

IMDb

<https://bit.ly/2TXtBhc>

เว็บ IMDb รวบรวมข้อมูลเกี่ยวกับภาพยนตร์ต่างๆ และตัดข้อมูลบางส่วนมาให้ใช้ทำการทดลองต่างๆ ได้ <https://www.imdb.com/interfaces/>

โจทย์นี้คัดกรอง [ข้อมูลบางส่วนใน IMDb](#) มา โดยโดยมีไฟล์นี้อยู่

filteredCast.tsv

- tconst (string) - alphanumeric unique identifier of the title
- ordering (integer) – a number to uniquely identify rows for a given titleId
- nconst (string) - alphanumeric unique identifier of the name/person
- category (string) - the category of job that person was in
- job (string) - the specific job title if applicable, else 'N'
- characters (string) - the name of the character played if applicable, else 'N'

filteredTitle.tsv

- tconst (string) - alphanumeric unique identifier of the title
- titleType (string) – the type/format of the title (e.g. movie, short, tvseries, tvepisode, video, etc)
- primaryTitle (string) – the more popular title / the title used by the filmmakers on promotional materials at the point of release
- originalTitle (string) - original title, in the original language
- isAdult (boolean) - 0: non-adult title; 1: adult title
- startYear (YYYY) – represents the release year of a title. In the case of TV Series, it is the series start year
- endYear (YYYY) – TV Series end year. 'N' for all other title types
- runtimeMinutes – primary runtime of the title, in minutes
- genres (string array) – includes up to three genres associated with the title

filteredStar.tsv

- nconst (string) - alphanumeric unique identifier of the name/person
- primaryName (string)– name by which the person is most often credited
- birthYear – in YYYY format
- deathYear – in YYYY format if applicable, else 'N'
- primaryProfession (array of strings)– the top-3 professions of the person
- knownForTitles (array of tconsts) – titles the person is known for

▼ โจทย์

1. มีภาพยนตร์กี่ประเภท (genres) แต่ละประเภทมีอย่างละกี่เรื่อง พร้อมพล็อตกราฟ

2. มีนักแสดงชายและหญิงอย่างละกี่คน(นับจากไฟล์ filteredStar)
3. นักแสดงที่ยังมีชีวิตมีอายุเท่าไรบ้าง แต่ละช่วงมีกี่คน
4. มีนักแสดงกี่คนที่เคยแสดงในภาพยนตร์ Action
5. มีนักแสดงกี่คนที่แสดงในภาพยนตร์มากกว่า 1 ประเภท
6. มีภาพยนตร์ทั้งหมดกี่เรื่องที่เข้าฉายในปีอธิกสุรทิน
7. มีผู้กำกับกี่คนที่เป็นนักแสดงด้วย
8. หาจำนวนภาพยนตร์แนวโรแมนติกในแต่ละปี พร้อมพล็อตกราฟ(นับเฉพาะปีที่มีตั้งแต่ 1 เรื่องขึ้นไป)

```
1 !unzip filteredIMDB-v2.zip
```

```
Archive: filteredIMDB-v2.zip
  inflating: filteredCast-v2.tsv
  inflating: filteredStar-v2.tsv
  inflating: filteredTitle-v2.tsv
```

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 cast = pd.read_csv('filteredCast-v2.tsv' , delimiter= '\t', encoding= 'utf-8')
6 star = pd.read_csv('filteredStar-v2.tsv' , delimiter= '\t', encoding= 'utf-8')
7 title = pd.read_csv('filteredTitle-v2.tsv' , delimiter= '\t', encoding= 'utf-8')
8
9 def genres_id(s):
10     s.count(',')
11     for i in range(s.count(',')+1):
12         s1 = s.split(',')[i]
13         l.append(s1)
14     return l
15
16 def primary_sex(s):
17     s.count(',')
18     for i in range(s.count(',')+1):
19         s1 = s.split(',')[i]
20         if s1 == 'actress':
21             actress = actress + 1
22         elif s1 == 'actor':
23             actor = actor + 1
24     return actress , actor
25
```

1. มีภาพยนตร์กี่ประเภท (genres) แต่ละประเภทมีอย่างละกี่เรื่อง พร้อมพล็อตกราฟ

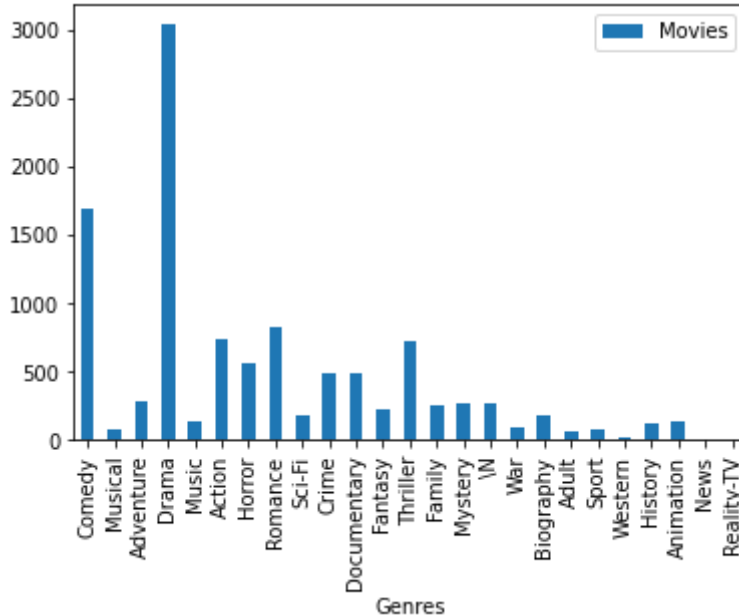
```
1 #solution
2 l=[]
```

```

3 for i in range(title.shape[0]):
4     s = title.loc[:, 'genres'][i]
5     l = genres_id(s)
6     seen = set()
7     uniq = [x for x in l if x not in seen and not seen.add(x)]
8     count_point = [l.count("Comedy"), l.count('Musical'), l.count('Adventure'), l.coun
9     df = pd.DataFrame(list(zip(uniq, count_point)),
10                        columns=['Genres', 'Movies'])
11 df.plot(kind='bar', x='Genres', y='Movies', )

```

↳ <matplotlib.axes._subplots.AxesSubplot at 0x7fa29b005198>



2. มีนักแสดงชายและหญิงอย่างละกี่คน(นับจากไฟล์ filteredStar)

```

1 #solution
2 actor = 0
3 actress = 0
4 for i in range(star.shape[0]):
5     s2 = star.loc[:, "primaryProfession"][i]
6     s2.count(',')
7     for i in range(s2.count(',')+1):
8         s2_1 = s2.split(',')[i]
9         if s2_1 == 'actress':
10             actress = actress + 1
11         elif s2_1 == 'actor':
12             actor = actor + 1
13 print('Numbers of Actor is' , actor)
14 print('Numbers of Actress is' , actress)

```

↳ Numbers of Actor is 4300
Numbers of Actress is 3537

▼ 3. นักแสดงที่ยังมีชีวิตมีอายุเท่าไรบ้าง แต่ละช่วงมีกี่คน

```
1 #solution
2
```

▼ 4. มีนักแสดงกี่คนที่เคยแสดงในภาพยนตร์ Action

```
1 #solution
```

▼ 5. มีนักแสดงกี่คนที่แสดงในภาพยนตร์มากกว่า 1 ประเภท

```
1 #solution
```

▼ 6. มีภาพยนตร์ทั้งหมดกี่เรื่องที่เข้าฉายในปีอธิกสุรทิน

```
1 #solution
2 count_a = 0
3 for i in range(title.shape[0]):
4     s6 = int(title.loc[:, 'startYear'][i])
5     if (s6%4 == 0) :
6         count_a += 1
7 print('Leap year count = ' , count_a)
8
```

```
↳ Leap year count = 1504
```

▼ 7. มีผู้กำกับกี่คนที่เป็นนักแสดงด้วย

```
1 #solution
2 num_sum = 0
3 num_act_direct = 0
4 for i in range(star.shape[0]):
5     s7 = star.loc[:, "primaryProfession"][i]
6     for i in range(s7.count(',') + 1):
7         num_act_direct = 0
8         s7_1 = s7.split(',')[i]
9         if (s7_1 == 'actress') | (s7_1 == 'actor') :
10             num_act_direct = num_act_direct + 1
11         elif (s7_1 == 'director') | (s7_1 == 'assistant_director') :
12             num_act_direct = num_act_direct + 1
```

```
13     if num_act_direct == 2 :  
14         num_sum += 1  
15 print('Profession of director and act is' , num_sum)  
16
```

➤ Profession of director and act is 0

8. หาจำนวนภาพยนตร์แนวโรแมนติกในแต่ละปี พร้อมพล็อตกราฟ(นับเฉพาะปีที่มีตั้งแต่ 1 เรืองขึ้นไป)

1 #solution