

```
In [1]: import matplotlib.pyplot as plt
import pandas as pd
```

```
In [2]: df=pd.read_csv('IMDBMovieData.csv')
df.head()
```

Out[2]:

	Rank	Title	Genre	Description	Director	Actors	Year
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016

You are a studio that wants to create the best drama every! What is the current best drama, and what can we learn from it?

```
In [13]: best_current_drama=df[(df['Metascore']==df['Metascore'].max()) & (str(df
['Genre']).find('Drama')!=-1)]
best_current_drama
```

Out[13]:

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	R (N)
656	657	Boyhood	Drama	The life of Mason, from early childhood to his...	Richard Linklater	Ellar Coltrane, Patricia Arquette, Ethan Hawke...	2014	165	7.9	286722	

Lets hire the director, Richard Linklater, for this project. Who should our leading actor be? Someone with a long carreer of highly rated movies.

```
In [4]: actor_dict={}
actorspro=df["Actors"].values.tolist()
actors=[]
for x in actorspro:
    bruh=x.split(',')
    for y in bruh:
        actors.append(y)
counter = 0
best_actor = actors[0]
for x in actors:
    freq=actors.count(x)
    if(freq> counter):
        counter = freq
        best_actor = x
best_actor
```

Out[4]: 'Mark Wahlberg'

How much money should we expect to make? Some combination of genre, runtime, and rating. Assume a above 95 metascore

```
In [5]: avg_rev_drama=df[df['Genre']=='Drama']['Revenue (Millions)'].mean()
avg_rev_runtime=df[(df['Runtime (Minutes)']>150)&(df['Runtime (Minutes)']<180)]['Revenue (Millions)'].mean()
avg_rev_rating=df[df['Metascore']>95]['Revenue (Millions)'].mean()
expected_rev= (avg_rev_drama+avg_rev_runtime+avg_rev_rating)/3
expected_rev
```

Out[5]: 92.38294418842804

Not too good. Maybe we would do better in awards season. Lets get a supporting cast for Mark.

```
In [6]: actor_dict={}
actorspro=df["Actors"].values.tolist()
actors=[]
for x in actorspro:
    bruh=x.split(',')
    for y in bruh:
        actors.append(y)
counter = 0
count_1=0
count_2=0
best_actor = actors[0]
support_1=actors[1]
support_2=actors[2]
for x in actors:
    freq=actors.count(x)
    if(freq> counter):
        counter = freq
        best_actor = x
    elif(freq<counter and freq>count_1):
        count_1=freq
        support_1=x
    elif(freq<count_1 and freq>count_2):
        count_2=freq
        support_2=x
print(support_1 + '\n' + support_2)
```

Christian Bale
Tom Hardy

We have put in all our work, lets let the writing staff take it over from here.... They are done. Lets see how the movie did

```
In [7]: description='Have you ever heard the tragedy of Darth Plageuis the Wise? I thought not.'
df2=pd.DataFrame({'Rank':[1001], 'Title':['Dramatic Movie'], 'Genre':['Drama, Thriller'], 'Description':[description], 'Director':['Richard Linklater'], 'Actors':[(best_actor+', '+support_1+', '+support_2)], 'Year':[2020], 'Runtime (Minutes)': [155], 'Rating':[8.9], 'Votes':[286722], 'Revenue (Millions)':[expected_rev], 'Metascore': [97]})
df.append(df2)
```

Out[7]:

	Rank	Title	Genre	Description	Director	Actors	Ye
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	20
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	20
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	20
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	20
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	20
...
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	20
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	20
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	20
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	20
0	1001	Dramatic Movie	Drama, Thriller	Have you ever heard the tragedy of Darth Plag...	Richard Linklater	Mark Wahlberg, Christian Bale, Tom Hardy	20

1001 rows × 12 columns

Come up with an audience popularity score

```
In [8]: def pop_score(s):  
        return s/1000
```

```
In [9]: df['Audience_Pop']=df['Votes']*df['Revenue (Millions)']/10000  
df
```

Out[9]:

	Rank	Title	Genre	Description	Director	Actors	Ye
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	20
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	20
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	20
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	20
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	20
...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	20
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	20
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	20
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	20
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	20

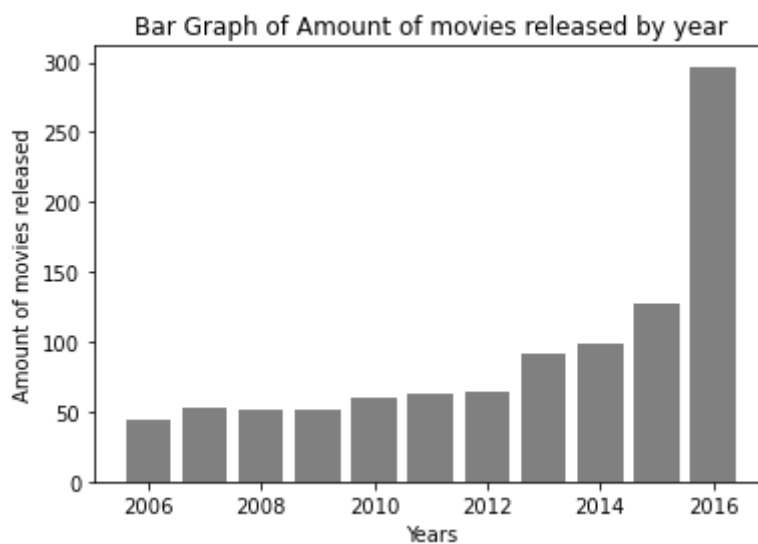
1000 rows × 13 columns

What year had the most movie

```
In [10]: x=df['Year'].value_counts().keys()
y=df['Year'].value_counts()
plt.bar(x, y, color = 'gray')

plt.title('Bar Graph of Amount of movies released by year')
plt.xlabel('Years')
plt.ylabel('Amount of movies released')
```

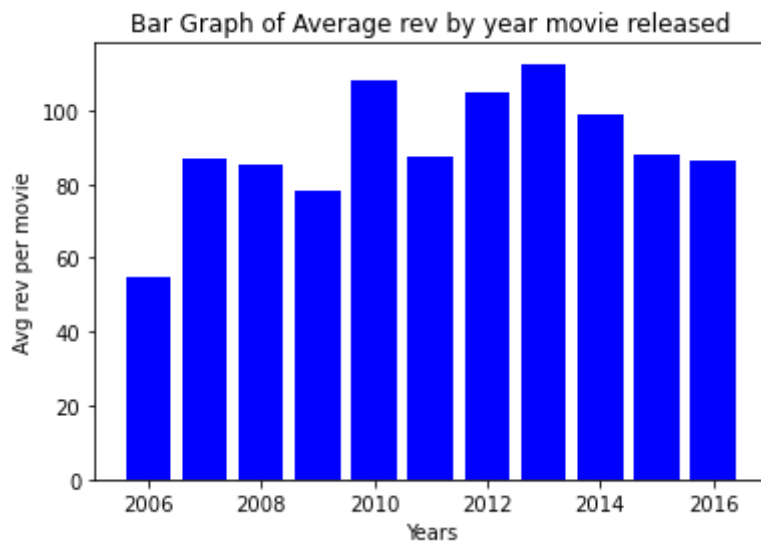
```
Out[10]: Text(0, 0.5, 'Amount of movies released')
```



What year had the highest grossing movies

```
In [11]: y= df.groupby('Year')['Revenue (Millions)'].mean()  
plt.bar(x, y, color = 'blue')  
  
plt.title('Bar Graph of Average rev by year movie released')  
plt.xlabel('Years')  
plt.ylabel('Avg rev per movie')
```

```
Out[11]: Text(0, 0.5, 'Avg rev per movie')
```



```
In [ ]:
```