```
[28]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45466 entries, 0 to 45465
Data columns (total 24 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   adult                  45466 non-null   object
 1   belongs_to_collection  4494 non-null    object
 2   budget                 45466 non-null   object
 3   genres                 45466 non-null   object
 4   homepage               7782 non-null    object
 5   id                     45466 non-null   object
 6   imdb_id                45449 non-null   object
 7   original_language      45455 non-null   object
 8   original_title         45466 non-null   object
 9   overview               44512 non-null   object
 10  popularity             45461 non-null   object
 11  poster_path            45080 non-null   object
 12  production_companies   45463 non-null   object
 13  production_countries   45463 non-null   object
 14  release_date           45379 non-null   object
 15  revenue                45460 non-null   float64
 16  runtime                45203 non-null   float64
 17  spoken_languages       45460 non-null   object
 18  status                 45379 non-null   object
 19  tagline                20412 non-null   object
 20  title                  45460 non-null   object
 21  video                  45460 non-null   object
 22  vote_average           45460 non-null   float64
 23  vote_count             45460 non-null   float64
dtypes: float64(4), object(20)
memory usage: 8.3+ MB
```
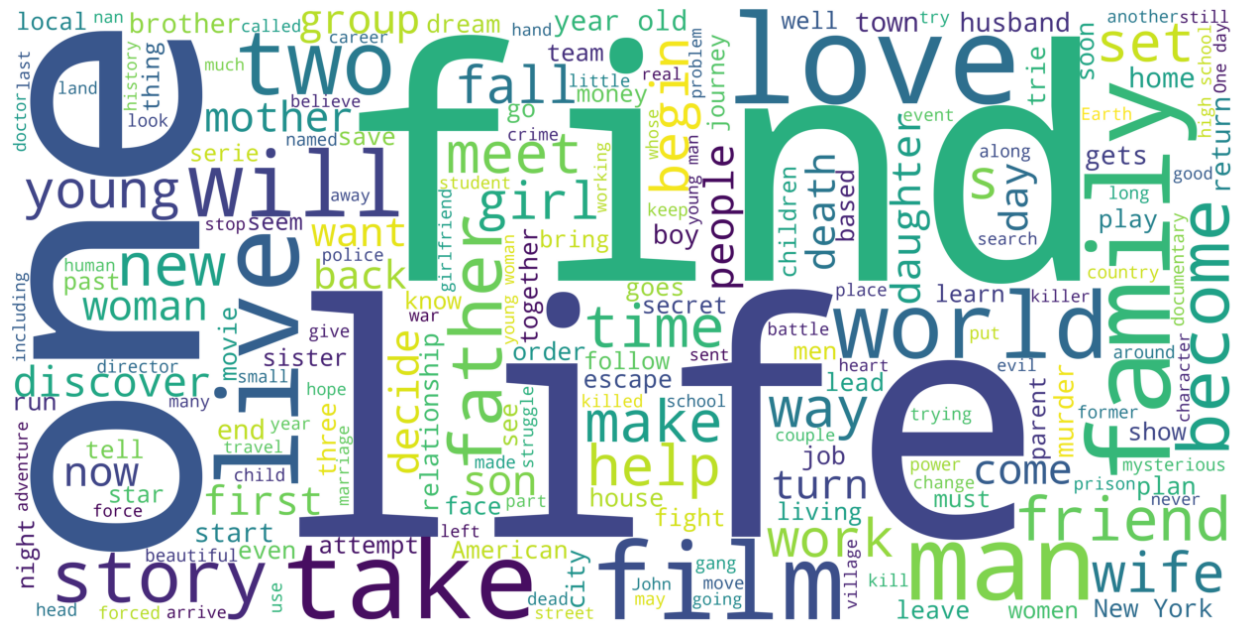
Title

Overview

Production Countries for the MovieLens Movies (Apart from US)

	belongs_to_collection	count	mean	sum
552	Harry Potter Collection	8	9.634209e+08	7.707367e+09
1160	Star Wars Collection	8	9.293118e+08	7.434495e+09
646	James Bond Collection	26	2.733450e+08	7.106970e+09
1317	The Fast and the Furious Collection	8	6.406373e+08	5.125099e+09
968	Pirates of the Caribbean Collection	5	9.043154e+08	4.521577e+09
1550	Transformers Collection	5	8.732202e+08	4.366101e+09
325	Despicable Me Collection	4	9.227676e+08	3.691070e+09
1491	The Twilight Collection	5	6.684215e+08	3.342107e+09
610	Ice Age Collection	5	6.433417e+08	3.216709e+09
666	Jurassic Park Collection	4	7.578710e+08	3.031484e+09

	belongs_to_collection	count	mean	sum
646	James Bond Collection	26	2.733450e+08	7.106970e+09
473	Friday the 13th Collection	12	3.874155e+07	4.648985e+08
976	Pokémon Collection	11	6.348189e+07	6.983008e+08
552	Harry Potter Collection	8	9.634209e+08	7.707367e+09
540	Halloween Collection	8	3.089601e+07	2.471681e+08
29	A Nightmare on Elm Street Collection	8	4.544894e+07	3.635916e+08
1317	The Fast and the Furious Collection	8	6.406373e+08	5.125099e+09
1432	The Pink Panther (Original) Collection	8	2.055978e+07	1.644782e+08
1160	Star Wars Collection	8	9.293118e+08	7.434495e+09
977	Police Academy Collection	7	4.352046e+07	3.046432e+08

```
]: com_pivot.sort_values('Total', ascending=False).head(10)
```
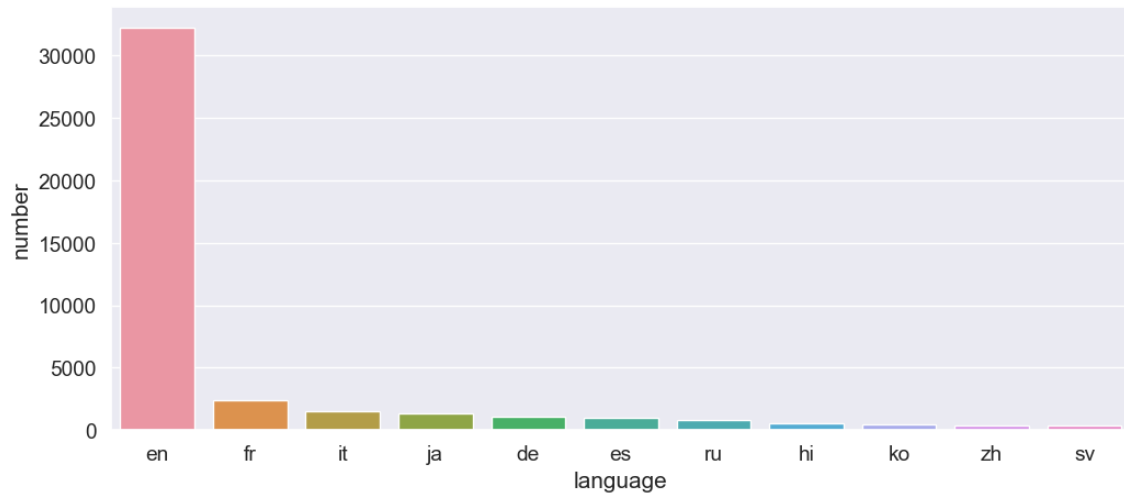
	Total	Average	Number
companies			
Warner Bros.	6.352519e+10	1.293792e+08	491
Universal Pictures	5.525919e+10	1.193503e+08	463
Paramount Pictures	4.880819e+10	1.235650e+08	395
Twentieth Century Fox Film Corporation	4.768775e+10	1.398468e+08	341
Walt Disney Pictures	4.083727e+10	2.778046e+08	147
Columbia Pictures	3.227974e+10	1.367785e+08	236
New Line Cinema	2.217339e+10	1.119868e+08	198
Amblin Entertainment	1.734372e+10	2.550547e+08	68
DreamWorks SKG	1.547575e+10	1.984071e+08	78
Dune Entertainment	1.500379e+10	2.419966e+08	62

```
[59]: com_pivot[com_pivot['Number'] >= 15].sort_values('Average', ascending=False).head(10)
```
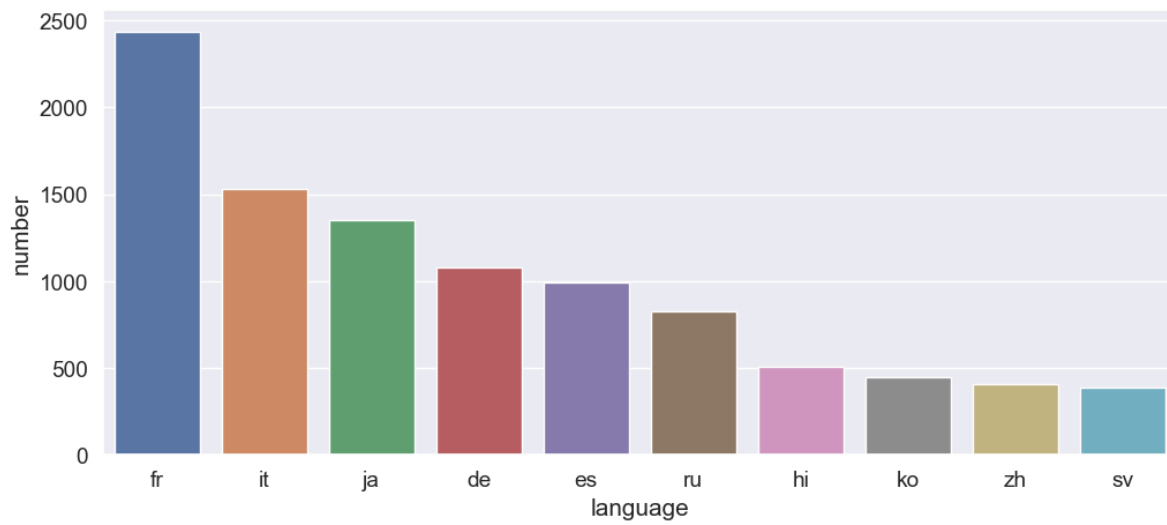
[59]:

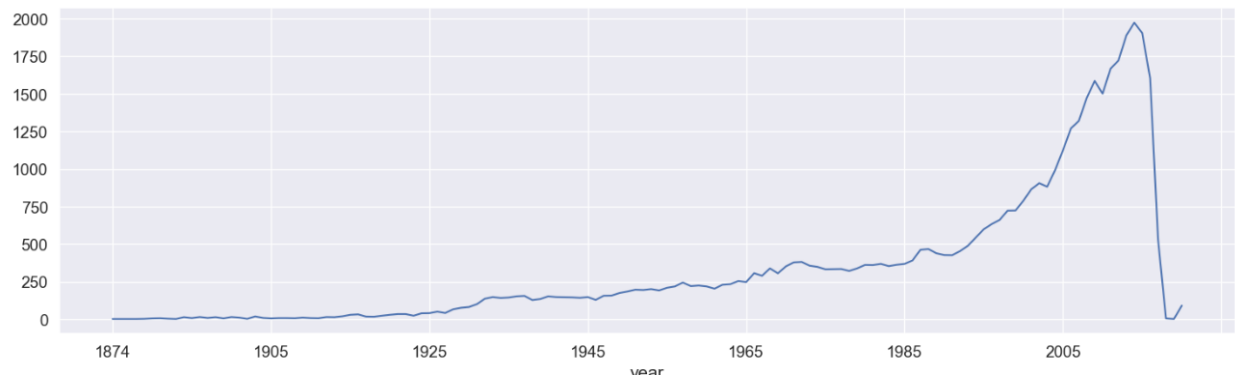|	Total	Average	Number
companies			
Pixar Animation Studios	1.118853e+10	6.215852e+08	18
Marvel Studios	1.169964e+10	6.157703e+08	19
Revolution Sun Studios	8.120339e+09	5.413559e+08	15
Lucasfilm	9.898421e+09	4.499282e+08	22
DreamWorks Animation	1.370752e+10	4.031622e+08	34
DC Entertainment	6.212609e+09	3.882880e+08	16
Dentsu	6.853205e+09	3.807336e+08	18
Jerry Bruckheimer Films	8.957441e+09	3.732267e+08	24
Marvel Enterprises	6.538067e+09	3.441088e+08	19
Legendary Pictures	1.346866e+10	3.367166e+08	40

```
plt.figure(figsize=(12,5))
sns.barplot(x='language', y='number', data=lang_df.iloc[0:11])
plt.show()
```



```
plt.figure(figsize=(12,5))
sns.barplot(x='language', y='number', data=lang_df.iloc[1:11])
plt.show()
```
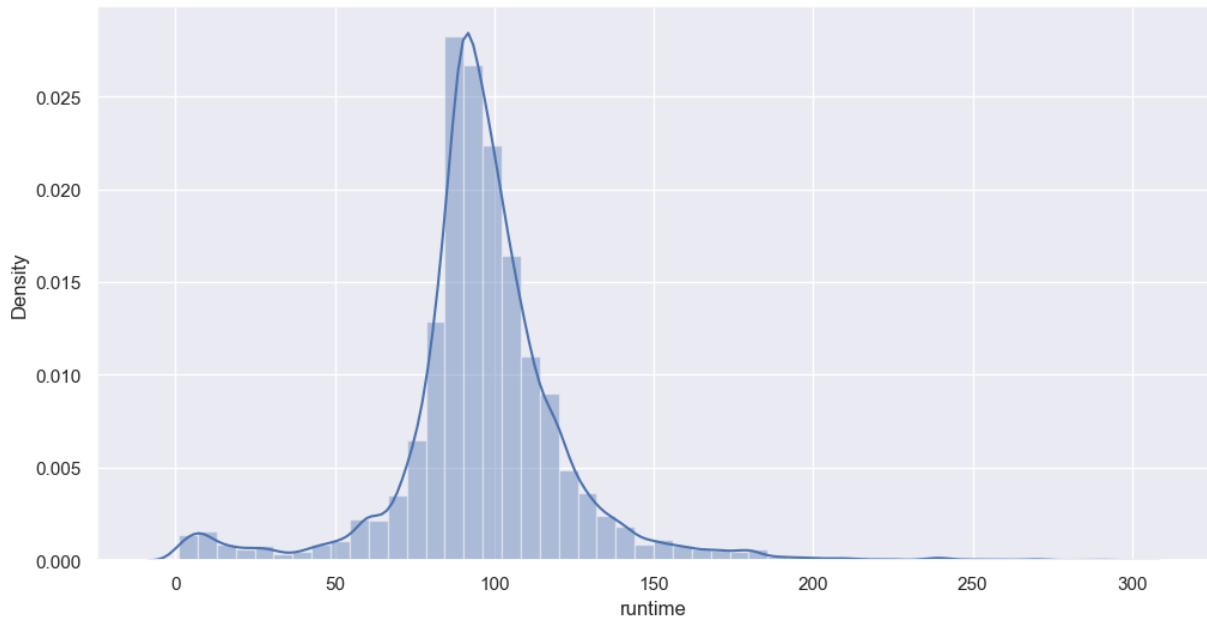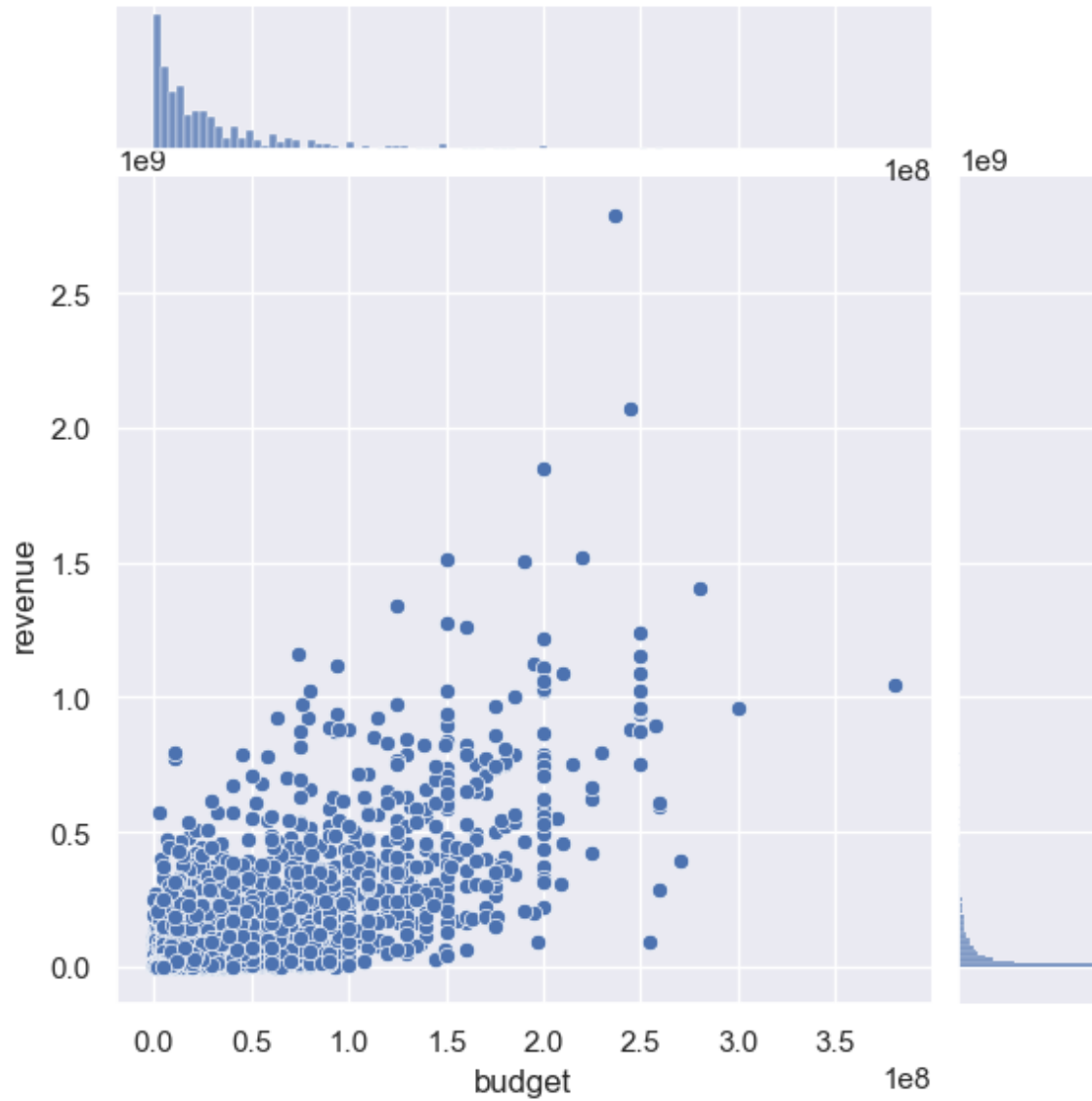
<Axes: xlabel='year'>

