionBudget_y
domesticBoxoffice_y  \
1            7610   30.0  Mar 26, 2010       $165,000,000
$217,581,232
2           12368   15.0   May 7, 2010       $170,000,000
$312,433,331
5           22186   38.0  Jul 16, 2010       $160,000,000
$292,576,195
6            4229   17.0  Feb 12, 2010        $95,000,000
$88,768,303
7           18676    1.0  Dec 18, 2009       $425,000,000
$760,507,625
...           ...    ...           ...                ...
...
```

```
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
24597          156   20.0  Jan 22, 2019        $77,500,000
$0
25388           54  100.0   Feb 2, 2018        $30,000,000
$490,973
26207            1   66.0   Nov 6, 2009        $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
...                        ...
24538               $6,089,100
24575              $21,709,539
24597               $1,521,672
25388                 $648,599
26207              $34,356,760

[1110 rows x 26 columns]
```

Dropped the domesticBoxOffice columns.

```
df['domesticBoxoffice_y'] == df['domesticBoxoffice_x']
df = df.drop(columns = ['domesticBoxoffice_y'], axis = 1)
df

                                                        movie
production_year  \
1                                  How to Train Your Dragon
2010.0
2                                                Iron Man 2
2010.0
5                                                 Inception
2010.0
6       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
25388                          Bilal: A New Breed of Hero
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      122040100.0         425000000.0          760507625.0   
...            ...                 ...                  ...   
24538  246820100.0          10000000.0            4286367.0   
24575  281630100.0          30000000.0           20868638.0   
24597  227610100.0          77500000.0                  0.0   
25388  265190100.0          30000000.0             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                5.398259e+08  PG-13       Science Fiction   
6                1.342826e+08     PG               Fantasy   
7                2.015838e+09  PG-13       Science Fiction   
...                       ...    ...                   ...   
24538            1.802733e+06      R         Dramatization   
24575            8.409010e+05      R  Contemporary Fiction   
24597            1.521672e+06  PG-13  Contemporary Fiction   
25388            1.576260e+05  PG-13         Dramatization   
26207            1.128992e+07  PG-13               Fantasy   

           production_method              genre  ... 
original_language  \
1          Digital Animation          Adventure  ... 
en   
2                Live Action             Action  ... 
en   
5      Animation/Live Action  Thriller/Suspense  ... 
en   
6                Live Action          Adventure  ... 
en   
7      Animation/Live Action             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
...           ...          ...
24538      10.034   2018-06-15
24575       9.371   2018-01-12
24597       9.022   2018-12-21
25388       2.707   2018-02-02
26207       0.840   2018-03-04

                                                   title  vote_average
\
1                               How to Train Your Dragon           7.7

2                                             Iron Man 2           6.8

5                                              Inception           8.3

6      Percy Jackson & the Olympians: The Lightning T...           6.1

7                                                 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


      vote_count   id_y  release_date productionBudget_y  \
1           7610   30.0  Mar 26, 2010       $165,000,000
2          12368   15.0   May 7, 2010       $170,000,000
5          22186   38.0  Jul 16, 2010       $160,000,000
```

```
6           4229   17.0  Feb 12, 2010        $95,000,000
7          18676    1.0  Dec 18, 2009       $425,000,000
...          ...    ...           ...                ...
24538        231   64.0  Jun 15, 2018        $10,000,000
24575        259   50.0  Jan 12, 2018        $30,000,000
24597        156   20.0  Jan 22, 2019        $77,500,000
25388         54  100.0   Feb 2, 2018        $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
...                        ...
24538               $6,089,100
24575              $21,709,539
24597               $1,521,672
25388                 $648,599
26207              $34,356,760

[1110 rows x 25 columns]
```

```
df = df.drop(columns = ['productionBudget_y'], axis = 1)
df
```

```
                                                   movie
production_year  \
1                               How to Train Your Dragon
2010.0
2                                             Iron Man 2
2010.0
5                                              Inception
2010.0
6      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
25388                         Bilal: A New Breed of Hero
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      122040100.0         425000000.0          760507625.0
...            ...                 ...                  ...
24538  246820100.0          10000000.0            4286367.0
24575  281630100.0          30000000.0           20868638.0
24597  227610100.0          77500000.0                  0.0
25388  265190100.0          30000000.0             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                5.398259e+08  PG-13       Science Fiction
6                1.342826e+08     PG               Fantasy
7                2.015838e+09  PG-13       Science Fiction
...                       ...    ...                   ...
24538            1.802733e+06      R         Dramatization
24575            8.409010e+05      R  Contemporary Fiction
24597            1.521672e+06  PG-13  Contemporary Fiction
25388            1.576260e+05  PG-13         Dramatization
26207            1.128992e+07  PG-13               Fantasy

           production_method              genre  ...    id_x  \
1          Digital Animation          Adventure  ...   10191
2                Live Action             Action  ...   10138
5      Animation/Live Action  Thriller/Suspense  ...   27205
6                Live Action          Adventure  ...   32657
7      Animation/Live Action             Action  ...   19995
...                      ...                ...  ...     ...
24538            Live Action              Drama  ...  339103
24575            Live Action             Action  ...  442064
24597            Live Action             Action  ...  335788
25388      Digital Animation          Adventure  ...  332718
26207            Live Action  Thriller/Suspense  ...  509314

       original_language  popularity releaseDate  \
1                     en      28.734  2010-03-26
2                     en      28.515  2010-05-07
5                     en      27.920  2010-07-16
6                     en      26.691  2010-02-11
7                     en      26.526  2009-12-18
...                  ...         ...         ...
24538                 en      10.034  2018-06-15
```

```
24575                 en       9.371  2018-01-12
24597                 fr       9.022  2018-12-21
25388                 en       2.707  2018-02-02
26207                 en       0.840  2018-03-04

                                                   title vote_average
\
1                               How to Train Your Dragon          7.7

2                                             Iron Man 2          6.8

5                                              Inception          8.3

6      Percy Jackson & the Olympians: The Lightning T...          6.1

7                                                 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


       vote_count   id_y  release_date   internationalBoxOffice
1            7610   30.0  Mar 26, 2010             $494,870,992
2           12368   15.0   May 7, 2010             $621,156,389
5           22186   38.0  Jul 16, 2010             $835,524,642
6            4229   17.0  Feb 12, 2010             $223,050,874
7           18676    1.0  Dec 18, 2009           $2,776,345,279
...           ...    ...           ...                      ...
24538         231   64.0  Jun 15, 2018               $6,089,100
24575         259   50.0  Jan 12, 2018              $21,709,539
24597         156   20.0  Jan 22, 2019               $1,521,672
25388          54  100.0   Feb 2, 2018                 $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)
df
```

```
                                                   movie   movie_odid
\
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

...                                                    ...          ...

