



ML/DL Track - Task 3

🕒 Created	@June 19, 2024 5:11 PM
☑ Reviewed	✓

Netflix Movies Analysis Report

Introduction

This report provides an analysis of the duration of movies on Netflix, focusing particularly on short movies (those under 60 minutes). The analysis also investigates the genres and countries of origin for these short movies to identify any patterns or trends.

Data Preparation

We start by loading the dataset and filtering it based on the type of content: Movies and TV Shows. Then, we create a subset of the dataset containing only movies and further filter it to find movies with a duration of less than 60 minutes.

```
import pandas as pd

# Load the dataset
df = pd.read_csv('netflix_data.csv')
```

```
# Filter the dataset to include only movies
netflix_movies = df[df['type'] == 'Movie']

# Select relevant columns
netflix_movies = netflix_movies[['title', 'country', 'genre',
    'release_year', 'duration']]

# Filter movies with a duration of less than 60 minutes
short_movies = netflix_movies[netflix_movies['duration'] < 60]

# Display the short movies
print(short_movies)
```

Observations

From the output, we can see that most short movies are usually from the United States and predominantly from the genre of Documentaries.

Descriptive Analysis

To further validate these observations, we use the `describe` function on the relevant columns.

```
print("\nGenre Description:")
print(short_movies['genre'].describe())

print("\nCountry Description:")
print(short_movies['country'].describe())
```

Results

- **Genre Description:** Most short movies belong to the 'Documentaries' genre.
- **Country Description:** The United States is the most common country of origin for short movies.

Visualization

We create a scatter plot to visualize the duration of movies by their release year, using different colors to represent different genres.

```
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches

colors = []
for index, 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('yellow')

fig = plt.figure()
plt.scatter(netflix_movies['release_year'], netflix_movies['duration'], c=colors)
plt.xlabel('Release year')
plt.ylabel('Duration (min)')
plt.title('Movie Duration by Year of Release')

red_patch = mpatches.Patch(color='red', label='Children')
```

```
blue_patch = mpatches.Patch(color='blue', label='Documentaries')
green_patch = mpatches.Patch(color='green', label='Stand-Up')
yellow_patch = mpatches.Patch(color='yellow', label='Other')

plt.legend(handles=[red_patch, blue_patch, green_patch, yellow_patch], title="Genres")

plt.show()
```

Summary

- The majority of short movies (under 60 minutes) on Netflix are from the United States and belong to the 'Documentaries' genre.
- A scatter plot shows the distribution of movie durations by their release year, with genres color-coded for clarity.

Key Takeaways

- **Genre and Country:** The 'Documentaries' genre and the United States dominate the short movie category on Netflix.
- **Visualization Insights:** The scatter plot helps visualize how movie durations have varied over the years, with genre distinctions.

Conclusion

The analysis provides insights into the characteristics of short movies on Netflix. Although we have observed that a significant number of short movies are documentaries from the United States, we cannot conclusively say that movies are getting shorter over time based on the given data.

Answer: Are we certain that movies are getting shorter? **No.**

This report concludes that while there are patterns in the genres and origins of short movies, further data and analysis would be needed to determine trends in movie durations over time.
