top-1000-bollywood-movies-dataset

October 16, 2024

[3]:
```
#Python Project On Top 1000 Bollywood Movies And Their Box Office
import seaborn as sns
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

[4]:
```
#To watch the first five records
act=pd.read_csv("Top 1000 Bollywood Movies and their boxoffice.csv")
print(act.iloc[100:120])
```

```
     Unnamed: 0   SN                            Movie   Worldwide  \
100         100  101                     Bol Bachchan  1750000000
101         101  102                             Raid  1540000000
102         102  103                        2 States  1730000000
103         103  104                      Grand Masti  1450000000
104         104  105                           Race 2  1620000000
105         105  106                          Singham  1510000000
106         106  107                          Shivaay  1470000000
107         107  108                      Batla House  1270000000
108         108  109                             Zero  1780000000
109         109  110                     Welcome Back  1680000000
110         110  111                         Baaghi 3  1370000000
111         111  112                             Baby  1420000000
112         112  113  Manikarnika: The Queen of Jhansi  1420000000
113         113  114                      Luka Chuppi  1280000000
114         114  115                        Raajneeti  1450000000
115         115  116                          Talaash  1750000000
116         116  117                Satyameva Jayate  1210000000
117         117  118                           Bholaa  1230000000
118         118  119                 Singh Is Bliing  1460000000
119         119  120          Zindagi Na Milegi Dobara  1530000000

      India Net  India Gross   Overseas      Budget        Verdict
100  1030000000   1440000000  300000000   650000000            Hit
101  1030000000   1320000000  220000000   350000000       SuperHit
102  1020000000   1420000000  300000000   500000000       SuperHit
103  1020000000   1300000000  140000000   350000000       SuperHit
104  1010000000   1310000000  300000000   850000000            Hit
```

```
105  1000000000   1300000000  200000000   520000000       SuperHit
106  1000000000   1380000000   80000000  1000000000  Below Average
107   990000000   1160000000  110000000   500000000           Hit
108   960000000   1200000000  570000000  2000000000       Disaster
109   960000000   1300000000  380000000   900000000        Average
110   960000000   1140000000  230000000  1000000000  Below Average
111   950000000   1290000000  130000000   590000000           Hit
112   940000000   1160000000  250000000  1010000000        Average
113   940000000   1110000000  170000000   200000000       SuperHit
114   930000000   1260000000  190000000   600000000       SuperHit
115   930000000   1210000000  530000000   600000000           Hit
116   900000000   1160000000   50000000   450000000           Hit
117   900000000   1080000000  150000000  1000000000        Average
118   900000000   1200000000  260000000   760000000        Average
119   890000000   1210000000  310000000   550000000           Hit
```

[5]:
```
#Looking the rows and columns of the dataset
act.shape
```

[5]:
```
(1000, 9)
```

[6]:
```
#for showing all row and columns
act.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   Unnamed: 0   1000 non-null   int64 
 1   SN           1000 non-null   int64 
 2   Movie        1000 non-null   object
 3   Worldwide    1000 non-null   int64 
 4   India Net    1000 non-null   int64 
 5   India Gross  1000 non-null   int64 
 6   Overseas     1000 non-null   int64 
 7   Budget       1000 non-null   int64 
 8   Verdict      1000 non-null   object
dtypes: int64(7), object(2)
memory usage: 70.4+ KB
```

[7]:
```
#for showing all row and columns
act.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype 
```

```
---  ------       --------------  -----
 0   Unnamed: 0   1000 non-null   int64
 1   SN           1000 non-null   int64
 2   Movie        1000 non-null   object
 3   Worldwide    1000 non-null   int64
 4   India Net    1000 non-null   int64
 5   India Gross  1000 non-null   int64
 6   Overseas     1000 non-null   int64
 7   Budget       1000 non-null   int64
 8   Verdict      1000 non-null   object
dtypes: int64(7), object(2)
memory usage: 70.4+ KB
```

[8]:
```
#To Know The Null Values Of Dataset
act.isnull()
```

[8]:
```
     Unnamed: 0     SN  Movie  Worldwide  India Net  India Gross  Overseas  \
0         False  False  False      False      False        False     False
1         False  False  False      False      False        False     False
2         False  False  False      False      False        False     False
3         False  False  False      False      False        False     False
4         False  False  False      False      False        False     False
..          …    …    …        …          …            …         …
995       False  False  False      False      False        False     False
996       False  False  False      False      False        False     False
997       False  False  False      False      False        False     False
998       False  False  False      False      False        False     False
999       False  False  False      False      False        False     False

     Budget  Verdict
0     False    False
1     False    False
2     False    False
3     False    False
4     False    False
..      …      …
995   False    False
996   False    False
997   False    False
998   False    False
999   False    False

[1000 rows x 9 columns]
```

[9]:
```
#it tells null values present or not in all columns
act.isnull().sum()
```

[9]:
```
Unnamed: 0     0
SN             0
Movie          0
Worldwide      0
India Net      0
India Gross    0
Overseas       0
Budget         0
Verdict        0
dtype: int64
```

[10]:
```
#to know the duplicate values
act.duplicated()
```

[10]:
```
0      False
1      False
2      False
3      False
4      False
       …
995    False
996    False
997    False
998    False
999    False
Length: 1000, dtype: bool
```

[11]:
```
#to know the duplicate values as per column
act.duplicated().sum()
```

[11]:
```
0
```

[12]:
```
#to know the duplicate values as per specific column
act["Movie"].duplicated().sum()
```

[12]:
```
6
```

[13]:
```
#to drop duplicates from the column
print(act.drop_duplicates("Movie"))
```

```
     Unnamed: 0    SN                        Movie    Worldwide   India Net  \
0             0     1                      Pathaan  10500000000  5240000000
1             1     2  Baahubali 2 The Conclusion  17880000000  5100000000
2             2     3                KGF Chapter 2  12080000000  4350000000
3             3     4                       Dangal  20700000000  3740000000
4             4     5                         Sanju   5880000000  3420000000
..          …   …                            …           …          …
995         995   996                      Nikamma           0    20000000
```

```
996         996   997                  Saugandh     30000000     20000000
997         997   998                Akaash Vani     30000000     20000000
998         998   999                    Zwigato     10000000     20000000
999         999  1000                     FryDay     20000000     20000000

     India Gross     Overseas      Budget                Verdict
0     6570000000   3920000000  2500000000  All Time Blockbuster
1    14160000000   3710000000  2500000000  All Time Blockbuster
2    10000000000   2070000000  1000000000  All Time Blockbuster
3     5350000000  15350000000   700000000  All Time Blockbuster
4     4380000000   1500000000  1000000000  All Time Blockbuster
..           …            …           …                   …
995     20000000            0    10000000                   0
996     30000000            0    20000000             Average
997     20000000            0   100000000             Disaster
998     10000000            0    10000000                   0
999     20000000            0   150000000             Disaster

[994 rows x 9 columns]
```

[14]:
```
act.iloc[90:100]
```

[14]:
```
    Unnamed: 0   SN                       Movie   Worldwide   India Net  \
90          90   91          Bhaag Milkha Bhaag  1680000000  1080000000
91          91   92  Don 2: The Chase Continues  2100000000  1080000000
92          92   93                 Golmaal 3   1620000000  1060000000
93          93   94   Pushpa: The Rise (Part 1)  3500000000  1060000000
94          94   95                    ABCD 2   1660000000  1060000000
95          95   96                 Ek Villain  1550000000  1050000000
96          96   97             Son Of Sardaar  1500000000  1050000000
97          97   98                    Kaabil   1780000000  1040000000
98          98   99             De De Pyaar De  1390000000  1040000000
99          99  100                         83  1930000000  1040000000

    India Gross   Overseas      Budget      Verdict
90   1460000000  210000000   300000000     SuperHit
91   1440000000  650000000   760000000          Hit
92   1430000000  180000000   400000000  Blockbuster
93   3130000000  360000000  1500000000     SuperHit
94   1460000000  190000000   650000000     SuperHit
95   1420000000  120000000   390000000     SuperHit
96   1410000000   80000000   400000000     SuperHit
97   1450000000  320000000   350000000          Hit
98   1130000000  260000000   750000000          Hit
99   1290000000  640000000  2200000000     Disaster
```

[15]:
```
act.loc[(act["Budget"]==0),"Verdict"], "High Flop"
```

[15]:
```
(887    All Time Blockbuster
 933               Blockbuster
 946    All Time Blockbuster
 968               Blockbuster
 976               Blockbuster
 Name: Verdict, dtype: object,
 'High Flop')
```

[16]:
```
act.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Unnamed: 0   1000 non-null   int64
 1   SN           1000 non-null   int64
 2   Movie        1000 non-null   object
 3   Worldwide    1000 non-null   int64
 4   India Net    1000 non-null   int64
 5   India Gross  1000 non-null   int64
 6   Overseas     1000 non-null   int64
 7   Budget       1000 non-null   int64
 8   Verdict      1000 non-null   object
dtypes: int64(7), object(2)
memory usage: 70.4+ KB
```

[17]:
```
ver2=act["Verdict"]=act["Verdict"].replace(0,"Flop")
print(ver2)
```

```
0      All Time Blockbuster
1      All Time Blockbuster
2      All Time Blockbuster
3      All Time Blockbuster
4      All Time Blockbuster
               …
995                       0
996                 Average
997                Disaster
998                       0
999                Disaster
Name: Verdict, Length: 1000, dtype: object
```

[19]:
```
act["Movie With Verdict"] = act["Movie"]+ " " + act["Verdict"]
print(act)
```

```
     Unnamed: 0    SN                      Movie    Worldwide   India Net  \
0             0     1                    Pathaan  10500000000  5240000000
```

```
1             1     2  Baahubali 2 The Conclusion  17880000000  5100000000
2             2     3                KGF Chapter 2  12080000000  4350000000
3             3     4                       Dangal  20700000000  3740000000
4             4     5                        Sanju   5880000000  3420000000
..          …   …                              …           …         …
995         995   996                     Nikamma           0    20000000
996         996   997                    Saugandh     30000000    20000000
997         997   998                 Akaash Vani     30000000    20000000
998         998   999                     Zwigato     10000000    20000000
999         999  1000                      FryDay     20000000    20000000

     India Gross     Overseas      Budget               Verdict  \
0     6570000000   3920000000  2500000000  All Time Blockbuster
1    14160000000   3710000000  2500000000  All Time Blockbuster
2    10000000000   2070000000  1000000000  All Time Blockbuster
3     5350000000  15350000000   700000000  All Time Blockbuster
4     4380000000   1500000000  1000000000  All Time Blockbuster
..           …           …         …                    …
995     20000000           0    10000000                    0
996     30000000           0    20000000              Average
997     20000000           0   100000000             Disaster
998     10000000           0    10000000                    0
999     20000000           0   150000000             Disaster

                                Movie With Verdict
0                       Pathaan All Time Blockbuster
1    Baahubali 2 The Conclusion All Time Blockbuster
2                  KGF Chapter 2 All Time Blockbuster
3                         Dangal All Time Blockbuster
4                          Sanju All Time Blockbuster
..                                               …
995                                      Nikamma 0
996                                 Saugandh Average
997                              Akaash Vani Disaster
998                                      Zwigato 0
999                                   FryDay Disaster

[1000 rows x 10 columns]
```

[23]:
```
act.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Unnamed: 0          1000 non-null   int64
 1   SN                  1000 non-null   int64
```

```
 2   Movie                1000 non-null   object
 3   Worldwide            1000 non-null   int64
 4   India Net            1000 non-null   int64
 5   India Gross          1000 non-null   int64
 6   Overseas             1000 non-null   int64
 7   Budget               1000 non-null   int64
 8   Verdict              1000 non-null   object
 9   Movie With Verdict   1000 non-null   object
dtypes: int64(7), object(3)
memory usage: 78.3+ KB
```

[27]:
```
act["Verdict"].isnull().iloc[990:997]
```

[27]:
```
990    False
991    False
992    False
993    False
994    False
995    False
996    False
Name: Verdict, dtype: bool
```

[31]:
```
#To Addition India Worth And India Net That Store in India Worth New Column
act["India worth"]=act["India Net"] + act["India Gross"]
print(act)
```

```
     Unnamed: 0    SN                        Movie    Worldwide   India Net  \
0             0     1                      Pathaan  10500000000  5240000000
1             1     2  Baahubali 2 The Conclusion  17880000000  5100000000
2             2     3                KGF Chapter 2  12080000000  4350000000
3             3     4                       Dangal  20700000000  3740000000
4             4     5                        Sanju   5880000000  3420000000
..          …   …                            …           …          …
995         995   996                     Nikamma           0    20000000
996         996   997                    Saugandh    30000000    20000000
997         997   998                 Akaash Vani    30000000    20000000
998         998   999                     Zwigato    10000000    20000000
999         999  1000                      FryDay    20000000    20000000

     India Gross     Overseas      Budget                Verdict  \
0     6570000000   3920000000  2500000000  All Time Blockbuster
1    14160000000   3710000000  2500000000  All Time Blockbuster
2    10000000000   2070000000  1000000000  All Time Blockbuster
3     5350000000  15350000000   700000000  All Time Blockbuster
4     4380000000   1500000000  1000000000  All Time Blockbuster
..           …           …          …                     …
995     20000000           0    10000000                     0
996     30000000           0    20000000                Average
```

```
997     20000000            0   100000000              Disaster
998     10000000            0    10000000                     0
999     20000000            0   150000000              Disaster

                                  Movie With Verdict  India worth
0                        Pathaan All Time Blockbuster  11810000000
1    Baahubali 2 The Conclusion All Time Blockbuster  19260000000
2                  KGF Chapter 2 All Time Blockbuster  14350000000
3                         Dangal All Time Blockbuster   9090000000
4                          Sanju All Time Blockbuster   7800000000
..                                                 …           …
995                                      Nikamma 0     40000000
996                                  Saugandh Average     50000000
997                               Akaash Vani Disaster     40000000
998                                        Zwigato 0     30000000
999                                    FryDay Disaster     40000000

[1000 rows x 11 columns]
```

[38]:
```
#to summarize the data using group by function
gp=act.groupby(["Movie","Verdict"]).agg({"Movie":"count"})[0:10]
print(gp)
```

```
                                        Movie
Movie                Verdict
102 Not Out          Hit                    1
1920                 Average                1
1920 - Evil Returns  Above Average          1
1920 London          Below Average          1
1921                 Above Average          1
2                    Average                1
2 States             SuperHit               1
3 Idiots             All Time Blockbuster   1
36 China Town        Above Average          1
7 Khoon Maaf         Flop                   1
```

[41]:
```
#to change the datatype into to float, float to int etc...
act['Budget']=act['Budget'].astype('float')
```

[43]:
```
act.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Unnamed: 0          1000 non-null   int64
 1   SN                  1000 non-null   int64
```

```
 2   Movie                1000 non-null   object
 3   Worldwide            1000 non-null   int64
 4   India Net            1000 non-null   int64
 5   India Gross          1000 non-null   int64
 6   Overseas             1000 non-null   int64
 7   Budget               1000 non-null   float64
 8   Verdict              1000 non-null   object
 9   Movie With Verdict   1000 non-null   object
 10  India worth          1000 non-null   int64
dtypes: float64(1), int64(7), object(3)
memory usage: 86.1+ KB
```

[47]:
```
act['Budget']=act['Budget'].astype("int")
```

[49]:
```
act.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Unnamed: 0           1000 non-null   int64
 1   SN                   1000 non-null   int64
 2   Movie                1000 non-null   object
 3   Worldwide            1000 non-null   int64
 4   India Net            1000 non-null   int64
 5   India Gross          1000 non-null   int64
 6   Overseas             1000 non-null   int64
 7   Budget               1000 non-null   int32
 8   Verdict              1000 non-null   object
 9   Movie With Verdict   1000 non-null   object
 10  India worth          1000 non-null   int64
dtypes: int32(1), int64(7), object(3)
memory usage: 82.2+ KB
```

[51]:
```
#for changing the name of SN column to Serial No., inplace is used for saving
act.rename(columns={"SN":"Serial No."},inplace=True)
act.head(50)
```

[51]:

	Unnamed: 0	Serial No.	Movie	Worldwide	\
0	0	1	Pathaan	10500000000	
1	1	2	Baahubali 2 The Conclusion	17880000000	
2	2	3	KGF Chapter 2	12080000000	
3	3	4	Dangal	20700000000	
4	4	5	Sanju	5880000000	
5	5	6	PK	7920000000	
6	6	7	Tiger Zinda Hai	5580000000	

7	7	8	Bajrangi Bhaijaan	9220000000
8	8	9	War	4710000000
9	9	10	Padmaavat	5850000000
10	10	11	Sultan	6270000000
11	11	12	Kabir Singh	3770000000
12	12	13	Tanhaji: The Unsung Warrior	3610000000
13	13	14	RRR	12300000000
14	14	15	Dhoom 3	5580000000
15	15	16	The Kashmir Files	3410000000
16	16	17	Brahmastra Part One: Shiva	4310000000
17	17	18	URI The Surgical Strike	3410000000
18	18	19	Simmba	3900000000
19	19	20	Drishyam 2	3420000000
20	20	21	Kick	3780000000
21	21	22	Krrish 3	3740000000
22	22	23	Chennai Express	4220000000
23	23	24	Bharat	3210000000
24	24	25	Housefull 4	2960000000
25	25	26	Prem Ratan Dhan Payo	4050000000
26	26	27	Golmaal Again	3080000000
27	27	28	Good Newwz	3160000000
28	28	29	Mission Mangal	2900000000
29	29	30	3 Idiots	4600000000
30	30	31	Happy New Year	3970000000
31	31	32	Ek Tha Tiger	3200000000
32	32	33	Sooryavanshi	2930000000
33	33	34	2	7440000000
34	34	35	Yeh Jawaani Hai Deewani	3180000000
35	35	36	Bhool Bhulaiyaa 2	2650000000
36	36	37	Bajirao Mastani	3620000000
37	37	38	Bang Bang!	3400000000
38	38	39	Race 3	3000000000
39	39	40	Baaghi 2	2570000000
40	40	41	Raees	2850000000
41	41	42	Kesari	2080000000
42	42	43	Total Dhamaal	2320000000
43	43	44	Dabangg 2	2650000000
44	44	45	Chhichhore	2110000000
45	45	46	Tanu Weds Manu Returns	2580000000
46	46	47	Bodyguard	2300000000
47	47	48	Dilwale	3880000000
48	48	49	Super 30	2100000000
49	49	50	Tu Jhoothi Main Makkar	2230000000

	India Net	India Gross	Overseas	Budget	Verdict	\
0	5240000000	6570000000	3920000000	-2147483648	All Time Blockbuster	
1	5100000000	14160000000	3710000000	-2147483648	All Time Blockbuster	

2	4350000000	10000000000	2070000000	1000000000	All Time	Blockbuster
3	3740000000	5350000000	15350000000	700000000	All Time	Blockbuster
4	3420000000	4380000000	1500000000	1000000000	All Time	Blockbuster
5	3400000000	4890000000	3030000000	850000000	All Time	Blockbuster
6	3390000000	4330000000	1240000000	2100000000		Blockbuster
7	3200000000	4320000000	4890000000	900000000	All Time	Blockbuster
8	3030000000	3750000000	960000000	1500000000		Blockbuster
9	3020000000	4000000000	1850000000	-2147483648		Blockbuster
10	3000000000	4210000000	2060000000	800000000		Blockbuster
11	2780000000	3300000000	470000000	550000000	All Time	Blockbuster
12	2770000000	3250000000	360000000	1500000000		Blockbuster
13	2720000000	9150000000	3140000000	-2147483648		Blockbuster
14	2710000000	3720000000	1860000000	1750000000	All Time	Blockbuster
15	2520000000	2950000000	460000000	200000000	All Time	Blockbuster
16	2490000000	3180000000	1130000000	-2147483648		Hit
17	2440000000	2930000000	480000000	700000000	All Time	Blockbuster
18	2400000000	2950000000	950000000	900000000		Blockbuster
19	2390000000	2820000000	590000000	500000000	All Time	Blockbuster
20	2310000000	3100000000	680000000	1000000000		SuperHit
21	2310000000	3130000000	610000000	940000000		Blockbuster
22	2270000000	3100000000	1120000000	1150000000		SuperHit
23	2120000000	2510000000	700000000	1650000000		Hit
24	2100000000	2470000000	490000000	1750000000		SuperHit
25	2100000000	3100000000	950000000	1800000000		SuperHit
26	2050000000	2620000000	450000000	800000000		Blockbuster
27	2050000000	2420000000	740000000	600000000		Blockbuster
28	2030000000	2360000000	540000000	400000000		Blockbuster
29	2020000000	2740000000	1860000000	550000000	All Time	Blockbuster
30	1990000000	2950000000	1020000000	1500000000		Hit
31	1980000000	2630000000	570000000	750000000		Blockbuster
32	1950000000	2300000000	620000000	1800000000		SuperHit
33	1900000000	5510000000	1930000000	-2147483648		Average
34	1880000000	2420000000	760000000	750000000		Blockbuster
35	1840000000	2160000000	490000000	800000000		Blockbuster
36	1840000000	2630000000	990000000	1450000000		Hit
37	1740000000	2700000000	700000000	1600000000		Hit
38	1690000000	2170000000	830000000	1700000000		Average
39	1650000000	2110000000	450000000	750000000		Blockbuster
40	1640000000	2270000000	570000000	920000000		Hit
41	1550000000	1830000000	250000000	1000000000		Hit
42	1550000000	1830000000	480000000	1050000000		SuperHit
43	1550000000	2260000000	390000000	840000000		Blockbuster
44	1530000000	1800000000	310000000	700000000		Blockbuster
45	1500000000	2140000000	440000000	390000000		Blockbuster
46	1480000000	1890000000	410000000	600000000		Blockbuster
47	1480000000	2120000000	1760000000	1650000000		Average
48	1470000000	1710000000	390000000	600000000		SuperHit

```
49  1470000000   1750000000    480000000   700000000                  Hit

                                Movie With Verdict  India worth
0                       Pathaan All Time Blockbuster  11810000000
1   Baahubali 2 The Conclusion All Time Blockbuster  19260000000
2                 KGF Chapter 2 All Time Blockbuster  14350000000
3                        Dangal All Time Blockbuster   9090000000
4                         Sanju All Time Blockbuster   7800000000
5                            PK All Time Blockbuster   8290000000
6                       Tiger Zinda Hai Blockbuster   7720000000
7             Bajrangi Bhaijaan All Time Blockbuster   7520000000
8                                    War Blockbuster   6780000000
9                              Padmaavat Blockbuster   7020000000
10                               Sultan Blockbuster   7210000000
11                  Kabir Singh All Time Blockbuster   6080000000
12           Tanhaji: The Unsung Warrior Blockbuster   6020000000
13                                   RRR Blockbuster  11870000000
14                      Dhoom 3 All Time Blockbuster   6430000000
15            The Kashmir Files All Time Blockbuster   5470000000
16                    Brahmastra Part One: Shiva Hit   5670000000
17      URI The Surgical Strike All Time Blockbuster   5370000000
18                                   Simmba Blockbuster   5350000000
19                  Drishyam 2 All Time Blockbuster   5210000000
20                                     Kick SuperHit   5410000000
21                               Krrish 3 Blockbuster   5440000000
22                          Chennai Express SuperHit   5370000000
23                                       Bharat Hit   4630000000
24                               Housefull 4 SuperHit   4570000000
25                    Prem Ratan Dhan Payo SuperHit   5200000000
26                        Golmaal Again Blockbuster   4670000000
27                           Good Newwz Blockbuster   4470000000
28                       Mission Mangal Blockbuster   4390000000
29                    3 Idiots All Time Blockbuster   4760000000
30                                 Happy New Year Hit   4940000000
31                           Ek Tha Tiger Blockbuster   4610000000
32                              Sooryavanshi SuperHit   4250000000
33                                       2 Average   7410000000
34                Yeh Jawaani Hai Deewani Blockbuster   4300000000
35                     Bhool Bhulaiyaa 2 Blockbuster   4000000000
36                                 Bajirao Mastani Hit   4470000000
37                                      Bang Bang! Hit   4440000000
38                                      Race 3 Average   3860000000
39                                Baaghi 2 Blockbuster   3760000000
40                                       Raees Hit   3910000000
41                                     Kesari Hit   3380000000
42                               Total Dhamaal SuperHit   3380000000
43                                Dabangg 2 Blockbuster   3810000000
```

```
44                          Chhichhore Blockbuster   3330000000
45              Tanu Weds Manu Returns Blockbuster   3640000000
46                           Bodyguard Blockbuster   3370000000
47                                 Dilwale Average   3600000000
48                              Super 30 SuperHit   3180000000
49                     Tu Jhoothi Main Makkar Hit   3220000000
```

[53]:
```
#for describing the specific columns
act[["India Net","India Gross","Budget"]].describe()
```

[53]:
```
          India Net   India Gross        Budget
count  1.000000e+03  1.000000e+03  1.000000e+03
mean   4.079600e+08  6.126900e+08  2.285455e+08
std    5.783055e+08  9.814085e+08  5.357887e+08
min    2.000000e+07  0.000000e+00 -2.147484e+09
25%    9.000000e+07  1.400000e+08  8.000000e+07
50%    2.000000e+08  3.100000e+08  2.000000e+08
75%    4.600000e+08  6.825000e+08  3.800000e+08
max    5.240000e+09  1.416000e+10  2.100000e+09
```

[55]:
```
act[["Verdict"]].describe()
```

[55]:
```
       Verdict
count     1000
unique      10
top       Flop
freq       197
```

[57]:
```
act.columns
```

[57]:
```
Index(['Unnamed: 0', 'Serial No.', 'Movie', 'Worldwide', 'India Net',
       'India Gross', 'Overseas', 'Budget', 'Verdict', 'Movie With Verdict',
       'India worth'],
      dtype='object')
```

[59]:
```
act["Budget"].describe()
```

[59]:
```
count    1.000000e+03
mean     2.285455e+08
std      5.357887e+08
min     -2.147484e+09
25%      8.000000e+07
50%      2.000000e+08
75%      3.800000e+08
max      2.100000e+09
Name: Budget, dtype: float64
```

[61]:
```
#Removing Negative Value From Budget Column
act=act[(act["Budget"]>=0)]
```

[63]:
```
act.head(50)
```

[63]:

	Unnamed: 0	Serial No.	Movie	Worldwide	\
2	2	3	KGF Chapter 2	12080000000	
3	3	4	Dangal	20700000000	
4	4	5	Sanju	5880000000	
5	5	6	PK	7920000000	
6	6	7	Tiger Zinda Hai	5580000000	
7	7	8	Bajrangi Bhaijaan	9220000000	
8	8	9	War	4710000000	
10	10	11	Sultan	6270000000	
11	11	12	Kabir Singh	3770000000	
12	12	13	Tanhaji: The Unsung Warrior	3610000000	
14	14	15	Dhoom 3	5580000000	
15	15	16	The Kashmir Files	3410000000	
17	17	18	URI The Surgical Strike	3410000000	
18	18	19	Simmba	3900000000	
19	19	20	Drishyam 2	3420000000	
20	20	21	Kick	3780000000	
21	21	22	Krrish 3	3740000000	
22	22	23	Chennai Express	4220000000	
23	23	24	Bharat	3210000000	
24	24	25	Housefull 4	2960000000	
25	25	26	Prem Ratan Dhan Payo	4050000000	
26	26	27	Golmaal Again	3080000000	
27	27	28	Good Newwz	3160000000	
28	28	29	Mission Mangal	2900000000	
29	29	30	3 Idiots	4600000000	
30	30	31	Happy New Year	3970000000	
31	31	32	Ek Tha Tiger	3200000000	
32	32	33	Sooryavanshi	2930000000	
34	34	35	Yeh Jawaani Hai Deewani	3180000000	
35	35	36	Bhool Bhulaiyaa 2	2650000000	
36	36	37	Bajirao Mastani	3620000000	
37	37	38	Bang Bang!	3400000000	
38	38	39	Race 3	3000000000	
39	39	40	Baaghi 2	2570000000	
40	40	41	Raees	2850000000	
41	41	42	Kesari	2080000000	
42	42	43	Total Dhamaal	2320000000	
43	43	44	Dabangg 2	2650000000	
44	44	45	Chhichhore	2110000000	
45	45	46	Tanu Weds Manu Returns	2580000000	
46	46	47	Bodyguard	2300000000	

47	47	48	Dilwale	3880000000
48	48	49	Super 30	2100000000
49	49	50	Tu Jhoothi Main Makkar	2230000000
50	50	51	Dabangg 3	2180000000
53	53	54	Dream Girl	1950000000
54	54	55	Singham Returns	2160000000
55	55	56	Dabangg	2150000000
56	56	57	Gully Boy	2350000000
57	57	58	Judwaa 2	2270000000

	India Net	India Gross	Overseas	Budget	Verdict	\
2	4350000000	10000000000	2070000000	1000000000	All Time Blockbuster	
3	3740000000	5350000000	15350000000	700000000	All Time Blockbuster	
4	3420000000	4380000000	1500000000	1000000000	All Time Blockbuster	
5	3400000000	4890000000	3030000000	850000000	All Time Blockbuster	
6	3390000000	4330000000	1240000000	2100000000	Blockbuster	
7	3200000000	4320000000	4890000000	900000000	All Time Blockbuster	
8	3030000000	3750000000	960000000	1500000000	Blockbuster	
10	3000000000	4210000000	2060000000	800000000	Blockbuster	
11	2780000000	3300000000	470000000	550000000	All Time Blockbuster	
12	2770000000	3250000000	360000000	1500000000	Blockbuster	
14	2710000000	3720000000	1860000000	1750000000	All Time Blockbuster	
15	2520000000	2950000000	460000000	200000000	All Time Blockbuster	
17	2440000000	2930000000	480000000	700000000	All Time Blockbuster	
18	2400000000	2950000000	950000000	900000000	Blockbuster	
19	2390000000	2820000000	590000000	500000000	All Time Blockbuster	
20	2310000000	3100000000	680000000	1000000000	SuperHit	
21	2310000000	3130000000	610000000	940000000	Blockbuster	
22	2270000000	3100000000	1120000000	1150000000	SuperHit	
23	2120000000	2510000000	700000000	1650000000	Hit	
24	2100000000	2470000000	490000000	1750000000	SuperHit	
25	2100000000	3100000000	950000000	1800000000	SuperHit	
26	2050000000	2620000000	450000000	800000000	Blockbuster	
27	2050000000	2420000000	740000000	600000000	Blockbuster	
28	2030000000	2360000000	540000000	400000000	Blockbuster	
29	2020000000	2740000000	1860000000	550000000	All Time Blockbuster	
30	1990000000	2950000000	1020000000	1500000000	Hit	
31	1980000000	2630000000	570000000	750000000	Blockbuster	
32	1950000000	2300000000	620000000	1800000000	SuperHit	
34	1880000000	2420000000	760000000	750000000	Blockbuster	
35	1840000000	2160000000	490000000	800000000	Blockbuster	
36	1840000000	2630000000	990000000	1450000000	Hit	
37	1740000000	2700000000	700000000	1600000000	Hit	
38	1690000000	2170000000	830000000	1700000000	Average	
39	1650000000	2110000000	450000000	750000000	Blockbuster	
40	1640000000	2270000000	570000000	920000000	Hit	
41	1550000000	1830000000	250000000	1000000000	Hit	

42	1550000000	1830000000	480000000	1050000000	SuperHit
43	1550000000	2260000000	390000000	840000000	Blockbuster
44	1530000000	1800000000	310000000	700000000	Blockbuster
45	1500000000	2140000000	440000000	390000000	Blockbuster
46	1480000000	1890000000	410000000	600000000	Blockbuster
47	1480000000	2120000000	1760000000	1650000000	Average
48	1470000000	1710000000	390000000	600000000	SuperHit
49	1470000000	1750000000	480000000	700000000	Hit
50	1460000000	1780000000	400000000	900000000	Average
53	1410000000	1630000000	320000000	300000000	SuperHit
54	1400000000	2000000000	160000000	1050000000	SuperHit
55	1400000000	2070000000	70000000	420000000	Blockbuster
56	1390000000	1640000000	710000000	600000000	SuperHit
57	1380000000	1920000000	350000000	650000000	SuperHit

	Movie With Verdict	India worth
2	KGF Chapter 2 All Time Blockbuster	14350000000
3	Dangal All Time Blockbuster	9090000000
4	Sanju All Time Blockbuster	7800000000
5	PK All Time Blockbuster	8290000000
6	Tiger Zinda Hai Blockbuster	7720000000
7	Bajrangi Bhaijaan All Time Blockbuster	7520000000
8	War Blockbuster	6780000000
10	Sultan Blockbuster	7210000000
11	Kabir Singh All Time Blockbuster	6080000000
12	Tanhaji: The Unsung Warrior Blockbuster	6020000000
14	Dhoom 3 All Time Blockbuster	6430000000
15	The Kashmir Files All Time Blockbuster	5470000000
17	URI The Surgical Strike All Time Blockbuster	5370000000
18	Simmba Blockbuster	5350000000
19	Drishyam 2 All Time Blockbuster	5210000000
20	Kick SuperHit	5410000000
21	Krrish 3 Blockbuster	5440000000
22	Chennai Express SuperHit	5370000000
23	Bharat Hit	4630000000
24	Housefull 4 SuperHit	4570000000
25	Prem Ratan Dhan Payo SuperHit	5200000000
26	Golmaal Again Blockbuster	4670000000
27	Good Newwz Blockbuster	4470000000
28	Mission Mangal Blockbuster	4390000000
29	3 Idiots All Time Blockbuster	4760000000
30	Happy New Year Hit	4940000000
31	Ek Tha Tiger Blockbuster	4610000000
32	Sooryavanshi SuperHit	4250000000
34	Yeh Jawaani Hai Deewani Blockbuster	4300000000
35	Bhool Bhulaiyaa 2 Blockbuster	4000000000
36	Bajirao Mastani Hit	4470000000

```
37                                Bang Bang! Hit   4440000000
38                                Race 3 Average   3860000000
39                            Baaghi 2 Blockbuster   3760000000
40                                     Raees Hit   3910000000
41                                   Kesari Hit   3380000000
42                           Total Dhamaal SuperHit   3380000000
43                            Dabangg 2 Blockbuster   3810000000
44                           Chhichhore Blockbuster   3330000000
45              Tanu Weds Manu Returns Blockbuster   3640000000
46                           Bodyguard Blockbuster   3370000000
47                                 Dilwale Average   3600000000
48                               Super 30 SuperHit   3180000000
49                     Tu Jhoothi Main Makkar Hit   3220000000
50                               Dabangg 3 Average   3240000000
53                            Dream Girl SuperHit   3040000000
54                      Singham Returns SuperHit   3400000000
55                           Dabangg Blockbuster   3470000000
56                            Gully Boy SuperHit   3030000000
57                              Judwaa 2 SuperHit   3300000000
```

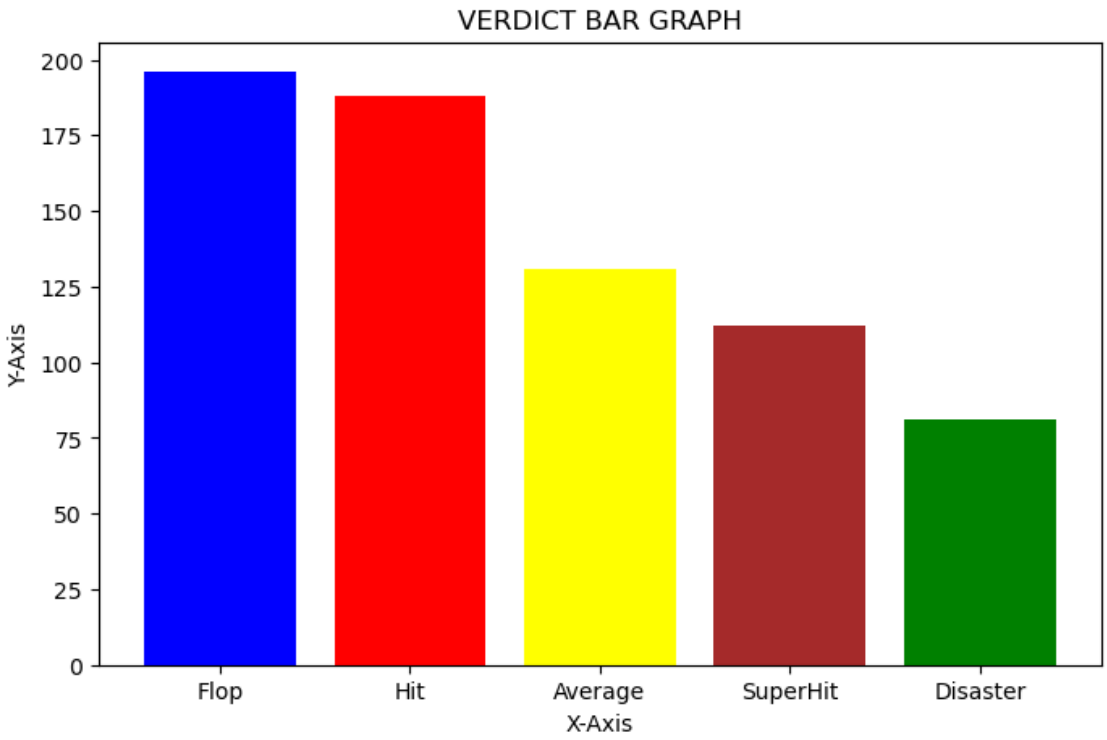
[65]:
```
#Getting the frequency of Verdict Column
act["Verdict"].value_counts()
```

[65]:
```
Verdict
Flop                   196
Hit                    188
Average                131
SuperHit               112
Disaster                81
Blockbuster             71
Below Average           66
Above Average           55
0                       44
All Time Blockbuster    26
Name: count, dtype: int64
```

[67]:
```
#Getting the frequency of Verdict Column Top 10
act['Verdict'].value_counts()[0:5]
```

[67]:
```
Verdict
Flop        196
Hit         188
Average     131
SuperHit    112
Disaster     81
Name: count, dtype: int64
```

[69]:
```
#ploting the 5 verdict countplot chart with values in graph or say␣
 ↪labels=values her 1st half upto comma "," is for keys/Names and after comma␣
 ↪it is for values
plt.figure(figsize=(8,5))
plt.title("VERDICT BAR GRAPH")
plt.xlabel("X-Axis")
plt.ylabel("Y-Axis")
plt.bar(act["Verdict"].value_counts()[0:5].keys(),(act["Verdict"].
 ↪value_counts()[0:5]),color=["blue","red","yellow","brown","green"])
plt.show()
```

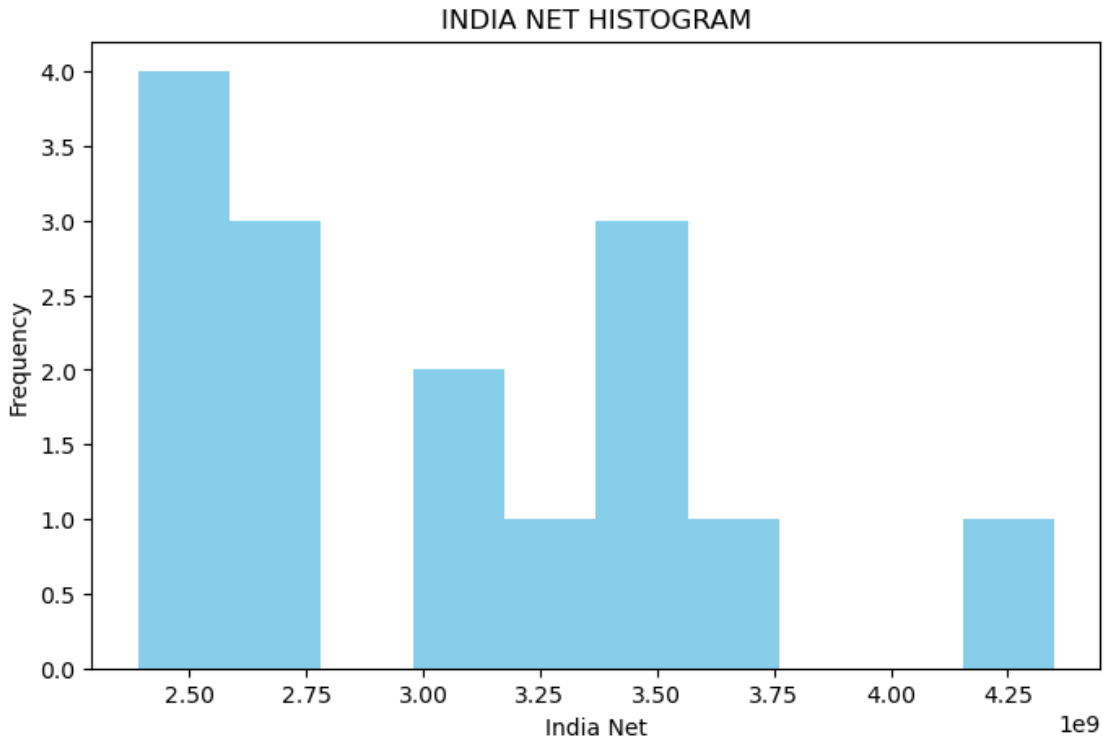


[71]:
```
#For Showing Counts Of India Net
act["India Net"].value_counts()[0:15]
```

[71]:
```
India Net
100000000    36
50000000     35
20000000     34
30000000     33
60000000     33
70000000     32
80000000     31
110000000    28
```

