```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
modf=pd.read_csv('/content/drive/MyDrive/Copy of moviesData (1).csv')
modf
```

	budget	genres	homepage	id	
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	"I
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id "oce
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id "s
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id "dc
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id "ba
...	...	...	...	...	
4798	220000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	NaN	9367	' state
4799	9000	[{"id": 35, "name": "Comedy"}, {"id": 10749, "...	NaN	72766	
4800	0	[{"id": 35, "name": "Comedy"}, {"id": 18, "nam...	http://www.hallmarkchannel.com/signedsealeddel...	231617	[{"id "da
4801	0	[]	http://shanghaicalling.com/	126186	
4802	0	[{"id": 99, "name": "Documentary"}]	NaN	25975	"ob

4803 rows × 20 columns

```
modf.columns
```

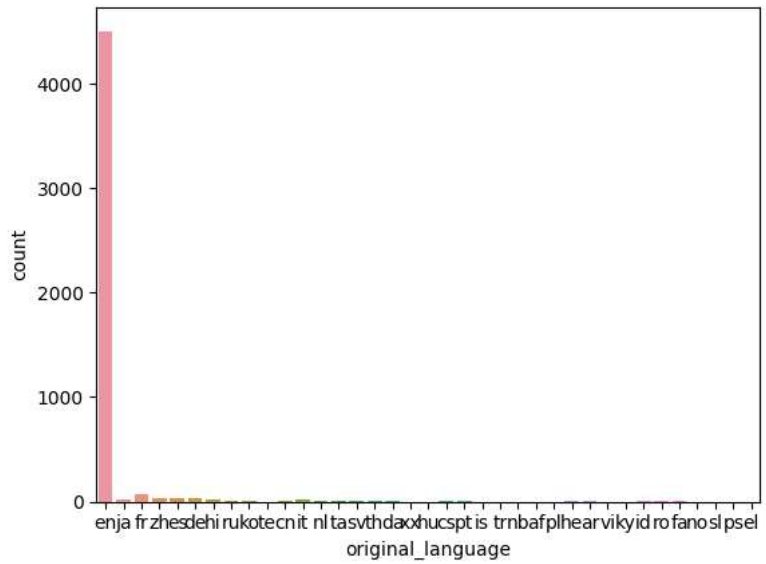
```
Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
       'original_title', 'overview', 'popularity', 'production_companies',
       'production_countries', 'release_date', 'revenue', 'runtime',
```

```
       'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
       'vote_count'],
      dtype='object')
```

```
modf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   budget                4803 non-null   int64  
 1   genres                4803 non-null   object 
 2   homepage              1712 non-null   object 
 3   id                    4803 non-null   int64  
 4   keywords              4803 non-null   object 
 5   original_language     4803 non-null   object 
 6   original_title        4803 non-null   object 
 7   overview              4800 non-null   object 
 8   popularity            4803 non-null   float64
 9   production_companies  4803 non-null   object 
 10  production_countries  4803 non-null   object 
 11  release_date          4802 non-null   object 
 12  revenue               4803 non-null   int64  
 13  runtime               4801 non-null   float64
 14  spoken_languages      4803 non-null   object 
 15  status                4803 non-null   object 
 16  tagline               3959 non-null   object 
 17  title                 4803 non-null   object 
 18  vote_average          4803 non-null   float64
 19  vote_count            4803 non-null   int64  
dtypes: float64(3), int64(4), object(13)
memory usage: 750.6+ KB
```

```
sb.countplot(x='original_language',data=modf)
```




```
sb.countplot(x='original_language',data=modf)
plt.xticks(rotation='vertical')
```