|: <Axes: xlabel='runtime', ylabel='Density'>

```python
sns.jointplot(x='budget',y='revenue',data=df[df['return'].notnull()])
```

<seaborn.axisgrid.JointGrid at 0x174d71c3e10>

```
[117]:  df[(df['return'].notnull()) & (df['budget'] > 5e6)][['title', 'budget', 'revenue', 'return', 'year']].sort_values('return', ascending=False).head(10)
```

[117]:

	title	budget	revenue	return	year
1065	E.T. the Extra-Terrestrial	10500000.0	792965326.0	75.520507	1982
256	Star Wars	11000000.0	775398007.0	70.490728	1977
1338	Jaws	7000000.0	470654000.0	67.236286	1975
1888	The Exorcist	8000000.0	441306145.0	55.163268	1973
352	Four Weddings and a Funeral	6000000.0	254700832.0	42.450139	1994
834	The Godfather	6000000.0	245066411.0	40.844402	1972
4492	Look Who's Talking	7500000.0	296000000.0	39.466667	1989
24258	Annabelle	6500000.0	255273813.0	39.272894	2014
1056	Dirty Dancing	6000000.0	213954274.0	35.659046	1987
1006	The Sound of Music	8200000.0	286214286.0	34.904181	1965

```
[118]:  df[(df['return'].notnull()) & (df['budget'] > 5e6) & (df['revenue'] > 10000)][['title', 'budget', 'revenue', 'return', 'year']].sort_values('return').head(10)
```

[118]:

	title	budget	revenue	return	year
11159	Chaos	20000000.0	10289.0	0.000514	2005
19027	5 Days of War	20000000.0	17479.0	0.000874	2011
21034	Special Forces	10000000.0	10759.0	0.001076	2011
25732	Foodfight!	65000000.0	73706.0	0.001134	2012
38388	Term Life	16500000.0	21256.0	0.001288	2016
19505	Laurence Anyways	9500000.0	12250.0	0.001289	2012
12038	The Good Night	15000000.0	20380.0	0.001359	2007
3966	Cherry 2000	10000000.0	14000.0	0.001400	1987
22097	Twice Born	13000000.0	18295.0	0.001407	2012
5651	All The Queen's Men	15000000.0	23000.0	0.001533	2001

Stacked Bar Chart of Movie Proportions by Genre