

File Edit View Insert Cell Kernel Widgets Help

Run Cell Code

Explanatory Data Analysis & Data Presentation Using Movies Dataset

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
pd.options.display.max_columns = 30
pd.options.display.float_format = '{:.2f}'.format
```

```
In [2]: df = pd.read_csv("movies_complete.csv", parse_dates =["release_date"])
```

```
In [3]: df
```

| | id | title | tagline | release_date | genres | belongs_to_collection | original_language | budget_musd | revenue_musd | production_companies |
|-------|--------|-----------------------------|-----------------------------------------------------------|--------------|--------------------------|--------------------------------|-------------------|-------------|--------------|----------------------|
| 0 | 862 | Toy Story | NaN | 1995-10-30 | Animation Comedy Family | Toy Story Collection | en | 30.00 | 373.55 | Pixar Anim |
| 1 | 8844 | Jumanji | Roll the dice and unlock the excitement! | 1995-12-15 | Adventure Fantasy Family | NaN | en | 65.00 | 262.80 | TriStar Pi Fil |
| 2 | 15602 | Grumpier Old Men | Still Yelling, Still Fighting, Still Ready for... | 1995-12-22 | Romance Comedy | Grumpy Old Men Collection | en | NaN | NaN | Warner Brt |
| 3 | 31357 | Waiting to Exhale | Friends are the people who let you be yourself... | 1995-12-22 | Comedy Drama Romance | NaN | en | 16.00 | 81.45 | Twentieth Film |
| 4 | 11862 | Father of the Bride Part II | Just When You Thought It's Back To Normal, Normal He's... | 1995-02-10 | Comedy | Father of the Bride Collection | en | NaN | 76.58 | Production |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 44686 | 439050 | Subdue | Rising and falling between a man and woman | NaN | Drama Family | NaN | fa | NaN | NaN | |
| 44687 | 111109 | Century of Birthing | NaN | 2011-11-17 | Drama | NaN | ti | NaN | NaN | |
| 44688 | 67758 | Betrayal | A deadly game of wits. | 2003-08-01 | Action Drama Thriller | NaN | en | NaN | NaN | American W |
| 44689 | 227506 | Satan Triumphant | NaN | 1917-10-21 | NaN | NaN | en | NaN | NaN | |
| 44690 | 461257 | Queerama | NaN | 2017-06-09 | NaN | NaN | en | NaN | NaN | |

44691 rows x 22 columns

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44691 entries, 0 to 44690
Data columns (total 22 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   id               44691 non-null   int64  
 1   title             44691 non-null   object 
 2   tagline           20284 non-null   object 
 3   release_date      44657 non-null   datetime64[ns]
 4   genres            42586 non-null   object 
 5   belongs_to_collection 4463 non-null   object 
 6   original_language 44681 non-null   object 
 7   budget_musd       8854 non-null   float64 
 8   revenue_musd      73356 non-null   float64 
 9   production_companies 30835 non-null   object 
 10  production_countries 44691 non-null   float64 
 11  vote_count         44691 non-null   float64 
 12  vote_average        42077 non-null   float64 
 13  popularity          44691 non-null   float64 
 14  runtime             43179 non-null   float64 
 15  overview            43740 non-null   object 
 16  spoken_languages    41094 non-null   object 
 17  poster_path          44467 non-null   object 
 18  cast                42502 non-null   object 
 19  cast_size            44691 non-null   int64  
 20  crew_size            44691 non-null   int64  
 21  director             43960 non-null   object 
dtypes: datetime64[ns](1), float64(6), int64(3), object(12)
memory usage: 7.5+ MB
```

```
In [5]: df.genres[1]
```

```
Out[5]: 'Adventure|Fantasy|Family'
```

```
In [6]: df.cast[1]
```

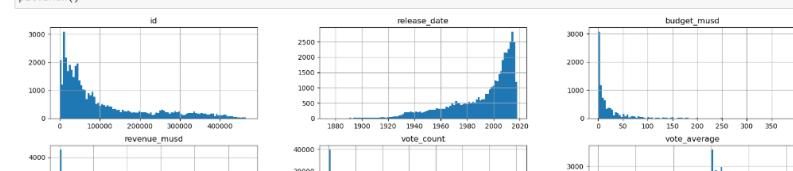
```
Out[6]: 'Robin Williams|Jonathan Hyde|Kirsten Dunst|Bradley Pierce|Bonnie Hunt|Bebe Neuwirth|David Alan Grier|Patricia Clarkson|Adam Hann-Byrd|Laura Bell Bundy|James Handy|Gillian Barber|Brandon Obray|Cyrus Thiedeke|Gary Joseph Thorup|Leonard Zola|Lloyd Berry|Macolm Stewart|Annabel Kershaw|Darryl Henriques|Robyn Driscoll|Peter Bryant|Sarah Gilson|Florica Vlad|June Lion|Brenda Lockmuller'
```

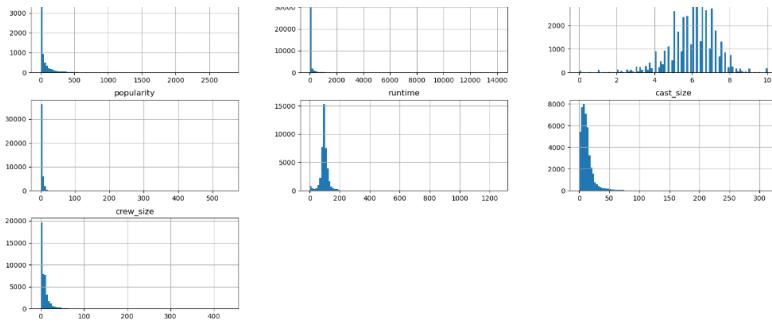
```
In [7]: df.describe()
```

```
Out[7]:
```

| | id | budget_musd | revenue_musd | vote_count | vote_average | popularity | runtime | cast_size | crew_size |
|-------|-----------|-------------|--------------|------------|--------------|------------|----------|-----------|-----------|
| count | 44691.00 | 8854.00 | 7385.00 | 44691.00 | 42077.00 | 44691.00 | 43179.00 | 44691.00 | 44691.00 |
| mean | 107186.24 | 21.67 | 68.97 | 111.65 | 6.00 | 2.96 | 97.57 | 12.48 | 10.31 |
| std | 111806.36 | 34.36 | 146.61 | 495.32 | 1.28 | 6.04 | 34.65 | 12.12 | 15.89 |
| min | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 25% | 26033.50 | 2.00 | 2.41 | 3.00 | 5.30 | 0.40 | 86.00 | 6.00 | 2.00 |
| 50% | 59110.00 | 8.20 | 16.87 | 10.00 | 6.10 | 1.15 | 95.00 | 10.00 | 6.00 |
| 75% | 154251.00 | 25.00 | 67.64 | 35.00 | 6.80 | 3.77 | 107.00 | 15.00 | 12.00 |
| max | 469172.00 | 380.00 | 2787.97 | 14075.00 | 10.00 | 547.49 | 1256.00 | 313.00 | 435.00 |

```
In [8]: df.hist(figsize = (20, 12), bins=100)
plt.show()
```





```
In [9]: df.budget_musd.value_counts(dropna = False).head(20)
```

```
Out[9]:
```

| | budget_musd |
|-------|-------------|
| NaN | 35837 |
| 5.00 | 286 |
| 10.00 | 258 |
| 20.00 | 243 |
| 2.00 | 241 |
| 15.00 | 226 |
| 3.00 | 220 |
| 25.00 | 206 |
| 1.00 | 195 |
| 30.00 | 189 |
| 4.00 | 188 |
| 6.00 | 173 |
| 12.00 | 171 |
| 40.00 | 166 |
| 8.00 | 155 |
| 35.00 | 141 |
| 0.50 | 141 |
| 7.00 | 137 |
| 58.00 | 124 |
| 1.50 | 119 |

Name: budget_musd, dtype: int64

```
In [10]: df.revenue_musd.value_counts(dropna = False).head(20)
```

```
Out[10]:
```

| | revenue_musd |
|-------|--------------|
| NaN | 37306 |
| 12.00 | 20 |
| 10.00 | 19 |
| 11.00 | 19 |
| 2.00 | 18 |
| 6.00 | 17 |
| 5.00 | 14 |
| 0.50 | 13 |
| 8.00 | 13 |
| 0.00 | 12 |
| 14.00 | 12 |
| 7.00 | 11 |
| 3.00 | 10 |
| 1.00 | 10 |
| 20.00 | 10 |
| 0.00 | 9 |
| 1.50 | 9 |
| 4.00 | 9 |
| 30.00 | 8 |
| 4.10 | 8 |

Name: revenue_musd, dtype: int64

```
In [11]: df.vote_average.value_counts()
```

```
Out[11]:
```

| | vote_average |
|------|--------------|
| 6.00 | 2421 |
| 5.00 | 1969 |
| 7.00 | 1859 |
| 6.50 | 1705 |
| 6.30 | 1583 |
| | ... |
| 9.40 | 3 |
| 9.60 | 1 |
| 9.80 | 1 |
| 0.70 | 1 |
| 1.10 | 1 |

Name: vote_average, Length: 92, dtype: int64

```
In [12]: df.vote_count.value_counts()
```

```
Out[12]:
```

| | vote_count |
|---------|------------|
| 1.00 | 3186 |
| 2.00 | 3056 |
| 3.00 | 2729 |
| 0.00 | 2614 |
| 4.00 | 2442 |
| | ... |
| 2755.00 | 1 |
| 1187.00 | 1 |
| 4200.00 | 1 |
| 3322.00 | 1 |
| 2712.00 | 1 |

Name: vote_count, Length: 1820, dtype: int64

```
In [13]: df.describe(include = "object")
```

```
Out[13]:
```

| | title | tagline | genres | belongs_to_collection | original_language | production_companies | production_countries | overview | spoken_languages |
|--------|------------|------------------------|--------|-----------------------|-------------------|---------------------------|--------------------------|--------------------------------------|---------------------|
| count | 44691 | 20284 | 42586 | | 44681 | 33356 | 38835 | 43740 | 41094 |
| unique | 41605 | 20171 | 4044 | | 1691 | 89 | 22537 | 2377 | 43715 |
| top | Cinderella | Based on a true story. | Drama | The Bowery Boys | en | Metro-Goldwyn-Mayer (MGM) | United States of America | Adaptation of the Jane Austen novel. | English src=http:// |
| freq | 11 | 7 | 4935 | | 29 | 31755 | 741 | 17723 | 3 |

```
In [14]: df[df.title=="Cinderella"]
```

```
Out[14]:
```

| | id | title | tagline | release_date | genres | belongs_to_collection | original_language | budget_musd | revenue_musd | popularity |
|-------|--------|------------|------------------------------------|--------------|----------------------------------|------------------------------------|-------------------|-------------|--------------|------------|
| 984 | 11224 | Cinderella | The greatest love story ever told. | 1950-03-04 | Family Fantasy Animation Romance | Cinderella Collection | en | 2.90 | 263.59 | 3000 |
| 12988 | 42884 | Cinderella | | NaN | 1997-11-02 | Movie Family Fantasy Music Romance | en | 2.90 | 263.59 | 3000 |
| 23254 | 92349 | Cinderella | | NaN | 1914-12-28 | Fantasy Drama | en | 2.90 | 263.59 | 3000 |
| 23265 | 105875 | Cinderella | The version children level | 2002-08-06 | Animation Family Fantasy | | en | 2.90 | 263.59 | 3000 |
| 28073 | 261985 | Cinderella | | NaN | 2011-10-30 | Family TV Movie | en | 2.90 | 263.59 | 3000 |
| 28340 | 150689 | Cinderella | Midnight is just the | 2015-03-12 | Romance Fantasy Family Drama | | en | 95.00 | 543.51 | 3000 |

