# EXPLORATORY DATA ANALYSIS FOR MICROSOFT'S NEW MOVIE STUDIO

Insights & Recommendation For Successful Movie Production

PRESENTED BY :
ISRAEL WASIKE KAHAYI

on 28th July 2024

# BUSINESS PROBLEM

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. To ensure the success of their new venture, they need to understand the current market trends and determine what types of films are performing best at the box office.

# OBJECTIVE

Provide insights on what types of films are doing well at the box office.
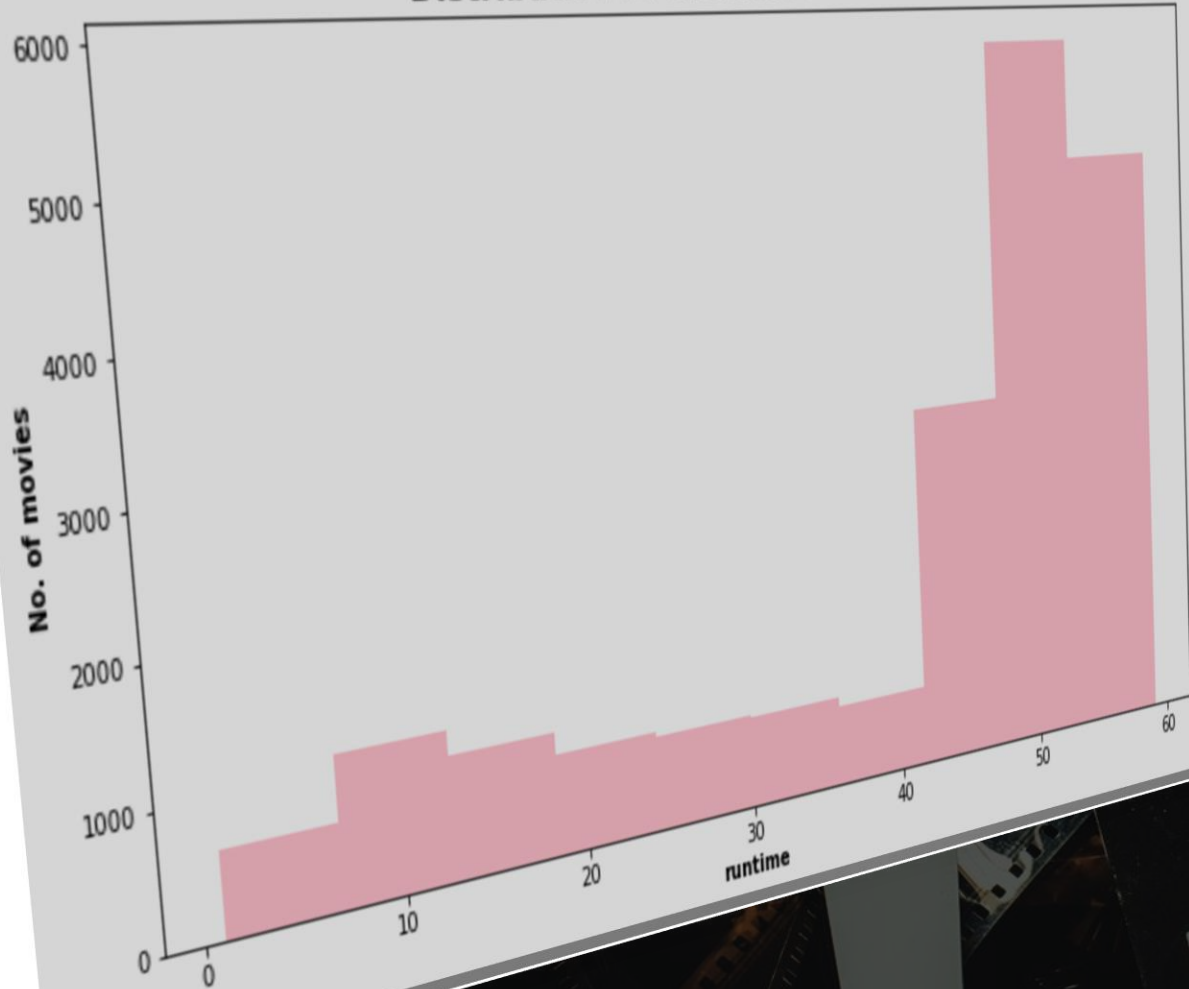
# KEY QUESTIONS :

- What genres are the most successful?

- What runtime is typical for successful movies?

- Are there any trends in movie releases over the years?

# DATA ANALYSIS PROCESS

1. **Data cleaning** – Loading data, inspecting for missing values and clean the dataframe by handling missing values.

2. **Data analysis** – Exploring the movie genres & determining the most common genres with the highest number of movies.

3. **Runtime analysis-** Analyzing the distribution of movie runtimes.

4. **Trends overtime-** Examining trends in the number of movies released per year.

