Abstract

Python Project : Movie DataSet

Author : Piyush Goyal

we have a dataset, and we have to analyze this dataset and perform many query on this.

- there is no need to import OS tool.
- we don't implement any OS function in these query

In this project I have used many terminologies-

- File handling
- data spliting
- type casting without storing data
- read and write format
- optimization techniques

advantages of this project:

- how to perform various operation on file data using file handling.
- We don't need the pandas module to perform many things.
- we don't need any other heavier module to handle all stuff.

keynotes :

- try this using exception handling.
- make this code more optimize

language : Python

QUERY-1 : Find the number of movies released between 1950 and 1960.

```python
In [2]: fh = open('movie.txt','r')
        count = 0
        for line in fh:
            l = line.split(",")
            if(int(l[2]) > 1950 and int(l[2]) < 1960):
                count += 1
        print("Number of movies between year 1950 and 1960 : ",count)
        fh.close()
```

```
Number of movies between year 1950 and 1960 :  414
```

Q2 - Find the number of movies having rating more than 4.

```python
In [2]: fh = open('movie.txt','r')
        count = 0
        for line in fh:
            l = line.split(",")
            if(l[3] > '4.0'):
                count += 1
        print(count)
        fh.close()
```

```
897
```

Q3 - Find the movies whose rating are between 3 and 4.

```
In [3]: fh = open('movie.txt','r')
        count = 0
        a = []
        for line in fh:
            l = line.split(",")
            if(l[3] > '3.0' and l[3] < '4.0'):
                a.append(l[1])
                count += 1
        print(count)
        fh.close()
```

7161

Q4 - Find the number of movies with duration more than 2 hours (7200 second).

```
In [8]: fh = open('movie.txt','r')
        count = 0
        for line in fh:
            l = line.split(",")
            if(l[4]!='' and l[4] > '7200'):
                count += 1
        print(count)
        fh.close()
```

1042

Q5 - Find the list of years and number of movies released each year.

```
In [5]: fh = open('movie.txt','r')
        count = 0
        a = []
        for line in fh:
            l = line.split(",")
            a.append(l[2])
        p = list(set(a))
        for i in range(len(p)):
            x = a.count(p[i])
            print("In ", p[i] ," Number of Movies are",x)
        fh.close()
```

```
In  2003  Number of Movies are 1399
In  2000  Number of Movies are 902
In  1953  Number of Movies are 17
In  1937  Number of Movies are 4
In  1985  Number of Movies are 334
In  1989  Number of Movies are 421
In  2014  Number of Movies are 1
In  1936  Number of Movies are 7
In  1994  Number of Movies are 517
In  1991  Number of Movies are 364
In  1924  Number of Movies are 5
In  1958  Number of Movies are 73
In  1956  Number of Movies are 60
In  1932  Number of Movies are 4
In  2011  Number of Movies are 5511
In  2012  Number of Movies are 4339
In  1973  Number of Movies are 168
In  1925  Number of Movies are 5
In  1961  Number of Movies are 119
In  1945  Number of Movies are 9
In  1914  Number of Movies are 20
In  1923  Number of Movies are 4
In  1959  Number of Movies are 87
In  1979  Number of Movies are 140
In  1919  Number of Movies are 3
In  1942  Number of Movies are 3
In  1931  Number of Movies are 3
In  1982  Number of Movies are 153
In  2004  Number of Movies are 1381
In  1993  Number of Movies are 564
In  1926  Number of Movies are 2
In  1992  Number of Movies are 479
In  1927  Number of Movies are 4
In  2001  Number of Movies are 1173
In  1965  Number of Movies are 104
In  1976  Number of Movies are 118
In  1980  Number of Movies are 107
In  1978  Number of Movies are 231
In  2013  Number of Movies are 981
In  1972  Number of Movies are 166
In  1949  Number of Movies are 9
In  1930  Number of Movies are 5
In  1920  Number of Movies are 6
In  1998  Number of Movies are 843
In  1968  Number of Movies are 173
In  1922  Number of Movies are 2
In  1915  Number of Movies are 1
In  1974  Number of Movies are 178
In  1951  Number of Movies are 33
In  1935  Number of Movies are 11
In  1981  Number of Movies are 112
In  1943  Number of Movies are 7
In  1999  Number of Movies are 1181
In  1944  Number of Movies are 10
In  1970  Number of Movies are 141
In  2009  Number of Movies are 4451
In  1928  Number of Movies are 2
```

```
In  1918  Number of Movies are 1
In  1988  Number of Movies are 334
In  1938  Number of Movies are 5
In  1990  Number of Movies are 470
In  1952  Number of Movies are 15
In  2010  Number of Movies are 5107
In  1995  Number of Movies are 592
In  1946  Number of Movies are 6
In  1913  Number of Movies are 3
In  1983  Number of Movies are 270
In  1971  Number of Movies are 131
In  1955  Number of Movies are 14
In  1941  Number of Movies are 7
In  1996  Number of Movies are 688
In  1934  Number of Movies are 8
In  1950  Number of Movies are 10
In  1966  Number of Movies are 103
In  1947  Number of Movies are 9
In  1948  Number of Movies are 13
In  1929  Number of Movies are 5
In  2005  Number of Movies are 1937
In  1997  Number of Movies are 788
In  1921  Number of Movies are 2
In  2002  Number of Movies are 1117
In  1957  Number of Movies are 98
In  2007  Number of Movies are 2892
In  1969  Number of Movies are 124
In  1975  Number of Movies are 134
In  1984  Number of Movies are 303
In  1940  Number of Movies are 9
In  1962  Number of Movies are 124
In  2006  Number of Movies are 2416
In  1963  Number of Movies are 88
In  1933  Number of Movies are 7
In  1977  Number of Movies are 136
In  1986  Number of Movies are 287
In  1954  Number of Movies are 17
In  1960  Number of Movies are 123
In  2008  Number of Movies are 3358
In  1916  Number of Movies are 1
In  1967  Number of Movies are 279
In  1987  Number of Movies are 280
In  1964  Number of Movies are 126
In  1939  Number of Movies are 6
```

Q6 - Find total number of movies in dataset

In [10]:
```python
print("Total Number of Movies : ",sum(1 for l in open('movie.txt','r')))
```

```
Total Number of Movies :  49590
```

In [ ]: