KANIGALPULA GEETHA SAI THRILOKYA
Y20CS078

Lab cycle-5

This data set contains some important statistics from a large sample of movies. The data includes the movie budget and revenue from different sources as well as ratings from Rotten Tomatoes and IMDB.

```
import pandas as pd
import datetime as dt
from dateutil.parser import parse
df=pd.read_csv("/content/movies_1_.csv")
```

preprocessing:

```
k=df[df["Release Date"].str.contains("TBD")==True]
ind=k.index
df=df.drop(ind)
for i in range(len(df)):
  df.iloc[i,5]=parse(df.iloc[i,5])
df['year'] = pd.DatetimeIndex(df['Release Date']).year
df['month'] = pd.DatetimeIndex(df['Release Date']).month
```

1 (i). Find out the no of movies released in every month of the year 1995.

```
k=df[df["year"]==1995]
g=k.groupby(["month"]).count()
g["Title"]
```

Output:

```
month
1       3
2       2
3       3
4       5
5       6
6       7
7       8
8       7
9       4
10      8
11      7
12     13
Name: Title, dtype: int64
```

1.ii). Find out the no. of movies released in every year from 1990 to 1998.

```
k=df[(df["year"]>=1990) & (df["year"]<=1998)]
g=k.groupby(["year"]).count()
g["Title"]
```

Output:
```
year
1990     28
1991     33
1992     28
1993     39
1994     52
1995     73
1996     99
1997     97
1998    144
Name: Title, dtype: int64
```

2.(i). Find no. of movies released under each genre given in the database.
```
l=df.groupby('Major Genre')['Title'].count()
l
```

Output:
```
Major Genre
Action                 420
Adventure              274
Black Comedy            36
Comedy                 675
Concert/Performance      5
Documentary             43
Drama                  789
Horror                 219
Musical                 53
Romantic Comedy        137
Thriller/Suspense      239
Western                 36
Name: Title, dtype: int64
```

2.(ii). Find the movies under each genre with 1MDB rating >7 and rotten tomatoes rating > 60.
```
d=df[(df['IMDB Rating']>7) & (df['Rotten Tomatoes Rating']>60.0)]
l=d.groupby(['Major Genre','Title'])['Title'].count()
l
```

Output:
```
Major Genre  Title
Action       A Bridge Too Far
1
             Aliens
1
             Apocalypse Now
1
             Avatar
1
             Batman
1

..
Western      Pale Rider
1
```

```
              The Assassination of Jesse James by the Coward Robert Ford
1
              The Ballad of Cable Hogue
1
              The Wild Bunch
1
              Tombstone
1
Name: Title, Length: 541, dtype: int64
```

3.(i). Find the movies released under each fiction with each director in the ascending order of release dates.
k=df.sort_values(by="Release Date")
g=k.groupby(["Creative Type","Director","Title"])["Title"].count()
g

Output:

```
Creative Type         Director       Title
Contemporary Fiction  Adam McKay     Talladega Nights: The Ballad of Ricky
Bobby    1
                      Adam Shankman  A Walk to Remember
1
                                     Bringing Down the House
1
                                     Cheaper by the Dozen 2
1
                                     The Pacifier
1

..
Super Hero            Tim Burton     Batman Returns
1
                      Tim Story      Fantastic Four
1
                                     Fantastic Four: Rise of the Silver
Surfer      1
                      Warren Beatty  Dick Tracy
1
                      Zack Snyder    Watchmen
1
Name: Title, Length: 1695, dtype: int64
```

3.(ii). Find movies released under each distributor in the order of genre and director
p=df.sort_values(by=['Major Genre','Director'])
q=p.groupby(['Distributor','Title'])['Title'].count()
q

Output:

```
Distributor       Title
20th Century Fox  12 Rounds                  1
                  28 Weeks Later             1
                  A Good Year                1
                  AVP: Alien Vs. Predator    1
                  Alexander's Ragtime Band   1
                                            ..
Yash Raj Films    Veer-Zaara                 1
Zeitgeist         AimÈe & Jaguar             1
```

KANIGALPULA GEETHA SAI THRILOKYA
Y20CS078

```
                  Following                    1
                  Nowhere in Africa            1
                  Travellers and Magicians     1
Name: Title, Length: 2962, dtype: int64
```

4.(i). Find the movies released world-wide and find out the revenue received world-wide other than US with their ratings.

```
df['US Gross'] = pd.to_numeric(df['US Gross'], downcast='integer', errors='coerce')
df['Worldwide Gross'] = pd.to_numeric(df['Worldwide Gross'], downcast='integer', errors='coerce')
df['res']=df['Worldwide Gross']-df['US Gross']
f=df[['Title','res','IMDB Rating']]
f
```

Output:

	Title	res	IMDB Rating
0	The Land Girls	0.0	6.1
1	First Love, Last Rites	0.0	6.9
2	I Married a Strange Person	0.0	6.8
3	Let's Talk About Sex	0.0	NaN
4	Slam	77702.0	3.4
...	...	...	...
3196	Zack and Miri Make a Porno	5398360.0	7.0
3197	Zodiac	50000000.0	NaN
3198	Zoom	516860.0	3.4
3199	The Legend of Zorro	95900000.0	5.7
3200	The Mask of Zorro	139871255.0	6.7

3194 rows × 3 columns

4.(ii). Find the movies with loss & profit released in each year with genre and ratings.

```
df["profit & loss"]=df["Worldwide Gross"]-df["Production Budget"]
k=df.groupby(["year"])
for i,j in k:
  print(i)
  print(j[["Title","Major Genre","IMDB Rating","profit & loss"]])
```

Output:

```
1929
                     Title Major Genre  IMDB Rating  profit & loss
114  The Broadway Melody     Musical          6.7      3979000.0
1930
               Title Major Genre  IMDB Rating  profit & loss
404  Hell's Angels         NaN          7.9            NaN
1931
         Title Major Genre  IMDB Rating  profit & loss
```

```
572  Mata Hari         NaN          2.2       342000.0
1934
                Title      Major Genre  IMDB Rating  profit & loss
951  It Happened One Night  Romantic Comedy          8.3      2175000.0
1938
                      Title Major Genre  IMDB Rating  profit & loss
51     Alexander's Ragtime Band       Drama          NaN      2000000.0
1050  You Can't Take It With You         NaN          8.0      2356000.0
1939
                       Title Major Genre  IMDB Rating  profit & loss
623  Mr. Smith Goes To Washington       Drama          8.2      7500000.0
1940
        Title Major Genre  IMDB Rating  profit & loss
115  Boom Town         NaN          7.1      7172000.0
754    Rebecca       Drama          8.4      4712000.0
1942
         Title Major Genre  IMDB Rating  profit & loss
213  Casablanca       Drama          8.8      9512500.0
1944
             Title Major Genre  IMDB Rating  profit & loss
141  Bathing Beauty     Musical          6.2      1139000.0
1945
                    Title Major Genre  IMDB Rating  profit & loss
548         The Lost Weekend         NaN          8.2      9750000.0
884               Spellbound         NaN          7.7      5500000.0
1004  The Valley of Decision         NaN          7.3      6972000.0
1946
                  Title Major Genre  IMDB Rating  profit & loss
453  It's a Wonderful Life         NaN          8.7      3420000.0
661              Notorious         NaN          6.3     22464742.0
1947
                  Title Major Genre  IMDB Rating  profit & loss
383  Gentleman's Agreement         NaN          7.4      5800000.0
1948
         Title Major Genre  IMDB Rating  profit & loss
710  The Pirate         NaN          7.1      -744000.0
768   Red River         NaN          7.8      6012000.0
1949
                        Title Major Genre  IMDB Rating  profit & loss
920         Sands of Iwo Jima         NaN          7.1      6800000.0
926  She Wore a Yellow Ribbon         NaN          7.3      3800000.0
1952
        Title Major Genre  IMDB Rating  profit & loss
413  High Noon         NaN          8.3      7270000.0
1954
            Title Major Genre  IMDB Rating  profit & loss
285  The Egyptian         NaN          6.2     10000000.0
1956
                Title Major Genre  IMDB Rating  profit & loss
541   Love Me Tender         NaN          5.9      8000000.0
582        Moby Dick   Adventure          7.4      5900000.0
977          Trapeze         NaN          6.7     10400000.0
1034   War and Peace         NaN          6.8      6500000.0

2072
                                  Title Major Genre  IMDB Rating  \
21                                 1776       Drama          7.0
244                         Deep Throat         NaN          5.2
295  Everything You Always Wanted to Know      Comedy          NaN
352                              Frenzy         NaN          7.5
369                       The Godfather         NaN          9.2
```

KANIGALPULA GEETHA SAI THRILOKYA
Y20CS078

```
433                 High Plains Drifter     Western          7.6
523          The Last House on the Left         NaN          6.7

     profit & loss
21      -4000000.0
244     44975000.0
295     16016290.0
352      9100000.0
369    261500000.0
433            0.0
523      3013000.0
```