# Project

##### Project Instructions

Your friend suspects that movies are getting shorter and they've found some initial evidence of this. Having peaked your interest, you will perform exploratory data analysis on the netflix\_data.csv data to understand what may be contributing to movies getting shorter over time. Your analysis will follow these steps:

* Load the CSV file and store as netflix\_df.
* Filter the data to remove TV shows and store as netflix\_subset.
* Investigate and subset the Netflix movie data, keeping only the columns "title", "country", "genre", "release\_year", "duration", and saving this into a new DataFrame called netflix\_movies.
* Filter netflix\_movies to find the movies that are strictly shorter than 60 minutes, saving the resulting DataFrame as short\_movies; inspect the result to find possible contributing factors.
* Using a for loop and if/elif statements, iterate through the rows of netflix\_movies and assign colors of your choice to four genre groups ("Children", "Documentaries", "Stand-Up", and "Other" for everything else). Save the results in a colors list. Initialize a matplotlib figure object called fig and create a scatter plot for movie duration by release year using the colors list to color the points and using the labels "Release year" for the x-axis, "Duration (min)" for the y-axis, and the title "Movie Duration by Year of Release".
* After inspecting the plot, answer the question "Are we certain that movies are getting shorter?" by assigning either "yes" or "no" to the variable answer.
* Click the "Submit Project" button to check your solution.

## Explanation

**Netflix**! What started in 1997 as a DVD rental service has since exploded into one of the largest entertainment and media companies.

Given the large number of movies and series available on the platform, it is a perfect opportunity to flex your exploratory data analysis skills and dive into the entertainment industry. Our friend has also been brushing up on their Python skills and has taken a first crack at a CSV file containing Netflix data. They believe that the average duration of movies has been declining. Using your friends initial research, you'll delve into the Netflix data to see if you can determine whether movie lengths are actually getting shorter and explain some of the contributing factors, if any.

You have been supplied with the dataset netflix\_data.csv , along with the following table detailing the column names and descriptions. This data does contain null values and some outliers, but handling these is out of scope for the project. Feel free to experiment after submitting!

### The data

#### **netflix\_data.csv**

| **Column** | **Description** |
| --- | --- |
| show\_id | The ID of the show |
| type | Type of show |
| title | Title of the show |
| director | Director of the show |
| cast | Cast of the show |
| country | Country of origin |
| date\_added | Date added to Netflix |
| release\_year | Year of Netflix release |
| duration | Duration of the show in minutes |
| description | Description of the show |
| genre | Show genre  CODE STARTING |

1

# Importing pandas and matplotlib

2

import pandas as pd

3

import matplotlib.pyplot as plt

4

# Start coding!

Follow the instructions and code

Happy coding ☺