beginning

| 33805 | 42651 | Cinderella | NaN | 1947-11-28 | Comedy Family Fantasy | NaN | ru | NaN | NaN |
|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|-----------------------------|-------|------------|---------------------------------------|------|-------|-----|-----|
| 35114 | 44459 | Cinderella | NaN | 1957-03-31 | Drama Romance | NaN | en | NaN | NaN |
| 35116 | 289673 | Cinderella | NaN | 2000-01-01 | NaN | NaN | en | NaN | NaN |
| 40439 | 114108 | Cinderella | NaN | 1899-10-01 | Fantasy Horror Science Fiction Family | NaN | fr | NaN | NaN |
| 44633 | 98604 | Cinderella | NaN | 2012-02-14 | Comedy Romance | NaN | ru | NaN | NaN |
| | | | | | | | | | |
| from IPython.display import HTML | | | | | | | | | |
| df_best = df[["poster_path","title","budget_musd", "revenue_musd","vote_count", "vote_average","popularity"]].copy() | | | | | | | | | |
| df_best | | | | | | | | | |
| | | | | | | | | | |
| poster_path | | | | | | | | | |
| 0 | | Toy Story | 30.00 | 373.55 | 5415.00 | 7.70 | 21.95 | 343.55 | 12.45 |
| 1 | | Grimpier Old Men | NaN | NaN | 92.00 | 6.50 | 11.71 | NaN | NaN |
| 3 | | Father of the Bride Part II | NaN | 76.58 | 173.00 | 5.70 | 8.39 | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 44686 | | Century of Birthing | NaN | NaN | 3.00 | 9.00 | 0.18 | NaN | NaN |
| 44688 | | Satan Triumphant | NaN | NaN | 0.00 | NaN | 0.00 | NaN | NaN |
| 44690 | | 30.00 | 373.55 | 5415.00 | 7.70 | 21.95 | 343.55 |
| | Jumanji | | NaN | NaN | 92.00 | 6.50 | 11.71 | NaN |
| | Waiting to Exhale | | NaN | 76.58 | 173.00 | 5.70 | 8.39 | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | Subdue | | NaN | NaN | 3.00 | 9.00 | 0.18 | NaN |
| | Betrayal | | NaN | NaN | 0.00 | NaN | 0.00 | NaN |
| | Queerama |
```

```
In [24]: subset = df_best.iloc[:5, :2]
subset
```

Out[24]:

| Title                       | Budget       |
|-----------------------------|--------------|
| Toy Story                   | \$30,000,000 |
| Jumanji                     | \$65,000,000 |
| Grumpier Old Men            | NaN          |
| Waiting to Exhale           | \$10,000,000 |
| Father of the Bride Part II | \$10,000,000 |

```
In [25]: HTML(subset_to_html(escape=False))
```

Out[25]:

Budget

|                             |  |       |
|-----------------------------|--|-------|
| Toy Story                   |  | 30.00 |
| Jumanji                     |  | 65.00 |
| Grumpier Old Men            |  | NaN   |
| Waiting to Exhale           |  | 16.00 |
| Father of the Bride Part II |  | NaN   |

```
In [26]: df_best.sort_values(by = "Average Rating", ascending = False)
```

```
Out[26]:
```

| Title                                      | Budget                                              | Revenue | Votes | Average Rating | Popularity | Profit | ROI |
|--------------------------------------------|-----------------------------------------------------|---------|-------|----------------|------------|--------|-----|
| Portrait of a Young Man in Three Movements | NaN                                                 | NaN     | NaN   | 1.00           | 10.00      | 0.04   | NaN |
| Brave Revolutionary                        | <img src='http://image.tmdbs.org/t/p/w185/lzAb2...' | NaN     | NaN   | 1.00           | 10.00      | 0.32   | NaN |
| Other Voices Other Rooms                   | <img src='http://image.tmdbs.org/t/p/w185/4fFp...'  | NaN     | NaN   | 1.00           | 10.00      | 0.04   | NaN |
| The Lion of Thebes                         | <img src='http://image.tmdbs.org/t/p/w185/ldOc...'  | NaN     | NaN   | 1.00           | 10.00      | 1.78   | NaN |
| Katt Williams: Priceless: Afterlife        | <img src='http://image.tmdbs.org/t/p/w185/wkHr...'  | NaN     | NaN   | 2.00           | 10.00      | 0.48   | NaN |
| ...                                        | ...                                                 | ...     | ...   | ...            | ...        | ...    | ... |
| Altar of Fire                              | <img src='http://image.tmdbs.org/t/p/w185/lj78...'  | NaN     | NaN   | 0.00           | NaN        | 0.00   | NaN |
| The Wonders of Aladdin                     | <img src='http://image.tmdbs.org/t/p/w185/AvfX...'  | NaN     | NaN   | 0.00           | NaN        | 0.09   | NaN |
| Deep Hearts                                | <img src='http://image.tmdbs.org/t/p/w185/l4J4...'  | NaN     | NaN   | 0.00           | NaN        | 0.01   | NaN |
| Satan Triumphant                           | <img src='http://image.tmdbs.org/t/p/w185/aoB...'   | NaN     | NaN   | 0.00           | NaN        | 0.00   | NaN |
| Queerama                                   | <img src='http://image.tmdbs.org/t/p/w185/oxFE...'  | NaN     | NaN   | 0.00           | NaN        | 0.16   | NaN |

44691 rows x 8 columns

```
In [27]: df_best.sort_values(by = "ROI", ascending = False)
```

```
Out[27]:
```

| Title                   | Budget                                             | Revenue | Votes  | Average Rating | Popularity | Profit | ROI    |
|-------------------------|----------------------------------------------------|---------|--------|----------------|------------|--------|--------|
| Less Than Zero          | <img src='http://image.tmdbs.org/t/p/w185/1GY0...' | 0.00    | 12.40  | 77.00          | 6.10       | 4.03   | 12.40  |
| Modern Times            | <img src='http://image.tmdbs.org/t/p/w185/7uoI...' | 0.00    | 8.50   | 881.00         | 8.10       | 8.16   | 8.50   |
| Welcome to Dongmakgol   | <img src='http://image.tmdbs.org/t/p/w185/5iGV...' | 0.00    | 33.58  | 49.00          | 7.70       | 4.22   | 33.58  |
| Aqui Entre Nos          | <img src='http://image.tmdbs.org/t/p/w185/0ftx...' | 0.00    | 2.76   | 3.00           | 6.00       | 0.23   | 2.76   |
| The Karate Kid, Part II | <img src='http://image.tmdbs.org/t/p/w185/mSne...' | 0.00    | 115.10 | 457.00         | 5.90       | 9.23   | 115.10 |
| ...                     | ...                                                | ...     | ...    | ...            | ...        | ...    | ...    |
| Subdue                  | <img src='http://image.tmdbs.org/t/p/w185/pfC8...' | NaN     | NaN    | 1.00           | 4.00       | 0.07   | NaN    |
| Century of Birth        | <img src='http://image.tmdbs.org/t/p/w185/zXkm...' | NaN     | NaN    | 3.00           | 9.00       | 0.18   | NaN    |
| Betrayal                | <img src='http://image.tmdbs.org/t/p/w185/eGga...' | NaN     | NaN    | 6.00           | 3.80       | 0.90   | NaN    |
| Satan Triumphant        | <img src='http://image.tmdbs.org/t/p/w185/aoB...'  | NaN     | NaN    | 0.00           | NaN        | 0.00   | NaN    |
| Queerama                | <img src='http://image.tmdbs.org/t/p/w185/oxFE...' | NaN     | NaN    | 0.00           | NaN        | 0.16   | NaN    |

44691 rows x 8 columns

```
In [28]: df_best.loc[df_best.Budget >= 5].sort_values(by = "ROI", ascending = False)
```

```
Out[28]:
```

| Title                                       | Budget                                             | Revenue | Votes  | Average Rating | Popularity | Profit | ROI    |
|---------------------------------------------|----------------------------------------------------|---------|--------|----------------|------------|--------|--------|
| E.T. the Extra-Terrestrial                  | <img src='http://image.tmdbs.org/t/p/w185/cBfk...' | 10.50   | 792.97 | 3359.00        | 7.30       | 19.36  | 782.47 |
| My Big Fat Greek Wedding                    | <img src='http://image.tmdbs.org/t/p/w185/3TB2...' | 5.00    | 368.74 | 686.00         | 6.20       | 6.72   | 363.74 |
| Star Wars                                   | <img src='http://image.tmdbs.org/t/p/w185/6fIC...' | 11.00   | 775.40 | 6778.00        | 8.10       | 42.15  | 764.40 |
| Jaws                                        | <img src='http://image.tmdbs.org/t/p/w185/s2xc...' | 7.00    | 470.65 | 2628.00        | 7.50       | 19.73  | 463.65 |
| Crocodile Dundee                            | <img src='http://image.tmdbs.org/t/p/w185/kwO...'  | 5.00    | 328.20 | 512.00         | 6.30       | 7.79   | 323.20 |
| ...                                         | ...                                                | ...     | ...    | ...            | ...        | ...    | ...    |
| The Blonde with Bare Breasts                | <img src='http://image.tmdbs.org/t/p/w185/zBjp...' | 7.50    | NaN    | 3.00           | 4.00       | 0.88   | NaN    |
| Detroit                                     | <img src='http://image.tmdbs.org/t/p/w185/7APL...' | 34.00   | NaN    | 67.00          | 7.30       | 9.80   | NaN    |
| Good Guys Go to Heaven, Bad Guys Go to Hell | <img src='http://image.tmdbs.org/t/p/w185/cfVB...' | 5.40    | NaN    | 153.00         | 5.30       | 5.61   | NaN    |
| The Visitors: Bastille Day                  | <img src='http://image.tmdbs.org/t/p/w185/kBm...'  | 25.87   | NaN    | 167.00         | 4.00       | 7.29   | NaN    |
| House of the Long Shadows                   | <img src='http://image.tmdbs.org/t/p/w185/21ZG...' | 7.50    | NaN    | 18.00          | 6.10       | 1.19   | NaN    |

5449 rows x 8 columns

```
In [29]: df_best.Budget.fillna(0, inplace = True)
```

```
df_best.Votes.fillna(0, inplace = True)
```

```
In [30]: df_best.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 44691 entries, Toy Story to Queerama
Data columns (total 8 columns):
 # Column Non-Null Count Dtype

 0 Budget 44691 non-null float64
 1 Revenue 7385 non-null float64
 2 Votes 44691 non-null float64
 3 Average Rating 42077 non-null float64
 4 Popularity 44691 non-null float64
 5 Profit 5371 non-null float64
 6 ROI 5371 non-null float64
 7 dTypes: float64(7), object(1)
memory usage: 3.1+ MB
```

```
In [31]: def best_worst(n, by, ascending = False, min_bud = 0, min_votes = 0):
 df2 = df_best.loc[(df_best.Budget >= min_bud) & (df_best.Votes >= min_votes),[" ",by]].sort_values(by = by, ascending = ascending)
 return HTML(df2.to_html(escape=False))
```

Movies Top 5 - Highest Revenue

