

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
from google.colab import files
uploaded = files.upload()
```



Choose Files IMDB-Movie-Data.csv

- **IMDB-Movie-Data.csv**(application/vnd.ms-excel) - 309767 bytes, last modified: 5/30/2021 - 100% done
Saving IMDB-Movie-Data.csv to IMDB-Movie-Data (1).csv

```
import io
df = pd.read_csv(io.BytesIO(uploaded['IMDB-Movie-Data.csv']))
# Dataset is now stored in a Pandas Dataframe
```

```
df
```

	Rank	Title	Genre	Director	Actors	Year	Run (Minu	
df.head()								
	Rank	Title	Genre	Director	Actors	Year	Runti (Minute	
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	1	
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	1	
2	3	Split	Horror,Thriller	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	1	
3	4	Sing	Animation,Comedy,Family	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth	2016	1	
df.tail()								
	Rank	Title	Genre	Description	Director	Actors	Year	(I
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	
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	
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	
df.shape								
(1000, 12)								
df.ndim								
2								

```
df.dtypes
```

```
Rank          int64
Title         object
Genre         object
Description   object
Director      object
Actors        object
Year          int64
Runtime (Minutes)  int64
Rating        float64
Votes         int64
Revenue (Millions) float64
Metascore     float64
dtype: object
```

```
df.index
```

```
RangeIndex(start=0, stop=1000, step=1)
```

```
df.values
```

```
array([[1, 'Guardians of the Galaxy', 'Action,Adventure,Sci-Fi', ...,
       757074, 333.13, 76.0],
       [2, 'Prometheus', 'Adventure,Mystery,Sci-Fi', ..., 485820, 126.46,
       65.0],
       [3, 'Split', 'Horror,Thriller', ..., 157606, 138.12, 62.0],
       ...,
       [998, 'Step Up 2: The Streets', 'Drama,Music,Romance', ..., 70699,
       58.01, 50.0],
       [999, 'Search Party', 'Adventure,Comedy', ..., 4881, nan, 22.0],
       [1000, 'Nine Lives', 'Comedy,Family,Fantasy', ..., 12435, 19.64,
       11.0]], dtype=object)
```

```
df.isnull()
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False

```
df.notnull()
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes
0	True	True	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	True	True
...
995	True	True	True	True	True	True	True	True	True	True
996	True	True	True	True	True	True	True	True	True	True
997	True	True	True	True	True	True	True	True	True	True
998	True	True	True	True	True	True	True	True	True	True
999	True	True	True	True	True	True	True	True	True	True

1000 rows × 12 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Rank                                  1000 non-null   int64
1   Title                                1000 non-null   object
2   Genre                                1000 non-null   object
3   Description                           1000 non-null   object
4   Director                             1000 non-null   object
5   Actors                               1000 non-null   object
6   Year                                  1000 non-null   int64
7   Runtime (Minutes)                    1000 non-null   int64
8   Rating                               1000 non-null   float64
9   Votes                                1000 non-null   int64
10  Revenue (Millions)                   872 non-null    float64
11  Metascore                            936 non-null    float64
dtypes: float64(3), int64(4), object(5)
memory usage: 93.9+ KB
```

```
df.columns
```

```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')
```

```
df.nunique()
```

```
Rank          1000
Title          999
Genre          207
Description    1000
Director        644
Actors          996
Year            11
Runtime (Minutes)  94
Rating          55
Votes          997
Revenue (Millions)  814
Metascore       84
dtype: int64
```

```
df.loc[2]
```

```
Rank          3
Title          Split
Genre          Horror,Thriller
Description    Three girls are kidnapped by a man with a diag...
Director        M. Night Shyamalan
Actors          James McAvoy, Anya Taylor-Joy, Haley Lu Richar...
Year            2016
Runtime (Minutes)  117
Rating          7.3
Votes          157606
Revenue (Millions)  138.12
Metascore       62
Name: 2, dtype: object
```

```
type(df.loc[2])
```

```
pandas.core.series.Series
```

```
df.T
```

	0	1	2	3
Rank	1	2	3	4
Title	Guardians of the Galaxy	Prometheus	Split	Sing
Genre	Action,Adventure,Sci-Fi	Adventure,Mystery,Sci-Fi	Horror,Thriller	Animation,Comedy,Family
Description	A group of intergalactic criminals are forced ...	Following clues to the origin of mankind, a te...	Three girls are kidnapped by a man with a diag...	In a city of humanoid animals, a hustling thea...
Director	James Gunn	Ridley Scott	M. Night Shyamalan	Christophe Lourdelet
Actors	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	Noomi Rapace, Logan Marshall-Green, Michael Fa...	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	Matthew McConaughey, Reese Witherspoon, Seth Ma...
Year	2014	2012	2016	2016
Runtime (Minutes)	121	124	117	108
Rating	8.1	7	7.3	7.2
Votes	757074	485820	157606	60545

Revenue

df['Title'].count()

1000

12 rows × 1000 columns

df['Rating'].unique()

```
array([8.1, 7. , 7.3, 7.2, 6.2, 6.1, 8.3, 6.4, 7.1, 7.5, 7.8, 7.9, 7.7,
       6.6, 8.2, 6.7, 8. , 6.5, 5.3, 6.8, 4.7, 5.9, 6.3, 5.6, 8.6, 7.6,
       6.9, 2.7, 3.7, 5.8, 9. , 7.4, 4.1, 8.5, 8.8, 5.4, 6. , 5.7, 8.4,
       5.2, 5.5, 4.8, 3.9, 4.9, 5.1, 5. , 4.3, 4.6, 4.2, 4. , 3.2, 4.5,
       1.9, 3.5, 4.4])
```

df.Genre.value_counts()

```
Action,Adventure,Sci-Fi    50
Drama                      48
```

```

Comedy,Drama,Romance      35
Comedy                     32
Drama,Romance              31
..
Animation,Drama,Romance    1
Biography,Drama,Mystery    1
Action,Horror,Romance      1
Animation,Comedy,Drama      1
Mystery,Romance,Thriller    1
Name: Genre, Length: 207, dtype: int64

```

```
df.describe()
```

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Meta
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	872.000000	936.0
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.956376	58.9
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	103.253540	17.1
min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.000000	11.0
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	13.270000	47.0
50%	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	47.985000	59.5
75%	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	113.715000	72.0
max	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.630000	100.0

```
df['Rank']
```

```

0      1
1      2
2      3
3      4
4      5
...
995    996
996    997
997    998
998    999
999   1000
Name: Rank, Length: 1000, dtype: int64

```

```
df.loc[:, 'Votes']
```

```

0      757074
1      485820
2      157606
3       60545
4      393727
...
995     27585

```

```
996      73152
997      70699
998      4881
999      12435
Name: Votes, Length: 1000, dtype: int64
```

```
df.iloc[:,4]
```

```
0      James Gunn
1      Ridley Scott
2      M. Night Shyamalan
3      Christophe Lourdelet
4      David Ayer
...
995     Billy Ray
996     Eli Roth
997     Jon M. Chu
998     Scot Armstrong
999     Barry Sonnenfeld
Name: Director, Length: 1000, dtype: object
```

```
df.loc[:, "Title": "Rating"]
```


Title

Genre

Description

Director

Actors

Year

df.loc[(df["Year"]>2014),["Title","Rating"]]

	Title	Rating
2	Split	7.3
3	Sing	7.2
4	Suicide Squad	6.2
5	The Great Wall	6.1
6	La La Land	8.3
...
978	It's Only the End of the World	7.0
980	Miracles from Heaven	7.0
984	Max	6.8
995	Secret in Their Eyes	6.2
999	Nine Lives	5.3

424 rows × 2 columns

or df...

df.where(df["Year"]>2014)

	Rank	Title	Genre	Description	Director	Actors
0	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN
2	3.0	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...

```
df.query("Rating>7")
```

Rank	Title	Genre	Description	Director	Actors		
df.groupby("Genre").sum()							
Guardians		A group of intergalactic		James	Chris Pratt, Vir		
	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Genre							
Action	705	4032	176	12.3	17237	131.56	37
Action,Adventure	999	6033	373	19.2	1003377	671.22	172
Action,Adventure,Biography	568	4031	263	14.0	97560	33.00	125
Action,Adventure,Comedy	5700	28199	1535	92.3	2709033	1340.27	661
Action,Adventure,Crime	2379	12074	699	38.6	1201937	543.72	290
...
Romance,Sci-Fi,Thriller	707	2011	106	7.1	208632	62.45	0
Sci-Fi	801	4032	213	9.9	1453	20.76	53
Sci-Fi,Thriller	513	2008	91	5.0	170897	64.51	34
Thriller	6133	18133	857	53.6	305642	0.32	418
Thriller,War	161	2016	106	6.0	5926	0.00	40
207 rows × 7 columns							
The Skin I				prasad	Pedro	Antonio Banderas	
df.groupby("Rating").count()							

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Votes	Re (Mill
Rating										
1.9	1	1	1	1	1	1	1	1	1	
2.7	2	2	2	2	2	2	2	2	2	
3.2	1	1	1	1	1	1	1	1	1	
3.5	2	2	2	2	2	2	2	2	2	
3.7	2	2	2	2	2	2	2	2	2	
3.9	3	3	3	3	3	3	3	3	3	
4.0	1	1	1	1	1	1	1	1	1	
4.1	1	1	1	1	1	1	1	1	1	
4.2	2	2	2	2	2	2	2	2	2	
4.3	4	4	4	4	4	4	4	4	4	
4.4	1	1	1	1	1	1	1	1	1	
4.5	1	1	1	1	1	1	1	1	1	
4.6	5	5	5	5	5	5	5	5	5	
4.7	6	6	6	6	6	6	6	6	6	
4.8	4	4	4	4	4	4	4	4	4	
4.9	7	7	7	7	7	7	7	7	7	
5.0	4	4	4	4	4	4	4	4	4	
5.1	5	5	5	5	5	5	5	5	5	
5.2	11	11	11	11	11	11	11	11	11	
5.3	12	12	12	12	12	12	12	12	12	
5.4	12	12	12	12	12	12	12	12	12	
5.5	14	14	14	14	14	14	14	14	14	
5.6	17	17	17	17	17	17	17	17	17	
5.7	21	21	21	21	21	21	21	21	21	
5.8	26	26	26	26	26	26	26	26	26	
5.9	19	19	19	19	19	19	19	19	19	
6.0	26	26	26	26	26	26	26	26	26	
6.1	31	31	31	31	31	31	31	31	31	
6.2	37	37	37	37	37	37	37	37	37	

6.3	44	44	44	44	44	44	44	44	44
6.4	35	35	35	35	35	35	35	35	35
6.5	40	40	40	40	40	40	40	40	40
6.6	42	42	42	42	42	42	42	42	42
6.7	48	48	48	48	48	48	48	48	48
6.8	37	37	37	37	37	37	37	37	37
6.9	31	31	31	31	31	31	31	31	31
7.0	46	46	46	46	46	46	46	46	46
7.1	52	52	52	52	52	52	52	52	52
7.2	42	42	42	42	42	42	42	42	42
7.3	42	42	42	42	42	42	42	42	42
7.4	33	33	33	33	33	33	33	33	33
7.5	35	35	35	35	35	35	35	35	35
7.6	27	27	27	27	27	27	27	27	27
7.7	27	27	27	27	27	27	27	27	27

```
df.groupby(["Genre", "Year"])["Rating"].mean()
```

```

Genre      Year
Action      2016    6.150000
Action,Adventure  2007    6.200000
              2012    7.000000
              2014    6.000000
Action,Adventure,Biography  2015    6.900000
...
Sci-Fi,Thriller  2008    5.000000
Thriller        2007    7.100000
              2014    6.700000
              2016    5.685714
Thriller,War    2016    6.000000
Name: Rating, Length: 583, dtype: float64

```

```
df.sort_values("Year")
```

	Rank	Title	Genre	Description	Director	Actors
773	774	Lady in the Water	Drama,Fantasy,Mystery	Apartment building superintendent Cleveland He...	M. Night Shyamalan	Paul Giamatti, Bryce Dallas Howard, Jeffrey Wr...
445	446	Silent Hill	Adventure,Horror,Mystery	A woman, Rose, goes in search for her adopted ...	Christophe Gans	Radha Mitchell, Laurie Holden, Sean Bean,Debor...
99	100	The Departed	Crime,Drama,Thriller	An undercover cop and a mole in the police att...	Martin Scorsese	Leonardo DiCaprio, Matt Damon, Jack Nicholson,...
550	551	The Break-Up	Comedy,Drama,Romance	In a bid to keep their luxurious condo from th...	Peyton Reed	Jennifer Aniston, Vince Vaughn, Jon Favreau, J...
885	886	Babel	Drama	Tragedy strikes a married couple on vacation i...	Alejandro González Iñárritu	Brad Pitt, Cate Blanchett, Gael García Bernal,...
...
				The story of a		Annette

```
df.sort_index(axis=1)
```

	Actors	Description	Director	Genre	Metascore	Rank
0	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	A group of intergalactic criminals are forced ...	James Gunn	Action,Adventure,Sci-Fi	76.0	1
1	Noomi Rapace, Logan Marshall-Green, Michael Fa...	Following clues to the origin of mankind, a te...	Ridley Scott	Adventure,Mystery,Sci-Fi	65.0	2
2	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	Horror,Thriller	62.0	3
3	Matthew McConaughey,Reese Witherspoon, Seth Ma...	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Animation,Comedy,Family	59.0	4
	Will Smith. Jared	A secret government				

df.isna()

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...
995	False	False	False	False	False	False	False	False	False	False
996	False	False	False	False	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False	False	False
998	False	False	False	False	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False	False	False

1000 rows × 12 columns

df.notna()

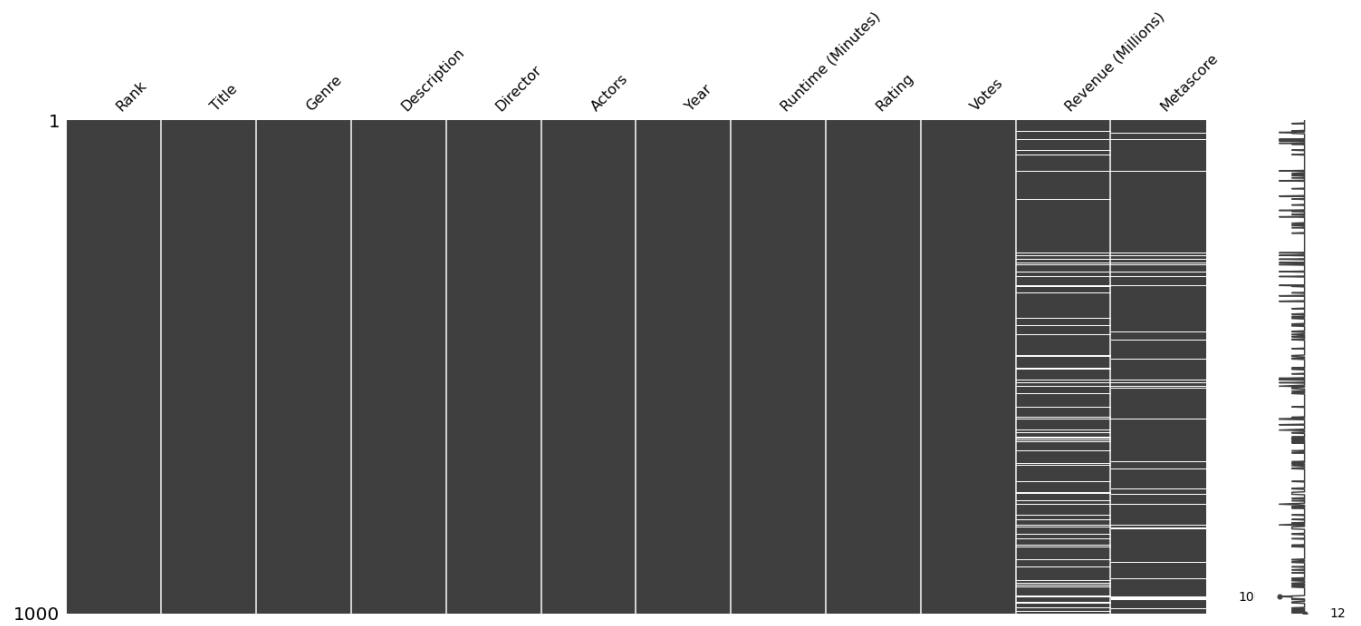
	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes
0	True	True	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	True	True
...
995	True	True	True	True	True	True	True	True	True	True
996	True	True	True	True	True	True	True	True	True	True
997	True	True	True	True	True	True	True	True	True	True
998	True	True	True	True	True	True	True	True	True	True
999	True	True	True	True	True	True	True	True	True	True

1000 rows × 12 columns

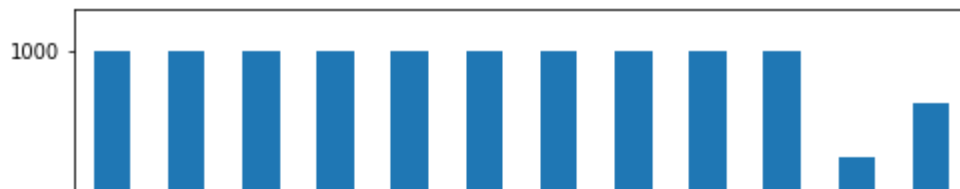
```
df.isna().count()
```

```
Rank          1000
Title         1000
Genre         1000
Description    1000
Director      1000
Actors        1000
Year          1000
Runtime (Minutes) 1000
Rating        1000
Votes         1000
Revenue (Millions) 1000
Metascore     1000
dtype: int64
```

```
import missingno
missingno.matrix(df)
plt.show()
```

```
plt.figure(figsize=(8,8))
df.notnull().sum().plot.bar()
plt.show()
```



```
(df.where(df["Rating"]<6.5)).dropna()  
    .set_index("Rating","Advanced Rating")  
    .rename(columns={"Title","Good movie Title"})
```

Rank

Title

Genre

Description

Director

Actors

(df.where(df["Rating"]<6.5)).set_index("Rating","Advanced Rating")

	Rank	Title	Genre	Description	Director	Actors	
Rating							
NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	
6.2	5.0	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	20
...
6.2	996.0	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	20
5.5	997.0	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	20
6.2	998.0	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
							Milli...

(df.where(df["Rating"]<6.5).rename(columns={"Title":"Good movie Title"}))

	Rank	Good movie Title	Genre	Description	Director	Actors	Yea
0	NaN	NaN	NaN	NaN	NaN	NaN	Na
1	NaN	NaN	NaN	NaN	NaN	NaN	Na
2	NaN	NaN	NaN	NaN	NaN	NaN	Na
3	NaN	NaN	NaN	NaN	NaN	NaN	Na
4	5.0	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016.
...
995	996.0	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015.
		Hostel		Three American college		Lauren German, Heather	

```
df.drop("Description",axis=1,inplace=True)
df.head()
```

	Rank	Title	Genre	Director	Actors	Year	Runti (Minute)
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	1
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	1
2	3	Split	Horror,Thriller	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	1
3	4	Sing	Animation,Comedy,Family	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth	2016	1

```
def split(c):
    return c.split(',')

def add(n):
    return n+10
```

```
df["Genre"].apply(split)
```

```

0      [Action, Adventure, Sci-Fi]
1      [Adventure, Mystery, Sci-Fi]
2      [Horror, Thriller]
3      [Animation, Comedy, Family]
4      [Action, Adventure, Fantasy]
...
995     [Crime, Drama, Mystery]
996     [Horror]
997     [Drama, Music, Romance]
998     [Adventure, Comedy]
999     [Comedy, Family, Fantasy]
Name: Genre, Length: 1000, dtype: object
```

```
df.Rating.apply(add)
```

```

0      18.1
1      17.0
2      17.3
3      17.2
4      16.2
...
995     16.2
996     15.5
997     16.2
998     15.6
999     15.3
Name: Rating, Length: 1000, dtype: float64
```

```
df[["Rating", "Votes", "Metascore"]].apply(add)
```

Rating Votes Metascore

df.corr()

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Rank	1.000000	-0.261605	-0.221739	-0.219555	-0.283876	-0.271592	-0.191869
Year	-0.261605	1.000000	-0.164900	-0.211219	-0.411904	-0.126790	-0.079305
Runtime (Minutes)	-0.221739	-0.164900	1.000000	0.392214	0.407062	0.267953	0.211978
Rating	-0.219555	-0.211219	0.392214	1.000000	0.511537	0.217654	0.631897
Votes	-0.283876	-0.411904	0.407062	0.511537	1.000000	0.639661	0.325684
Revenue (Millions)	-0.271592	-0.126790	0.267953	0.217654	0.639661	1.000000	0.142397

```
plt.figure(figsize=(12,12))
sns.heatmap(df.corr(),annot=True,fmt=".4f",linewidth=2,square=True,cmap="Blues")
plt.title("Correlation",size=15)
plt.show()
```

