

## Task 1

```
In [282]: ▶ import pandas as pd
netflix = pd.read_csv(r'netflix_data.csv')

#the data frame for movies only
netflix_df_movies_only = netflix.query('type == "Movie"')

#printing movies only with selected titles
netflix_movies_col_subset = netflix_df_movies_only[['title', 'country', 'genre', 'release_year', 'duration']]
netflix_movies_col_subset.head()
```

Out[282]:

	title	country	genre	release_year	duration
1	7:19	Mexico	Dramas	2016	93
2	23:59	Singapore	Horror Movies	2011	78
3	9	United States	Action	2009	80
4	21	United States	Dramas	2008	123
6	122	Egypt	Horror Movies	2019	95

```
In [287]: ▶ #data frame for Movies only and country USA
netflix_df_country = netflix_movies_col_subset.query('country == "United States"')

movie_country_df = netflix_df_country[['title', 'country', 'genre', 'release_year', 'duration']]
movie_country_df.head()
```

Out[287]:

	title	country	genre	release_year	duration
3	9	United States	Action	2009	80
4	21	United States	Dramas	2008	123
7	187	United States	Dramas	1997	119
10	1922	United States	Dramas	2017	103
14	3022	United States	Independent Movies	2019	91

## Task 2

```
In [289]: ▶ over50_movie_country = movie_country_df.query('duration > 50')

long_genre = over50_movie_country.groupby('genre').mean()

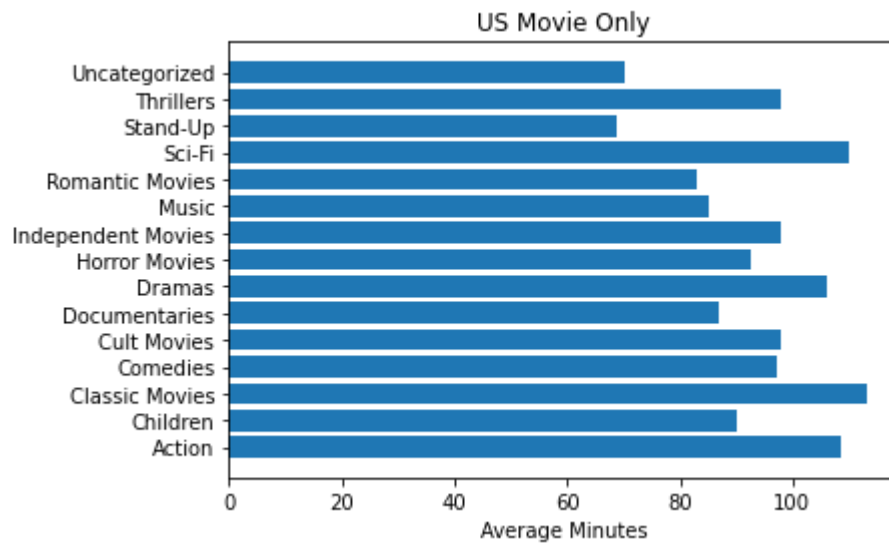
#Long_genre = Long_1[['genre', 'release_year', 'duration']]
long_genre
```

Out[289]:

	release_year	duration
genre		
Action	2008.889344	108.709016
Children	2011.111111	90.076923
Classic Movies	1971.648649	113.270270
Comedies	2012.384615	97.113846
Cult Movies	1990.111111	97.888889
Documentaries	2016.162465	86.859944
Dramas	2012.927614	105.924933
Horror Movies	2014.463636	92.690909
Independent Movies	2016.000000	98.000000
Music	2016.600000	85.000000
Romantic Movies	2017.500000	83.000000
Sci-Fi	2011.833333	109.833333
Stand-Up	2015.046154	68.656410
Thrillers	2013.300000	97.775000
Uncategorized	2014.250000	70.125000

## Task 3

```
In [290]: ▶ import matplotlib.pyplot as plt
%matplotlib inline
x = long_genre.index
y = long_genre['duration']
plt.xlabel("Average Minutes")
plt.title("US Movie Only")
plt.barh(x,y)
plt.show()
```



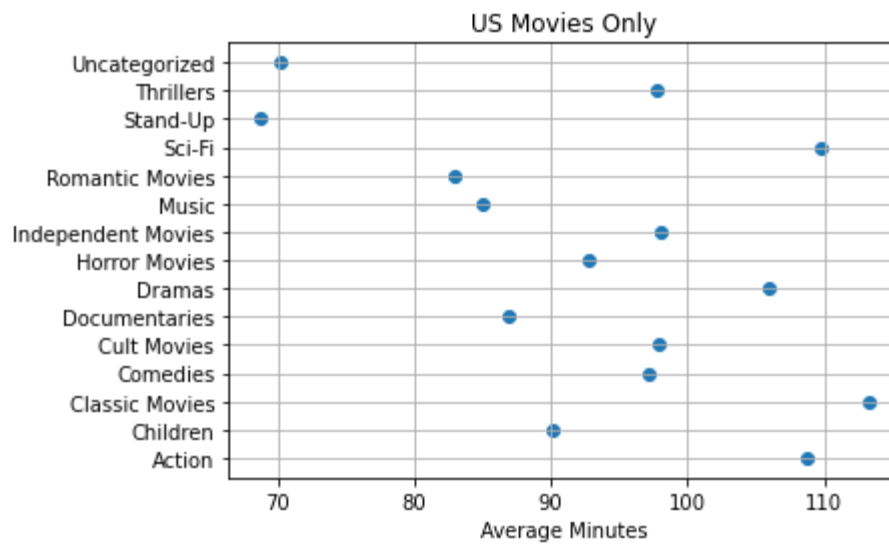
## Task 4

```
In [288]: import matplotlib.pyplot as plt
%matplotlib inline

y = long_genre.index
x = long_genre ['duration']

plt.title("US Movies Only")
plt.xlabel("Average Minutes")
plt.grid()
plt.scatter(x,y)
```

Out[288]: <matplotlib.collections.PathCollection at 0x2764f9a1550>



## Task 5

```
In [279]: movie_release = netflix.query('release_year >= 2016 & release_year <= 2020')
netflix_movie_release = movie_release[['title', 'country', 'genre', 'release_year', 'duration']]
netflix_movie_release
```

Out[279]:

	title	country	genre	release_year	duration
0	3%	Brazil	International TV	2020	4
1	7:19	Mexico	Dramas	2016	93
5	46	Turkey	International TV	2016	1
6	122	Egypt	Horror Movies	2019	95
8	706	India	Horror Movies	2019	118
...	...	...	...	...	...
7779	Zona Rosa	Mexico	International TV	2019	1
7780	Zoo	India	Dramas	2018	94
7784	Zulu Man in Japan	NaN	Documentaries	2019	44
7785	Zumbo's Just Desserts	Australia	International TV	2019	1
7786	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	United Kingdom	Documentaries	2019	90

4879 rows × 5 columns

## Task 6

```
In [280]: netflix1 = netflix_movie_release.groupby('country').count()  
netflix1
```

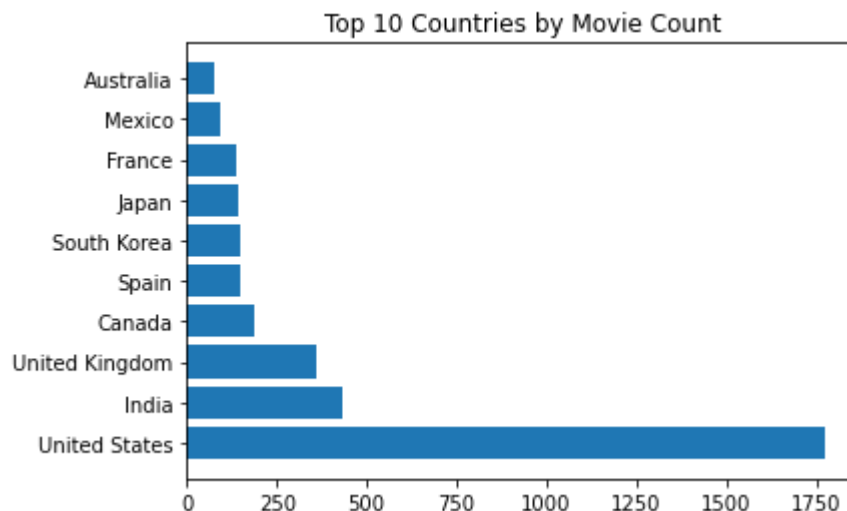
Out[280]:

	title	genre	release_year	duration
country				
Argentina	55	55	55	55
Australia	74	74	74	74
Austria	7	7	7	7
Bangladesh	2	2	2	2
Belarus	1	1	1	1
...	...	...	...	...
United States	1776	1776	1776	1776
Uruguay	8	8	8	8
Venezuela	1	1	1	1
Vietnam	3	3	3	3
Zimbabwe	1	1	1	1

78 rows × 4 columns

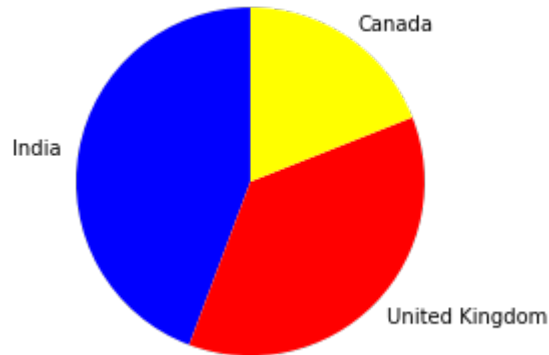
## Task 7

```
In [252]: sorted_netflix_title = netflix_movie_release.sort_values(by= 'title')  
only_10country = sorted_netflix_title ['country'].value_counts().head(10)  
  
plt.barh(only_10country.index , only_10country.values)  
plt.title("Top 10 Countries by Movie Count")  
plt.show()
```



## Task 8

```
In [253]: ▶ mylabels = only_10country.index  
mycolors = ['blue', 'red', 'yellow']  
  
#print(y)  
#print(mylabels)  
plt.pie(only_10country [1:4].values, labels = mylabels[1:4], colors = myco.  
plt.show()
```



## Task 9

In [291]:

```
import numpy as np
plt.pie(only_10country.values,
        labels = only_10country.keys(),
        textprops={'fontsize':10},

        autopct=('%1.1f%%'),

        colors=['#088F8F','yellow','#BF40BF','red','#61cb

        radius=3,
        shadow=True,
        wedgeprops = {'width': 1.7},
        startangle = 90)

circle = plt.Circle((0,0),0.6, color = 'white')
p = plt.gcf()
p.gca().add_artist(circle)
plt.title("Number of titles released by top 10 countries",fontdict={"fontst:
        pad=200)

plt.show()
```



Number of titles released by top 10 countries