```
In [32]: best_worst(n = 5, bv = "Revenue")
```

Out[32]:

| Revenue |                              |         |
|---------|------------------------------|---------|
|         | Title                        |         |
|         | Avatar                       | 2787.97 |
|         | Star Wars: The Force Awakens | 2068.22 |
|         | Titanic                      | 1845.03 |
|         | The Avengers                 | 1519.56 |
|         | Jurassic World               | 1513.53 |

Movies Top 5 - Highest Profit

In [33]: `best_worst(5, "Profit")`

Out[33]:

| Profit |                              |         |
|--------|------------------------------|---------|
|        | Title                        |         |
|        | Avatar                       | 2550.97 |
|        | Star Wars: The Force Awakens | 1823.22 |
|        | Titanic                      | 1645.03 |
|        | Jurassic World               | 1363.53 |
|        | Furious 7                    | 1316.25 |

Movies Top 5 - Highest Budget

In [34]: `best_worst(5, "Budget")`

Out[34]:

| Budget |                                             |        |
|--------|---------------------------------------------|--------|
|        | Title                                       |        |
|        | Pirates of the Caribbean: On Stranger Tides | 380.00 |
|        | Pirates of the Caribbean: At World's End    | 300.00 |
|        | Avengers: Age of Ultron                     | 280.00 |
|        | Superman Returns                            | 270.00 |
|        | Transformers: The Last Knight               | 260.00 |

Movies Top 5 - Lowest Profit

In [35]: `best_worst(5, "Profit", ascending = True)`

Out[35]:

| Profit |                 |         |
|--------|-----------------|---------|
|        | Title           |         |
|        | The Lone Ranger | -165.71 |
|        | The Alamo       | -119.18 |
|        | Mars Needs Moms | -111.01 |

|  |                                             |         |
|--|---------------------------------------------|---------|
|  | Valerian and the City of a Thousand Planets | -107.45 |
|  | The 13th Warrior                            | -98.30  |

Movies Top 5 - Highest ROI

```
In [36]: best_worst(5, "ROI", min_bud = 0)
```

```
Out[36]: ROI
```

| Title | ROI                                |
|-------|------------------------------------|
|       | Less Than Zero 12396383.00         |
|       | Modern Times 8500000.00            |
|       | Welcome to Dongmakgol 4197476.62   |
|       | Aquí Entre Nos 2755584.00          |
|       | The Karate Kid, Part II 1018619.28 |

Moviest Top 5 - Lowest ROI

```
In [37]: best_worst(5, "ROI", ascending = True, min_bud = 1)
```

```
Out[37]: ROI
```

| Title | ROI                          |
|-------|------------------------------|
|       | Chasing Liberty 0.00         |
|       | The Cookout 0.00             |
|       | Never Talk to Strangers 0.00 |
|       | To Rob a Thief 0.00          |
|       | Deadfall 0.00                |

Moviest Top 5 - Most Votes

```
In [38]: best_worst(5, "Votes")
```

```
Out[38]: Votes
```

| Title | Votes                    |
|-------|--------------------------|
|       | Inception 14075.00       |
|       | The Dark Knight 12269.00 |
|       | Avatar 12114.00          |
|       | The Avengers 12000.00    |
|       | Deadpool 11444.00        |

Movies Top 5 - Highest Rating

```
In [39]: best_worst(5, "Average_Rating", min_votes = 100)
```

```
Out[39]: Average_Rating
```

```
Title
```

|                             |  |      |
|-----------------------------|--|------|
| Dilwale Dulhania Le Jayenge |  | 9.10 |
| Planet Earth                |  | 8.80 |
| Your Name.                  |  | 8.50 |
| The Shawshank Redemption    |  | 8.50 |
| The Godfather               |  | 8.50 |

Movies Top 5 - Lowest Rating

```
In [40]: best_worst(5, "Average Rating", ascending = True, min_votes = 100)
```

```
Out[40]:
```

| Title                |  |      |
|----------------------|--|------|
| House of the Dead    |  | 2.80 |
| Dragonball Evolution |  | 2.90 |
| Battlefield Earth    |  | 3.00 |
| Disaster Movie       |  | 3.10 |
| Alone in the Dark    |  | 3.10 |

Movies Top 5 - Popularity

```
In [41]: best_worst(5, "Popularity")
```

```
Out[41]:
```

| Title                |  |        |
|----------------------|--|--------|
| Minions              |  | 547.49 |
| Wonder Woman         |  | 294.34 |
| Beauty and the Beast |  | 287.25 |
| Baby Driver          |  | 228.03 |
| Big Hero 6           |  | 213.85 |

Find your next movie

Search 1: Sciene Fiction Action Movie with Bruce Willis (High Rating)

```
In [42]: df.genres[0]
```

```
Out[42]: 'Animation|Comedy|Family'
```

```
In [43]: mask_genres = df.genres.str.contains("Action") & df.genres.str.contains("Science Fiction")
```

```
mask_genres
```

```
Out[43]: 0 False
1 False
2 False
3 False
4 False
...
44686 False
44687 False
44688 False
44689 False
44690 False
Name: genres, Length: 44691, dtype: bool
```

```
In [44]: df.cast[0]
```

```
Out[44]: 'Tom Hanks|Tim Allen|Don Rickles|Jim Varney|Wallace Shawn|John Ratzenberger|Annie Potts|John Morris|Erik von Detten|Laurie Metcalf|R. Lee Ermey|Sarah Freeman|Penn Jillette'
```

```
In [45]: mask_actor = df.cast.str.contains("Bruce Willis")
mask_actor
```

```
Out[45]: 0 False
1 False
2 False
3 False
4 False
...
44686 False
44687 False
44688 False
44689 False
44690 NaN
Name: cast, Length: 44691, dtype: object
```

```
In [46]: df.loc[mask_actor & mask_genres, ["title", "vote_average"]].sort_values(by = "vote_average", ascending = False)
```

```
Out[46]: title vote_average
1448 The Fifth Element 7.30
19218 Looper 6.60
1786 Armageddon 6.50
14135 Surrogates 5.90
20333 G.I. Joe: Retaliation 5.40
27619 Vice 4.10
```

```
In [47]: bruce = df.loc[mask_actor & mask_genres, ["title", "poster_path", "vote_average"]].sort_values(by = "vote_average", ascending = False)
HTML(bruce.to_html(escape=False))
```

```
Out[47]: title poster_path vote_average
1448 The Fifth Element /img/.../fifth_element.jpg 7.30
19218 Looper /img/.../looper.jpg 6.60
1786 Armageddon /img/.../armageddon.jpg 6.50
14135 Surrogates /img/.../surrogates.jpg 5.90
20333 G.I. Joe: Retaliation /img/.../g_i_joe_retaliation.jpg 5.40
27619 Vice /img/.../vice.jpg 4.10
```

Search 2: Movies with Uma Thurman and directed by Quentin Tarantino (low runtime)

```
In [48]: df.director
```

```
Out[48]: 0 John Lasseter
1 Joe Johnston
2 Howard Deutch
3 Forest Whitaker
4 Charles Shyer
...
44686 Hamid Nematollah
44687 Lev Diaz
44688 Mark L. Lester
44689 Yakov Protazanov
44690 Daisy Asquith
Name: director, Length: 44691, dtype: object
```

```
In [49]: mask_director = df.director == "Quentin Tarantino"
```

```
In [50]: mask_actor = df.cast.str.contains("Uma Thurman")
```

```
In [51]: quentin = df.loc[mask_director & mask_actor, ["title", "poster_path", "runtime"]].sort_values(by = "runtime").set_index("title")
```

```
In [52]: HTML(quentin.to_html(escape=False))
```

```
Out[52]: poster_path runtime
title
Kill Bill: Vol. 1 /img/.../kill_bill_v1.jpg 111.00
Kill Bill: Vol. 2 /img/.../kill_bill_v2.jpg 136.00
Pulp Fiction /img/.../pulp_fiction.jpg 154.00
```

Search 3: Most Successful Pixar Studio Movies between 2010 and 2015 (High Revenue)

```
In [53]: df.production_companies[1]
```

```
Out[53]: 'TriStar Pictures|Teitler Film|Interscope Communications'
```

```
In [54]: mask_studio = df.production_companies.str.contains("Pixar").fillna(False)
```

```
In [55]: df.release_date
```

```
Out[55]: 0 1995-10-30
1 1995-12-15
2 1995-12-22
3 1995-12-22
4 1995-02-18
...
44686 ...
44687 ...
44688 ...
44689 ...
44690 NaT
```

```
44081 2011-11-1
44688 2003-08-01
44689 1917-10-21
44690 2017-06-09
Name: release_date, Length: 44691, dtype: datetime64[ns]
```

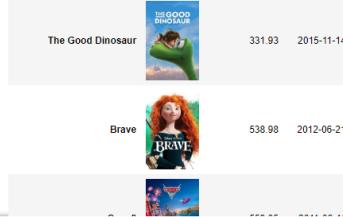
```
In [56]: mask_time = df.release_date.between("2010-01-01", "2015-12-31")
```

```
In [57]: pixar = df.loc[mask_studio & mask_time, ["title", "poster_path", "revenue_musd", "release_date"]].sort_values(by = "revenue_musd")
```

```
In [58]: HTML(pixel.to_html(escape=False))
```

Out[58]: poster\_path revenue\_musd release\_date

tit



```
In [59]: from wordcloud import WordCloud
```

In

```
Out[60]: id title tagline release_date genres belongs_to_collection original_language budget_musd revenue_musd production_musd
0 1 Star Wars: Episode I ... 1999-05-19T00:00:00 Science Fiction Star Wars English 100.0 180.0 100.0
```

|       |        |                             |                                                         |            |                          |                                |     |       |        |                 |
|-------|--------|-----------------------------|---------------------------------------------------------|------------|--------------------------|--------------------------------|-----|-------|--------|-----------------|
| 0     | 862    | Toy Story                   | NaN                                                     | 1995-10-30 | Animation Comedy Family  | Toy Story Collection           | en  | 30.00 | 373.55 | Pixar Anim      |
| 1     | 8844   | Jumanji                     | Roll the dice and unleash the excitement!               | 1995-12-15 | Adventure Fantasy Family | NaN                            | en  | 65.00 | 262.80 | TriStar Pic Fil |
| 2     | 15602  | Grumpier Old Men            | Still Yelling.<br>Still Fighting.<br>Still Ready for... | 1995-12-22 | Romance Comedy           | Grumpy Old Men Collection      | en  | NaN   | NaN    | Warner Br       |
| 3     | 31357  | Waiting to Exhale           | Friends are the people who let you be yourself...       | 1995-12-22 | Comedy Drama Romance     | NaN                            | en  | 16.00 | 81.45  | Twentieth Film  |
| 4     | 11862  | Father of the Bride Part II | Just When His World Is Back To Normal... He's...        | 1995-02-10 | Comedy                   | Father of the Bride Collection | en  | NaN   | 76.58  | Production      |
| ...   | ...    | ...                         | ...                                                     | ...        | ...                      | ...                            | ... | ...   | ...    | ...             |
| 44686 | 439050 | Subdue                      | Rising and falling between a man and woman              | NaT        | Drama Family             | NaN                            | fa  | NaN   | NaN    |                 |
| 44687 | 111109 | Century of Birthing         | NaN                                                     | 2011-11-17 | Drama                    | NaN                            | ti  | NaN   | NaN    |                 |
| 44688 | 67758  | Betrayal                    | A deadly game of wits.                                  | 2003-08-01 | Action Drama Thriller    | NaN                            | en  | NaN   | NaN    | American W      |
| 44689 | 227506 | Satan Triumphant            | NaN                                                     | 1917-10-21 | NaN                      | NaN                            | en  | NaN   | NaN    |                 |
| 44690 | 461257 | Queerama                    | NaN                                                     | 2017-06-09 | NaN                      | NaN                            | en  | NaN   | NaN    |                 |

[View Details](#)

```
Out[61]: 'Roll the dice and unleash the excitement!'

In [62]: df.overview[1]

Out[62]: "When siblings Judy and Peter discover an enchanted board game that opens the door to a magical world, they unwittingly invite Alan -- an adult who's been trapped inside the game for 26 years -- into their living room. Alan's only hope for freedom is to finish the game, which proves risky as all three find themselves running from giant rhinoceroses, evil monkeys and other terrif
```

```
In [63]: title = df.title.dropna()
overview = df.overview.dropna()
```

```
In [64]: title
Out[64]: 0 Toy Story
 1 Jumanji
 2 Grumpier Old Men
 3 Waiting to Exhale
 4 Father of the Bride Part II
 ...
44686 Subdue
44687 Century of Birthring
44688 Betrayal
44689 Satan Triumphant
44690 Queerama
```

```
In [65]: ''.join(title)
```

**Out[65]:** 'Toy StoryJumanjiGrumpier Old MenWaiting to ExhaleFather of the Bride Part IIHeatSabrinaTom and HuckSudden DeathGoldenEyeThe AmericanPresidentDracula: Dead and Loving ItBalticInnCuthroat IslandCasinoSense and SensibilityFour RoomsAdventure: When Nature CallsMoney TrainGet ShortyCopycatAssassinsPowderLeaving Las VegasGet HellNow and ThenPersuasionThe City of Lost ChildrenShanghaiTriadDangerous MindsTwelve MonkeysWings of CourageBad BoysCarvingtonDead Man WalkingAcross the Sea of TimeIt Takes TwoLieutenant Scry, the Beloved CountyRichard IIIDead PresidentsRestorationMortal KombatTo Die ForHow To Make An American QuiltSeven Pocahontaswhen Night Is FallingThe Usual SuspectsGuardian AngelMighty AphroditeAmerica The Big Green Georgiakids of the Round TableHome for the HolidaysPostmanThe ConfessionalsThe Indian in the CupboardEye for an EyeHr...h's OpusDon't Be A Menace to South Central While Drinking Your Juice in the HoodToo If by Seabio-DomeLawnmower Man 2: Beyond CyberdineTwo BitsFrench TwistFriday From Dusk Till DawnFairy GameKicking and ScreamingLes MisérablesBed of RosesBig BullyScreamersNico IconThe CrossingGuardThe JurorThe White BalloonThings to Do in Denver When You're Deadonata's LineOnce Upon a Time... When We Were ColoredLast Summer in the HamptonsAngels and InsectsWhite SquallDunston Checks InBlack Sheepnick of TimeThe Journey of AugustKingLarry ReillyVampire in BrooklynBeautiful GirlsBroken ArrowA Midwinter's TaleLa HaineShoppingSpree Fleiss: Hollywood Nadam City HallBottle RocketWrong!Unforgettable Happy GilmoreThe Bridges of Madison CountyYobbody Loves Melleput Treasure IslandCatwalkHeadless Body in Topless BarBraveheartTaxi DriverRumble in the BronxBefore and AfterMargaret's MuseumHappiness Is in the FieldAnne FrankRememberedThe Young Poisons' HandbookLucy FerrellDie Big Steel Little Rascals the SunThe Boys of St. Vince ntBoomerangChungking ExpressThe Star MakerFlirting with DisasterThe Neverending Story III: Escape from FantasiaThe Silences of the PalaceJuupiter's WifeLife in the SkyAngelinaFrankie StarlightJade WebulaYolonic StoriesDown PeriscopeFrom the Journals of Jean SebergPlan of the YearThe Neon BibleGetup Close & PersonalThe BirdcageVirgin MaryThe Brothers McMullenBad BoysThe Amazing Panda AdventureThe Basketball DiariesAn Awfully Big AdventureAmateurApollo 13Rob RoyThe AddictionBatman ForeverEverbelle de Jonge'.

```
In [66]: title_corpus = ' '.join(title)
overview_corpus = ' '.join(overview)
tagline_corpus = ' '.join(tagline)
```

```
In [67]: tagline corpus
```

**VUOLI: NO!** Well the nice and unlease the excitement! Still yelling. Still fighting. Still ready for love, friends are the people who tell you to be yourself... and never let you forget it. Just When His World Is Back To Normal... He's In For The Surprise Of His Life! A Los Angeles Crime Saga You are cordially invited to the most surprising merger of the year. The Original Bad Boys. Terry goes into overtime. No limits. No fears. No substitutes. Why can't the most powerful man in the world have the one thing he wants most? Part Dog. Part Wolf. All Hero. Triumphant in Victory. Bitter in Defeat. He changed the world, But Lost a Nation. The Course Has Been Set. There is No Turning Back. Prepare Your Weapons. Summon Your Courage. Discover the Adventure of a Lifetime! No one stays at the top forever. Lose your heart and come to your senses. Twelve outrageous guests. Four scandalous requests. And one lone bellhop, in his first day on the job, who's in for the wildest New Year's Eve of his life. New animal is. New adventures. Same hair. Get on, or GET OUT THE WAY! The mob is tough, but it's nothing like show business. One man is copying the most notorious killers in history one at a time. Together, two women must stop him from killing again. Or they're next. In the shadows of life, In the business of death, One man found a reason to live... An extraordinary encounter with a another human being. I Love You... The Way You Are. Envy, greed, jealousy and love. In every woman there is the girl she left behind. Where happily ever after is just a dream. In 1930's Shanghai violence was not the problem. It was the solution. She broke the rules, and changed their lives. The future is history. A little pig goes a long way. A love so unusual it has to be True Two identical strangers. Two different worlds. One perfect match. Sex. Clothes. Popularity. Is there a problem here? I can smile, and murmur while I smile In this daring heist, the only color that counts is green. Nothing In This World Has Prepared You For This. All she wanted was a little attention. There's beauty in the patterns of life. Seven deadly sins. Seven ways to die. An American legend comes to life. Five Criminals. One Line Up. No Coincidence. She's no angel of mercy. The biggest kick of the year! The toughest act to follow was their dreams. We'll do it every year... until we get it right what do you think?

```
In [68]: title_wordcloud = WordCloud(background_color='white', height=2000, width=4000, max_words= 200).generate(title_corpus)
title_wordcloud
```

```
In [69]: plt.figure(figsize=(16,8))
plt.imshow(title_wordcloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```



```
In [70]: tagline_wordcloud = WordCloud(background_color="white", height=2000, width=4000, max_words= 200).generate(tagline_corpus)
plt.figure(figsize=(16,8))
plt.imshow(tagline_wordcloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```



## Are Franchises more successful?

```
In [71]: df.belongs_to_collection
Out[71]: 0 Toy Story Collection
1 NaN
2 Grumpy Old Men Collection
3 NaN
4 Father of the Bride Collection
...
44686 NaN
44687 NaN
44688 NaN
44689 NaN
44690 NaN
Name: belongs_to_collection, Length: 44691, dtype: object
```

```
In [72]: df["Franchise"] = df.belongs_to_collection.notna()
```

```
In [73]: df.Franchise
```

```
Out[74]: False 40228
 True 4463
 Name: Franchise, dtype: int64
```

```
Franchise vs Stand-alone Average Revenue

In [75]: df.groupby("Franchise").revenue_musd.mean()

Out[75]: Franchise
False 44.74
True 165.71
Name: revenue_musd, dtype: float64
```

## Franchise vs Stand-alone: Return on Investment / Profitability

```
In [76]: df['ROI'] = df.revenue_musd.div(df.budget_musd)
```

```
In [77]: df.groupby("Franchise").ROI.median()
```

```
Out[77]:
Franchise
False 1.62
True 3.71
Name: ROI, dtype: float64
```

## Franchise vs Stand-alone Average Budget

```
In [78]: df.groupby("Franchise").budget_musd.mean()
```

```
Out[78]:
Franchise
False 18.05
True 38.32
Name: budget_musd, dtype: float64
```

## Most Successful Franchises

```
In [79]: df.belongs_to_collection
```

```
Out[79]:
0 Toy Story Collection
1 NaN
2 Grumpy Old Men Collection
3 NaN
4 Father of the Bride Collection
...
44686 NaN
44687 NaN
44688 NaN
44689 NaN
44690 NaN
Name: belongs_to_collection, Length: 44691, dtype: object
```

```
In [80]: df.belongs_to_collection.value_counts()
```

```
Out[80]:
The Bowery Boys 29
Totò Collection 27
Zatōichi: The Blind Swordsman 26
James Bond Collection 26
The Carry On Collection 25
...
Salt and Pepper Collection 1
Deadpool Collection 1
Ant-Man Collection 1
Elvira Collection 1
Red Lotus Collection 1
Name: belongs_to_collection, Length: 1691, dtype: int64
```

```
In [81]: franchises = df.groupby("belongs_to_collection").agg({"title": "count", "budget_musd": ["sum", "mean"],
 "revenue_musd": ["sum", "mean"],
 "vote_average": "mean", "popularity": "mean",
 "ROI": "median",
 "vote_count": "mean"})
```

```
In [82]: franchises
```

```
Out[82]:
title budget_musd revenue_musd vote_average popularity ROI vote_count
count sum mean sum mean mean median mean
belongs_to_collection
...
... Has Fallen Collection 2 130.00 65.00 366.78 183.39 6.00 13.01 2.88 2333.00
00 Schneider Filmreihe 1 0.00 NaN 0.00 NaN 6.50 1.93 NaN 16.00
08/15 Collection 1 0.00 NaN 0.00 NaN 5.90 0.63 NaN 4.00
100 Girls Collection 2 0.00 NaN 0.00 NaN 5.15 3.08 NaN 64.00
101 Dalmatians (Animated) Collection 2 4.00 4.00 215.88 215.88 6.25 13.06 53.97 937.00
...
...
Сказки Чуковского 1 0.00 NaN 0.00 NaN 3.00 0.73 NaN 3.00
Чебурашка и крокодил Гена 1 0.00 NaN 0.00 NaN 6.70 0.88 NaN 7.00
Что Творят мужчины! (Коллекция) 2 2.00 2.00 0.00 NaN 3.15 1.30 NaN 5.50
男はつらいよ シリーズ 3 0.00 NaN 0.00 NaN 7.00 0.04 NaN 0.67
...
식객 시리즈 2 0.00 NaN 0.00 NaN 4.95 0.16 NaN 3.50
```

1691 rows x 9 columns

```
In [83]: franchises.nlargest(20, ("title", "count"))
```

```
Out[83]:
title budget_musd revenue_musd vote_average popularity ROI vote_count
count sum mean sum mean mean median mean
belongs_to_collection
...
The Bowery Boys 29 0.00 NaN 0.00 NaN 6.67 0.20 NaN 0.72
Totò Collection 27 0.00 NaN 0.00 NaN 6.84 1.05 NaN 18.04
James Bond Collection 26 1539.65 59.22 7106.97 273.35 6.34 13.45 6.13 1284.31
Zatōichi: The Blind Swordsman 26 0.00 NaN 0.00 NaN 6.40 1.10 NaN 11.19
The Carry On Collection 25 0.00 NaN 0.00 NaN 6.17 3.22 NaN 21.04
Charlie Chan (Sidney Toler) Collection 21 0.00 NaN 0.00 NaN 6.61 0.63 NaN 5.81
Pokémon Collection 20 250.72 50.14 601.87 66.87 6.05 4.37 4.46 71.10
Godzilla (Showa) Collection 16 2.81 0.56 0.00 NaN 5.97 3.42 NaN 39.75
Charlie Chan (Warner Oland) Collection 15 0.00 NaN 0.00 NaN 6.66 0.66 NaN 6.87
Dragon Ball Z (Movie) Collection 15 5.00 5.00 112.12 56.06 6.61 6.18 12.35 133.33
Uuno Turhapuro 15 0.00 NaN 0.00 NaN 5.53 0.17 NaN 3.87
Monster High Collection 14 0.50 0.50 0.00 NaN 6.49 3.59 NaN 12.57
The Land Before Time Collection 14 12.30 12.30 84.46 84.46 6.16 4.91 6.87 96.07
George Carlin Comedy Collection 13 0.00 NaN 0.00 NaN 7.47 1.30 NaN 12.54
Sharpe Collection 13 0.00 NaN 0.00 NaN 6.74 1.71 NaN 13.23
Friday the 13th Collection 12 84.40 7.03 464.90 38.74 5.42 9.76 6.67 326.58
Johan Falk GSI Collection 12 0.00 NaN 0.00 NaN 6.50 1.52 NaN 9.25
Sherlock Holmes (1939 series) 12 0.00 NaN 0.00 NaN 6.92 2.88 NaN 21.58
The Amityville Horror Collection 12 21.02 5.25 213.35 53.34 4.36 7.18 5.69 91.92
An American Girl Collection 10 6.80 6.80 0.00 NaN 5.25 1.94 NaN 8.70
```

```
In [84]: franchises.nlargest(20, ("revenue_musd", "sum"))
```

```
Out[84]:
title budget_musd revenue_musd vote_average popularity ROI vote_count
count sum mean sum mean mean median mean
belongs_to_collection
...
Harry Potter Collection 8 1280.00 160.00 7707.37 963.42 7.54 26.25 6.17 5983.25
Star Wars Collection 8 854.35 106.79 7434.49 929.31 7.38 23.41 8.24 5430.38
James Bond Collection 26 1539.65 59.22 7106.97 273.35 6.34 13.45 6.13 1284.31
The Fast and the Furious Collection 8 1009.00 126.12 5125.10 640.64 6.66 10.80 4.94 3197.00
Pirates of the Caribbean Collection 5 1250.00 250.00 4521.58 904.32 6.88 53.97 3.45 5016.00
Transformers Collection 5 965.00 193.00 4366.10 873.22 6.14 14.43 5.20 3046.40
Despicable Me Collection 6 299.00 74.75 3691.07 922.77 6.78 106.72 12.76 3041.33
...
The Twilight Collection 5 206.00 77.00 3247.11 480.47 6.04 70.40 10.77 5776.76
```

| title                            | budget_musd | revenue_musd | vote_average | popularity | ROI     | vote_count |        |       |          |
|----------------------------------|-------------|--------------|--------------|------------|---------|------------|--------|-------|----------|
| Ice Age Collection               | 5           | 429.00       | 85.80        | 3216.71    | 643.34  | 6.38       | 16.08  | 8.26  | 2643.80  |
| Jurassic Park Collection         | 4           | 379.00       | 94.75        | 3031.48    | 757.87  | 6.50       | 10.77  | 7.03  | 4608.75  |
| Shrek Collection                 | 5           | 535.00       | 133.75       | 2955.81    | 738.95  | 6.46       | 12.97  | 5.56  | 2344.20  |
| The Hunger Games Collection      | 4           | 490.00       | 122.50       | 2944.16    | 736.04  | 6.88       | 54.77  | 6.27  | 6543.50  |
| The Hobbit Collection            | 3           | 750.00       | 250.00       | 2935.52    | 978.51  | 7.23       | 25.21  | 3.83  | 5981.33  |
| The Avengers Collection          | 2           | 500.00       | 250.00       | 2924.95    | 1462.48 | 7.35       | 63.63  | 5.96  | 9454.00  |
| The Lord of the Rings Collection | 3           | 266.00       | 88.67        | 2916.54    | 972.18  | 8.03       | 30.27  | 11.73 | 8253.00  |
| X-Men Collection                 | 6           | 983.00       | 163.83       | 2808.83    | 468.14  | 6.82       | 9.71   | 3.02  | 4593.83  |
| Avatar Collection                | 1           | 237.00       | 237.00       | 2787.97    | 2787.97 | 7.20       | 185.07 | 11.76 | 12114.00 |
| Mission: Impossible Collection   | 5           | 650.00       | 130.00       | 2778.98    | 555.80  | 6.60       | 16.51  | 4.55  | 2801.00  |
| Spider-Man Collection            | 3           | 597.00       | 199.00       | 2496.35    | 832.12  | 6.47       | 22.62  | 3.92  | 4505.67  |
| The Dark Knight Collection       | 3           | 585.00       | 195.00       | 2463.72    | 821.24  | 7.80       | 57.42  | 4.34  | 9681.00  |

In [85]: franchises.nlargest(20, ("revenue\_musd", "mean"))

Out[85]:

| title                               | budget_musd | revenue_musd | vote_average | popularity | ROI     | vote_count |        |       |          |
|-------------------------------------|-------------|--------------|--------------|------------|---------|------------|--------|-------|----------|
| belongs_to_collection               | count       | sum          | mean         | sum        | mean    | mean       |        |       |          |
| Avatar Collection                   | 1           | 237.00       | 237.00       | 2787.97    | 2787.97 | 7.20       | 185.07 | 11.76 | 12114.00 |
| The Avengers Collection             | 2           | 500.00       | 250.00       | 2924.95    | 1462.48 | 7.35       | 63.63  | 5.96  | 9454.00  |
| Frozen Collection                   | 2           | 150.00       | 150.00       | 1274.22    | 1274.22 | 7.10       | 16.88  | 8.49  | 3035.00  |
| Finding Nemo Collection             | 2           | 294.00       | 147.00       | 1985.91    | 984.45  | 7.20       | 19.99  | 7.57  | 5312.50  |
| The Hobbit Collection               | 3           | 750.00       | 250.00       | 2935.52    | 978.51  | 7.23       | 25.21  | 3.83  | 5981.33  |
| The Lord of the Rings Collection    | 3           | 266.00       | 88.67        | 2915.54    | 972.18  | 8.03       | 30.27  | 11.73 | 8253.00  |
| Harry Potter Collection             | 8           | 1280.00      | 160.00       | 7707.37    | 963.42  | 7.54       | 26.25  | 6.17  | 5983.25  |
| Star Wars Collection                | 8           | 854.35       | 106.79       | 7434.49    | 929.31  | 7.38       | 23.41  | 8.24  | 5430.38  |
| Despicable Me Collection            | 6           | 299.00       | 74.75        | 3691.07    | 922.77  | 6.78       | 106.72 | 12.76 | 3041.33  |
| Pirates of the Caribbean Collection | 5           | 1250.00      | 250.00       | 4521.58    | 904.32  | 6.88       | 53.97  | 3.45  | 5016.00  |
| The Secret Life of Pets Collection  | 1           | 75.00        | 75.00        | 875.46     | 875.46  | 5.90       | 11.01  | 11.67 | 3536.00  |
| Transformers Collection             | 5           | 965.00       | 193.00       | 4366.10    | 873.22  | 6.14       | 14.43  | 5.20  | 3046.40  |
| Spider-Man Collection               | 3           | 597.00       | 199.00       | 2498.35    | 832.12  | 6.47       | 22.62  | 3.92  | 4505.67  |
| The Dark Knight Collection          | 3           | 585.00       | 195.00       | 2463.72    | 821.24  | 7.80       | 57.42  | 4.34  | 9681.00  |
| Wonder Woman Collection             | 1           | 149.00       | 149.00       | 8205.98    | 8205.98 | 7.20       | 294.34 | 5.51  | 5025.00  |
| Guardians of the Galaxy Collection  | 2           | 370.00       | 185.00       | 1636.74    | 818.37  | 7.75       | 119.31 | 4.43  | 7436.00  |
| Fantastic Beasts Collection         | 1           | 180.00       | 180.00       | 809.34     | 809.34  | 7.20       | 41.05  | 4.50  | 5612.00  |
| Iron Man Collection                 | 3           | 540.00       | 180.00       | 2424.55    | 808.18  | 6.93       | 21.63  | 4.18  | 8290.33  |
| The Lion King Collection            | 4           | 45.00        | 45.00        | 788.24     | 788.24  | 6.75       | 11.48  | 17.52 | 1827.50  |
| Deadpool Collection                 | 1           | 58.00        | 58.00        | 783.11     | 783.11  | 7.40       | 187.86 | 13.50 | 11444.00 |

In [86]: franchises.nlargest(20, ("budget\_musd", "sum"))

Out[86]:

| title                                   | budget_musd | revenue_musd | vote_average | popularity | ROI     | vote_count |       |      |         |
|-----------------------------------------|-------------|--------------|--------------|------------|---------|------------|-------|------|---------|
| belongs_to_collection                   | count       | sum          | mean         | sum        | mean    | mean       |       |      |         |
| James Bond Collection                   | 28          | 1539.65      | 59.22        | 7106.97    | 273.35  | 6.34       | 13.45 | 6.13 | 1284.31 |
| Harry Potter Collection                 | 8           | 1280.00      | 160.00       | 7707.37    | 963.42  | 7.54       | 26.25 | 6.17 | 5983.25 |
| Pirates of the Caribbean Collection     | 5           | 1250.00      | 250.00       | 4521.58    | 904.32  | 6.88       | 53.97 | 3.45 | 5016.00 |
| The Fast and the Furious Collection     | 8           | 1009.00      | 126.12       | 5125.10    | 640.64  | 6.66       | 10.80 | 4.94 | 3197.00 |
| X-Men Collection                        | 6           | 983.00       | 163.83       | 2808.83    | 468.14  | 6.82       | 9.71  | 3.02 | 4593.83 |
| Transformers Collection                 | 5           | 965.00       | 193.00       | 4366.10    | 873.22  | 6.14       | 14.43 | 5.20 | 3046.40 |
| Star Wars Collection                    | 8           | 854.35       | 106.79       | 7434.49    | 929.31  | 7.38       | 23.41 | 8.24 | 5430.38 |
| The Hobbit Collection                   | 3           | 750.00       | 250.00       | 2935.52    | 978.51  | 7.23       | 25.21 | 3.83 | 5981.33 |
| The Terminator Collection               | 5           | 661.40       | 132.28       | 1845.33    | 369.07  | 6.54       | 21.89 | 2.84 | 3366.40 |
| Mission: Impossible Collection          | 5           | 650.00       | 130.00       | 2778.98    | 555.80  | 6.60       | 16.51 | 4.55 | 2801.00 |
| Spider-Man Collection                   | 3           | 597.00       | 199.00       | 2496.35    | 832.12  | 6.47       | 22.62 | 3.92 | 4505.67 |
| The Dark Knight Collection              | 3           | 585.00       | 195.00       | 2463.72    | 821.24  | 7.80       | 57.42 | 4.34 | 9681.00 |
| Captain America Collection              | 3           | 560.00       | 186.67       | 2238.64    | 746.21  | 7.10       | 61.31 | 4.20 | 6839.00 |
| The Chronicles of Narnia Collection     | 3           | 560.00       | 186.67       | 1594.14    | 528.05  | 6.40       | 14.34 | 2.68 | 1977.33 |
| Iron Man Collection                     | 3           | 540.00       | 180.00       | 2424.55    | 808.18  | 6.93       | 21.63 | 4.18 | 8290.33 |
| Shrek Collection                        | 5           | 535.00       | 133.75       | 2955.81    | 738.95  | 6.46       | 12.97 | 5.56 | 2344.20 |
| Star Trek: Alternate Reality Collection | 3           | 525.00       | 175.00       | 1196.52    | 398.84  | 7.13       | 17.27 | 2.46 | 3896.33 |
| Madagascar Collection                   | 4           | 502.00       | 125.50       | 2257.05    | 564.26  | 6.42       | 16.21 | 4.59 | 2105.00 |
| The Avengers Collection                 | 2           | 500.00       | 250.00       | 2924.95    | 1462.48 | 7.35       | 63.63 | 5.96 | 9454.00 |
| Cars Collection                         | 3           | 495.00       | 165.00       | 1372.01    | 457.34  | 6.33       | 17.25 | 2.80 | 2265.67 |

In [87]: franchises[franchises[("vote\_count", "mean")] >= 1000].nlargest(20, ("vote\_average", "mean"))

Out[87]:

| title                               | budget_musd | revenue_musd | vote_average | popularity | ROI     | vote_count |        |       |          |
|-------------------------------------|-------------|--------------|--------------|------------|---------|------------|--------|-------|----------|
| belongs_to_collection               | count       | sum          | mean         | sum        | mean    | mean       |        |       |          |
| The Lord of the Rings Collection    | 3           | 266.00       | 88.67        | 2916.54    | 972.18  | 8.03       | 30.27  | 11.73 | 8253.00  |
| The Godfather Collection            | 3           | 73.00        | 24.33        | 429.38     | 143.13  | 7.97       | 31.64  | 3.66  | 3677.00  |
| Blade Runner Collection             | 1           | 28.00        | 28.00        | 33.14      | 33.14   | 7.90       | 96.27  | 1.18  | 3833.00  |
| The Man With No Name Collection     | 3           | 2.00         | 0.67         | 35.50      | 11.83   | 7.83       | 14.17  | 25.00 | 1422.67  |
| The Dark Knight Collection          | 3           | 585.00       | 195.00       | 2463.72    | 821.24  | 7.80       | 57.42  | 4.34  | 9681.00  |
| Guardians of the Galaxy Collection  | 2           | 370.00       | 185.00       | 1636.74    | 818.37  | 7.75       | 119.31 | 4.43  | 7436.00  |
| Kill Bill Collection                | 2           | 60.00        | 30.00        | 333.11     | 166.55  | 7.70       | 23.40  | 5.55  | 4576.00  |
| Kingsman Collection                 | 1           | 81.00        | 81.00        | 414.35     | 414.35  | 7.60       | 28.22  | 5.12  | 6069.00  |
| How to Train Your Dragon Collection | 2           | 310.00       | 155.00       | 1104.00    | 552.00  | 7.55       | 13.34  | 3.80  | 3741.00  |
| Harry Potter Collection             | 8           | 1280.00      | 160.00       | 7707.37    | 963.42  | 7.54       | 26.25  | 6.17  | 5983.25  |
| Toy Story Collection                | 3           | 320.00       | 106.67       | 1937.89    | 645.96  | 7.53       | 18.82  | 5.53  | 4679.67  |
| Back to the Future Collection       | 3           | 99.00        | 33.00        | 957.64     | 319.21  | 7.50       | 17.50  | 8.30  | 4381.00  |
| Trainspotting Collection            | 2           | 22.00        | 11.00        | 57.90      | 28.95   | 7.45       | 18.50  | 3.21  | 1766.00  |
| Deadpool Collection                 | 1           | 58.00        | 58.00        | 783.11     | 783.11  | 7.40       | 187.88 | 13.50 | 11444.00 |
| The Incredibles Collection          | 1           | 92.00        | 92.00        | 631.44     | 631.44  | 7.40       | 22.22  | 6.86  | 5290.00  |
| Star Wars Collection                | 8           | 854.35       | 106.79       | 7434.49    | 929.31  | 7.38       | 23.41  | 8.24  | 5430.38  |
| The Avengers Collection             | 2           | 500.00       | 250.00       | 2924.95    | 1462.48 | 7.35       | 63.63  | 5.96  | 9454.00  |
| The Lego Movie Collection           | 2           | 140.00       | 70.00        | 781.11     | 390.56  | 7.35       | 16.74  | 5.86  | 2300.00  |
| Tangled Collection                  | 2           | 260.00       | 260.00       | 591.79     | 591.79  | 7.25       | 12.32  | 2.28  | 1901.00  |
| The Space Odyssey Series            | 2           | 38.50        | 19.25        | 109.10     | 54.55   | 7.25       | 14.99  | 3.99  | 1682.50  |

In [88]: df.director

Out[88]:

|       |                  |
|-------|------------------|
| 0     | John Lasseter    |
| 1     | Joe Johnston     |
| 2     | Howard Deutch    |
| 3     | Forest Whitaker  |
| 4     | Charles Shyer    |
| 44686 | Hanif Nematollah |
| 44687 | Lav Diaz         |
| 44688 | Mark L. Lester   |

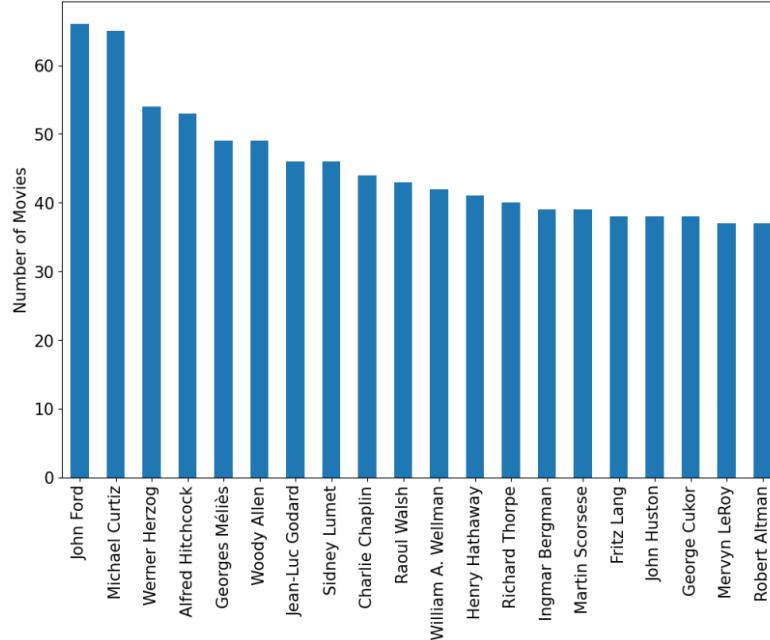
```
44689 Yakov Protazanov
44690 Daisy Asquith
Name: director, Length: 44691, dtype: object
```

