

Investigating Netflix Movies

November 24, 2023

1 Project Description :

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 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 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 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", "no", or "maybe" to the variable `answer`. Click the "Submit Project" button to check your solution.

```
[26]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[6]: # read the netflix series as a DataFrame
```

```
[4]: netflix_df = pd.read_csv(r"C:\Users\Asad Raja\Downloads\netflix_data (1).csv")
```

```
[5]: netflix_df.head()
```

```
[5]:
```

| | show_id | type | title | director | \ |
|---|---------|---------|-------|-------------------|---|
| 0 | s1 | TV Show | 3% | NaN | |
| 1 | s2 | Movie | 7:19 | Jorge Michel Grau | |
| 2 | s3 | Movie | 23:59 | Gilbert Chan | |
| 3 | s4 | Movie | 9 | Shane Acker | |
| 4 | s5 | Movie | 21 | Robert Luketic | |

| | cast | country \ |
|---|---|---------------|
| 0 | João Miguel, Bianca Comparato, Michel Gomes, R... | Brazil |
| 1 | Demián Bichir, Héctor Bonilla, Oscar Serrano, ... | Mexico |
| 2 | Tedd Chan, Stella Chung, Henley Hii, Lawrence ... | Singapore |
| 3 | Elijah Wood, John C. Reilly, Jennifer Connelly... | United States |
| 4 | Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar... | United States |

| | date_added | release_year | duration \ |
|---|-------------------|--------------|------------|
| 0 | August 14, 2020 | 2020 | 4 |
| 1 | December 23, 2016 | 2016 | 93 |
| 2 | December 20, 2018 | 2011 | 78 |
| 3 | November 16, 2017 | 2009 | 80 |
| 4 | January 1, 2020 | 2008 | 123 |

| | description | genre |
|---|---|------------------|
| 0 | In a future where the elite inhabit an island ... | International TV |
| 1 | After a devastating earthquake hits Mexico Cit... | Dramas |
| 2 | When an army recruit is found dead, his fellow... | Horror Movies |
| 3 | In a postapocalyptic world, rag-doll robots hi... | Action |
| 4 | A brilliant group of students become card-coun... | Dramas |

```
[7]: #subset the dataframe for type "Movies"
```

```
[8]: netflix_subset = netflix_df[netflix_df["type"] == "Movie"]
```

```
[9]: print(netflix_subset)
```

| | show_id | type | title \ |
|------|---------|-------|---|
| 1 | s2 | Movie | 7:19 |
| 2 | s3 | Movie | 23:59 |
| 3 | s4 | Movie | 9 |
| 4 | s5 | Movie | 21 |
| 6 | s7 | Movie | 122 |
| ... | ... | ... | ... |
| 7781 | s7782 | Movie | Zoom |
| 7782 | s7783 | Movie | Zozo |
| 7783 | s7784 | Movie | Zubaan |
| 7784 | s7785 | Movie | Zulu Man in Japan |
| 7786 | s7787 | Movie | ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS |

| | director | cast \ |
|-----|-------------------|---|
| 1 | Jorge Michel Grau | Demián Bichir, Héctor Bonilla, Oscar Serrano, ... |
| 2 | Gilbert Chan | Tedd Chan, Stella Chung, Henley Hii, Lawrence ... |
| 3 | Shane Acker | Elijah Wood, John C. Reilly, Jennifer Connelly... |
| 4 | Robert Luketic | Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar... |
| 6 | Yasir Al Yasiri | Amina Khalil, Ahmed Dawood, Tarek Lotfy, Ahmed... |
| ... | ... | ... |

| | | |
|------|--------------|---|
| 7781 | Peter Hewitt | Tim Allen, Courteney Cox, Chevy Chase, Kate Ma... |
| 7782 | Josef Fares | Imad Creidi, Antoinette Turk, Elias Gergi, Car... |
| 7783 | Mozez Singh | Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan... |
| 7784 | NaN | Nasty C |
| 7786 | Sam Dunn | NaN |

| | country | date_added | release_year | duration | \ |
|------|----------------|--------------------|--------------|----------|---|
| 1 | Mexico | December 23, 2016 | 2016 | 93 | |
| 2 | Singapore | December 20, 2018 | 2011 | 78 | |
| 3 | United States | November 16, 2017 | 2009 | 80 | |
| 4 | United States | January 1, 2020 | 2008 | 123 | |
| 6 | Egypt | June 1, 2020 | 2019 | 95 | |
| ... | ... | ... | ... | ... | |
| 7781 | United States | January 11, 2020 | 2006 | 88 | |
| 7782 | Sweden | October 19, 2020 | 2005 | 99 | |
| 7783 | India | March 2, 2019 | 2015 | 111 | |
| 7784 | NaN | September 25, 2020 | 2019 | 44 | |
| 7786 | United Kingdom | March 1, 2020 | 2019 | 90 | |

| | description | genre |
|------|---|---------------|
| 1 | After a devastating earthquake hits Mexico Cit... | Dramas |
| 2 | When an army recruit is found dead, his fellow... | Horror Movies |
| 3 | In a postapocalyptic world, rag-doll robots hi... | Action |
| 4 | A brilliant group of students become card-coun... | Dramas |
| 6 | After an awful accident, a couple admitted to ... | Horror Movies |
| ... | ... | ... |
| 7781 | Dragged from civilian life, a former superhero... | Children |
| 7782 | When Lebanon's Civil War deprives Zozo of his ... | Dramas |
| 7783 | A scrappy but poor boy worms his way into a ty... | Dramas |
| 7784 | In this documentary, South African rapper Nast... | Documentaries |
| 7786 | This documentary delves into the mystique behi... | Documentaries |

[5377 rows x 11 columns]

```
[10]: # select only the columns of interest
```

```
[12]: netflix_movies = netflix_subset[["title", "country", "genre", "release_year", "
↳ "duration"]]
```

```
[13]: netflix_movies.head()
```

```
[13]:
```

| | 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 |

```
[14]: # filter for duration of 60 minutes
```

```
[15]: short_movies = netflix_movies[netflix_movies["duration"] < 60]
```

```
[16]: print(short_movies)
```

| | title | country | |
|------|---|---------------|--|
| 35 | #Rucker50 | United States | |
| 55 | 100 Things to do Before High School | United States | |
| 67 | 13TH: A Conversation with Oprah Winfrey & Ava ... | NaN | |
| 101 | 3 Seconds Divorce | Canada | |
| 146 | A 3 Minute Hug | Mexico | |
| ... | ... | ... | |
| 7679 | WWII: Report from the Aleutians | United States | |
| 7692 | Ya no estoy aquí: Una conversación entre Guill... | NaN | |
| 7718 | Yoo Byung Jae: Discomfort Zone | South Korea | |
| 7771 | Zion | United States | |
| 7784 | Zulu Man in Japan | NaN | |

| | genre | release_year | duration |
|------|---------------|--------------|----------|
| 35 | Documentaries | 2016 | 56 |
| 55 | Uncategorized | 2014 | 44 |
| 67 | Uncategorized | 2017 | 37 |
| 101 | Documentaries | 2018 | 53 |
| 146 | Documentaries | 2019 | 28 |
| ... | ... | ... | ... |
| 7679 | Documentaries | 1943 | 45 |
| 7692 | Documentaries | 2020 | 15 |
| 7718 | Stand-Up | 2018 | 54 |
| 7771 | Documentaries | 2018 | 12 |
| 7784 | Documentaries | 2019 | 44 |

[420 rows x 5 columns]

```
[17]: # define an empty list
```

```
[20]: colors = []
```

```
[19]: # iterate over rows of netflix
```

```
[22]: for label, row in netflix_movies.iterrows():  
  
    if row["genre"] == "Children":  
        colors.append("red")  
  
    elif row["genre"] == "Documentaries":  
        colors.append("blue")
```

```
elif row["genre"] == "Stand-Up":  
    colors.append("green")  
  
else:  
    colors.append('black')
```

```
[23]: # inspect the first 10 values of list
```

```
[24]: colors[:10]
```

```
[24]: ['black',  
      'black',  
      'black',  
      'black',  
      'black',  
      'black',  
      'black',  
      'black',  
      'black',  
      'black']
```

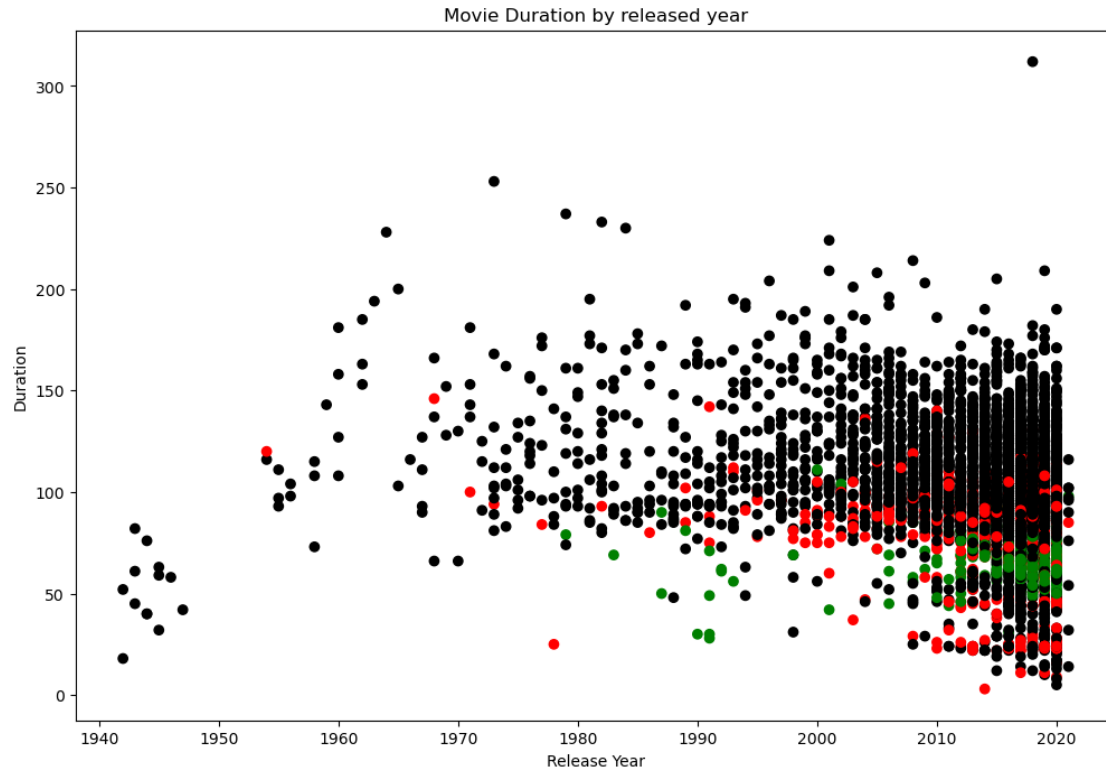
```
[25]: # Visualize the Plot
```

```
[28]: fig = plt.figure(figsize = (12,8))
```

<Figure size 1200x800 with 0 Axes>

```
[29]: # create the scatter plot
```

```
[34]: fig = plt.figure(figsize = (12,8))  
  
plt.scatter(netflix_movies.release_year, netflix_movies.duration , c = colors)  
  
plt.title("Movie Duration by released year")  
  
plt.xlabel("Release Year")  
  
plt.ylabel("Duration")  
  
plt.show()
```



```
[35]: # Are we certain that movies are getting shorter  
      answer = "may be"
```

2 Datacamp Guided by Project