PROJECT: INVESTIGATING NETFLIX MOVIES

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.

You work for a production company that specializes in nostalgic styles. You want to do some research on movies released in the 1990's. You'll delve into Netflix data and perform exploratory data analysis to better understand this awesome movie decade!

You have been supplied with the dataset `netflix_data.csv`, along with the following table detailing the column names and descriptions. Feel free to experiment further 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

```
# Importing pandas and matplotlib
import pandas as pd
import matplotlib.pyplot as plt

# Read in the Netflix CSV as a DataFrame
netflix_df = pd.read_csv("netflix_data.csv")

# Assuming 'release_year' is a column in the DataFrame
```

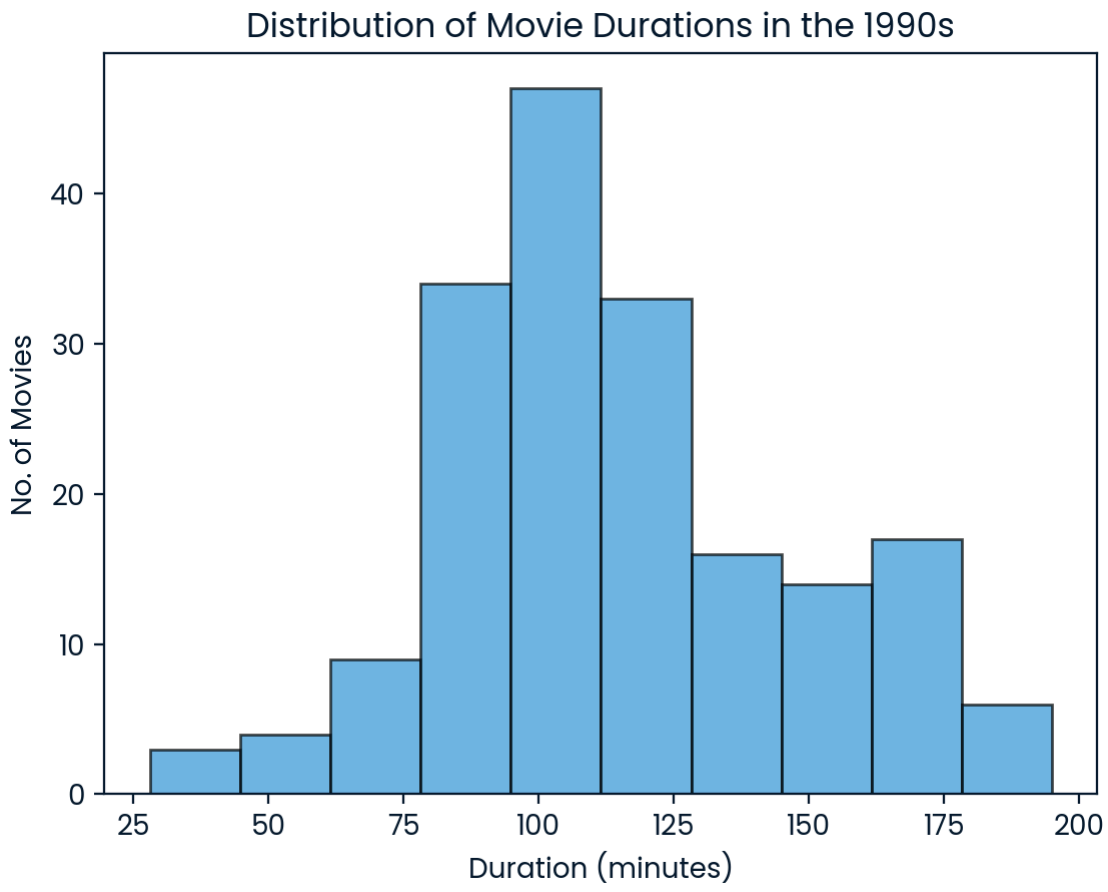
```
# Subset the DataFrame for type "Movie"
netflix_subset = netflix_df[netflix_df["type"] == "Movie"]

# Filter to keep only movies released in the 1990s
new_subset = netflix_subset[(netflix_subset["release_year"] >= 1990)]
movies_1990s = new_subset[(new_subset["release_year"] < 2000)]

# Visualizing the distribution with a custom color
plt.hist(movies_1990s["duration"], color='#3498db', edgecolor='black', alpha=0.7)
 # Blue color
plt.title('Distribution of Movie Durations in the 1990s')
plt.xlabel('Duration (minutes)')
plt.ylabel('No. of Movies')
plt.show()

# Counter for short movies (<90 min)
short_movie_count = 0
short_movie_count = 0
for label, row in action_movies_1990s.iterrows():
    if row["duration"] < 90:
        short_movie_count = short_movie_count + 1
    else:
        short_movie_count = short_movie_count  # This line does nothing!

print(f"Number of short movies duration <90 min in the 1990s:
{short_movie_count}")
```



```
Number of short movies duration <90 min in the 1990s: 7
```