24538                                                Gotti  246820100.0

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
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
...                   ...                  ...                     ... 
...
24538          10000000.0            4286367.0            1.802733e+06 
R
24575          30000000.0           20868638.0            8.409010e+05 
R
24597          77500000.0                  0.0            1.521672e+06 
PG-13
25388          30000000.0             490973.0            1.576260e+05 
PG-13
26207          25000000.0           15051977.0            1.128992e+07 
PG-13

              creative_type      production_method              genre 
sequel  \
1                   Fantasy      Digital Animation          Adventure
0.0
2                Super Hero            Live Action             Action
1.0
5           Science Fiction  Animation/Live Action  Thriller/Suspense
0.0
6                   Fantasy            Live Action          Adventure
0.0
7           Science Fiction  Animation/Live Action             Action
0.0
...                     ...                    ...                ...
...
24538         Dramatization            Live Action              Drama
0.0
24575  Contemporary Fiction            Live Action             Action
0.0
24597  Contemporary Fiction            Live Action             Action
0.0
25388         Dramatization      Digital Animation          Adventure
0.0
26207               Fantasy            Live Action  Thriller/Suspense
0.0

       ...    id_x  original_language popularity  releaseDate  \
1      ...   10191                 en     28.734   2010-03-26
2      ...   10138                 en     28.515   2010-05-07
5      ...   27205                 en     27.920   2010-07-16
6      ...   32657                 en     26.691   2010-02-11
7      ...   19995                 en     26.526   2009-12-18
...    ...     ...                ...        ...          ...
24538  ...  339103                 en     10.034   2018-06-15
24575  ...  442064                 en      9.371   2018-01-12
24597  ...  335788                 fr      9.022   2018-12-21
25388  ...  332718                 en      2.707   2018-02-02
26207  ...  509314                 en      0.840   2018-03-04

                                                   title  vote_average
\
1                               How to Train Your Dragon           7.7

2                                             Iron Man 2           6.8

5                                              Inception           8.3

6      Percy Jackson & the Olympians: The Lightning T...           6.1

7                                                 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

      vote_count   id_y  release_date   internationalBoxOffice
1           7610   30.0  Mar 26, 2010             $494,870,992
2          12368   15.0   May 7, 2010             $621,156,389
5          22186   38.0  Jul 16, 2010             $835,524,642
6           4229   17.0  Feb 12, 2010             $223,050,874
7          18676    1.0  Dec 18, 2009           $2,776,345,279
...          ...    ...           ...                      ...
24538        231   64.0  Jun 15, 2018               $6,089,100
24575        259   50.0  Jan 12, 2018              $21,709,539
24597        156   20.0  Jan 22, 2019               $1,521,672
25388         54  100.0   Feb 2, 2018                 $648,599
26207          1   66.0   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)
df
```

```
                                                   movie
productionBudget_x  \
1                               How to Train Your Dragon
165000000.0
2                                             Iron Man 2
170000000.0
5                                              Inception
160000000.0
6      Percy Jackson & the Olympians: The Lightning T...
95000000.0
7                                                 Avatar
425000000.0
...                                                  ...
...
24538                                              Gotti
10000000.0
24575                                         Proud Mary
```

```
30000000.0
24597                                          Renegades
77500000.0
25388                     Bilal: A New Breed of Hero
30000000.0
26207                                            The Box
25000000.0

       domesticBoxoffice_x  internationalBoxOffice rating  \
1              217581232.0            2.772898e+08     PG
2              312433331.0            3.087231e+08  PG-13
5              292576195.0            5.398259e+08  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                  0.0            1.521672e+06  PG-13
25388             490973.0            1.576260e+05  PG-13
26207           15051977.0            1.128992e+07  PG-13

           production_method              genre  sequel  running_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


       Unnamed: 0            genre_ids  popularity releaseDate
vote_average  \
1               1  [14, 12, 16, 10751]      28.734  2010-03-26
7.7
2               2        [12, 28, 878]      28.515  2010-05-07
```

```
6.8
5                4        [28, 878, 12]      27.920  2010-07-16
8.3
6                5      [12, 14, 10751]      26.691  2010-02-11
6.1
7                6    [28, 12, 14, 878]      26.526  2009-12-18
7.4
...           ...                  ...         ...         ...
...
24538       24168     [80, 18, 36, 53]      10.034  2018-06-15
5.2
24575       24212         [53, 28, 80]       9.371  2018-01-12
5.5
24597       24239             [53, 28]       9.022  2018-12-21
5.8
25388       25148         [28, 12, 16]       2.707  2018-02-02
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           18676          $2,776,345,279
...           ...                     ...
24538         231              $6,089,100
24575         259             $21,709,539
24597         156              $1,521,672
25388          54                $648,599
26207           1             $34,356,760

[1110 rows x 16 columns]
```

```
df = df.drop(columns = ['Unnamed: 0'], axis = 1)
df
df.isnull().sum()
df = df.dropna()
df
```

```
                                                   movie
productionBudget_x  \
1                                How to Train Your Dragon
165000000.0
2                                              Iron Man 2
170000000.0
5                                               Inception
160000000.0
```

```
6      Percy Jackson & the Olympians: The Lightning T...
95000000.0
7                                                 Avatar
425000000.0
...                                                  ...
...
24538                                              Gotti
10000000.0
24575                                         Proud Mary
30000000.0
24597                                          Renegades
77500000.0
25388                         Bilal: A New Breed of Hero
30000000.0
26207                                            The Box
25000000.0

       domesticBoxoffice_x  internationalBoxOffice rating  \
1              217581232.0            2.772898e+08     PG
2              312433331.0            3.087231e+08  PG-13
5              292576195.0            5.398259e+08  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                  0.0            1.521672e+06  PG-13
25388             490973.0            1.576260e+05  PG-13
26207           15051977.0            1.128992e+07  PG-13

           production_method              genre  sequel  running_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

                 genre_ids  popularity releaseDate  vote_average
vote_count  \
1      [14, 12, 16, 10751]      28.734  2010-03-26           7.7
7610
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
...                    ...         ...         ...           ...
...
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                   []       0.840  2018-03-04           8.0
1

       internationalBoxOffice
1                $494,870,992
2                $621,156,389
5                $835,524,642
6                $223,050,874
7              $2,776,345,279
...                       ...
24538              $6,089,100
24575             $21,709,539
24597              $1,521,672
25388                $648,599
26207             $34,356,760

[1110 rows x 15 columns]
```

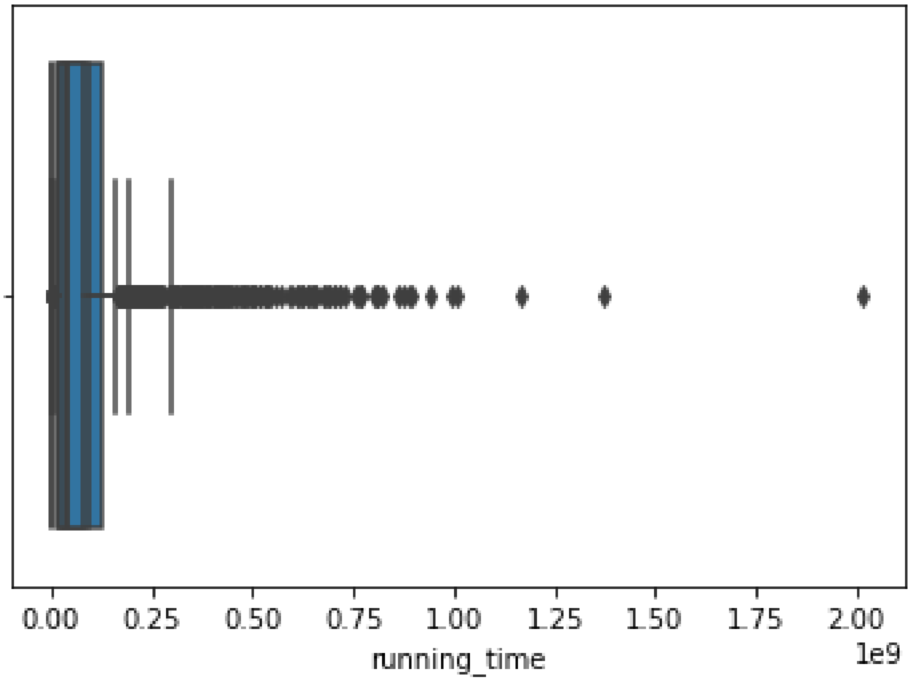
Exploratory Data Analysis (EDA)

- 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'])
```

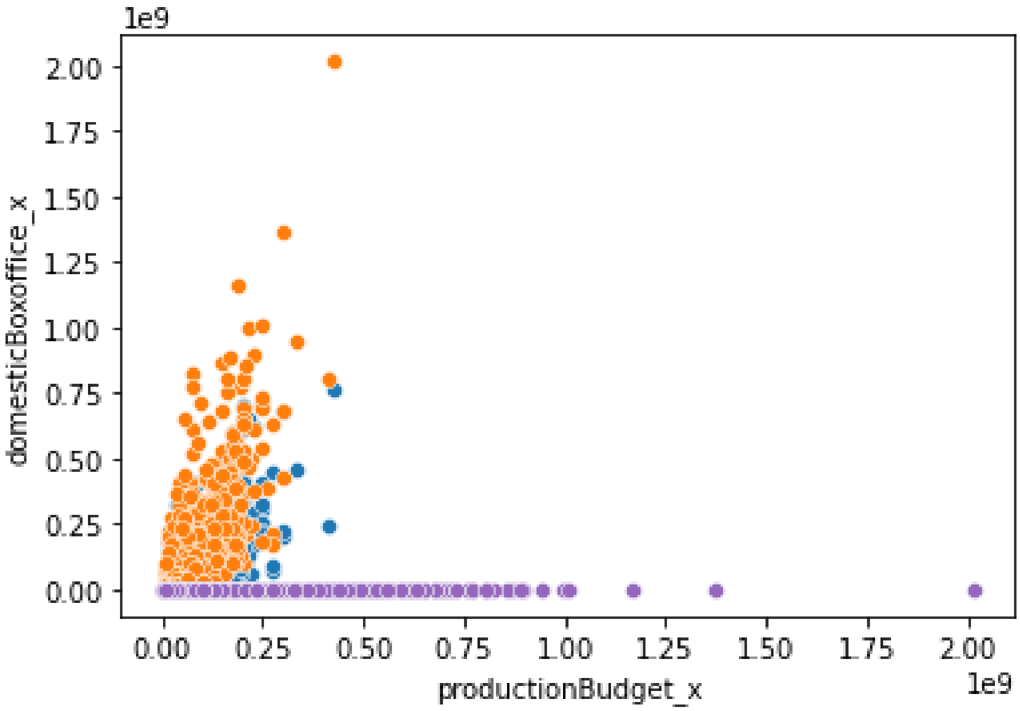
```
c:\Users\pc\anaconda3\envs\learn-env\lib\site-packages\seaborn\
_decorators.py:36: FutureWarning: Pass the following variable as a
keyword arg: x. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit
keyword will result in an error 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'])
```

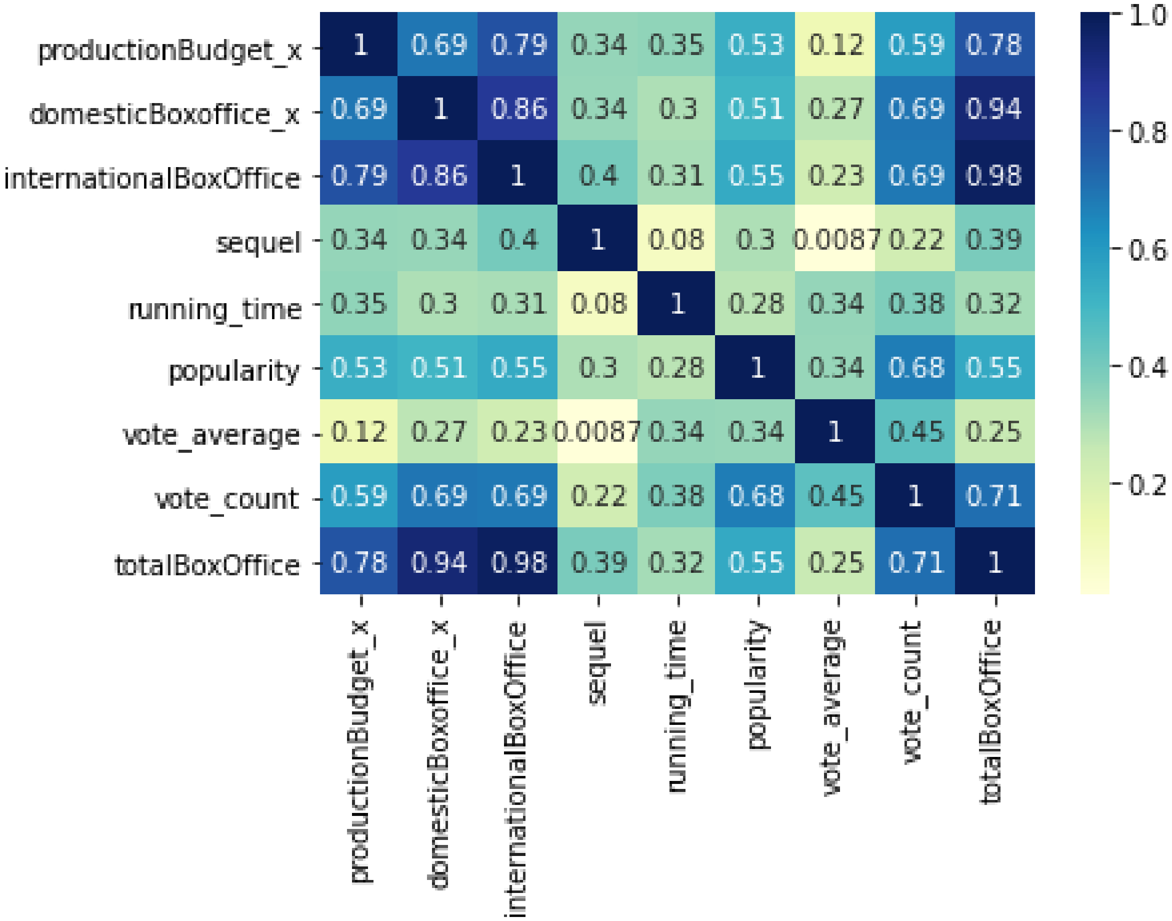
```
<AxesSubplot:xlabel='productionBudget_x',
ylabel='domesticBoxoffice_x'>
```



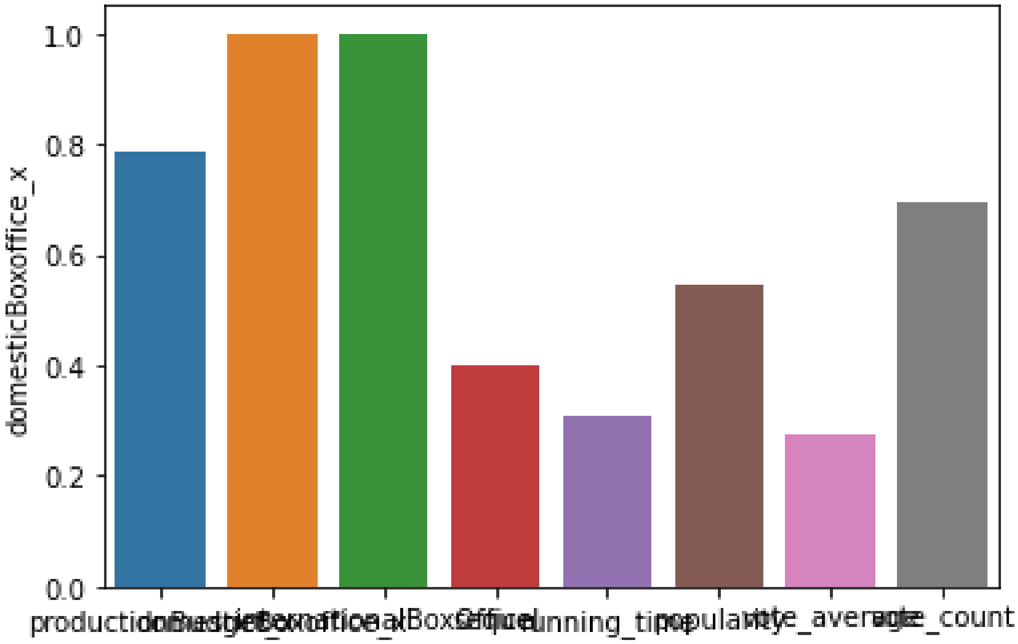
Correlation Analysis

- 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()
```

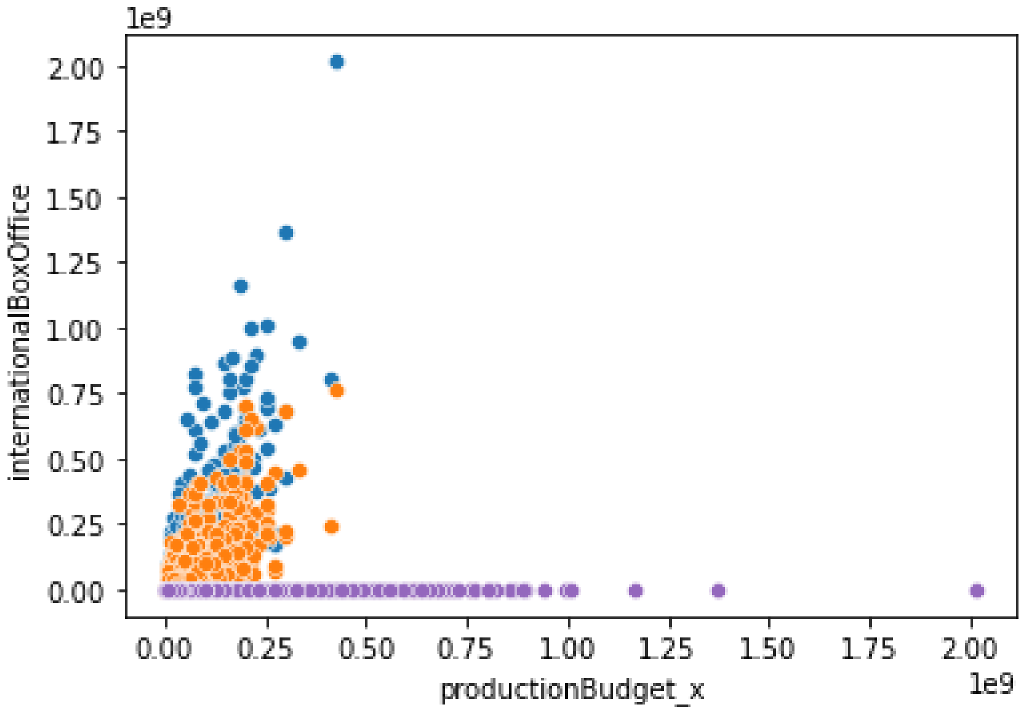


```
# bar chart of correlation

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'])
```

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



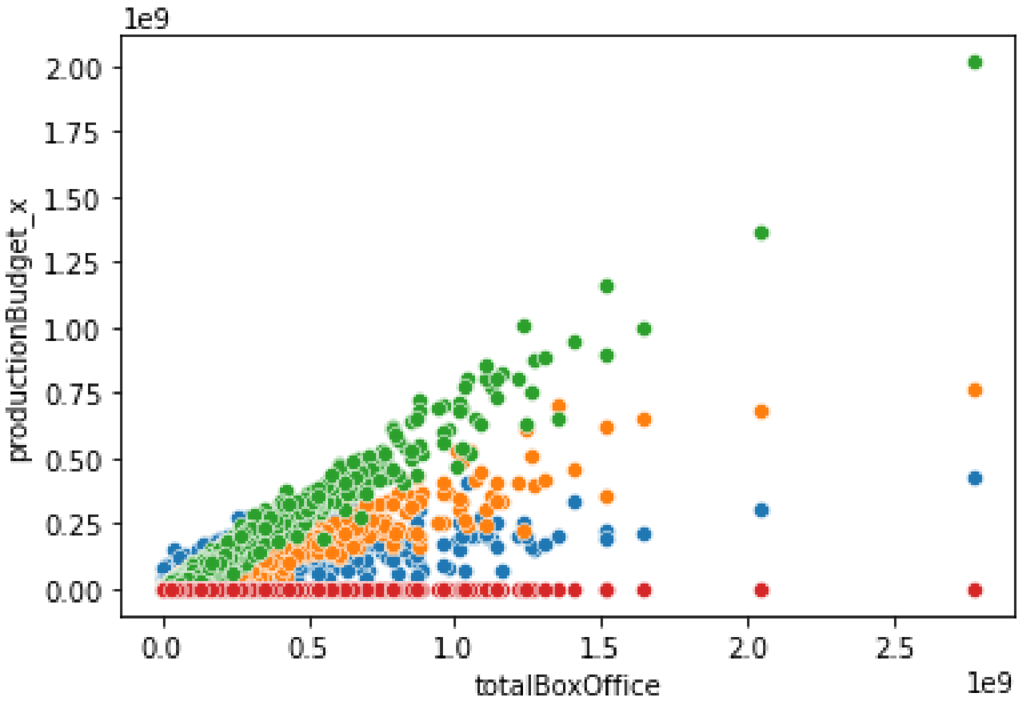
```
# missing values and duplicates

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
genre                      0
sequel                     0
running_time               0
genre_ids                  0
popularity                 0
releaseDate                0
vote_average               0
vote_count                 0
 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  \
1                               How to Train Your Dragon
165000000.0
2                                             Iron Man 2
170000000.0
5                                              Inception
160000000.0
6      Percy Jackson & the Olympians: The Lightning T...
95000000.0
7                                                 Avatar
425000000.0
...                                                  ...
...
24538                                              Gotti
10000000.0
24575                                         Proud Mary
30000000.0
24597                                          Renegades
77500000.0
25388                         Bilal: A New Breed of Hero
30000000.0
26207                                            The Box
25000000.0

       domesticBoxoffice_x  internationalBoxOffice rating  \
1             217581232.0            2.772898e+08     PG
2             312433331.0            3.087231e+08  PG-13
5             292576195.0            5.398259e+08  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                 0.0            1.521672e+06  PG-13
25388            490973.0            1.576260e+05  PG-13
26207          15051977.0            1.128992e+07  PG-13

           production_method              genre  sequel  running_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


                 genre_ids  popularity releaseDate  vote_average
vote_count  \
1      [14, 12, 16, 10751]      28.734  2010-03-26           7.7
7610
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
...                    ...         ...         ...           ...
...
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                   []       0.840  2018-03-04           8.0
1

      internationalBoxOffice  totalBoxOffice
1               $494,870,992    4.948710e+08
2               $621,156,389    6.211564e+08
5               $835,524,642    8.324021e+08
6               $223,050,874    2.230509e+08
7             $2,776,345,279    2.776345e+09
...                      ...             ...
24538             $6,089,100    6.089100e+06
```

```
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 x 16 columns]
```

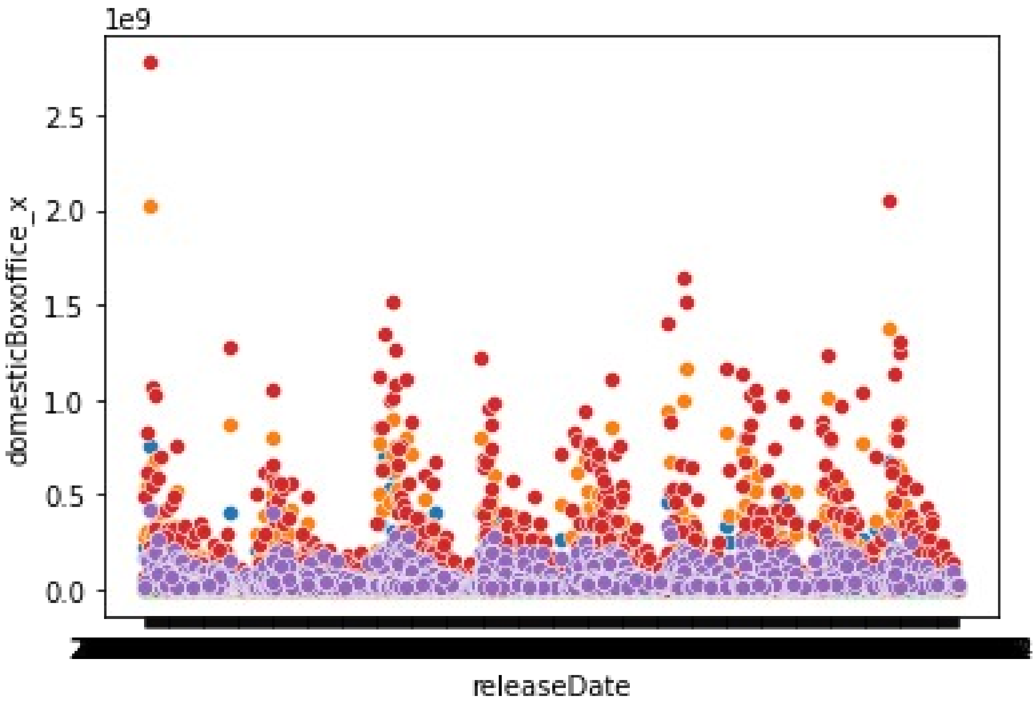
```
# check for correlation between total movie collection and other
variables

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

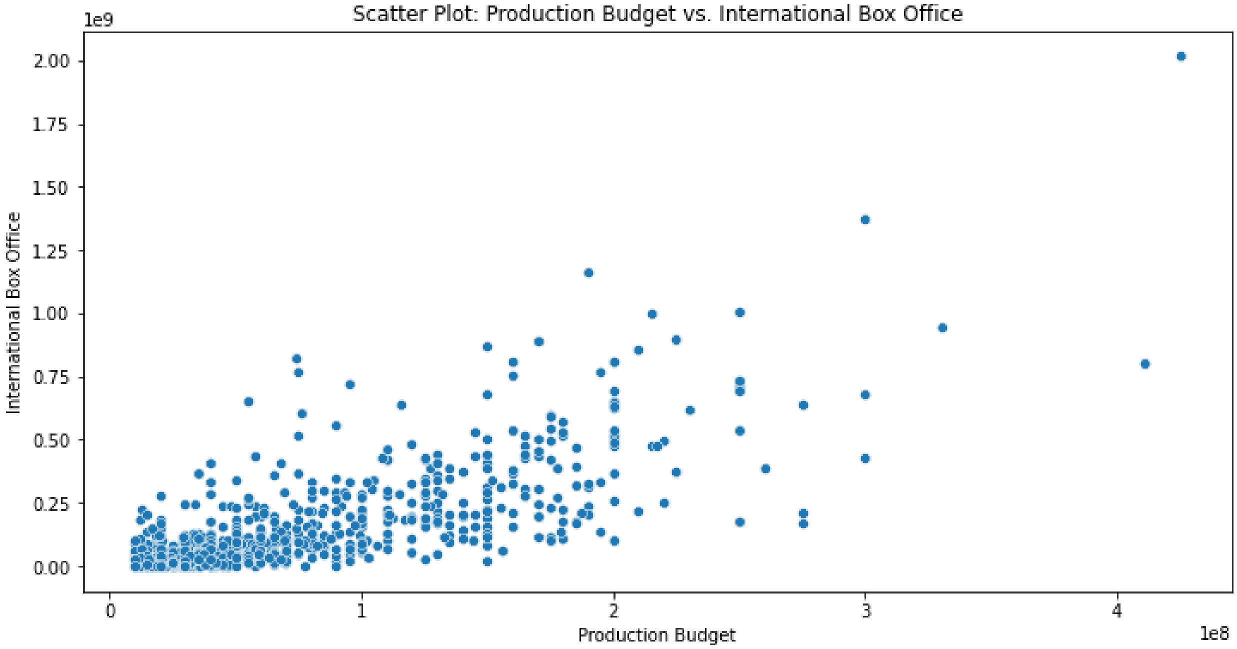


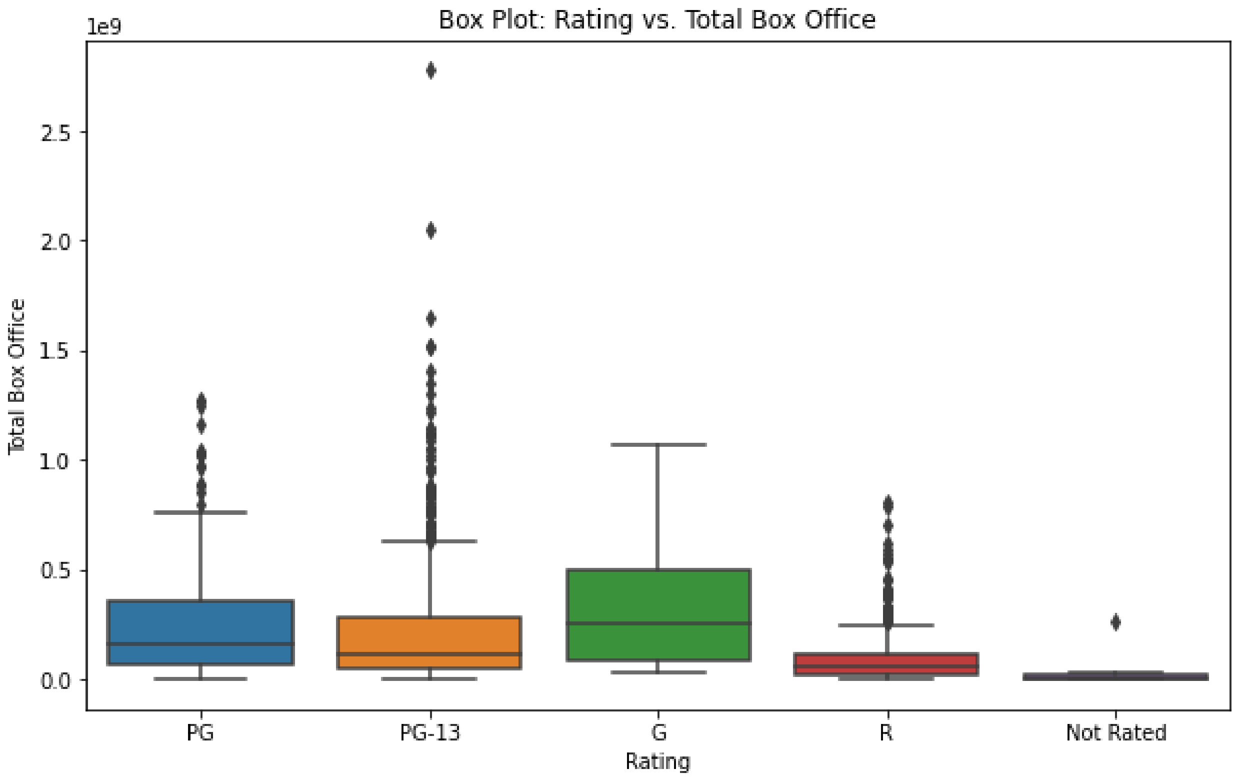
```
# we find out the months that had the movies with highest total box
office

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
```

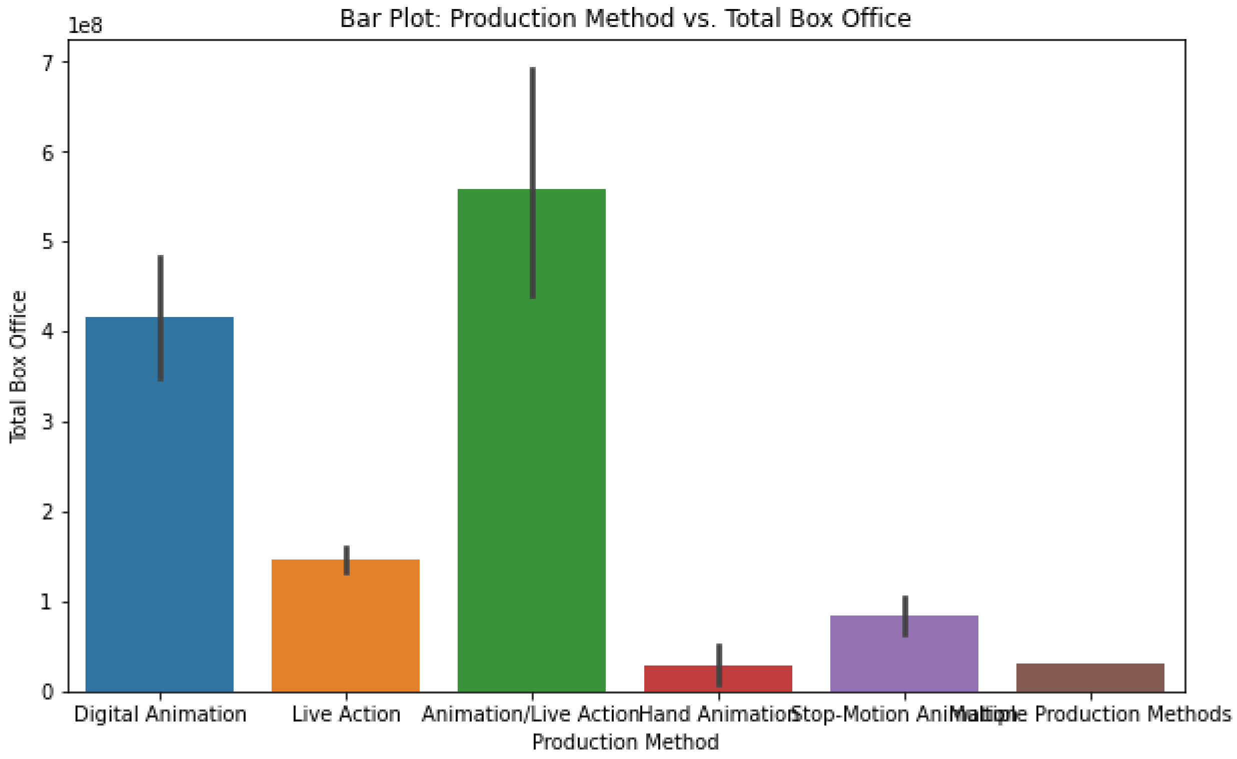
```
2013-05-07    7.370600e+04
Name: totalBoxOffice, Length: 633, dtype: float64
```

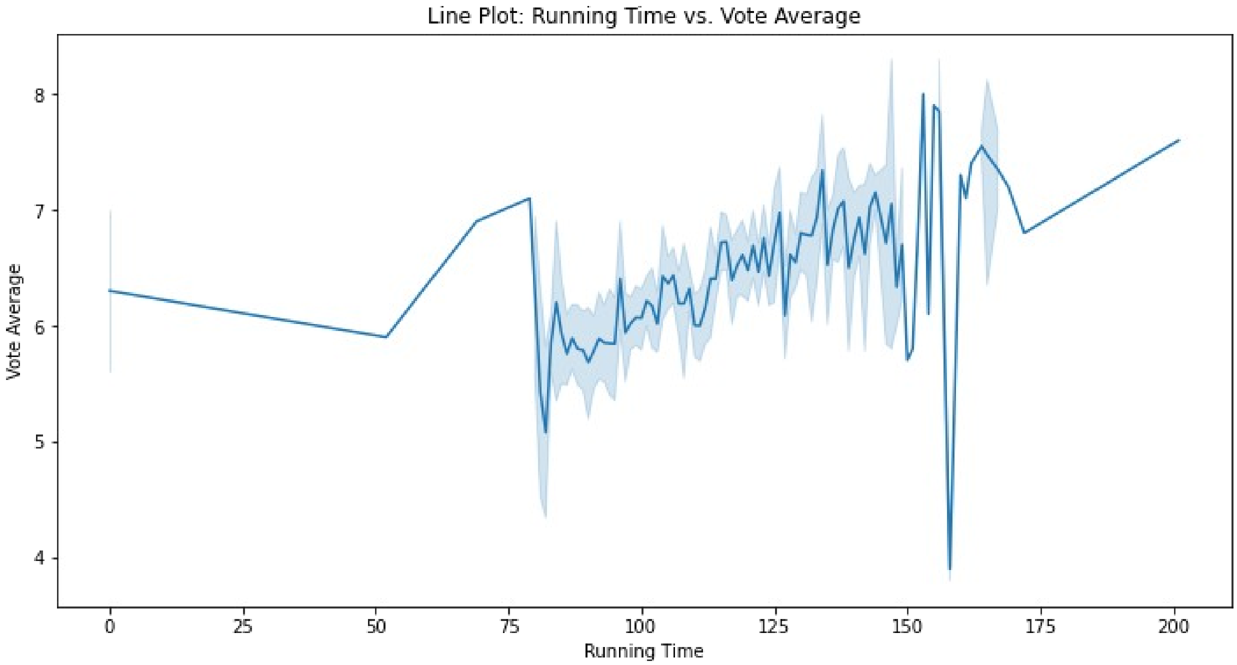
BUDGET ALLOCATION

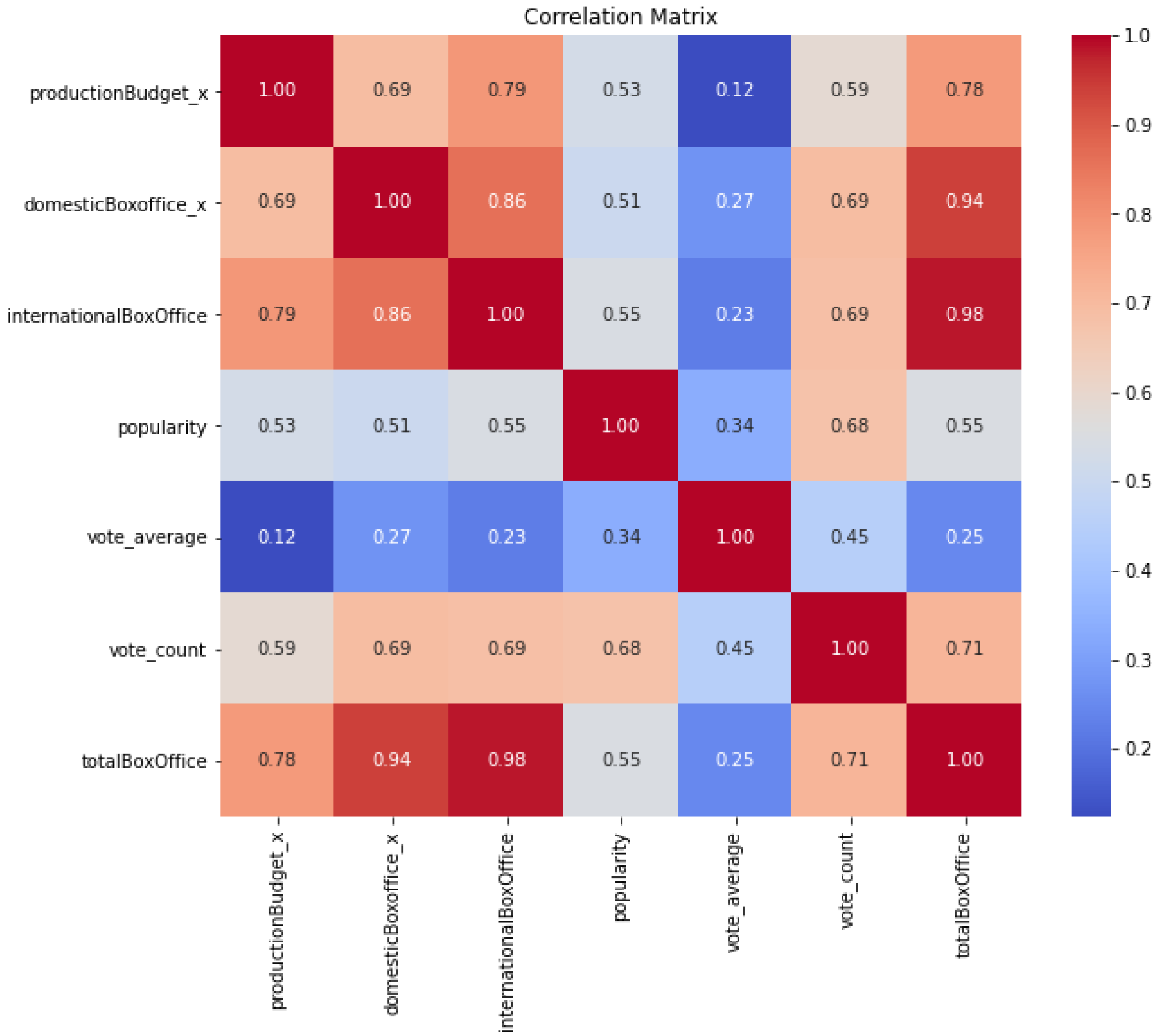
```
# budget allocation recommendation for the movie

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()
```




Box Plot: Rating vs. Total Box Office
1e9
2.5
2.0
1.5
1.0
0.5
0.0
Total Box Office
PG
PG-13
G
R
Not Rated
Rating


Bar Plot: Production Method vs. Total Box Office
1e8
7
6
5
4
3
2
1
0
Total Box Office
Digital Animation
Live Action
Animation/Live Action
Hand Animation
Stop-Motion Animation
Multiple Production Methods
Production Method


Line Plot: Running Time vs. Vote Average
Vote Average
4
5
6
7
8
Running Time
0
25
50
75
100
125
150
175
200



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.

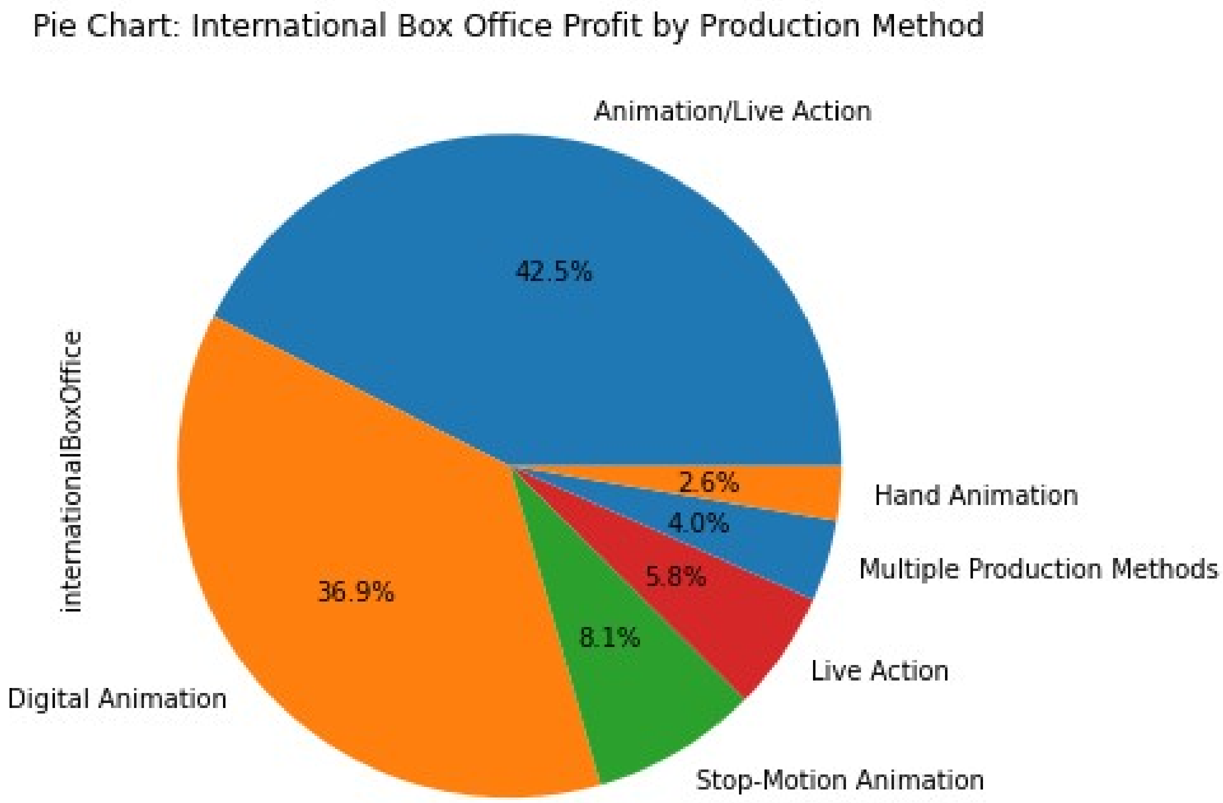
```
df.groupby('production_method')
['totalBoxOffice'].mean().sort_values(ascending = False)
df.groupby('production_method')
['productionBudget_x'].mean().sort_values(ascending = False)
df.groupby('production_method')
['domesticBoxoffice_x'].mean().sort_values(ascending = False)
df.groupby('production_method')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

```
['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
Multiple Production Methods    1.000000e+07
Name: productionBudget_x, dtype: float64

df.head()

                                               movie
productionBudget_x  \
1                           How to Train Your Dragon
165000000.0
2                                         Iron Man 2
170000000.0
5                                          Inception
160000000.0
6  Percy Jackson & the Olympians: The Lightning T...
95000000.0
7                                             Avatar
425000000.0

   domesticBoxoffice_x  internationalBoxOffice rating
production_method  \
1          217581232.0            2.772898e+08     PG       Digital
Animation
2          312433331.0            3.087231e+08  PG-13          Live
Action
5          292576195.0            5.398259e+08  PG-13  Animation/Live
Action
6           88768303.0            1.342826e+08     PG          Live
```

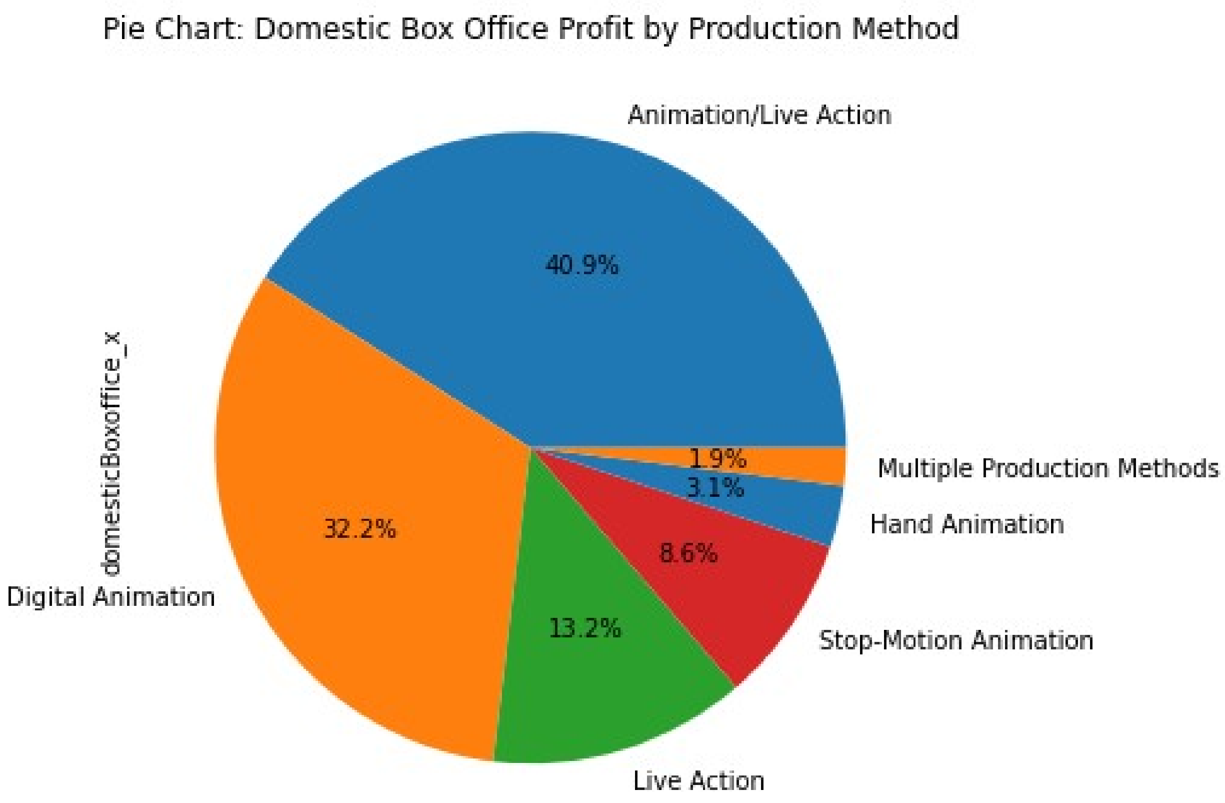
```
Action
7         760507625.0            2.015838e+09  PG-13  Animation/Live
Action

               genre  sequel  running_time            genre_ids
popularity  \
1          Adventure     0.0          91.0  [14, 12, 16, 10751]
28.734
2             Action     1.0         125.0        [12, 28, 878]
28.515
5  Thriller/Suspense     0.0         147.0        [28, 878, 12]
27.920
6          Adventure     0.0         119.0      [12, 14, 10751]
26.691
7             Action     0.0         162.0    [28, 12, 14, 878]
26.526

  releaseDate  vote_average  vote_count  internationalBoxOffice  \
1  2010-03-26           7.7        7610            $494,870,992
2  2010-05-07           6.8       12368            $621,156,389
5  2010-07-16           8.3       22186            $835,524,642
6  2010-02-11           6.1        4229            $223,050,874
7  2009-12-18           7.4       18676          $2,776,345,279

   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  \
24538                       Gotti          10000000.0
4286367.0
24575                  Proud Mary          30000000.0
20868638.0
24597                   Renegades          77500000.0
0.0
25388  Bilal: A New Breed of Hero          30000000.0
490973.0
26207                     The Box          25000000.0
15051977.0

       internationalBoxOffice rating  production_method
genre  \
24538               1802733.0      R        Live Action
Drama
```

```
24575                840901.0      R        Live Action
Action
24597               1521672.0  PG-13        Live Action
Action
25388                157626.0  PG-13  Digital Animation
Adventure
26207              11289919.0  PG-13        Live Action
Thriller/Suspense

       sequel  running_time         genre_ids  popularity releaseDate
\
24538     0.0         110.0  [80, 18, 36, 53]      10.034  2018-06-15

24575     0.0          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           5.2         231              $6,089,100
6089100.0
24575           5.5         259             $21,709,539
21709539.0
24597           5.8         156              $1,521,672
1521672.0
25388           6.8          54                $648,599
648599.0
26207           8.0           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()
```



```
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()
```



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 high-quality 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