In [6]: import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns

data= pd.read\_csv("D:\DATA ANALYST AND DATA SCIENCE\PYTHON\pandas project\IMDB-Movie-Data.csv") data





#### Out[6]:

: _	Rank		Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metasco, NoteG
	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	121	8.1	757074	333.13	76.0
	1	2	Prometheus	Adventure, Mystery, Sci-Fi	Following clues to the origin of mankind, a te $\label{eq:condition} te$	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa	2012	124	7.0	485820	126.46	65.0
	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	117	7.3	157606	138.12	62.0
	3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma	2016	108	7.2	60545	270.32	59.0
	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	123	6.2	393727	325.02	40.0
	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	2015	111	6.2	27585	NaN	45.0
	996	997	Hostel: Part II	Horror	Three American college students studying abroa	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli	2007	94	5.5	73152	17.54	46.0
	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,	2008	98	6.2	70699	58.01	50.0
	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	2014	93	5.6	4881	NaN	22.0
	999	1000	Nine Lives	Comedy, Family, Fantasy	A stuffy businessman finds himself trapped ins	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch	2016	87	5.3	12435	19.64	11.0

1000 rows × 12 columns

#### **DISPLAY TOP 10 ROWS OF THE DATASET**

In [8]: data.head(10)

Out[8]:





# CHECK LAST 10 ROWS OF THE DATASET

In [10]: data.tail(10)

١:	
_	

10]:		Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
	990	991	Underworld: Rise of the Lycans	Action, Adventure, Fantasy	An origins story centered on the centuries-old	Patrick Tatopoulos	Rhona Mitra, Michael Sheen, Bill Nighy, Steven	2009	92	6.6	129708	45.80	44.0
	991	992	Taare Zameen Par	Drama, Family, Music	An eight-year-old boy is thought to be a lazy	Aamir Khan	Darsheel Safary, Aamir Khan, Tanay Chheda, Sac	2007	165	8.5	102697	1.20	42.0
	992	993	Take Me Home Tonight	Comedy,Drama,Romance	Four years after graduation, an awkward high s	Michael Dowse	Topher Grace, Anna Faris, Dan Fogler, Teresa P	2011	97	6.3	45419	6.92	NaN
	993	994	Resident Evil: Afterlife	Action,Adventure,Horror	While still out to destroy the evil Umbrella $$C_{\cdots}$$	Paul W.S. Anderson	Milla Jovovich, Ali Larter, Wentworth Miller,K	2010	97	5.9	140900	60.13	37.0
	994	995	Project X	Comedy	3 high school seniors throw a birthday party t	Nima Nourizadeh	Thomas Mann, Oliver Cooper, Jonathan Daniel Br	2012	88	6.7	164088	54.72	48.0
	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	2015	111	6.2	27585	NaN	45.0
	996	997	Hostel: Part II	Horror	Three American college students studying abroa	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli		94	5.5	73152	17.54	46.0
	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,	2008	98	6.2	70699	58.01	50.0
	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	2014	93	5.6	4881	NaN	22.0
	999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch	2016	87	5.3	12435	19.64	11.0

# FIND SHAPE OF OUR DATASET(NUMBER OF ROWS AND NUMBER OF COLUMNS)



# GETTING INFORMATION ABOUT OUR DATASET LIKE TOTAL NUMBER OF COLUMNS, DATATYPES OF EACH COLUMN AND MEMORY REQUIREMENT



```
In [13]: data.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 1000 entries, 0 to 999
        Data columns (total 12 columns):
                              Non-Null Count Dtype
         # Column
                              _____
                              1000 non-null int64
             Rank
             Title
                              1000 non-null object
         1
             Genre
                              1000 non-null object
                              1000 non-null object
             Description
             Director
                               1000 non-null object
             Actors
                               1000 non-null
                                             object
             Year
                               1000 non-null
                                             int64
             Runtime (Minutes) 1000 non-null
                                            int64
             Rating
                               1000 non-null
                                            float64
                                            int64
             Votes
                               1000 non-null
```

10 Revenue (Millions) 872 non-null

dtypes: float64(3), int64(4), object(5)

936 non-null

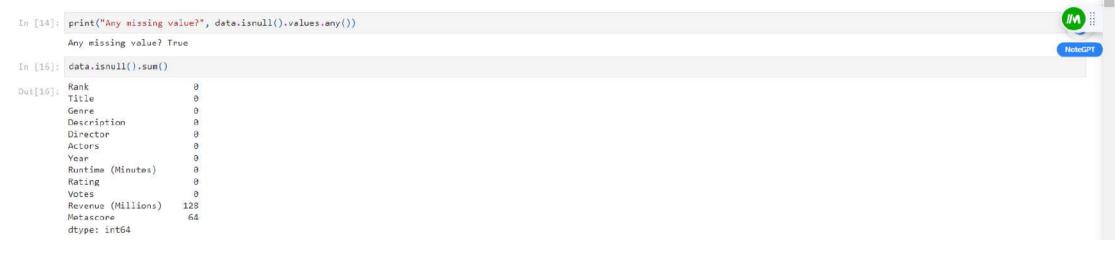
11 Metascore

memory usage: 93.9+ KB

float64

float64

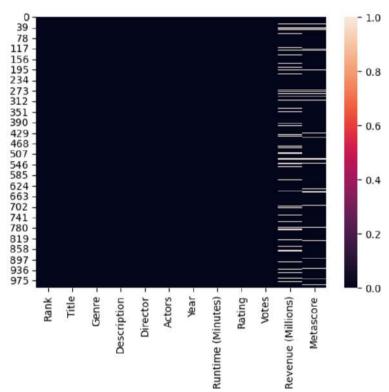
## CHECK MISSING VALUES IN THE DATASET



acype. Inco-

sns.heatmap(data.isnull()) In [17]:

<Axes: > Out[17]:





œ œ

NoteGPT

```
per_missing
                              0.0
        Rank
Out[18]:
        Title
                             0.0
        Genre
                             0.0
        Description
                             0.0
        Director
                             0.0
                             0.0
        Actors
        Year
                             0.0
        Runtime (Minutes)
                             0.0
                             0.0
        Rating
        Votes
                             0.0
```

Revenue (Millions)

Metascore

dtype: float64

In [18]: per\_missing= data.isnull().sum()\*100/len(data)

12.8

6.4

# DROP ALL THE MISSING VALUES



In [19]: data.dropna(axis=0)

Out[19]:

]:	ı	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
	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	121	8.1	757074	333.13	76.0
	1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te $\label{eq:condition} \text{te}$	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa		124	7.0	485820	126.46	65.0
	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		117	7.3	157606	138.12	62.0
	3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma	2016	108	7.2	60545	270.32	59.0
	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	123	6.2	393727	325.02	40.0
9	993	994	Resident Evil: Afterlife	Action,Adventure,Horror	While still out to destroy the evil Umbrella $$C_{\cdots}$$	Paul W.S. Anderson	Milla Jovovich, Ali Larter, Wentworth Miller,K	2010	97	5.9	140900	60.13	37.0
9	994	995	Project X	Comedy	3 high school seniors throw a birthday party t	Nima Nourizadeh	Thomas Mann, Oliver Cooper, Jonathan Daniel Br	2012	88	6.7	164088	54.72	48.0
9	996	997	Hostel: Part II	Horror	Three American college students studying abroa	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli	2007	94	5.5	73152	17.54	46.0
9	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,	2008	98	6.2	70699	58.01	50.0
9	999	1000	Nine Lives	Comedy, Family, Fantasy	A stuffy businessman finds himself trapped ins	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch	2016	87	5.3	12435	19.64	11.0

# **CHECK FOR DUPLICATE DATA**



In [20]: data= data.drop\_duplicates()

Out[20]:

]:	ı	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
	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	121	8.1	757074	333.13	76.0
	1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te $\label{eq:condition} t$	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa	2012	124	7.0	485820	126.46	65.0
	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	117	7.3	157606	138.12	62.0
	3	4	Sing	Animation, Comedy, Family	In a city of humanoid animals, a hustling thea	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma	2016	108	7.2	60545	270.32	59.0
	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	123	6.2	393727	325.02	40.0
9	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		111	6.2	27585	NaN	45.0
9	996	997	Hostel: Part II	Horror	Three American college students studying abroa	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli	2007	94	5.5	73152	17.54	46.0
9	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,	2008	98	6.2	70699	58.01	50.0
9	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	2014	93	5.6	4881	NaN	22.0
9	999	1000	Nine Lives	Comedy, Family, Fantasy	A stuffy businessman finds himself trapped ins	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch	2016	87	5.3	12435	19.64	11.0

#### GET OVERALL STATISTICS ABOUT THE DATAFRAME

//M

0.15047

data.describe(include="all")

Out[21]:

Runtime Revenue Metascore Rank Title Genre Description Director Actors Year Rating Votes (Minutes) (Millions) count 1000.000000 1000 1000 1000 1000 1000.000000 1.000000e+03 872.000000 936.000000 1000 1000.000000 1000.000000 999 207 1000 644 996 NaN unique NaN NaN NaN NaN NaN NaN The Action, Adventure, Sci-A group of intergalactic criminals are Ridley Jennifer Lawrence, Josh Hutcherson, Liam NaN NaN NaN NaN NaN NaN NaN top Host Fi forced ... Scott Hemsw... freq NaN 2 50 1 8 2 NaN NaN NaN NaN NaN NaN 500.500000 NaN NaN NaN NaN NaN 2012.783000 6.723200 1.698083e+05 82.956376 58.985043 mean 113.172000 288.819436 NaN 3.205962 18.810908 0.945429 1.887626e+05 103.253540 17.194757 std NaN NaN NaN NaN NaN 2006.000000 0.000000 11.000000 1.000000 NaN NaN NaN NaN 66.000000 1.900000 6.100000e+01 min 25% 250.750000 NaN NaN NaN NaN NaN 2010.000000 100.000000 6.200000 3.630900e+04 13.270000 47.000000 500.500000 NaN NaN NaN 2014.000000 111.000000 6.800000 1.107990e+05 47.985000 59.500000 NaN NaN 75% 750.250000 NaN NaN NaN 2016.000000 123.000000 113.715000 72.000000 NaN NaN 7.400000 2.399098e+05 NaN 2016.000000 936.630000 100.000000 max 1000.000000 NaN NaN NaN NaN 191.000000 9.000000 1.791916e+06

# DISPLAY TITLE OF THE MOVIE HAVING RUNTIME GREATER THAN OR EQUAL TO 180 MINS



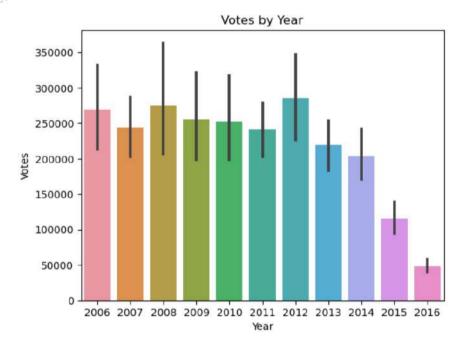


name. votes, utype. 110ato4

In [27]: sns.barplot(x="Year", y="Votes", data=data)
plt.title("Votes by Year")



Out[27]: Text(0.5, 1.0, 'Votes by Year')

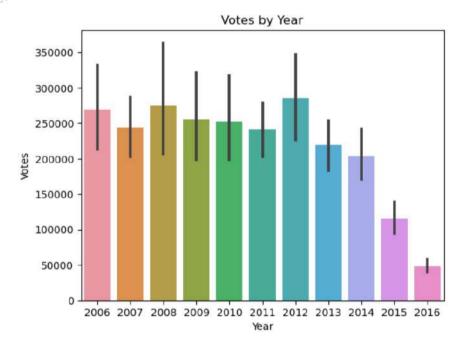


name. votes, utype. 110ato4

In [27]: sns.barplot(x="Year", y="Votes", data=data)
plt.title("Votes by Year")



Out[27]: Text(0.5, 1.0, 'Votes by Year')



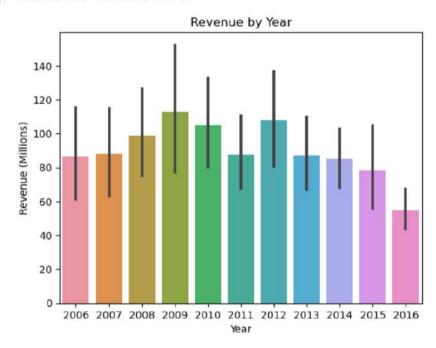
#### IN WHICH YEAR THERE WAS THE HIGHEST AVERAGE REVENUE?

```
In [28]: data.columns
Dut[28]: Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
                'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
                'Metascore'],
               dtype='object')
In [29]: data.groupby("Year")["Revenue (Millions)"].mean().sort_values(ascending=False)
         Year
Out[29]:
         2009
                 112.601277
         2012
                 107.973281
         2010
                 105.081579
         2008
                  99.082745
         2007
                  87.882245
         2011
                  87.612258
         2013
                  87.121818
         2006
                  86.296567
         2014
                  85.078723
         2015
                  78.355044
                  54.690976
         Name: Revenue (Millions), dtype: float64
```

manus ourselve quarassonery our particularies

In [30]: sns.barplot(x="Year", y="Revenue (Millions)", data=data)
plt.title("Revenue by Year")

Out[30]: Text(0.5, 1.0, 'Revenue by Year')





NoteGPT

#### FIND THE AVERAGE RATING FOR EACH DIRECTOR

```
In [31]: data.columns
         Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
                'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
                'Metascore'],
               dtype='object')
In [32]: data.groupby("Director")["Rating"].mean().sort_values(ascending=False)
         Director
Out[32]:
         Nitesh Tiwari
                             8.80
         Christopher Nolan
                             8.68
         Olivier Nakache
                             8.60
         Makoto Shinkai
                             8.60
         Aamir Khan
                             8.50
         Micheal Bafaro
                             3.50
         Jonathan Holbrook
                             3.20
         Shawn Burkett
                             2.70
         James Wong
                             2.70
         Jason Friedberg
                             1.90
         Name: Rating, Length: 644, dtype: float64
```

## **DISPLAY NUMBER OF MOVIES PER YEAR**

Out[43]: Runtime (Minutes)

Title	
Grindhouse	191
The Hateful Eight	187
The Wolf of Wall Street	180
La vie d'Adêle	180
Inland Empire	180
Cloud Atlas	172
3 Idiots	170
Interstellar	169
Pirates of the Caribbean: At World's End	169
The Hobbit: An Unexpected Journey	169

In [44]: sns.barplot(x="Runtime (Minutes)",y=top10\_len.index,data=top10\_len) <Axes: xlabel='Runtime (Minutes)', ylabel='Title'> Grindhouse The Hateful Eight -The Wolf of Wall Street La vie d'Adèle

Inland Empire

Cloud Atlas

interstellar

Pirates of the Caribbean: At World's End

The Hobbit: An Unexpected Journey

3 Idiots

25

50

75

100

Runtime (Minutes)

125

150

175

200

Title





# DISPLAY NUMBER OF MOVIES TITLE(HIGHEST REVENUE)

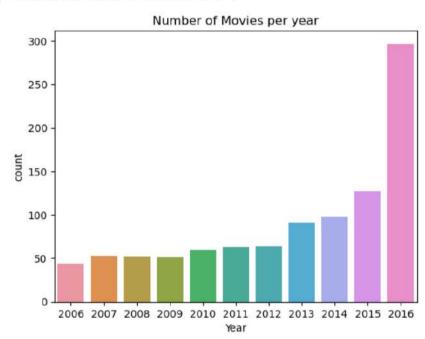


```
In [45]: data.columns
      'Metascore'],
          dtype='object')
In [45]: data["Year"].value_counts()
      2016
Out[46]:
      2015
            127
      2014
            98
      2013
            91
      2012
            64
      2011
            63
      2010
            60
      2007
            53
      2008
            52
      2009
            51
            44
      Name: Year, dtype: int64
```

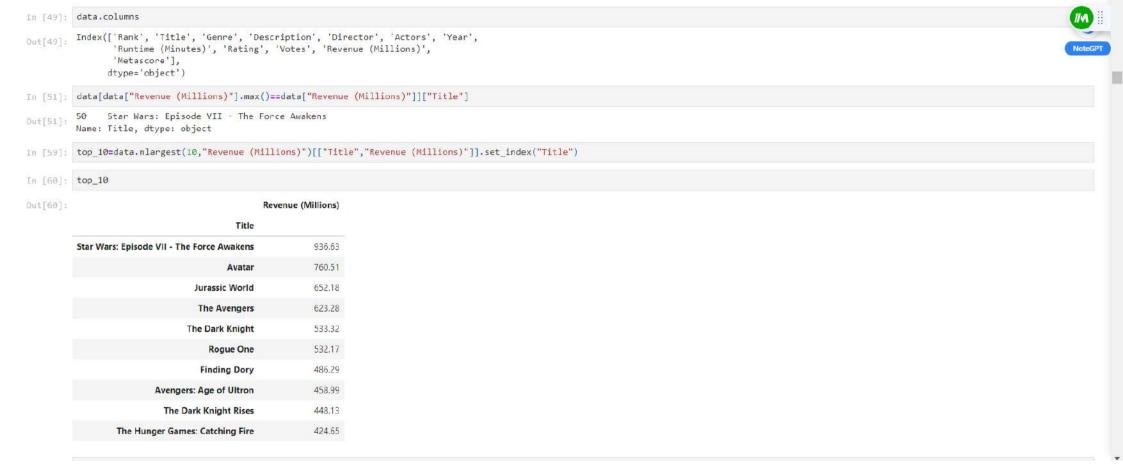
In [48]: sns.countplot(x="Year",data=data)
 plt.title("Number of Movies per year")



[Out[48]]: Text(0.5, 1.0, 'Number of Movies per year')



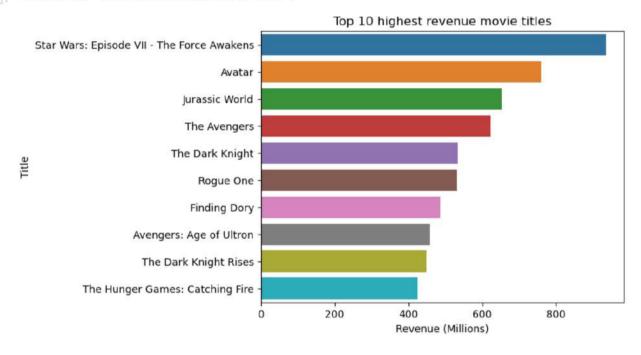
#### **DISPLAY TOP 10 HIGHEST REVENUE MOVIE TITLES**





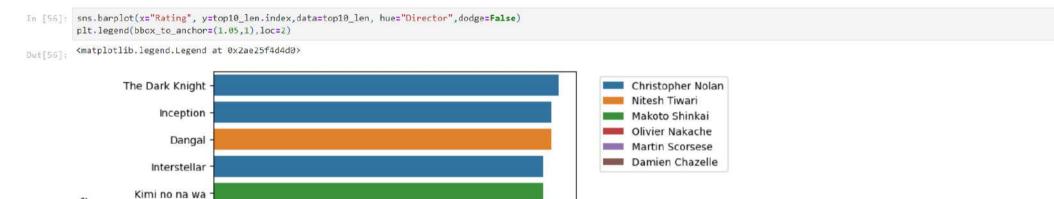


[62]: Text(0.5, 1.0, 'Top 10 highest revenue movie titles')



#### DISPLAY TOP 10 HIGHEST RATED MOVIE TITLES AND ITS DIRECTORS





Rating

Title

The Intouchables

The Prestige

The Departed

Whiplash

The Dark Knight Rises