```
40000000     27
150000000    25
200000000    25
170000000    24
130000000    23
120000000    22
90000000     21
Name: count, dtype: int64
```

[73]:
```
#histogram of India Net Column
plt.figure(figsize=(8,5))
plt.title("INDIA NET HISTOGRAM")
plt.xlabel("India Net")
plt.ylabel("Frequency")
plt.hist(act["India Net"][0:15],bins=10,color="SkyBlue")
plt.show()
```



[75]:
```
act["Verdict"].value_counts()
```

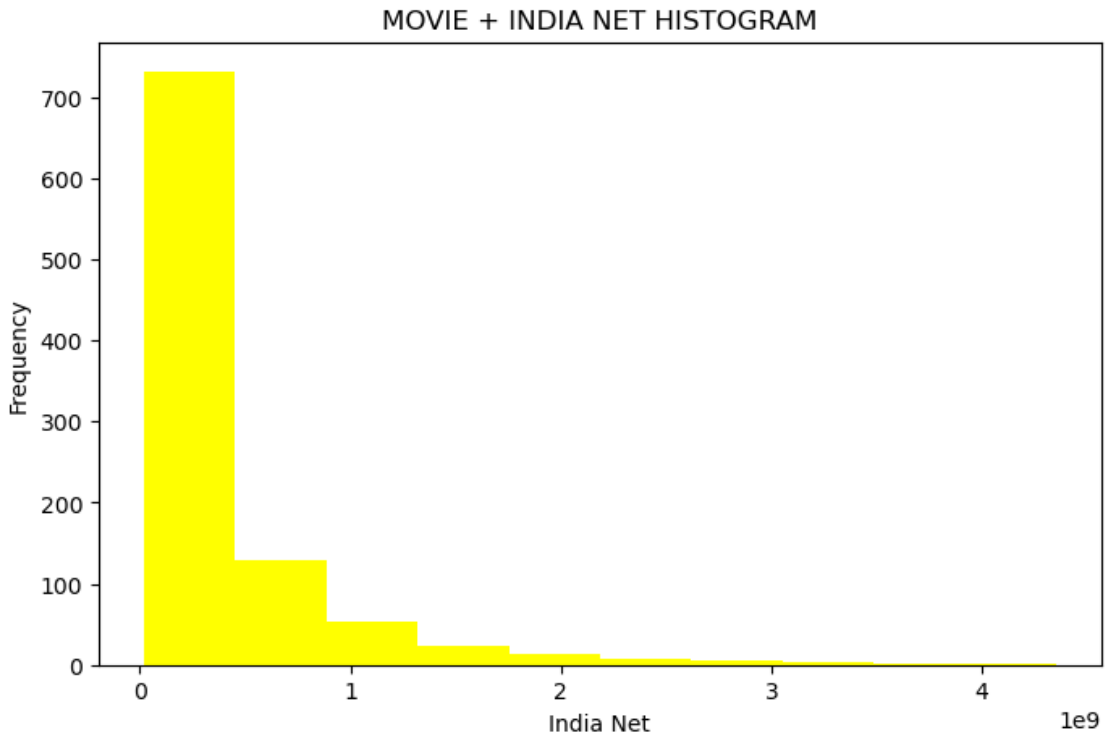
[75]:
```
Verdict
Flop       196
Hit        188
Average    131
```