Distribution of short movies runtime

The chart displays a spike, between the 50-to-60-minute mark implying that many short films fall within this time frame. Conversely there is a scarcity of movies lasting than 30 minutes hinting that shorter films usually exceed this duration.
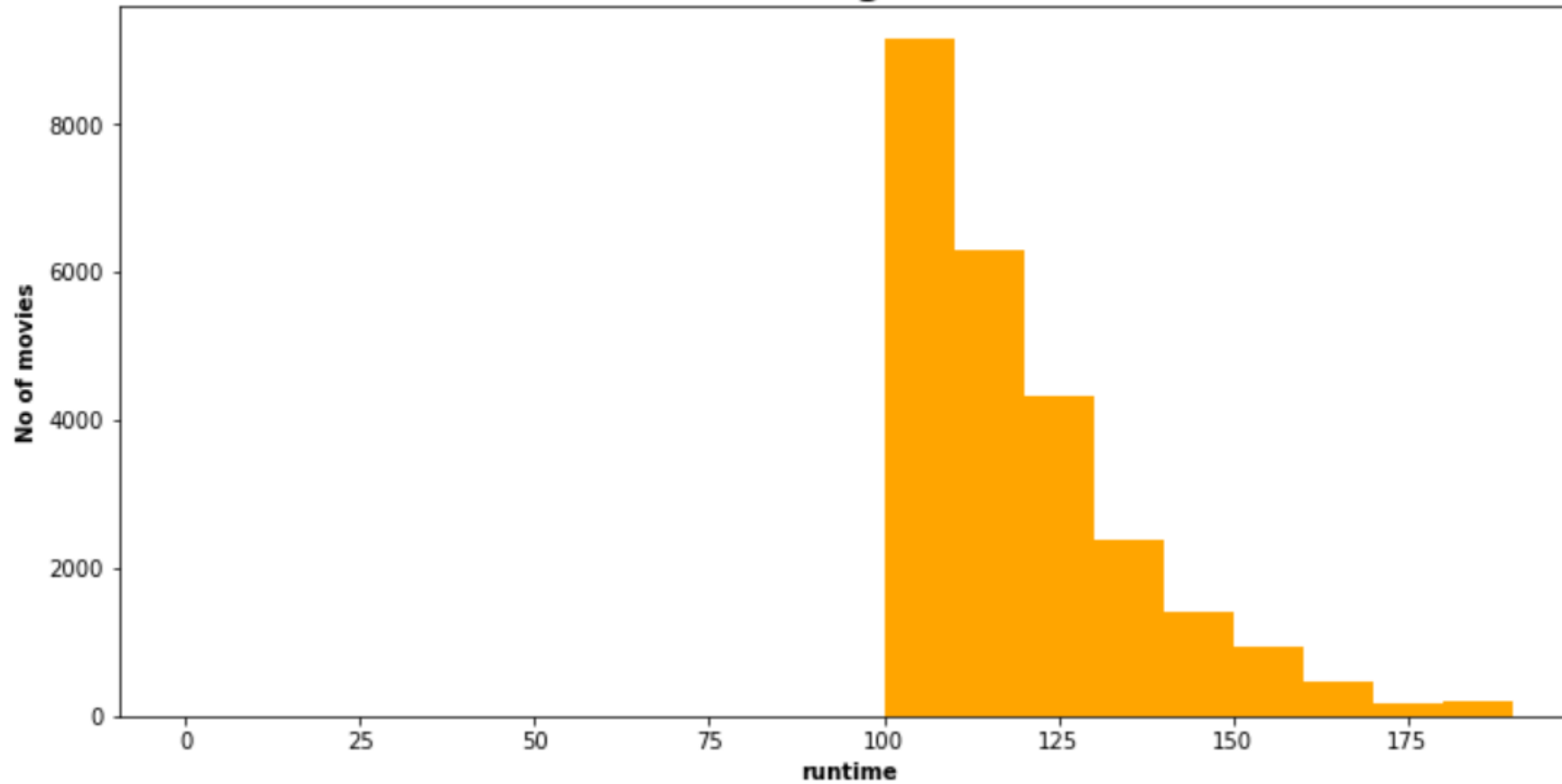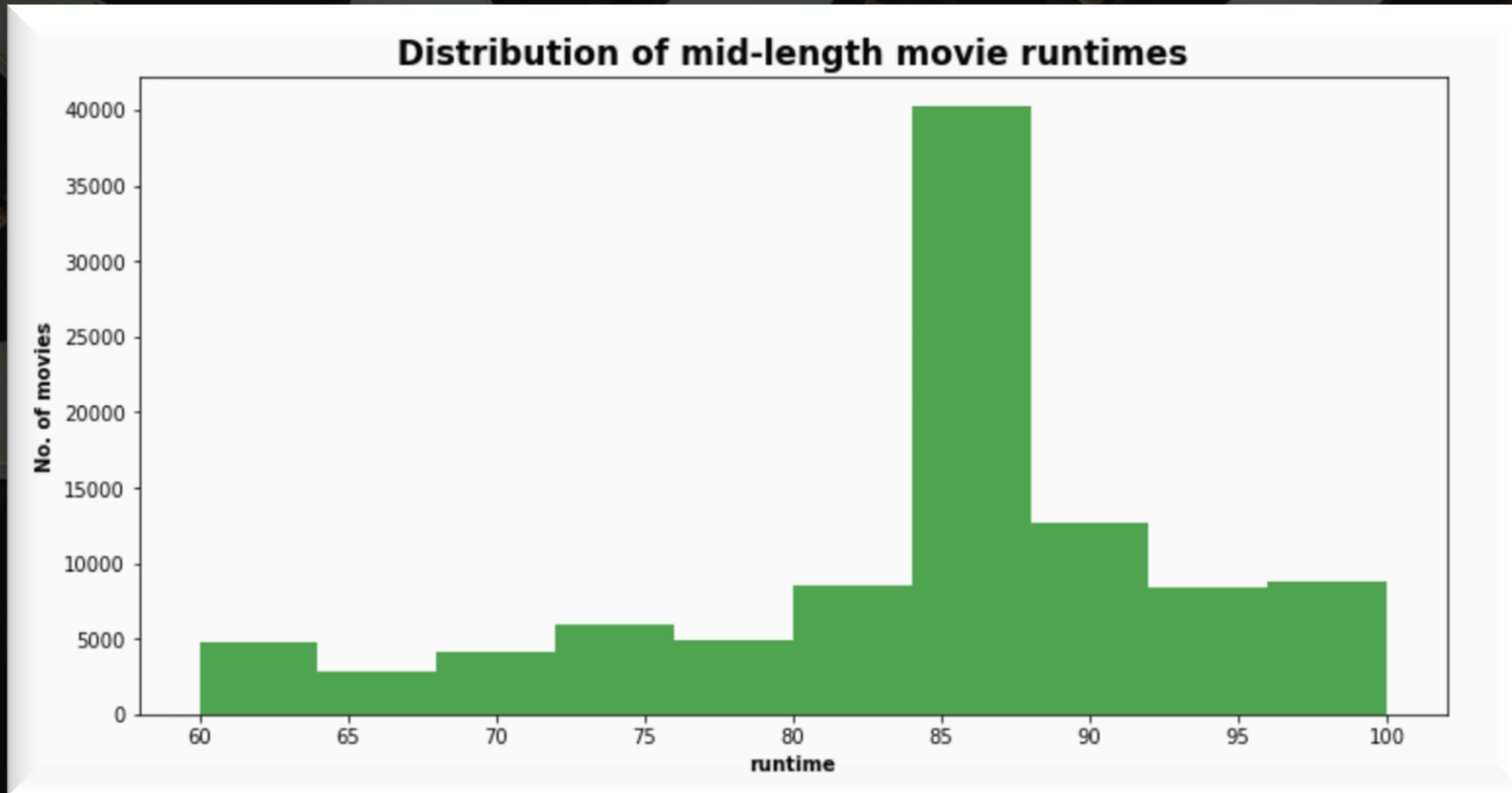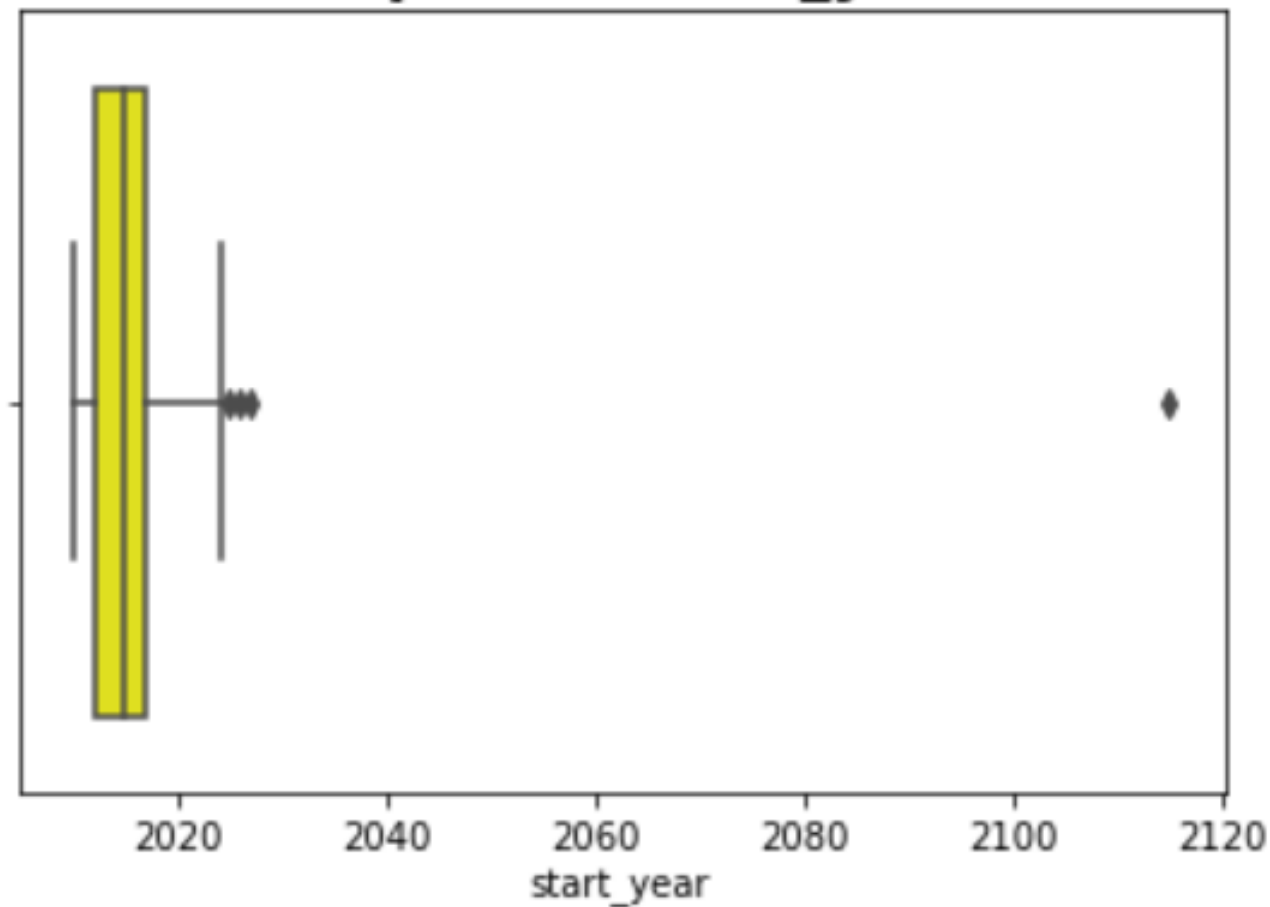
Distribution of long movies runtime

