File - C:\Users\jacob\Desktop\CSCE_110\Final Project\Netflix_Movie_Tracker.py

```
# Jacob Kastenschmidt - 328000135
# Ryan Holloway - 528007777
# Stephen Shell - 228004951
# CSCE 110 - Section 505
# Final Project | Netflix Movie Tracker |
# ---------------------------------------------------------------------------------------------------------------------
# ---------------------------------------------------------------------------------------------------------------------
# ---------------------------------------------------------------------------------------------------------------------
import csv
import matplotlib.pyplot as plt
import numpy as np
# ---------------------------------------------------------------------------------------------------------------------
file_name = "2016_movie_data.csv"
all_movies = []
with open(file_name, encoding="utf8", errors="ignore") as file:
    d = csv.reader(file, delimiter=",")
    for row in d:
        all_movies.append(row)

all_movies.pop(0)
movie_names = [movie[0] for movie in all_movies]
release_date = [movie[1] for movie in all_movies]
distributor = [movie[2] for movie in all_movies]
genre = [movie[3] for movie in all_movies]
mpaa = [movie[4] for movie in all_movies]
tickets_sold = [movie[5] for movie in all_movies]

tickets_sold_num = [int(i.replace(",", "")) for i in tickets_sold]
release_date_split = [i.split("/") for i in release_date]

month_names = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October",
               "November", "December"]
month_names_abrev = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
x_pos = np.arange(len(month_names))
# ---------------------------------------------------------------------------------------------------------------------
num_movies = len(movie_names)
num_genres = len(set(genre))
num_mpaa = len(set(mpaa))
num_distributors = len(set(distributor))
```

```
num_tickets = sum(tickets_sold_num)
# ------------------------------------------------------------------------------------------------------------------------
movies_jan = 0
movies_feb = 0
movies_mar = 0
movies_apr = 0
movies_may = 0
movies_jun = 0
movies_jul = 0
movies_aug = 0
movies_sep = 0
movies_oct = 0
movies_nov = 0
movies_dec = 0

for date in release_date_split:
    if date[0] == "1":
        movies_jan += 1
    elif date[0] == "2":
        movies_feb += 1
    elif date[0] == "3":
        movies_mar += 1
    elif date[0] == "4":
        movies_apr += 1
    elif date[0] == "5":
        movies_may += 1
    elif date[0] == "6":
        movies_jun += 1
    elif date[0] == "7":
        movies_jul += 1
    elif date[0] == "8":
        movies_aug += 1
    elif date[0] == "9":
        movies_sep += 1
    elif date[0] == "10":
        movies_oct += 1
    elif date[0] == "11":
        movies_nov += 1
    elif date[0] == "12":
```

```
        movies_dec += 1

movies_per_month = [movies_jan, movies_feb, movies_mar, movies_apr, movies_may, movies_jun, movies_jul, movies_aug,
                    movies_sep, movies_oct, movies_nov, movies_dec]
max_movies = max(movies_per_month)
max_movies_month = month_names[movies_per_month.index(max_movies)]
# ------------------------------------------------------------------------------------------------------------------
tickets_jan = 0
tickets_feb = 0
tickets_mar = 0
tickets_apr = 0
tickets_may = 0
tickets_jun = 0
tickets_jul = 0
tickets_aug = 0
tickets_sep = 0
tickets_oct = 0
tickets_nov = 0
tickets_dec = 0

for i in range(len(movie_names)):
    if release_date_split[i][0] == "1":
        tickets_jan += tickets_sold_num[i]
    elif release_date_split[i][0] == "2":
        tickets_feb += tickets_sold_num[i]
    elif release_date_split[i][0] == "3":
        tickets_mar += tickets_sold_num[i]
    elif release_date_split[i][0] == "4":
        tickets_apr += tickets_sold_num[i]
    elif release_date_split[i][0] == "5":
        tickets_may += tickets_sold_num[i]
    elif release_date_split[i][0] == "6":
        tickets_jun += tickets_sold_num[i]
    elif release_date_split[i][0] == "7":
        tickets_jul += tickets_sold_num[i]
    elif release_date_split[i][0] == "8":
        tickets_aug += tickets_sold_num[i]
    elif release_date_split[i][0] == "9":
        tickets_sep += tickets_sold_num[i]
```

```
    elif release_date_split[i][0] == "10":
        tickets_oct += tickets_sold_num[i]
    elif release_date_split[i][0] == "11":
        tickets_nov += tickets_sold_num[i]
    elif release_date_split[i][0] == "12":
        tickets_dec += tickets_sold_num[i]

tickets_per_month = [tickets_jan, tickets_feb, tickets_mar, tickets_apr, tickets_may, tickets_jun, tickets_jul,
                     tickets_aug, tickets_sep, tickets_oct, tickets_nov, tickets_dec]
max_tickets = max(tickets_per_month)
max_tickets_month = month_names[tickets_per_month.index(max_tickets)]
# ------------------------------------------------------------------------------------------------------------
percentage_distributors = {}

for d in set(distributor):
    percentage_distributors["{}".format(d)] = 0
    for i in range(len(movie_names)):
        if distributor[i] == "{}".format(d):
            percentage_distributors["{}".format(d)] += tickets_sold_num[i] / num_tickets

distributor_list_dp = []
percentage_list_dp = []

for d in percentage_distributors:
    distributor_list_dp.append(d)
    percentage_list_dp.append(percentage_distributors[d])

others = []
for i in percentage_list_dp:
    if i < .01:
        others.append(i)

for o in others:
    i = percentage_list_dp.index(o)
    distributor_list_dp.remove(distributor_list_dp[i])
    percentage_list_dp.remove(o)

percentage_list_dp.append(sum(others))
distributor_list_dp.append("Others")
```

```
percentage_list_dp, distributor_list_dp = (list(t) for t in zip(*sorted(zip(percentage_list_dp, distributor_list_dp))))
# ------------------------------------------------------------------------------------------------------------------
drama = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
horror = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
action = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
comedy = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

for i in range(len(movie_names)):
    if genre[i] == "Drama":
        if release_date_split[i][0] == "1":
            drama[0] += 1
        elif release_date_split[i][0] == "2":
            drama[1] += 1
        elif release_date_split[i][0] == "3":
            drama[2] += 1
        elif release_date_split[i][0] == "4":
            drama[3] += 1
        elif release_date_split[i][0] == "5":
            drama[4] += 1
        elif release_date_split[i][0] == "6":
            drama[5] += 1
        elif release_date_split[i][0] == "7":
            drama[6] += 1
        elif release_date_split[i][0] == "8":
            drama[7] += 1
        elif release_date_split[i][0] == "9":
            drama[8] += 1
        elif release_date_split[i][0] == "10":
            drama[9] += 1
        elif release_date_split[i][0] == "11":
            drama[10] += 1
        elif release_date_split[i][0] == "12":
            drama[11] += 1
    elif genre[i] == "Horror":
        if release_date_split[i][0] == "1":
            horror[0] += 1
        elif release_date_split[i][0] == "2":
            horror[1] += 1
```

```
        elif release_date_split[i][0] == "3":
            horror[2] += 1
        elif release_date_split[i][0] == "4":
            horror[3] += 1
        elif release_date_split[i][0] == "5":
            horror[4] += 1
        elif release_date_split[i][0] == "6":
            horror[5] += 1
        elif release_date_split[i][0] == "7":
            horror[6] += 1
        elif release_date_split[i][0] == "8":
            horror[7] += 1
        elif release_date_split[i][0] == "9":
            horror[8] += 1
        elif release_date_split[i][0] == "10":
            horror[9] += 1
        elif release_date_split[i][0] == "11":
            horror[10] += 1
        elif release_date_split[i][0] == "12":
            horror[11] += 1
    elif genre[i] == "Action":
        if release_date_split[i][0] == "1":
            action[0] += 1
        elif release_date_split[i][0] == "2":
            action[1] += 1
        elif release_date_split[i][0] == "3":
            action[2] += 1
        elif release_date_split[i][0] == "4":
            action[3] += 1
        elif release_date_split[i][0] == "5":
            action[4] += 1
        elif release_date_split[i][0] == "6":
            action[5] += 1
        elif release_date_split[i][0] == "7":
            action[6] += 1
        elif release_date_split[i][0] == "8":
            action[7] += 1
        elif release_date_split[i][0] == "9":
            action[8] += 1
```

```
        elif release_date_split[i][0] == "10":
            action[9] += 1
        elif release_date_split[i][0] == "11":
            action[10] += 1
        elif release_date_split[i][0] == "12":
            action[11] += 1
    elif genre[i] == "Comedy":
        if release_date_split[i][0] == "1":
            comedy[0] += 1
        elif release_date_split[i][0] == "2":
            comedy[1] += 1
        elif release_date_split[i][0] == "3":
            comedy[2] += 1
        elif release_date_split[i][0] == "4":
            comedy[3] += 1
        elif release_date_split[i][0] == "5":
            comedy[4] += 1
        elif release_date_split[i][0] == "6":
            comedy[5] += 1
        elif release_date_split[i][0] == "7":
            comedy[6] += 1
        elif release_date_split[i][0] == "8":
            comedy[7] += 1
        elif release_date_split[i][0] == "9":
            comedy[8] += 1
        elif release_date_split[i][0] == "10":
            comedy[9] += 1
        elif release_date_split[i][0] == "11":
            comedy[10] += 1
        elif release_date_split[i][0] == "12":
            comedy[11] += 1
# ------------------------------------------------------------------------------------------------------------------------
# ------------------------------------------------------------------------------------------------------------------------
# ------------------------------------------------------------------------------------------------------------------------
print("=========== Dataset Details ===========")
print()
print("Number of Movies: " + str(num_movies))
print("Number of different genres: " + str(num_genres))
print("Number of different MPAA: " + str(num_mpaa))
```

```
print("Number of different distributors: " + str(num_distributors))
print("Total number of tickets sold: " + str(num_tickets))
print()
print("========================================")
print()
print("Most number of movies released (" + str(max_movies) + ") in " + max_movies_month + ".")
print("Most amount of tickets sold (" + str(max_tickets) + ") in " + max_tickets_month + ".")
print()
print("========================================")
print()
print("===== Tickets sold by distributors =====")
print()
for i in range(len(percentage_list_dp)):
    print(distributor_list_dp[i] + ": %" + str(round(100 * percentage_list_dp[i], 2)))
print()
print("========================================")
# ------------------------------------------------------------------------------------------------------------------
fig = plt.figure(figsize=(20, 10))
plt.bar(x_pos, movies_per_month, align='center')
plt.xticks(x_pos, month_names)
plt.xlabel("Month")
plt.ylabel("Number of Movies released")
plt.title("Number of movies released in different months of 2016")

plt.tight_layout()
plt.show()
# ------------------------------------------------------------------------------------------------------------------
fig2 = plt.figure(figsize=(20, 10))
plt.plot(x_pos, tickets_per_month)
plt.xticks(x_pos, month_names)
plt.ylabel("Number of tickets sold")
plt.xlabel("Month")
plt.title("Number of tickets sold in different months of 2016")

plt.tight_layout()
plt.show()
# ------------------------------------------------------------------------------------------------------------------
fig3 = plt.figure(figsize=(20, 10))
colors = ["red", "green", "pink", "orange", "purple", "turquoise", "maroon", "grey", "blue", "lime"]
```

```
plt.pie(percentage_list_dp, labels=distributor_list_dp, colors=colors, startangle=180, autopct='%1.1f%%', radius=1.2)
plt.title("Percentage of tickets sold by different distributors")

plt.show()
# ------------------------------------------------------------------------------------------------------------------
fig4 = plt.figure(figsize=(20, 10))
plt.plot(x_pos, drama, color="red")
plt.plot(x_pos, horror, color="blue")
plt.plot(x_pos, action, color="green")
plt.plot(x_pos, comedy, color="purple")
plt.xticks(x_pos, month_names)
plt.ylabel("Number of Movies")
plt.xlabel("Month")
plt.title("Number of movies per genre released in different months of 2016")

plt.legend(["Drama", "Horror", "Action", "Comedy"])
plt.tight_layout()
plt.show()
# ------------------------------------------------------------------------------------------------------------------
```