```
SuperHit                112
Disaster                 81
Blockbuster              71
Below Average            66
Above Average            55
0                        44
All Time Blockbuster     26
Name: count, dtype: int64
```

[77]:
```
#histogram of India Net Column
ver=act[["Movie","India Gross"]][0:11]
print(ver)
```

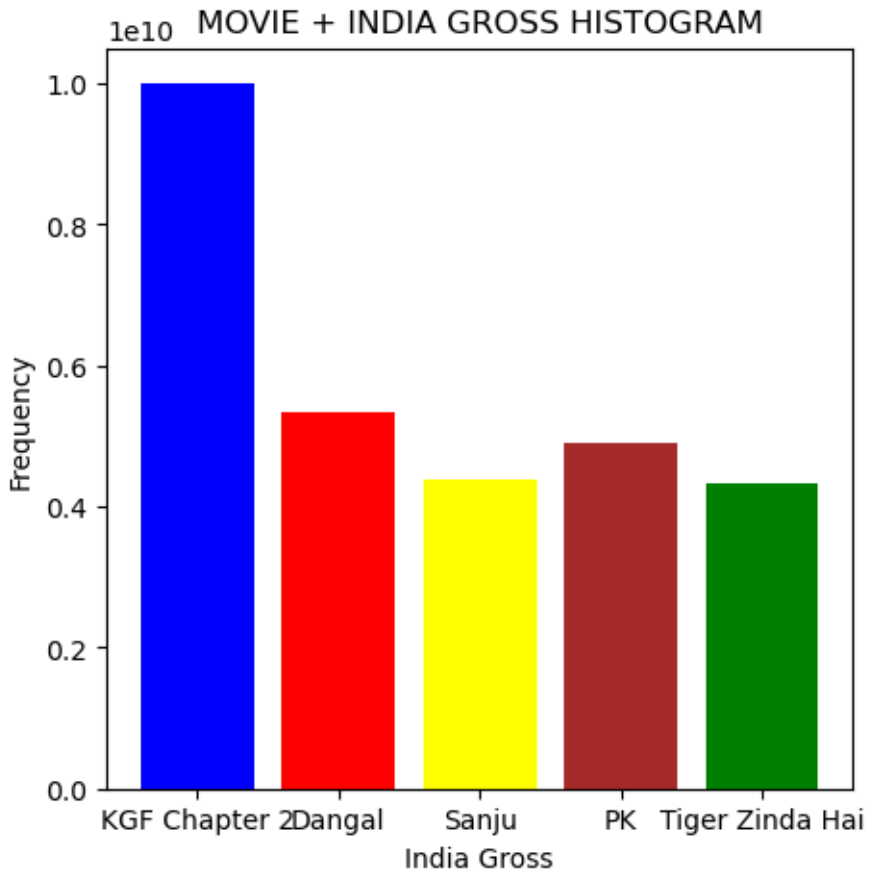
```
                         Movie  India Gross
2                KGF Chapter 2  10000000000
3                       Dangal   5350000000
4                        Sanju   4380000000
5                           PK   4890000000
6              Tiger Zinda Hai   4330000000
7           Bajrangi Bhaijaan    4320000000
8                          War   3750000000
10                       Sultan   4210000000
11                  Kabir Singh   3300000000
12  Tanhaji: The Unsung Warrior   3250000000
14                      Dhoom 3   3720000000
```

[79]:
```
plt.figure(figsize=(8,5))
plt.title("MOVIE + INDIA NET HISTOGRAM")
plt.xlabel("India Net")
plt.ylabel("Frequency")
plt.hist(act["India Net"],bins=10,color="yellow")
plt.show()
```



[81]:
```
plt.figure(figsize=(5,5))
plt.title("MOVIE + INDIA GROSS HISTOGRAM")
plt.xlabel("India Gross")
plt.ylabel("Frequency")
plt.bar(list(ver["Movie"][0:5]),list(ver["India Gross"])[0:
  ↪5],color=["blue","red","yellow","brown","green"])
```

[81]:
```
<BarContainer object of 5 artists>
```



[83]:
```
#to check following movie present in the column or not
#here true in war means war is present
ver["Movie"]=="War"[0:100]
```

[83]:
```
2     False
3     False
4     False
5     False
6     False
7     False
8      True
10    False
11    False
12    False
14    False
Name: Movie, dtype: bool
```

[85]:
```
#Dropping War From The List
ver.drop(8,inplace=True)
```

[99]:
```
#To Print The Updated List
#Output=False Means War is Dropped
ver=["Movie"]=="War"[0:100]
ver
```

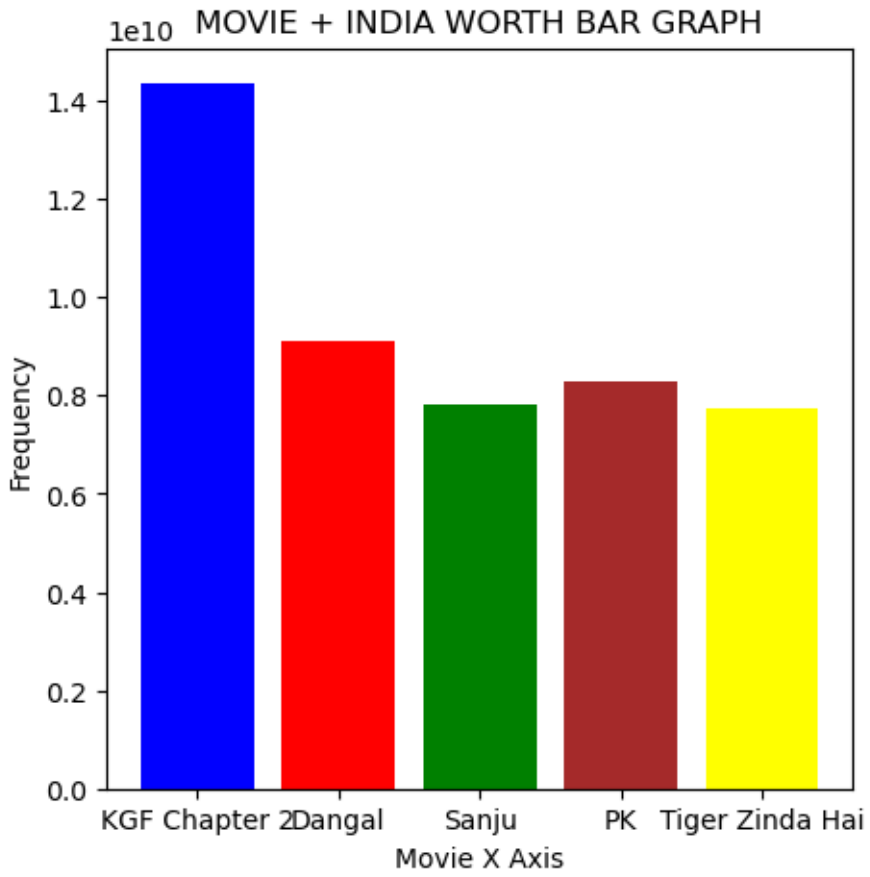
[99]:
```
False
```

[95]:
```
total=act[["Movie","India worth"]]
total.head(10)
```

[95]:

	Movie	India worth
2	KGF Chapter 2	14350000000
3	Dangal	9090000000
4	Sanju	7800000000
5	PK	8290000000
6	Tiger Zinda Hai	7720000000
7	Bajrangi Bhaijaan	7520000000
8	War	6780000000
10	Sultan	7210000000
11	Kabir Singh	6080000000
12	Tanhaji: The Unsung Warrior	6020000000

[318]:
```
#To Print Top 5 Movie With India Worth Graph
plt.figure(figsize=(5,5))
plt.title("MOVIE + INDIA WORTH BAR GRAPH")
plt.xlabel("Movie X Axis")
plt.ylabel("Frequency")
plt.bar(list(total["Movie"][0:5]),list(total["India worth"])[0:
 ↪5],color=["blue","red","green","brown","yellow"])
plt.show()
```



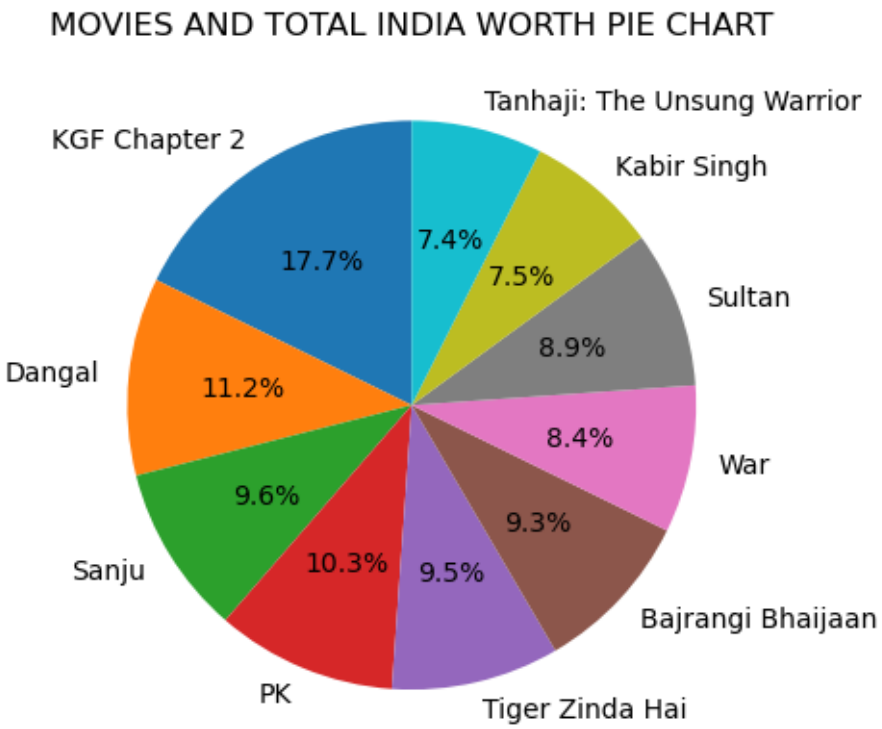
[110]:
```
#top 10 movies With India Total Worth Or Earning On India Basis
india_ttl=total.head(10)
india_ttl
```

[110]:

	Movie	India worth
2	KGF Chapter 2	14350000000
3	Dangal	9090000000
4	Sanju	7800000000
5	PK	8290000000
6	Tiger Zinda Hai	7720000000
7	Bajrangi Bhaijaan	7520000000
8	War	6780000000
10	Sultan	7210000000
11	Kabir Singh	6080000000
12	Tanhaji: The Unsung Warrior	6020000000

[112]:
```
#Plotting Pie Chart Of Top 10 Movies With Total India Worth Or Total Earning␣
 ↪from startangle 90 digree from 1 to 10 in order
plt.title("TOTAL INDIA WORTH WITH MOVIE PIE CHART")
```