```
In [89]: df.director.value_counts().head(20)
```

```
Out[89]: John Ford 66
Michael Curtiz 65
Werner Herzog 54
Alfred Hitchcock 53
Georges Méliès 49
Woody Allen 49
Jean-Luc Godard 46
Sidney Lumet 46
Charlie Chaplin 44
Raoul Walsh 43
William A. Wellman 42
Henry Hathaway 41
Richard Thorpe 40
Ingmar Bergman 39
Martin Scorsese 39
Fritz Lang 38
John Huston 38
George Cukor 38
Mervyn LeRoy 37
Robert Altman 37
Name: director, dtype: int64
```

```
In [90]: plt.figure(figsize = (12,8))
df.director.value_counts().head(20).plot(kind='bar', fontsize = 15)
plt.title("Most Active Directors", fontsize = 20)
plt.ylabel("Number of Movies", fontsize = 15)
plt.show()
```

Most Active Directors

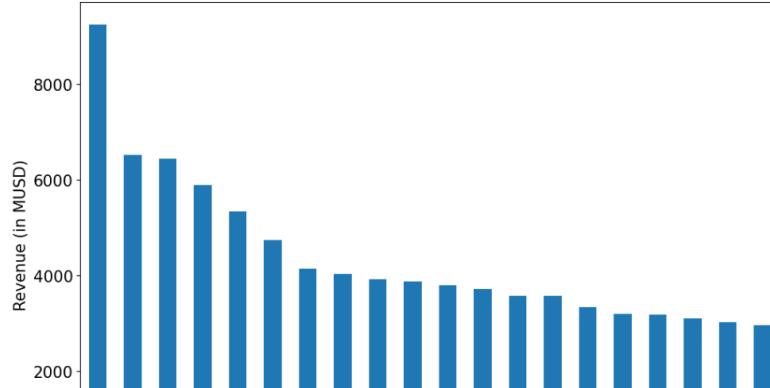


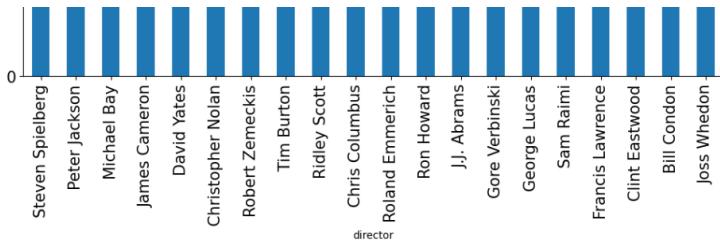
```
In [91]: df.groupby("director").revenue_musd.sum().nlargest(20)
```

```
Out[91]: director
Steven Spielberg 9256.62
Peter Jackson 6528.24
Michael Bay 6437.47
James Cameron 5900.61
David Yates 5334.56
Christopher Nolan 4747.41
Robert Zemeckis 4138.23
Tim Burton 4032.92
Ridley Scott 3917.53
Chris Columbus 3866.64
Roland Emmerich 3798.49
Ron Howard 3714.15
J.J. Abrams 3572.22
Gore Verbinski 3575.34
George Lucas 3341.55
Sam Raimi 3193.79
Francis Lawrence 3183.34
Clint Eastwood 3100.68
Bill Condon 3017.30
Joss Whedon 2963.83
Name: revenue_musd, dtype: float64
```

```
In [92]: plt.figure(figsize = (12,8))
df.groupby("director").revenue_musd.sum().nlargest(20).plot(kind='bar', fontsize = 15)
plt.title("Total Revenue", fontsize = 20)
plt.ylabel("Revenue (in MUSD)", fontsize = 15)
plt.show()
```

Total Revenue





```
In [93]: directors = df.groupby("director").agg({"title": "count", "vote_average": "mean", "vote_count": "sum"})
```

```
In [94]: directors
```

```
Out[94]:
 title vote_average vote_count
director
Dale Treville... 2 4.00 4.00
Davide Manuli 1 6.90 10.00
E.W. Swockhamer 1 5.90 5.00
Vitaliy Vorobyov 1 5.50 3.00
Yeon Sang-Ho 4 6.60 1039.00
...
Ярополк Лапшин 1 10.00 1.00
پژن معلقی 1 6.00 2.00
塙谷直義 1 7.20 40.00
杰森·莫玛 1 5.80 28.00
진모영 1 6.00 6.00
```

17349 rows × 3 columns

```
In [95]: directors[(directors.vote_count >= 10000) & (directors.title >= 10)].nlargest(20, "vote_average")
```

```
Out[95]:
 title vote_average vote_count
director
Hayao Miyazaki 14 7.70 14700.00
Christopher Nolan 11 7.62 67344.00
Quentin Tarantino 10 7.49 45910.00
Wes Anderson 10 7.37 11743.00
David Fincher 10 7.35 37588.00
Martin Scorsese 39 7.22 35541.00
Peter Jackson 13 7.14 47571.00
Joel Coen 17 7.02 18139.00
James Cameron 11 6.93 33736.00
Stanley Kubrick 16 6.91 18214.00
James Mangold 10 6.90 16607.00
Steven Spielberg 33 6.89 62266.00
Danny Boyle 14 6.87 16504.00
Guy Ritchie 10 6.80 19626.00
Robert Zemeckis 19 6.79 37666.00
Terry Gilliam 14 6.76 10049.00
Tim Burton 21 6.73 36922.00
John Lasseter 10 6.72 18337.00
Ang Lee 14 6.71 11164.00
Antoine Fuqua 12 6.71 15519.00
```

```
In [96]: df.genres = df.genres.astype(str)
```

```
In [97]: df.loc[df.genres.str.contains("Horror")].groupby("director").revenue_musd.sum().nlargest(20)
```

```
Out[97]:
director
Paul W.S. Anderson 982.29
James Wan 861.31
Wes Craven 834.93
Francis Lawrence 816.23
Ridley Scott 689.88
Hans Zimmer 514.07
Steven Spielberg 500.18
William Friedkin 466.40
Darren Lynn Bousman 456.34
M. Night Shyamalan 375.37
Henry Joost 349.07
David R. Ellis 348.74
Adrian Lyne 346.11
James DeMonaco 316.70
Stephen Sommers 311.46
Gore Verbinski 275.91
Guillermo del Toro 261.63
John R. Leonetti 255.27
Fede Alvarez 254.64
Jordan Peele 252.43
Name: revenue_musd, dtype: float64
```

Most Successful Actors