Movies tend to cluster around the 90 to 100-minute mark, which seems to be the sweet spot for longer films.After that, you see fewer and fewer movies as the runtime stretches on. Looks like filmmakers aren't too keen on keeping audiences in their seats for ages!

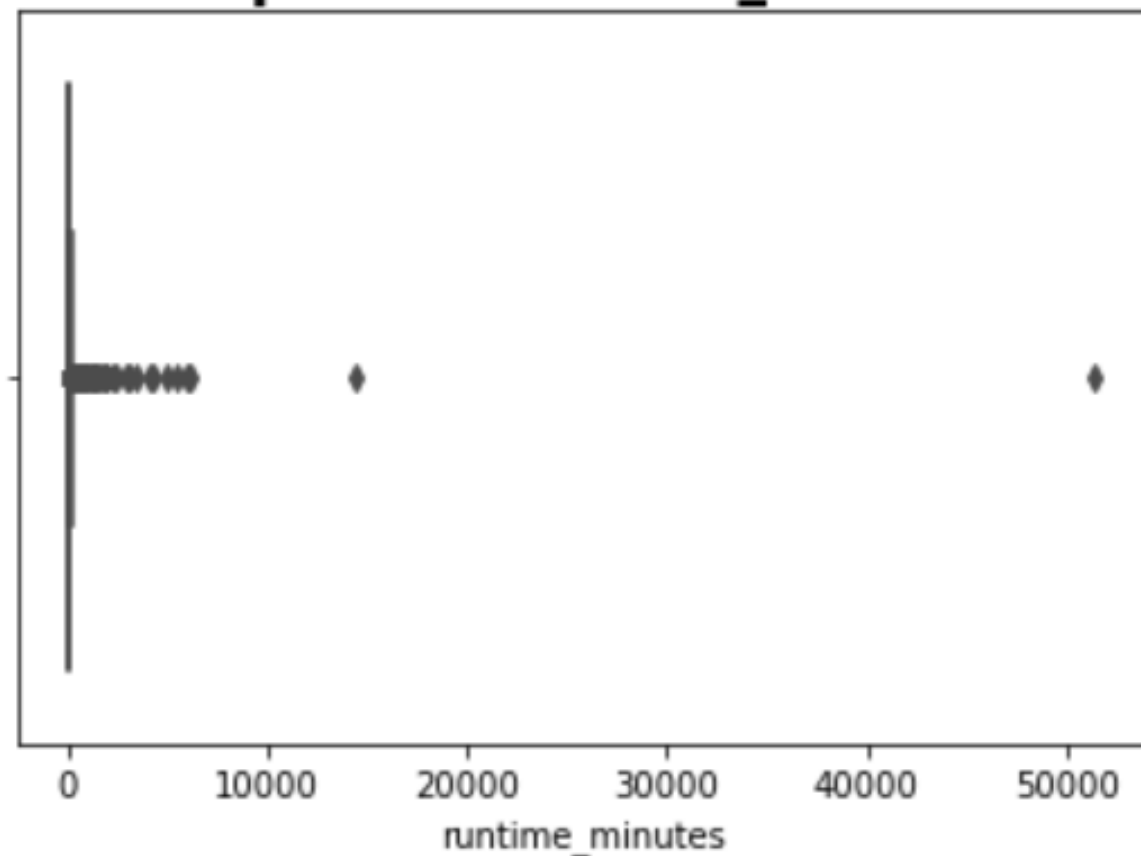**Distribution of mid-length movie runtimes**

The histogram indicates that movies of length (with durations falling between 60 and 100 minutes) tend to be 85 minutes long most often. This pattern implies that films of this duration are popular which could be due, to industry standards or audience preference.
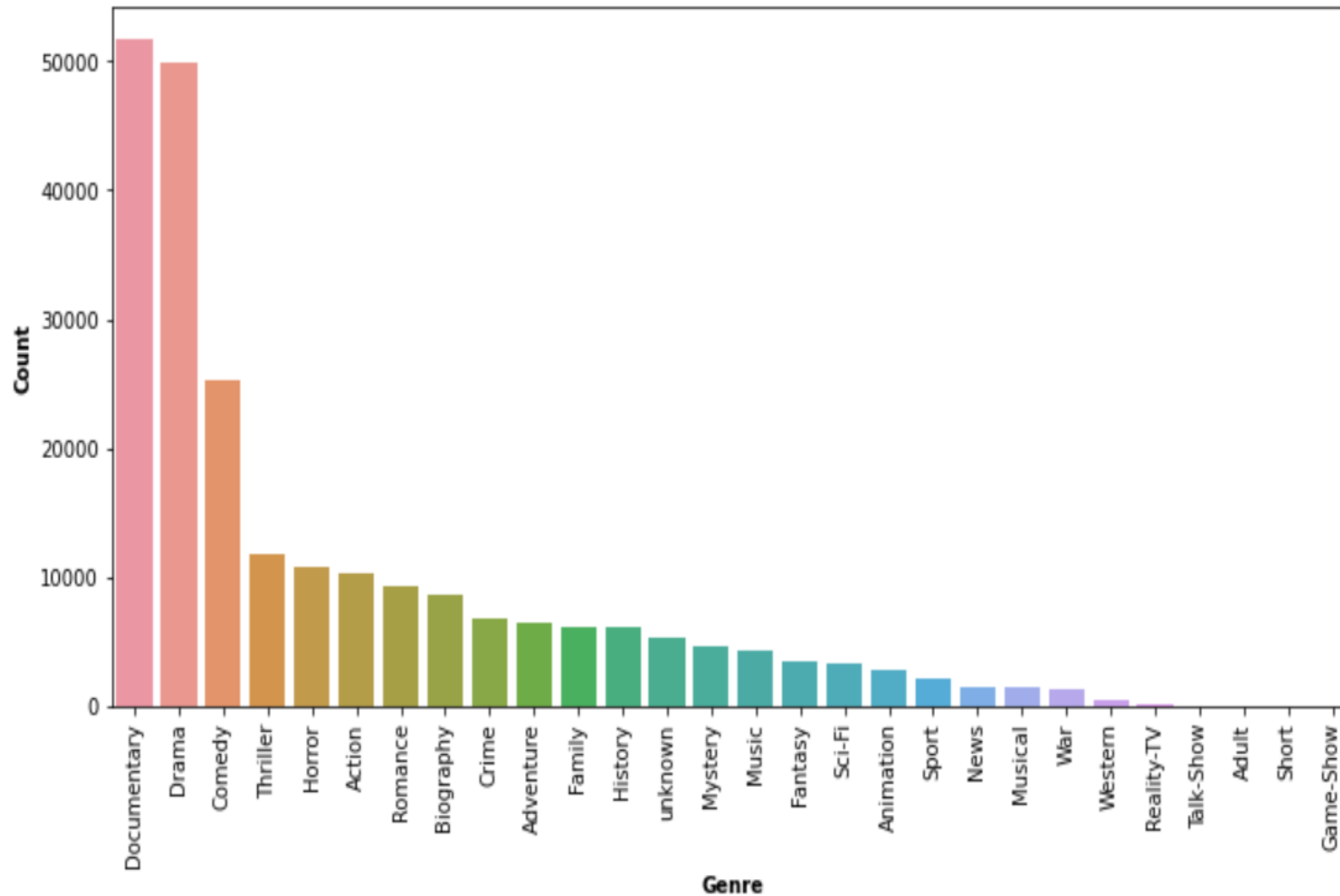
boxplot of start_year

The yellow box represents the interquartile range (IQR) where the middle 50% of the data lies. There is a clear outlier beyond the year 2120, indicating a start year significantly different from the rest of the data.
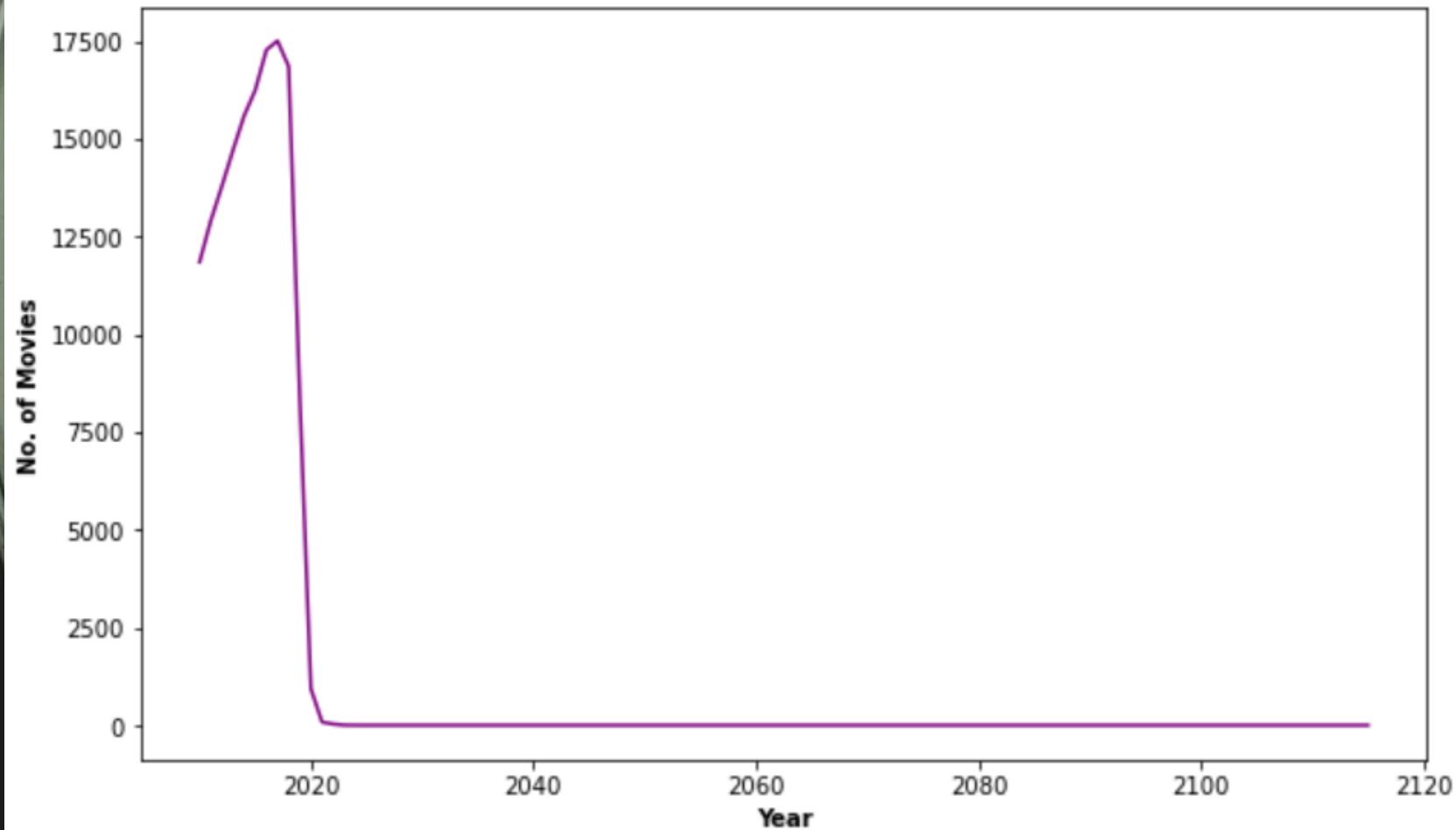
boxplot of runtime_minutes

runtime_minutes

The horizontal lines extending from the box represent the range within 1.5 times the IQR from the quartiles. There are several outliers beyond the typical range, with extreme values appearing around 20,000 and 50,000 minutes

**Movie Genre Distribution**
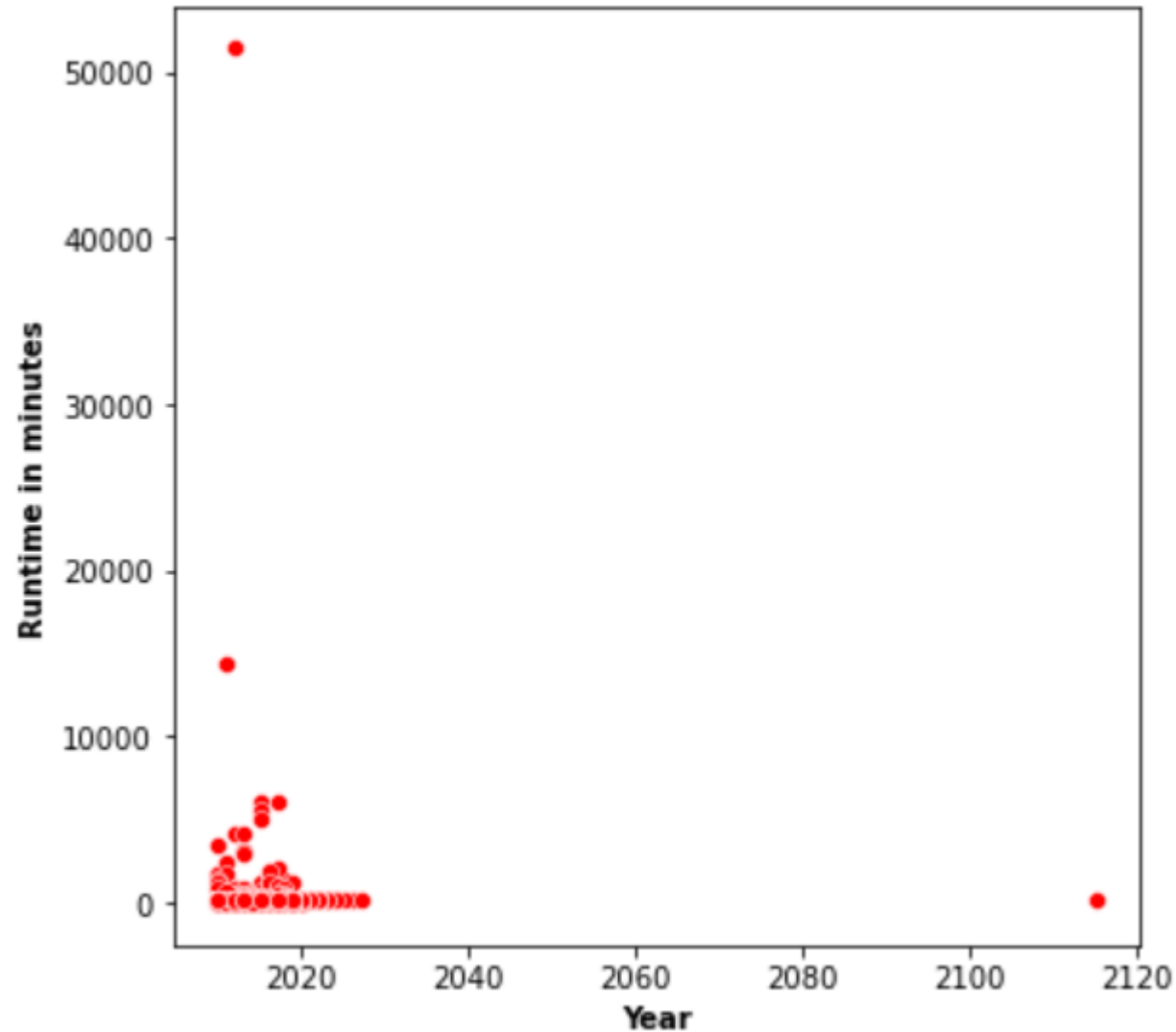
Documentaries take the lead with, over 50,000 movies making it the popular genre. Following behind is Drama with a number of films in its category. Comedy also holds a presence although not as abundant as drama. The chart showcases a range of genres like thriller, action, romance, biography, adventure, history, mystery, fantasy, animation and more each, with decreasing numbers of movies.

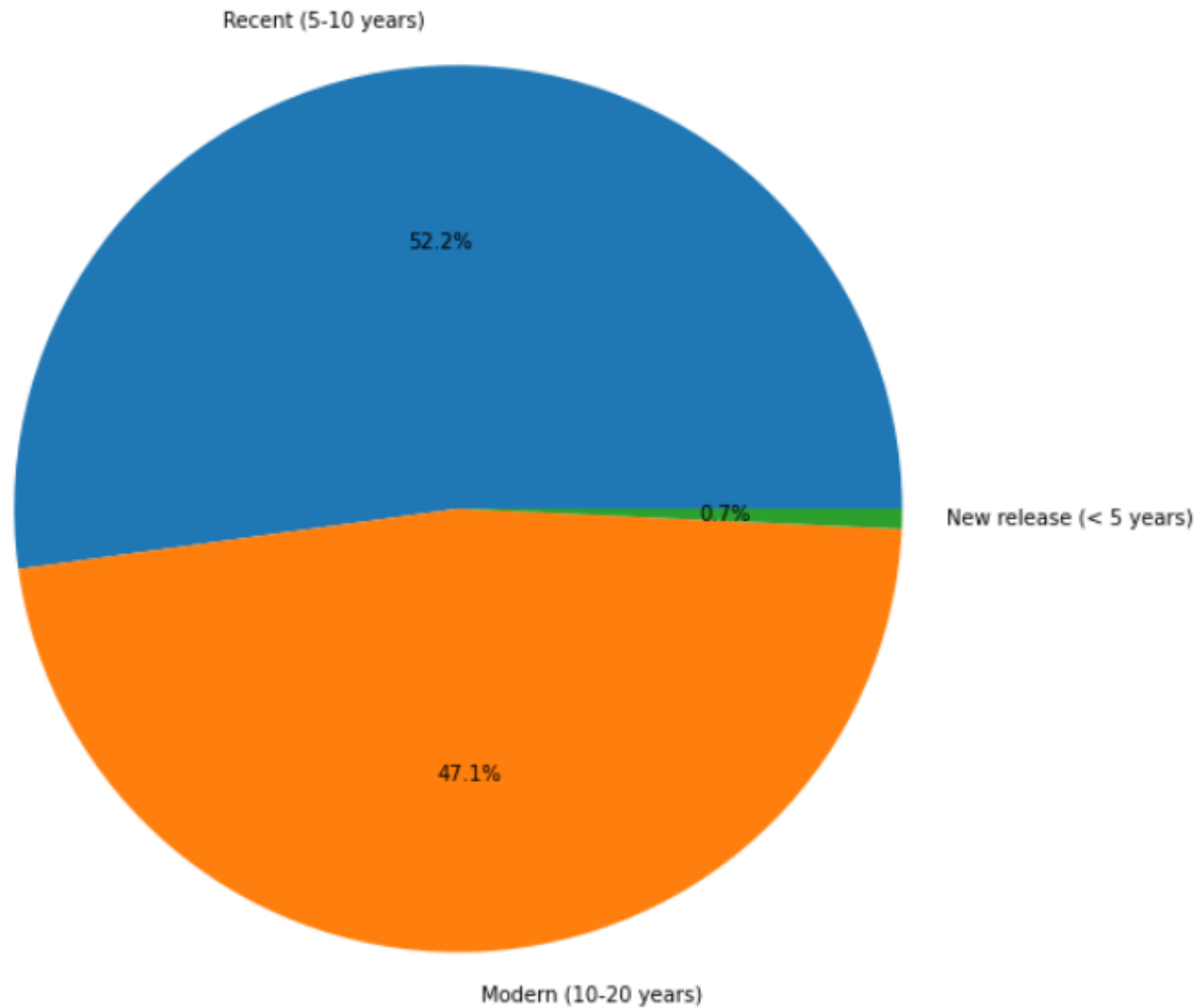**Number of Movies Released Over Time**

The number of films released in 2020 increased significantly. However, following this peak, there was a decline, indicating a decrease in movie releases in the following years

# Year & Runtime

The scatter plot does a job of showing how movie lengths vary over the years especially pointing out the presence of outliers movies that're much longer, than usual.

**Pie chart showing the distribution of movies by Age Category**

Recent (5-10 years)

52.2%

0.7%

New release (< 5 years)

47.1%

Modern (10-20 years)

The Pie chart shows that most films are categorized as "Recent" or "Modern," with a number falling under the label of "New releases"

# FINDINGS

- Most short movies have runtimes between 40 and 60 minutes, with the highest concentration in the 50–55-minute range. Out of 19,186 short films under 60 minutes, this duration bracket dominates viewership.

- Of the 25,757 long movies with runtimes of 100 minutes or more, the majority fall within the 100–125-minute range. This duration bracket represents the most common length for feature films in this category.

- The dataset contains 100,201 mid-length movies, with runtimes between 60 and 100 minutes. Within this category, the most common duration falls in the 85–90-minute range.

- For instance, majority of the movies have a runtime within 90- & 120-minutes range.

- Documentaries take the lead with, over 50,000 movies making it the popular genre. Following behind is Drama with a number of films in its category. Comedy also holds a presence although not as abundant as drama. The chart showcases a range of genres like thriller, action, romance, biography, adventure, history, mystery, fantasy, animation and more each, with decreasing numbers of movies.

# INSIGHTS & RECOMMENDATION

After analyzing the results here are a takeaways and practical recommendations.

- it's evident that audiences have a preference, for documentaries, dramas and comedies. These movie genres are gaining increasing popularity compared to others. If Microsoft is contemplating creating a film it may want to explore one of these categories. It might be wise to align with trends and offer viewers more of what they enjoy.

- In the past 10 years, there has been a rise, in the number of movies being released. It's important to expand production capabilities to keep up with the growth, in the industry.

- For instance the majority of films typically fall within the 90 to 120 minute range. It would be beneficial for Microsoft to focus on creating movies that fit within this timeframe to stay in line, with industry standards.