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
        34, 35, 36]),
 [Text(0, 0, 'en'),
  Text(1, 0, 'ja'),
  Text(2, 0, 'fr'),
  Text(3, 0, 'zh'),
  Text(4, 0, 'es'),
  Text(5, 0, 'de'),
  Text(6, 0, 'hi'),
  Text(7, 0, 'ru'),
  Text(8, 0, 'ko'),
  Text(9, 0, 'te'),
  Text(10, 0, 'cn'),
  Text(11, 0, 'it'),
  Text(12, 0, 'nl'),
  Text(13, 0, 'ta'),
  Text(14, 0, 'sv'),
  Text(15, 0, 'th'),
  Text(16, 0, 'da'),
  Text(17, 0, 'xx'),
  Text(18, 0, 'hu'),
  Text(19, 0, 'cs'),
  Text(20, 0, 'pt'),
  Text(21, 0, 'is'),
  Text(22, 0, 'tr'),
  Text(23, 0, 'nb'),
  Text(24, 0, 'af'),
  Text(25, 0, 'pl'),
  Text(26, 0, 'he'),
  Text(27, 0, 'ar'),
  Text(28, 0, 'vi'),
  Text(29, 0, 'ky'),
  Text(30, 0, 'id'),
  Text(31, 0, 'ro'),
  Text(32, 0, 'fa'),
  Text(33, 0, 'no'),
  Text(34, 0, 'sl'),
  Text(35, 0, 'ps'),
  Text(36, 0, 'el')])
```


```
df1=modf[modf['original_language']!='en']
sb.countplot(x='original_language',data=df1)
plt.xticks(rotation='vertical')
```

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
        34, 35]),
 [Text(0, 0, 'ja'),
  Text(1, 0, 'fr'),
  Text(2, 0, 'zh'),
  Text(3, 0, 'es'),
  Text(4, 0, 'de'),
  Text(5, 0, 'hi'),
  Text(6, 0, 'ru'),
  Text(7, 0, 'ko'),
  Text(8, 0, 'te'),
  Text(9, 0, 'cn'),
  Text(10, 0, 'it'),
  Text(11, 0, 'nl'),
  Text(12, 0, 'ta'),
  Text(13, 0, 'sv'),
  Text(14, 0, 'th'),
  Text(15, 0, 'da'),
  Text(16, 0, 'xx'),
  Text(17, 0, 'hu'),
  Text(18, 0, 'cs'),
  Text(19, 0, 'pt'),
  Text(20, 0, 'is'),
  Text(21, 0, 'tr'),
  Text(22, 0, 'nb'),
  Text(23, 0, 'af'),
  Text(24, 0, 'pl'),
  Text(25, 0, 'he'),
  Text(26, 0, 'ar'),
  Text(27, 0, 'vi'),
  Text(28, 0, 'ky'),
  Text(29, 0, 'id'),
  Text(30, 0, 'ro'),
  Text(31, 0, 'fa'),
  Text(32, 0, 'no'),
```

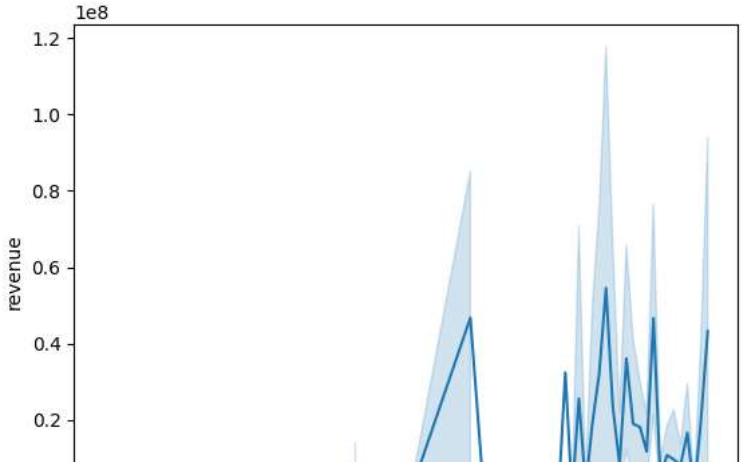
```
modf['release_date']=pd.to_datetime(df1['release_date'])
```

```
modf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype         
---  ------                --------------  -----         
 0   budget                4803 non-null   int64         
 1   genres                4803 non-null   object        
 2   homepage              1712 non-null   object        
 3   id                    4803 non-null   int64         
 4   keywords              4803 non-null   object        
 5   original_language     4803 non-null   object        
 6   original_title        4803 non-null   object        
 7   overview              4800 non-null   object        
 8   popularity            4803 non-null   float64       
 9   production_companies  4803 non-null   object        
 10  production_countries  4803 non-null   object        
 11  release_date          298 non-null    datetime64[ns]
 12  revenue               4803 non-null   int64         
 13  runtime               4801 non-null   float64       
 14  spoken_languages      4803 non-null   object        
 15  status                4803 non-null   object        
 16  tagline               3959 non-null   object        
 17  title                 4803 non-null   object        
 18  vote_average          4803 non-null   float64       
 19  vote_count            4803 non-null   int64         
 20  year                  298 non-null    float64       
dtypes: datetime64[ns](1), float64(4), int64(4), object(12)
memory usage: 788.1+ KB
```

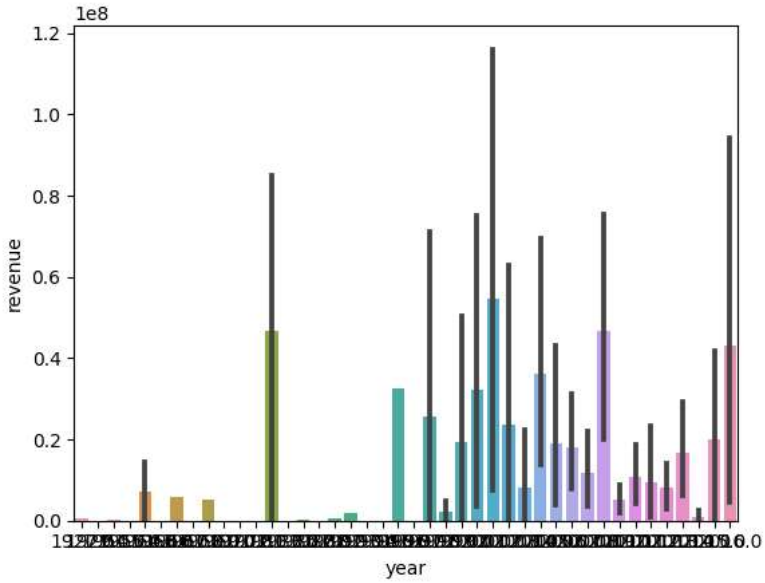
```
modf['year']=modf['release_date'].dt.year
sb.lineplot(x='year',y='revenue',data=modf)
```

```
<Axes: xlabel='year', ylabel='revenue'>
```




```
modf['year']=modf['release_date'].dt.year
sb.barplot(x='year',y='revenue',data=modf)
```

```
<Axes: xlabel='year', ylabel='revenue'>
```

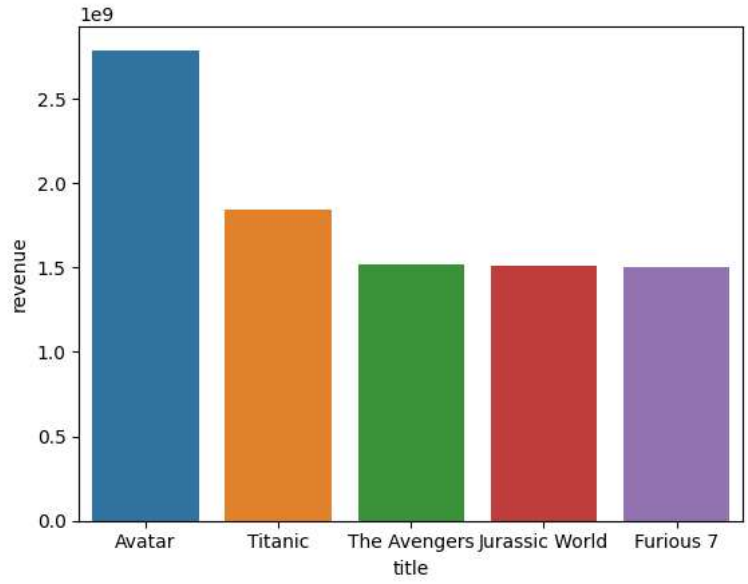



```
top5df=modf.nlargest(5,'revenue')
top5df
```

	budget	genres	homepage	id	keywords	origin
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
25	200000000	[{"id": 18, "name": "Drama"}, {"id":	http://www.titanicmovie.com	597	[{"id": 2580, "name": "shipwreck"},	

```
sb.barplot(x='title',y='revenue',data=top5df)
```

```
<Axes: xlabel='title', ylabel='revenue'>
```




```
modf['year']==2000
```

```
0       False
1       False
2       False
3       False
4       False
        ...
4798    False
4799    False
4800    False
4801    False
4802    False
Name: year, Length: 4803, dtype: bool
```

```
modf['release_date']=pd.to_datetime(df1['release_date'])
modf['year']=modf['release_date'].dt.year
```

```
modf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   budget                4803 non-null   int64
 1   genres                4803 non-null   object
 2   homepage              1712 non-null   object
 3   id                    4803 non-null   int64
 4   keywords              4803 non-null   object
 5   original_language     4803 non-null   object
 6   original_title        4803 non-null   object
 7   overview              4800 non-null   object
 8   popularity            4803 non-null   float64
 9   production_companies  4803 non-null   object
 10  production_countries  4803 non-null   object
 11  release_date          298 non-null    datetime64[ns]
 12  revenue               4803 non-null   int64
```

```
 13  runtime               4801 non-null   float64
 14  spoken_languages      4803 non-null   object
 15  status                4803 non-null   object
 16  tagline               3959 non-null   object
 17  title                 4803 non-null   object
 18  vote_average          4803 non-null   float64
 19  vote_count            4803 non-null   int64
 20  year                  298 non-null    float64
dtypes: datetime64[ns](1), float64(4), int64(4), object(12)
memory usage: 788.1+ KB
```

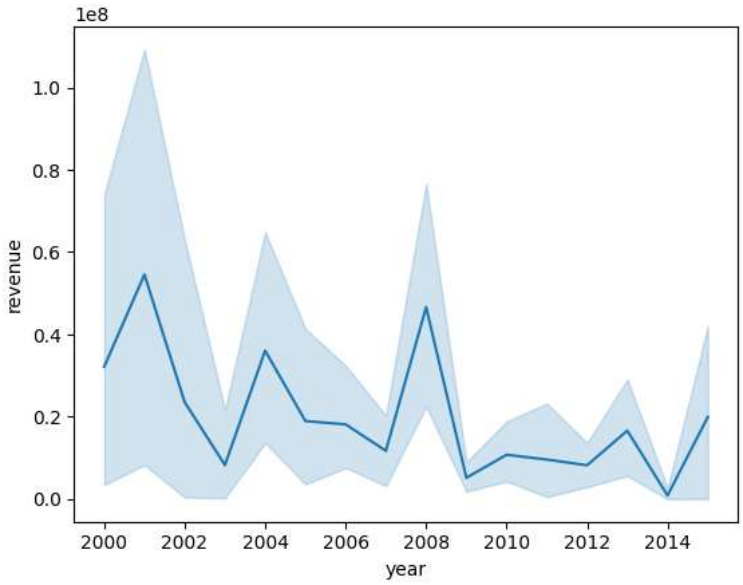
```
newdf=modf[(modf['year']>=2000) & (modf['year']<=2015)]
newdf
```

	budget	genres	homepage	id	
235	97250400	[{"id": 14, "name": "Fantasy"}, {"id": 12, "na...	http://www.asterixauxjeuxolympiques.com/index.php	2395	"c
317	94000000	[{"id": 18, "name": "Drama"}, {"id": 36, "name...	http://www.theflowersofwarmovie.com/	76758	[{ pl
474	0	[{"id": 9648, "name": "Mystery"}, {"id": 18, "...	NaN	330770	"l
492	8000000	[{"id": 35, "name": "Comedy"}, {"id": 16, "nam...	NaN	293644	[{ "l
678	65000000	[{"id": 28, "name": "Action"}, {"id": 18, "nam...	NaN	300168	
...	...	...	...	...	
4605	0	[{"id": 18, "name": "Drama"}]	http://www.dogtooth.gr/	38810	"r r
4677	0	[{"id": 10749, "name": "Romance"}, {"id": 18, ...	NaN	53256	
4684	0	[{"id": 27, "name": "Horror"}]	NaN	402515	"
4699	0	[{"id": 18, "name": "Drama"}]	NaN	344466	
4790	0	[{"id": 18, "name": "Drama"}, {"id": 10769, "n...	NaN	13898	

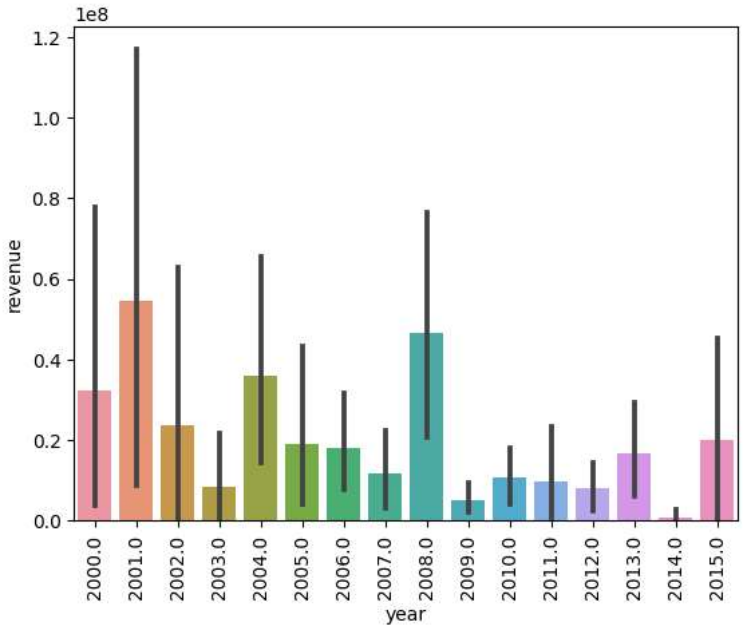
244 rows × 21 columns

```
newdf=modf[(modf['year']>=2000) & (modf['year']<=2015)]
newdf
```

```
sb.lineplot(x='year',y='revenue',data=newdf)
plt.show()
```




```
newdf=modf[(modf['year']>=2000) & (modf['year']<=2015)]
newdf
sb.barplot(x='year',y='revenue',data=newdf)
plt.xticks(rotation='vertical')
plt.show()
```




```
modf.columns
```

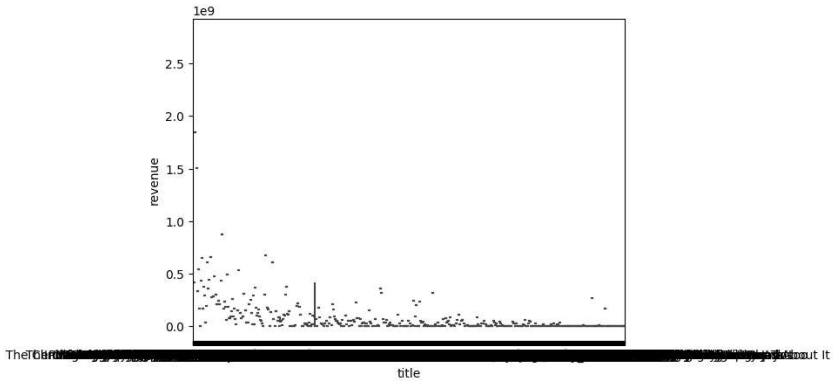
```
Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
       'original_title', 'overview', 'popularity', 'production_companies',
       'production_countries', 'release_date', 'revenue', 'runtime',
       'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
       'vote_count', 'year'],
      dtype='object')
```

```
sb.barplot
```

```
<function seaborn.categorical.barplot(data=None, *, x=None, y=None, hue=None, order=None, hue_order=None, estimator='mean', errorbar=
('ci', 95), n_boot=1000, units=None, seed=None, orient=None, color=None, palette=None, saturation=0.75, width=0.8, errcolor='.26',
errwidth=None, capsize=None, dodge=True, ci='deprecated', ax=None, **kwargs)>
```

```
sb.boxplot(x='title',y='revenue',data=modf)
```

```
<Axes: xlabel='title', ylabel='revenue'>
```




```
sb.heatmap(modf.isnull())
```

```
<Axes: >
```