```
In [98]: df.cast
```

```
Out[98]:
0 Tom Hanks|Tim Allen|Don Rickles|Jim Varney|Wal...
1 Robin Williams|Jonathan Hyde|Kirsten Dunst|Dra...
2 Walter Matthau|Jack Lemmon|Ann Margret|Sophia ...
3 Whitney Houston|Angela Bassett|Loretta Devine|...
4 Steve Martin|Diane Keaton|Martin Short|Kimberl...
...
44686 Leila Hatami|Kourosh Tahami|Elham Korda
44687 Angel Aquino|Perry Diaz|Hazel Orencio|Joel To...
44688 Erika Eleniak|Adam Baldwin|Julie du Page|James...
44689 Iwan Mosschuchin|Nathalie Lissenko|Pavel Pavlo...
44690 ...
Name: cast, Length: 44691, dtype: object
```

```
In [99]: df.set_index("id", inplace = True)
```

```
In [100]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 44691 entries, 862 to 461257
Data columns (total 23 columns):
 # Column Non-Null Count Dtype

 0 title 44691 non-null object
 1 tagline 20284 non-null object
 2 release_date 44657 non-null datetime64[ns]
 3 genres 44691 non-null object
 4 belongs_to_collection 4463 non-null object
 5 original_language 44681 non-null object
 6 budget_musd 8854 non-null float64
 7 revenue_musd 7385 non-null float64
```

```

8 production_companies 33356 non-null object
9 production_countries 38825 non-null object
10 vote_count 44691 non-null float64
11 vote_average 42077 non-null float64
12 popularity 44691 non-null float64
13 runtime 43179 non-null float64
14 overview 43748 non-null object
15 spoken_languages 41894 non-null object
16 poster_path 44467 non-null object
17 cast 42582 non-null object
18 cast_size 44691 non-null int64
19 crew_size 44691 non-null int64
20 director 43960 non-null object
21 Franchise 44691 non-null bool
22 ROI 5371 non-null float64
dtypes: bool(1), datetime64[ns](1), float64(7), int64(2), object(12)
memory usage: 7.9+ MB

```

In [102]: df.cast

```

Out[102]: id
862 Tom Hanks|Tim Allen|Don Rickles|Jim Varney|Wal...
8844 Robin Williams|Jonathan Hyde|Kirsten Dunst|Bra...
15602 Walter Matthau|Jack Lemmon|Ann-Margret|Sophia ...
31357 Whitney Houston|Angela Bassett|Loretta Devine|...
11862 Steve Martin|Diane Keaton|Martin Short|Kimberl...
...
439050 Leila Hatami|Kourosh Tahami|Elham Korda
111109 Angel Aquino|Perry Dizon|Hazel Orenco|Joel To...
67758 Erika Eleniak|Adam Baldwin|Julie du Page|James...
227506 Iwan Mosselchuchin|Natalie Lissenko|Pavel Pavlo...
461257 NaN
Name: cast, Length: 44691, dtype: object

```

In [104]: df.cast.str.split("|", expand = True)

```

Out[104]:
 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
id
862 Tom Hanks Tim Allen Don Rickles Jim Varney Wallace Shawn Ratzbenberger John Annie Potts John Morris Erik von Detten Laurie Metcalf R Lee Ermy Sarah Freeman Penn Jillette None N
8844 Robin Williams Jonathan Hyde Kirsten Dunst Bradley Pierce Bonnie Hunt Bebe Neuwirth David Alan Grier Patricia Clarkson Adam Hann-Byrd Laura Bell Bundy James Handy Gillian Barber Brandon Obray Cyrus Thiedeke C Jos Thr
15602 Walter Matthau Jack Lemmon Ann-Margret Sophia Loren Hannah Daryl Burgess Meredith Kevin Pollak None None None None None None None None None N
31357 Whitney Houston Angela Bassett Loretta Devine Kimberly Williams-Paisley Lela Rochon Gregory Hines Dennis Haystbert Michael Beach Mykelti Williamson Lamont Johnson Wesley Snipes None None None None None N
11862 Steve Martin Diane Keaton Martin Short Williams-Paisley George Newbern Kieran Culkin BD Wong Peter Michael Goetz Kate McGregor-Stewart Jane Adams Eugene Levy Lori Alan None None N
...
439050 Leila Hatami Kourosh Tahami Elham Korda None N
111109 Angel Aquino Perry Dizon Hazel Orenco Joel Torre Bart Guingona Soliman Cruz Roeder Angeli Bayani Dante Perez Betty Uy-Regala Modesta None None None N
67758 Erika Eleniak Adam Baldwin Julie du Page James Remar Damian Chapa Louis Mandylor Tom Wright Jeremy Lelliott James Quattrochi Jason Sabatino Joe Ellsworth Kiko Widener Don Swayze Peter Dobson Da Dubo
227506 Iwan Mosselchuchin Natalie Lissenko Pavel Pavlov Aleksandr Chabrov Vera Orlova None N
461257 NaN !
```

44691 rows × 313 columns

In [105]: act = df.cast.str.split("|", expand = True)

```

act
Out[105]:
 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
id
862 Tom Hanks Tim Allen Don Rickles Jim Varney Wallace Shawn Ratzbenberger John Annie Potts John Morris Erik von Detten Laurie Metcalf R Lee Ermy Sarah Freeman Penn Jillette None N
8844 Robin Williams Jonathan Hyde Kirsten Dunst Bradley Pierce Bonnie Hunt Bebe Neuwirth David Alan Grier Patricia Clarkson Adam Hann-Byrd Laura Bell Bundy James Handy Gillian Barber Brandon Obray Cyrus Thiedeke C Jos Thr
15602 Walter Matthau Jack Lemmon Ann-Margret Sophia Loren Hannah Daryl Burgess Meredith Kevin Pollak None None None None None None None None None N
31357 Whitney Houston Angela Bassett Loretta Devine Kimberly Williams-Paisley Lela Rochon Gregory Hines Dennis Haystbert Michael Beach Mykelti Williamson Lamont Johnson Wesley Snipes None None None None N
11862 Steve Martin Diane Keaton Martin Short Williams-Paisley George Newbern Kieran Culkin BD Wong Peter Michael Goetz Kate McGregor-Stewart Jane Adams Eugene Levy Lori Alan None None N
...
439050 Leila Hatami Kourosh Tahami Elham Korda None N
111109 Angel Aquino Perry Dizon Hazel Orenco Joel Torre Bart Guingona Soliman Cruz Roeder Angeli Bayani Dante Perez Betty Uy-Regala Modesta None None None N
67758 Erika Eleniak Adam Baldwin Julie du Page James Remar Damian Chapa Louis Mandylor Tom Wright Jeremy Lelliott James Quattrochi Jason Sabatino Joe Ellsworth Kiko Widener Don Swayze Peter Dobson Da Dubo
227506 Iwan Mosselchuchin Natalie Lissenko Pavel Pavlov Aleksandr Chabrov Vera Orlova None N
461257 NaN !
```

44691 rows × 313 columns

In [114]: act = act.stack().reset\_index(level = 1, drop=True).to\_frame()

In [115]: act.columns = ["Actor"]

In [116]: act

```

Out[116]:
 Actor
 id
862 Tom Hanks
862 Tim Allen
862 Don Rickles
862 Jim Varney
862 Wallace Shawn
...
227506 Iwan Mosselchuchin
227506 Natalie Lissenko
227506 Pavel Pavlov
227506 Aleksandr Chabrov
227506 Vera Orlova
557703 rows × 1 columns
```

In [117]: act = act.merge(df[["title", "revenue\_musd", "vote\_average", "popularity"]], how = "left", left\_index = True, right\_index = True)

In [118]: act

```

Out[118]:
 Actor title revenue_musd vote_average popularity
 id
```

|        |                    |                                 |     |      |      |
|--------|--------------------|---------------------------------|-----|------|------|
| 2      | Turo Pajala        | Ariel                           | NaN | 7.10 | 3.86 |
| 2      | Susanna Haavisto   | Ariel                           | NaN | 7.10 | 3.86 |
| 2      | Matti Pellonpää    | Ariel                           | NaN | 7.10 | 3.86 |
| 2      | Eetu Hilkamo       | Ariel                           | NaN | 7.10 | 3.86 |
| 3      | Matti Pellonpää    | Shadows in Paradise             | NaN | 7.10 | 2.29 |
| ...    | ...                | ...                             | ... | ...  | ...  |
| 469172 | Vasco Sequeira     | Manuel on the Island of Wonders | NaN | NaN  | 0.00 |
| 469172 | Armando Bacelar    | Manuel on the Island of Wonders | NaN | NaN  | 0.00 |
| 469172 | Rafael de Sousa    | Manuel on the Island of Wonders | NaN | NaN  | 0.00 |
| 469172 | José Antônio Gomes | Manuel on the Island of Wonders | NaN | NaN  | 0.00 |
| 469172 | Isabel Branco      | Manuel on the Island of Wonders | NaN | NaN  | 0.00 |

557703 rows × 5 columns

In [119]: `act.Actor.unique()`

Out[119]: 201501

In [120]: `act.Actor.unique()`

Out[120]: array(['Turo Pajala', 'Susanna Haavisto', 'Matti Pellonpää', ..., 'Armando Bacelar', 'Rafael de Sousa', 'José Antônio Gomes'], dtype=object)

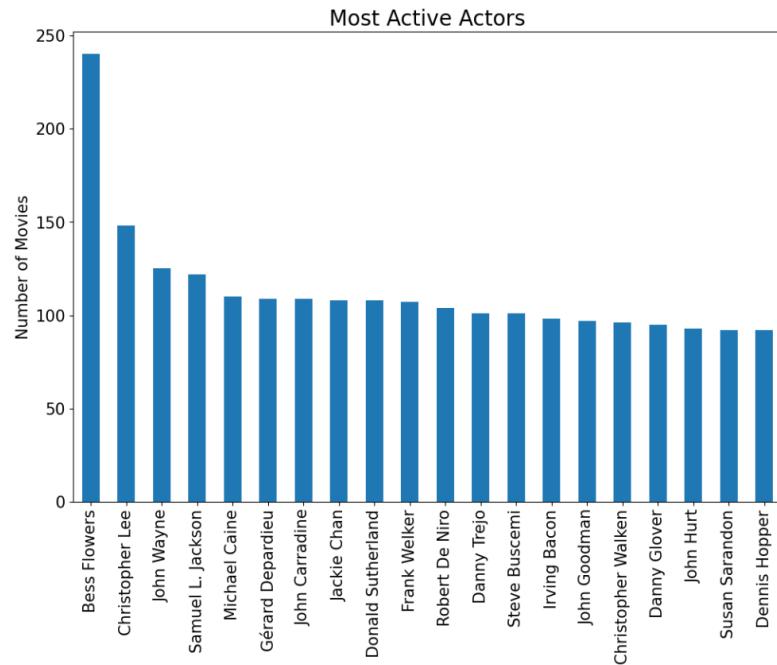
In [121]: `act.Actor.value_counts().head(20)`

Out[121]:

|                    |     |
|--------------------|-----|
| Bess Flowers       | 240 |
| Christopher Lee    | 148 |
| John Wayne         | 125 |
| Samuel L. Jackson  | 122 |
| Michael Caine      | 118 |
| Gérard Depardieu   | 109 |
| John Carradine     | 109 |
| Jackie Chan        | 108 |
| Donald Sutherland  | 108 |
| Frank Welker       | 107 |
| Robert De Niro     | 104 |
| Danny Trejo        | 101 |
| Steve Buscemi      | 101 |
| Irving Bacon       | 98  |
| John Goodman       | 97  |
| Christopher Walken | 96  |
| Danny Glover       | 95  |
| John Hurt          | 93  |
| Susan Sarandon     | 92  |
| Dennis Hopper      | 92  |

Name: Actor, dtype: int64

In [123]: `plt.figure(figsize = (12,8))  
act.Actor.value_counts().head(20).plot(kind="bar", fontsize = 15)  
plt.title("Most Active Actors", fontsize = 20)  
plt.xlabel("Number of Movies", fontsize = 15)  
plt.show()`



In [124]: `agg = act.groupby("Actor").agg(Total_Revenue = ("revenue_musd", "sum"),  
Mean_Revenue = ("revenue_musd", "mean"),  
Mean_Rating = ("vote_average", "mean"),  
Mean_Pop = ("popularity", "mean"),  
Total_Movies = ("Actor", "count"))`

In [125]: `agg.nlargest(10, "Total_Movies")`

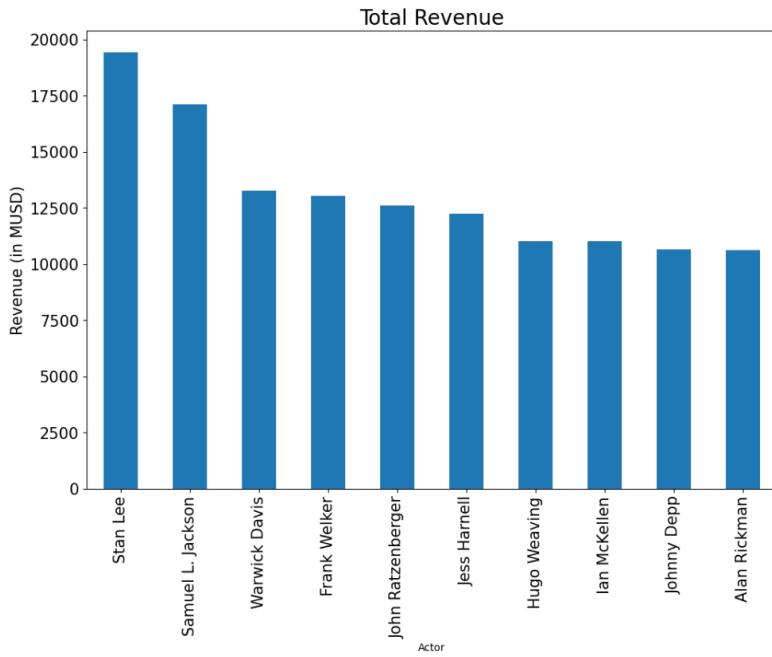
| Actor             | Total_Revenue | Mean_Revenue | Mean_Rating | Mean_Pop | Total_Movies |
|-------------------|---------------|--------------|-------------|----------|--------------|
| Bess Flowers      | 368.91        | 14.76        | 6.18        | 2.03     | 240          |
| Christopher Lee   | 9417.05       | 324.73       | 5.91        | 4.75     | 148          |
| John Wayne        | 236.09        | 11.24        | 5.71        | 3.09     | 125          |
| Samuel L. Jackson | 17109.62      | 213.87       | 6.27        | 11.70    | 122          |
| Michael Caine     | 8053.40       | 191.75       | 6.27        | 8.27     | 110          |
| Gérard Depardieu  | 1247.61       | 95.97        | 6.05        | 3.70     | 109          |
| John Carradine    | 255.84        | 19.68        | 5.55        | 2.43     | 109          |
| Donald Sutherland | 5390.77       | 138.22       | 6.23        | 7.00     | 108          |
| Jackie Chan       | 4699.19       | 146.85       | 6.28        | 5.86     | 108          |
| Frank Welker      | 13044.15      | 326.10       | 6.31        | 9.57     | 107          |

In [126]: `agg.nlargest(10, "Total_Revenue")`

| Actor    | Total_Revenue | Mean_Revenue | Mean_Rating | Mean_Pop | Total_Movies |
|----------|---------------|--------------|-------------|----------|--------------|
| Stan Lee | 19414.96      | 647.17       | 6.51        | 29.94    | 48           |
| ...      | ...           | ...          | ...         | ...      | ...          |
| ...      | ...           | ...          | ...         | ...      | ...          |
| ...      | ...           | ...          | ...         | ...      | ...          |

|                   |          |        |      |       |     |
|-------------------|----------|--------|------|-------|-----|
| Samuel L. Jackson | 17109.64 | £13.81 | 6.21 | 11.0  | 142 |
| Warwick Davis     | 13256.03 | 662.80 | 6.29 | 13.09 | 34  |
| Frank Welker      | 13044.15 | 326.10 | 6.31 | 9.57  | 107 |
| John Ratzenberger | 12596.13 | 449.86 | 6.48 | 10.96 | 46  |
| Jess Harnell      | 12234.61 | 611.73 | 6.44 | 10.92 | 35  |
| Hugo Weaving      | 11027.58 | 459.48 | 6.47 | 10.97 | 40  |
| Ian McKellen      | 11015.59 | 478.94 | 6.35 | 15.45 | 44  |
| Johnny Depp       | 10653.76 | 217.42 | 6.44 | 12.38 | 69  |
| Alan Rickman      | 10612.63 | 353.75 | 6.72 | 10.40 | 45  |

```
In [127]: plt.figure(figsize = (12,8))
agg.Total_Revenue.nlargest(10).plot(kind='bar', fontsize = 15)
plt.title("Total Revenue", fontsize = 20)
plt.ylabel("Revenue (in MUSD)", fontsize = 15)
plt.show()
```



```
In [128]: agg.Mean_Revenue.nlargest(10)
```

```
Out[128]: Actor
April Marie Thomas 2787.97
Ashley Jeffery 2787.97
Austin Wilson 2787.97
Brandon Jelkes 2787.97
Bravita A. Threatt 2787.97
Carvon Futrell 2787.97
Chris Mala 2787.97
Christa Oliver 2787.97
Christopher Nolen 2787.97
Colin Bleasdale 2787.97
Name: Mean_Revenue, dtype: float64
```

```
In [130]: act[act.Actor == "Tom Hardy"]
```

```
Out[130]:
 id Actor title revenue_musd vote_average popularity
0 201 Tom Hardy Star Trek: Nemesis 67.31 6.10 11.37
1 855 Tom Hardy Black Hawk Down 172.99 7.20 10.06
2 1887 Tom Hardy Marie Antoinette 60.47 6.50 8.85
3 4836 Tom Hardy Layer Cake 11.85 7.00 9.64
4 6309 Tom Hardy Flood NaN 4.50 6.50
5 13809 Tom Hardy RockNRolla 25.74 6.90 7.26
6 14979 Tom Hardy Thick as Thieves NaN 5.70 11.92
7 15069 Tom Hardy WAZ NaN 5.50 10.54
8 15473 Tom Hardy LD 50 Lethal Dose NaN 3.00 0.86
9 18533 Tom Hardy Bronson 2.26 6.90 9.52
10 18538 Tom Hardy Dot the I NaN 6.00 4.64
11 18558 Tom Hardy Scenes of a Sexual Nature NaN 5.20 3.03
12 25935 Tom Hardy The Reckoning NaN 6.40 3.53
13 27205 Tom Hardy Inception 825.53 8.10 29.11
14 36597 Tom Hardy Wuthering Heights NaN 6.90 8.75
15 37924 Tom Hardy Sweeney Todd NaN 6.30 4.00
16 49026 Tom Hardy The Dark Knight Rises 1084.94 7.60 20.58
17 49517 Tom Hardy Tinker Tailor Soldier Spy NaN 6.60 7.73
18 50032 Tom Hardy Stuart A Life Backwards NaN 7.70 4.82
19 59440 Tom Hardy Warrior 23.06 7.70 13.12
20 59958 Tom Hardy Gideon's Daughter NaN 6.00 1.17
21 59962 Tom Hardy This Means War 156.97 5.90 12.37
22 63971 Tom Hardy Deserter NaN 5.30 1.35
23 65096 Tom Hardy EMR NaN 5.00 0.88
24 76341 Tom Hardy Mad Max: Fury Road 378.86 7.30 29.36
25 82633 Tom Hardy Lawless 53.68 6.80 7.97
26 154400 Tom Hardy The Drop 19.05 6.70 11.70
27 172577 Tom Hardy The Take NaN 7.70 2.77
28 181283 Tom Hardy Child 44 3.32 6.10 17.61
29 210479 Tom Hardy Locke 4.64 6.90 12.79
30 276907 Tom Hardy Legend 42.97 6.70 11.19
31 281957 Tom Hardy The Revenant 532.95 7.30 23.50
32 328483 Tom Hardy London Road NaN 6.50 2.58
33 331214 Tom Hardy Band of Brothers NaN 8.20 7.90
34 374720 Tom Hardy Dunkirk 519.88 7.50 30.94
```

```
In [131]: agg[agg.Total_Movies >= 10].nlargest(10, "Mean_Revenue")
```

In [132]:

```
agg[agg.Total_Movies >= 10].nlargest(10, "Mean_Rating")
```

Out[132]:

| Actor            | Total_Revenue | Mean_Revenue | Mean_Rating | Mean_Pop | Total_Movies |
|------------------|---------------|--------------|-------------|----------|--------------|
| Gloria Stuart    | 1845.03       | 1845.03      | 6.37        | 3.48     | 18           |
| Keith Richards   | 2967.71       | 989.24       | 6.46        | 5.03     | 23           |
| Zoe Wanamaker    | 976.48        | 976.48       | 6.33        | 6.82     | 10           |
| James Cameron    | 1862.08       | 931.04       | 7.06        | 4.69     | 12           |
| Matthew Lewis    | 7915.31       | 879.48       | 7.37        | 23.10    | 11           |
| Luke de Wolfson  | 1720.67       | 860.34       | 5.72        | 8.77     | 11           |
| Yuri Lowenthal   | 1708.16       | 854.08       | 6.19        | 19.88    | 17           |
| Dominic Monaghan | 3289.61       | 822.40       | 6.05        | 10.62    | 11           |
| Philip Ng        | 821.71        | 821.71       | 5.92        | 6.35     | 10           |
| Peter Mayhew     | 4820.72       | 803.45       | 6.70        | 12.30    | 11           |

In [133]:

```
agg[agg.Total_Movies >= 10].nlargest(10, "Mean_Pop")
```

Out[133]:

| Actor              | Total_Revenue | Mean_Revenue | Mean_Rating | Mean_Pop | Total_Movies |
|--------------------|---------------|--------------|-------------|----------|--------------|
| Masao Hayama       | 0.00          | NaN          | 8.84        | 0.32     | 10           |
| David Attenborough | 0.00          | NaN          | 8.27        | 2.15     | 11           |
| Emil Jannings      | 0.00          | NaN          | 7.78        | 1.70     | 10           |
| Halit Akgatope     | 0.21          | 0.21         | 7.78        | 0.74     | 10           |
| Yo Oizumi          | 511.21        | 102.24       | 7.72        | 7.51     | 13           |
| Sener Sén          | 11.07         | 3.69         | 7.69        | 0.91     | 16           |
| Ayşen Gruda        | 0.91          | 0.46         | 7.68        | 0.74     | 10           |
| Akira Tani         | 0.33          | 0.16         | 7.65        | 5.04     | 12           |
| Joseph Oliveira    | 3543.44       | 354.34       | 7.64        | 34.45    | 10           |
| Bjork              | 51.33         | 25.67        | 7.62        | 2.36     | 10           |

In [ ]: