


A person is seen from behind, sitting on a couch and watching a large television. The room has a brick wall and a wooden desk in front of the TV. On the desk, there are two speakers, a small white device, and some books. The overall atmosphere is cozy and modern.

NETFLIX

INVESTIGATING NETFLIX MOVIES



OVERVIEW



Netflix which began in 1997 as a DVD rental service, has grown into one of the world's largest entertainment and media companies. With the vast selection of movies and series available on the platform, we can exploratory data analysis and explore the entertainment industry. We suspect that the average duration of movies has been decreasing. Building on their preliminary work, we determine whether movie lengths are really getting shorter and explore any potential contributing factors.




GOALS

01

EXPLORE MOVIE DURATIONS

Identify the most frequent
movie duration in the
1990s.



02

ANALYZE SHORT ACTION MOVIES

Count the number of
short action movies (less
than 90 minutes) released
in the 1990s.

DATASET

The dataset is provided on `netflix_data.csv`, along with a table outlining the column names and their descriptions. While the data does include some null values and outliers, dealing with these is beyond the scope of this project.

- **show_id** - Unique identifier for the show
- **type** - Show type (movie or series)
- **title** - Title of the show
- **director** - Director(s) of the show
- **cast** - Cast members of the show
- **country** - Country of origin
- **date_added** - Date the show was added to Netflix
- **release_year** - Year the show was released
- **duration** - Duration of the show in minutes
- **description** - Description of the show
- **genre** - Genre of the show



OUTPUTS

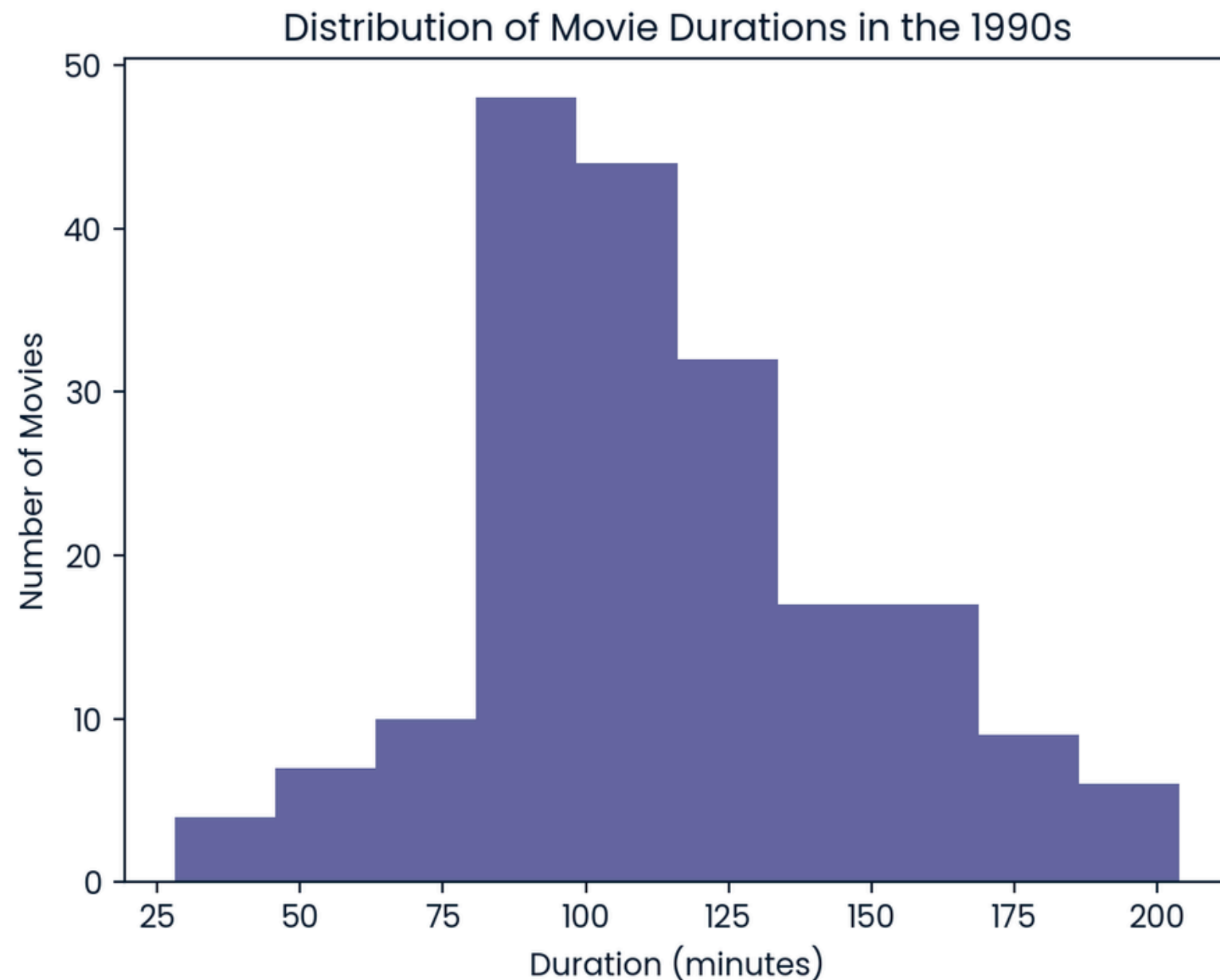
Conduct exploratory data analysis on the `netflix_data.csv` dataset to gain insights into movies from the 1990s.

01

What was the most common movie duration during the 1990s? Save as an integer variable named `duration` (using 1990 as the starting year of the decade).

02

A movie is classified as short if its duration is less than 90 minutes. Count the number of short action movies released in the 1990s and store this count as an integer variable named `short_movie_count`.



Possible Interpretations:

- **Audience Preferences:** The peak around 100 minutes might reflect the optimal viewing duration preferred by audiences during that time.
- **Production Constraints:** Factors like budget, production schedules, and target audience preferences may have influenced the distribution of movie durations.
- **Genre Variations:** Different genres of movies may have varying average durations. For example, action movies might tend to be longer than romantic comedies.

WHAT WAS THE MOST COMMON MOVIE DURATION DURING THE 1990S?

This histogram visualizes the distribution of movie durations in the 1990s. The x-axis represents the duration in minutes, while the y-axis shows the number of movies within each duration range (bin).

Key Observations:

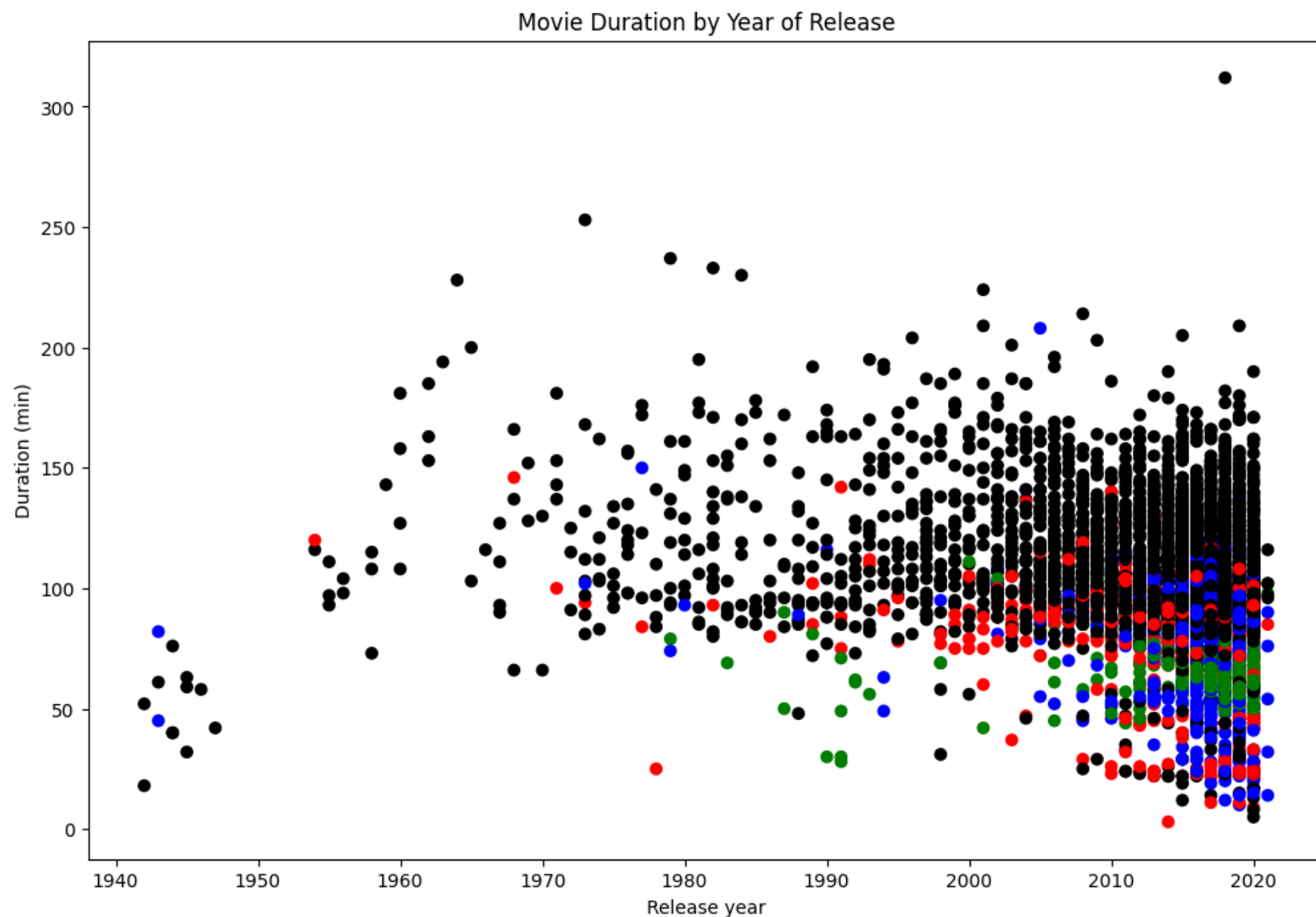
1. **Peak Duration:** The histogram shows a peak around the 100-minute mark, suggesting that a significant number of movies released in the 1990s had a duration of approximately 100 minutes.
2. **Right-Skewed Distribution:** The distribution is right-skewed, meaning there are more movies with shorter durations compared to longer ones. This is evident from the longer tail on the right side of the histogram.
3. **Range of Durations:** The histogram covers a wide range of durations, from short films to longer movies, indicating diversity in movie formats and genres.

ANALYZE SHORT ACTION MOVIES

Count the number of short action movies (less than 90 minutes) released in the 1990s.

The result is "**8**" indicates that there are **8 action movies with a duration of less than 90 minutes** that were released on Netflix during the **1990s**.

This means that out of all the action movies released on Netflix in the 1990s, only 8 of them were considered "short" based on the defined criteria of having a duration less than 90 minutes.



The scatter plot visualizes the relationship between the release year of a movie and its duration (in minutes). Different colors are used to represent different genres:

- **Black:** Other genres
- **Red:** Children's movies
- **Blue:** Documentaries
- **Green:** Stand-up Comedy

ANALYZING THE SCATTER PLOT OF MOVIE DURATION BY RELEASE YEAR

1. **Overall Trend:** There doesn't seem to be a clear, consistent trend of increasing or decreasing movie durations over time. The data points are scattered across the plot, indicating variability in movie lengths across different years.
2. **Genre-Specific Trends:**
 - **Children's Movies:** These tend to be shorter, often under 90 minutes, and are more concentrated in recent years.
 - **Documentaries:** Documentaries also tend to be shorter, with some exceptions.
 - **Stand-Up Comedy:** These are generally shorter, usually under 120 minutes.
 - **Other Genres:** This category includes a wide range of movie types, from short films to long epics, leading to a diverse distribution of durations.
3. **Outliers:** There are a few outliers, especially in the later years, with significantly longer durations. These could be long movies, miniseries, or other special formats.



INSIGHTS



01.

By analyzing movie data from the 1990s, we can gain insights into movie duration patterns during that era. This analysis can be extended to cover the entire dataset to explore movie duration trends across all release years.

02.

While the code doesn't confirm a shortening trend in movie durations, it allows for further exploration to identify such trends.

