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


print("Import Successful")
```

```
Import Successful
```

```
df1 = pd.read_csv("Data-Asset/archive/tmdb_5000_credits.csv")
df2 = pd.read_csv("Data-Asset/archive/tmdb_5000_movies.csv")

print("Data Read!!")
print("Shapes : {},{}".format(df1.shape,df2.shape))
```

```
Data Read!!
Shapes : (4803, 4),(4803, 20)
```

```
print(df1.columns)
print(df2.columns)
```

```
Index(['movie_id', 'title', 'cast', 'crew'], dtype='object')
Index(['budget', 'genres', 'homepage', 'id', 'keywords',
'original_language',
       'original_title', 'overview', 'popularity',
'production_companies',
       'production_countries', 'release_date', 'revenue', 'runtime',
       'spoken_languages', 'status', 'tagline', 'title',
'vote_average',
       'vote_count'],
      dtype='object')
```

```
df1.columns = ['id', 'title_', 'cast', 'crew']
df2 = df2.merge(df1, on='id')

C = df2["vote_average"].mean()
print(C)
```

```
6.092171559442016
```

```
m = df2["vote_count"].quantile(0.9)
print(m)
```

```
1838.4000000000015
```

```
q_movies = df2.copy().loc[df2["vote_count"] >= m]
q_movies.shape
```

```
(481, 23)
```

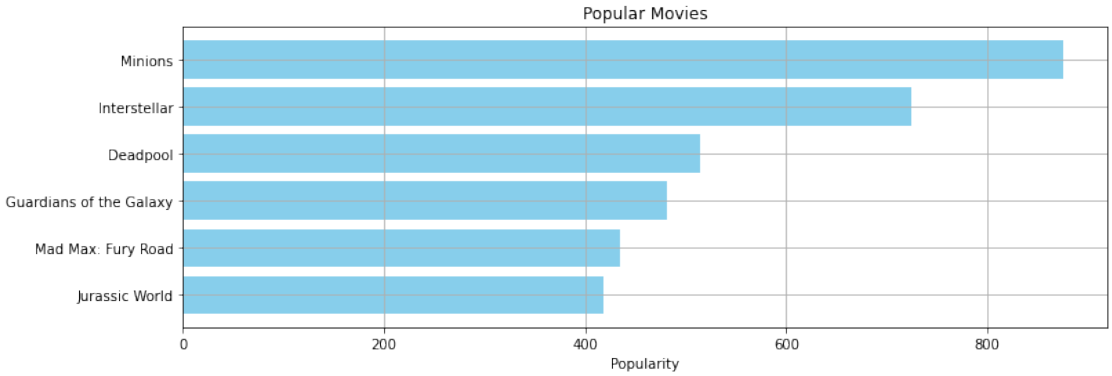
```
def weighted_rating(x,m=m,C=C):

    v = x["vote_count"]
```

```
    R = x["vote_average"]

    return round((v/(v+m)*R) + (m/(m+v)*C),2)

q_movies['score'] = q_movies.apply(weighted_rating,axis=1)

q_movies = q_movies.sort_values('score',
                                ascending=False)
q_movies[['title', 'vote_count', 'vote_average', 'score']].head(10)
```

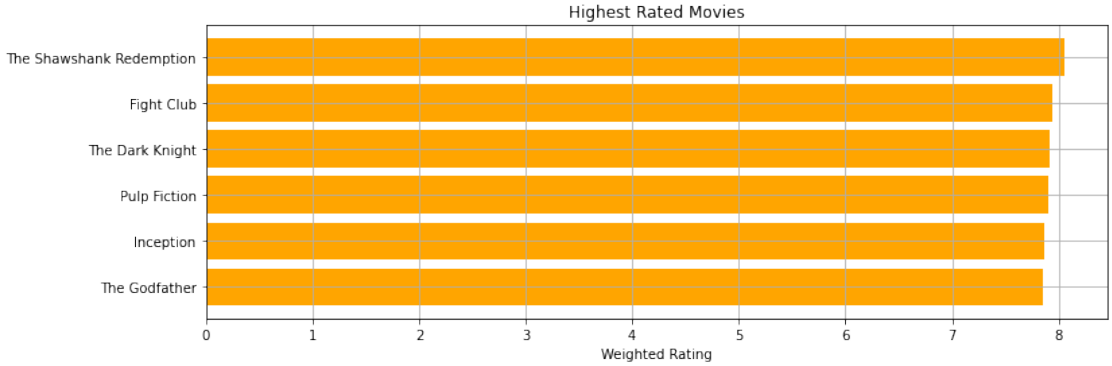
```
                                                 title  vote_count
vote_average  \
1881                         The Shawshank Redemption        8205
8.5
662                                       Fight Club        9413
8.3
65                                   The Dark Knight       12002
8.2
3232                                    Pulp Fiction        8428
8.3
96                                         Inception       13752
8.1
3337                                   The Godfather        5893
8.4
95                                      Interstellar       10867
8.1
809                                     Forrest Gump        7927
8.2
329   The Lord of the Rings: The Return of the King        8064
8.1
1990                          The Empire Strikes Back        5879
8.2


      score
1881   8.06
662    7.94
65     7.92
3232   7.90
96     7.86
3337   7.85
95     7.81
809    7.80
329    7.73
1990   7.70
```

```
pop= df2.sort_values('popularity', ascending=False)
plt.figure(figsize=(12,4))

plt.barh(pop['title'].head(6),pop['popularity'].head(6),
align='center',
        color='skyblue')
```

```
plt.gca().invert_yaxis()
plt.xlabel("Popularity")
plt.title("Popular Movies")
plt.grid()
```




```
plt.figure(figsize=(12,4))

plt.barh(q_movies['title'].head(6),q_movies['score'].head(6),
align='center',
        color='orange')
plt.gca().invert_yaxis()
plt.xlabel("Weighted Rating")
plt.title("Highest Rated Movies")
plt.grid()
```




```
msum = pop['popularity'].head(6).sum()
def normalise(x):

    return x['popularity']/msum

pop['popularity_score'] = pop.apply(normalise,
                                    axis=1)
pop[['title','popularity','popularity_score']].head(6)
```

	title	popularity	popularity_score
546	Minions	875.581305	0.253903
95	Interstellar	724.247784	0.210019

```
788                 Deadpool  514.569956          0.149216
94   Guardians of the Galaxy  481.098624          0.139510
127       Mad Max: Fury Road  434.278564          0.125933
28             Jurassic World  418.708552          0.121418
```

```
pop[['title','popularity']].to_csv('Data-Asset/pop.csv',index=False)
```

```
q_movies[['title','score']].to_csv('Data-Asset/wt-
rating.csv',index=False)
```