```
plt.pie(india_ttl["India worth"],labels=india_ttl["Movie"],autopct="%0.
  ↪1f%%",startangle=90)
plt.show()
```



[116]:
```
#To Concatinate New "Movie With Verdict Column"
ver=act[["Movie","India worth","Movie With Verdict"]]
per=ver.head(10)
per
```

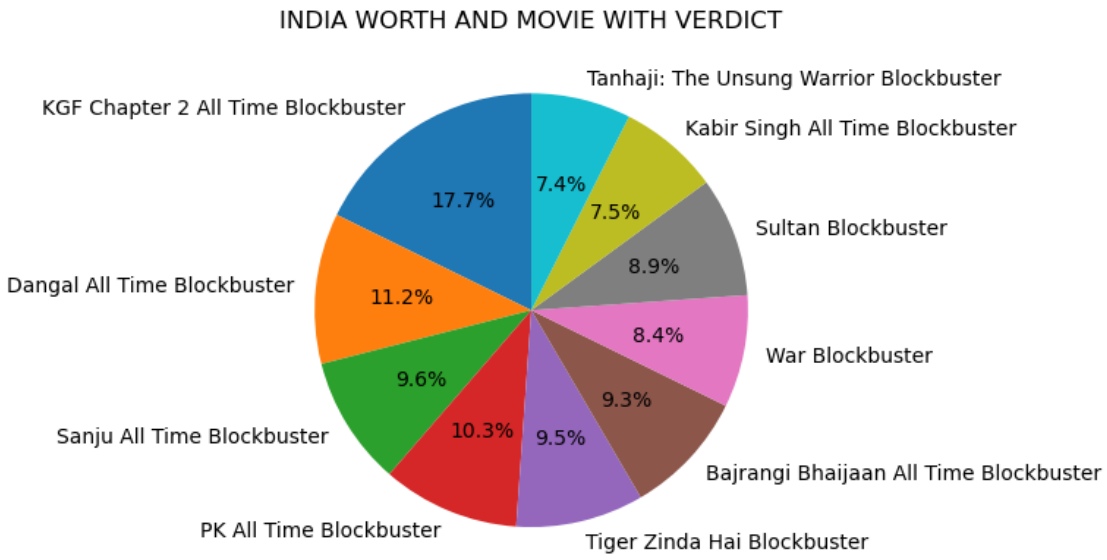
[116]:

	Movie	India worth	\
2	KGF Chapter 2	14350000000	
3	Dangal	9090000000	
4	Sanju	7800000000	
5	PK	8290000000	
6	Tiger Zinda Hai	7720000000	
7	Bajrangi Bhaijaan	7520000000	
8	War	6780000000	
10	Sultan	7210000000	
11	Kabir Singh	6080000000	
12	Tanhaji: The Unsung Warrior	6020000000	

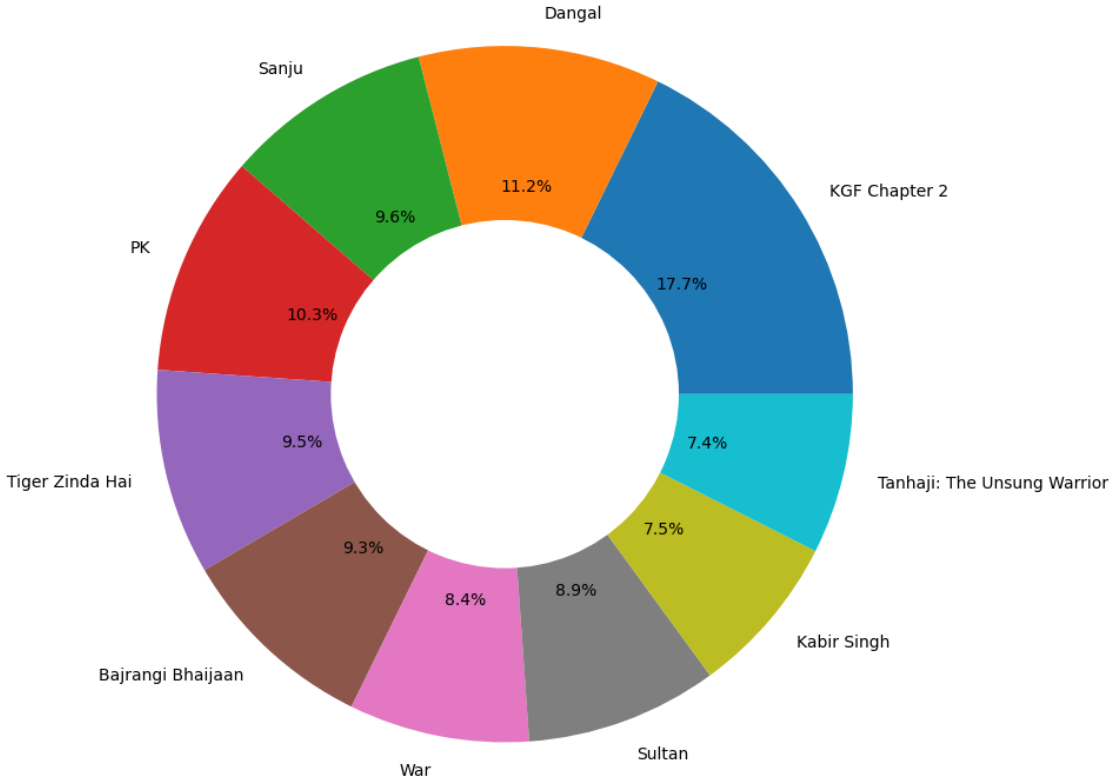
	Movie With Verdict

```
2        KGF Chapter 2 All Time Blockbuster
3               Dangal All Time Blockbuster
4                Sanju All Time Blockbuster
5                   PK All Time Blockbuster
6               Tiger Zinda Hai Blockbuster
7    Bajrangi Bhaijaan All Time Blockbuster
8                          War Blockbuster
10                      Sultan Blockbuster
11         Kabir Singh All Time Blockbuster
12  Tanhaji: The Unsung Warrior Blockbuster
```

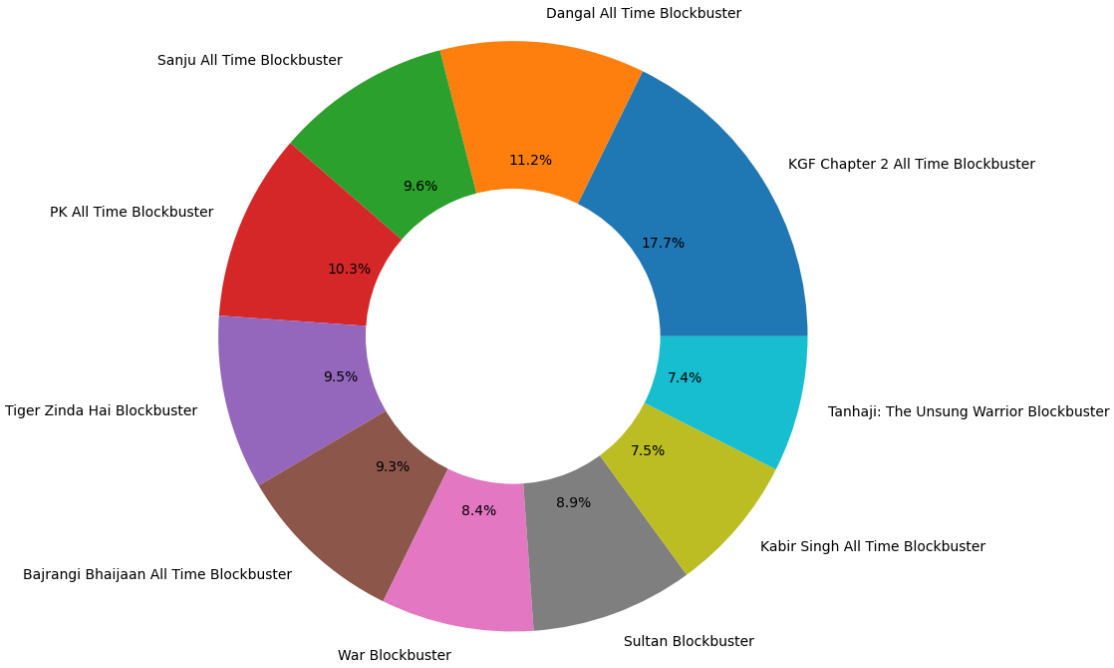
[118]:
```
#Plotting Pie Chart Of Top 10 Movies With Total India Worth Or Total Earning␣
 ↪from startangle 90 digree from 1 to 10 in order
plt.title("INDIA WORTH AND MOVIE WITH VERDICT")
plt.pie(per["India worth"],labels=per["Movie With Verdict"],autopct="%0.
 ↪1f%%",startangle=90)
plt.show()
```



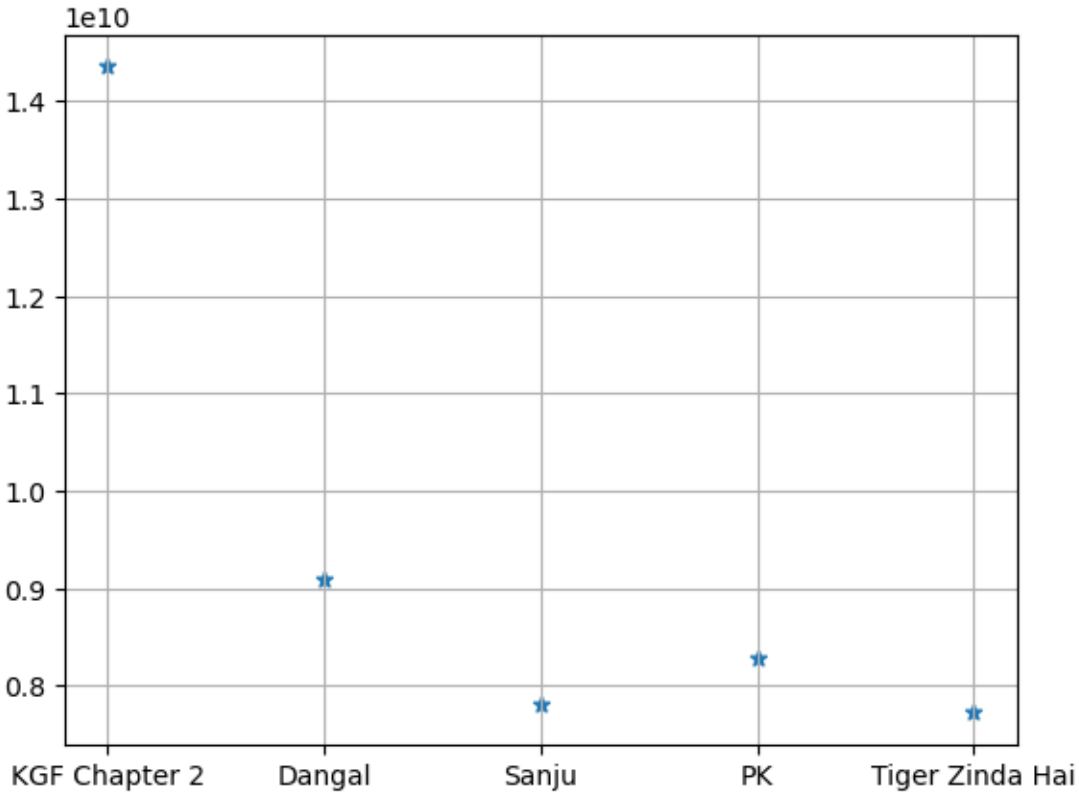
[198]:
```
#Doughnut Chart Of India Worth And Movies
plt.pie(per["India worth"],labels=per["Movie"],radius=2,autopct="%0.1f%%")
plt.pie([1],colors=["white"],radius=1)
plt.show()
```



[196]:
```
#Doughnut Chart Of India Worth And Movie With Verdict
plt.pie(per["India worth"],labels=per["Movie With␣
 ↪Verdict"],radius=2,autopct="%0.1f%%")
plt.pie([1],colors=["white"],radius=1)
plt.show()
```



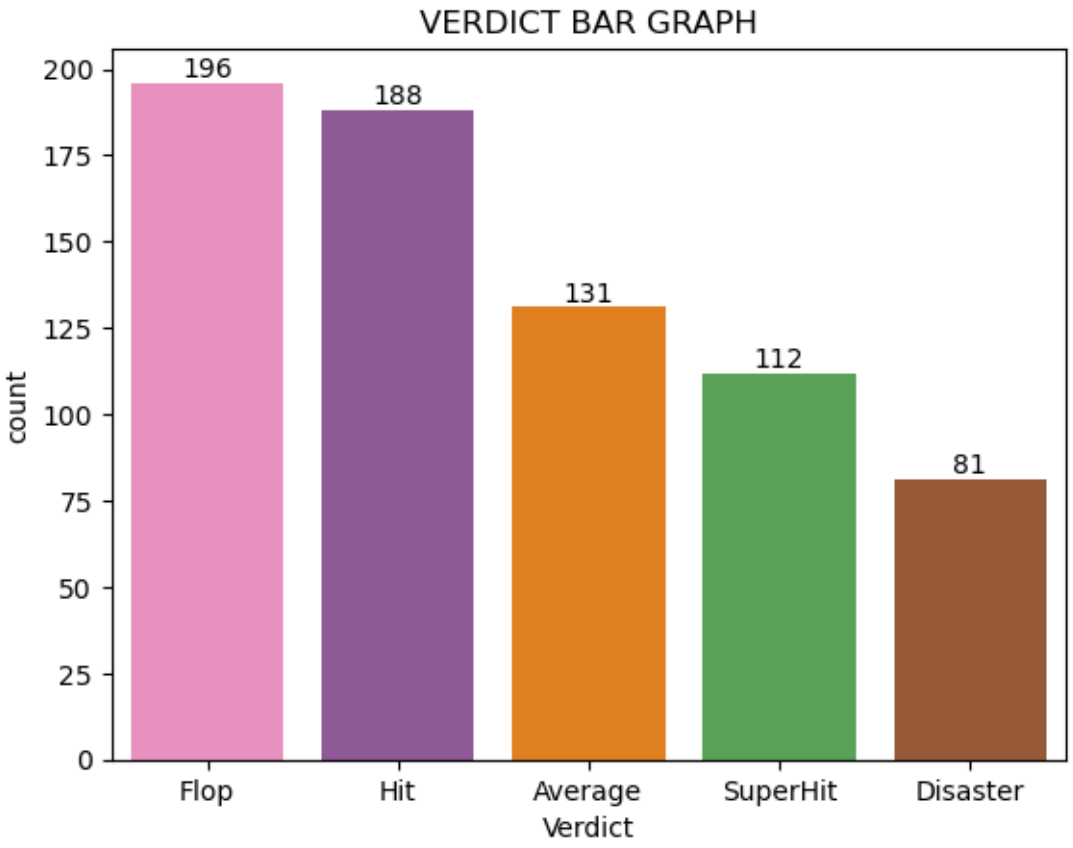
[218]:
```
#To Plot The Scatter Plot With Grids Of Top 5 Movies
plt.grid(True)
plt.scatter(ver["Movie"][0:5],list(ver["India worth"])[0:5],marker="*")
plt.show()
```



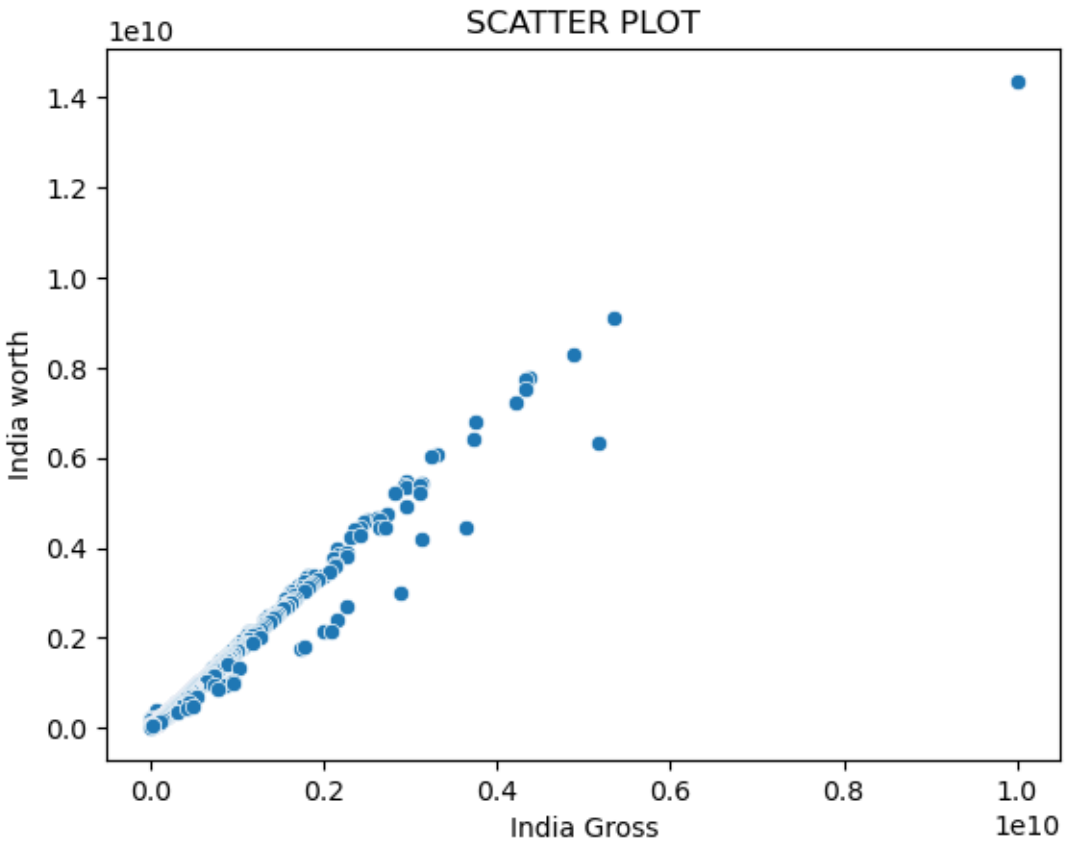
[328]:
```
#SeabornGraph Bar Graph
plt.title("VERDICT BAR GRAPH")
plot=sns.countplot(x="Verdict",data=act,hue="Verdict",order=act["Verdict"].
 ↪value_counts().nlargest(5).index, palette="Set1",legend=False)

#For Printing Values Only Contains In Seaborn
for bars in plot.containers:
    plt.bar_label(bars)

plt.show()
```



[356]:
```
#Plotting Scatter Graph Seaborn
plt.title("SCATTER PLOT")
sns.scatterplot(x="India Gross",y="India worth",data=act)
plt.show()
```



[360]:
```
#END OF PROJECT THE DATAA THAT WE USED!!!
act.head(10)
```

[360]:

	Unnamed: 0	Serial No.	Movie	Worldwide	India Net	India Gross	Overseas	Budget	Verdict
2	2	3	KGF Chapter 2	12080000000	4350000000	10000000000	2070000000	1000000000	All Time Blockbuster
3	3	4	Dangal	20700000000	3740000000	5350000000	15350000000	700000000	All Time Blockbuster
4	4	5	Sanju	5880000000	3420000000	4380000000	1500000000	1000000000	All Time Blockbuster
5	5	6	PK	7920000000	3400000000	4890000000	3030000000	850000000	All Time Blockbuster
6	6	7	Tiger Zinda Hai	5580000000					
7	7	8	Bajrangi Bhaijaan	9220000000					
8	8	9	War	4710000000					
10	10	11	Sultan	6270000000					
11	11	12	Kabir Singh	3770000000					
12	12	13	Tanhaji: The Unsung Warrior	3610000000					

```
6   3390000000   4330000000   1240000000  2100000000           Blockbuster
7   3200000000   4320000000   4890000000   900000000  All Time Blockbuster
8   3030000000   3750000000    960000000  1500000000           Blockbuster
10  3000000000   4210000000   2060000000   800000000           Blockbuster
11  2780000000   3300000000    470000000   550000000  All Time Blockbuster
12  2770000000   3250000000    360000000  1500000000           Blockbuster
```

```
                        Movie With Verdict  India worth
2        KGF Chapter 2 All Time Blockbuster  14350000000
3               Dangal All Time Blockbuster   9090000000
4                Sanju All Time Blockbuster   7800000000
5                   PK All Time Blockbuster   8290000000
6               Tiger Zinda Hai Blockbuster   7720000000
7    Bajrangi Bhaijaan All Time Blockbuster   7520000000
8                           War Blockbuster   6780000000
10                      Sultan Blockbuster   7210000000
11         Kabir Singh All Time Blockbuster   6080000000
12  Tanhaji: The Unsung Warrior Blockbuster   6020000000
```

[ ]: