

# CAMBRIDGE SPARK - PRACTICAL HACKATHON

## DFE - L4 DATA ANALYST BOOTCAMP - OCT 24 - COHORT 5

**COMPARATIVE ANALYSIS OF NETFLIX CONTENT ACROSS THE UNITED STATES, THE UNITED KINGDOM, AND TURKEY.**  
BY ERDI OZYURT



# THE NETFLIX DATASET (KAGGLE,2019)

- Exploring the Netflix Dataset: Netflix is one of the world's largest streaming platforms, with millions of users streaming movies and TV shows every day. As a result, Netflix has accumulated a massive amount of data that can be used to gain insights into user behavior, preferences, and more. In this project, students will explore a dataset containing information about movies and TV shows available on Netflix.
- This analysis outlines the steps for conducting a comparative analysis of Netflix content across the United States, the United Kingdom, and Turkey. It includes data preparation, exploration, analysis, and a concluding summary of findings and insights.
- Data Source: the Netflix dataset from the EDUKATE “Hackathon Data” notebook(Kaggle,2019).

# TASK

# DATA PREPARATION

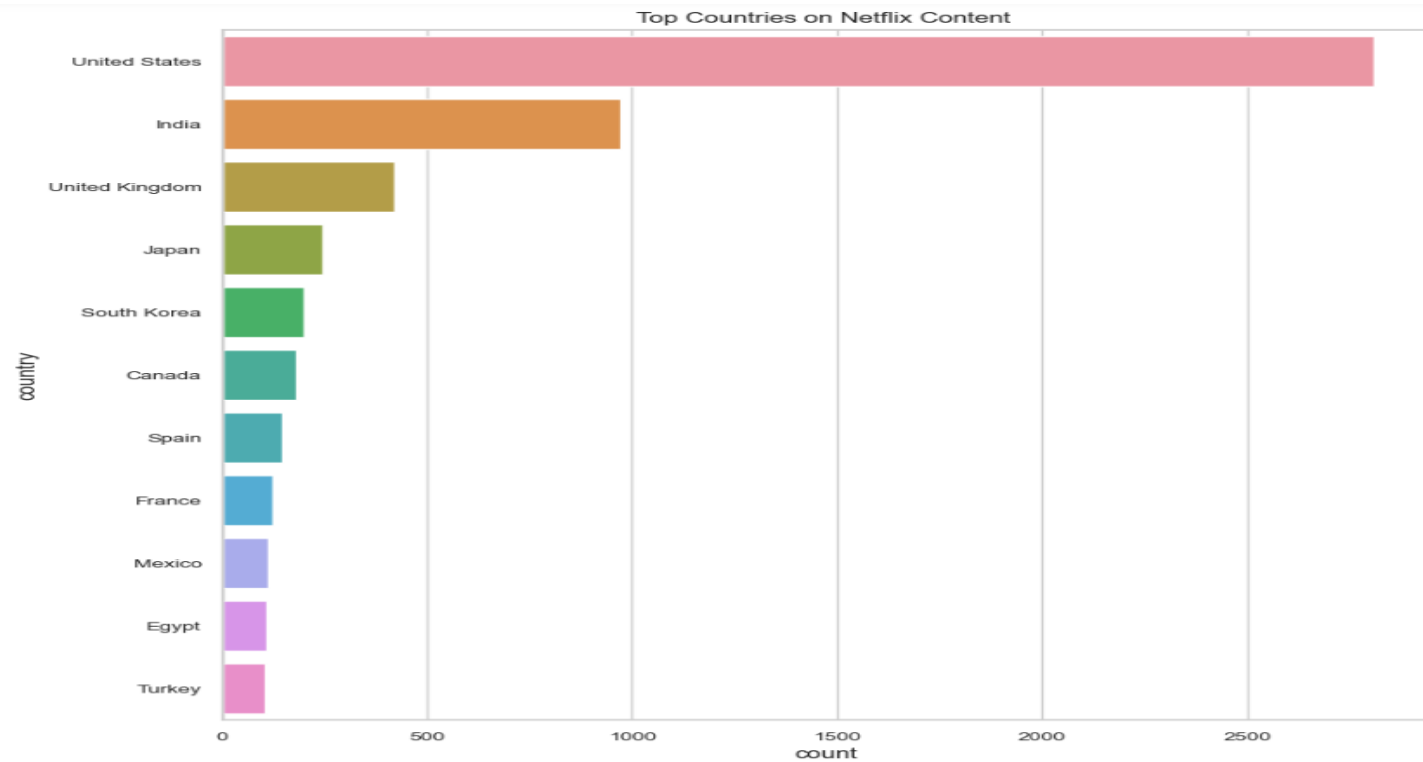
- 1.The 'date\_added' column is in 'DateTime' format.
- 2. the 'release\_year' column is in the correct numeric format.
- 3. Extract month and year from the 'date\_added' column for time-based analysis.
- 4.For numeric columns, filled with missing values with the mean.
- 5. For categorical columns 'rating', 'duration' and 'country', filled missing values with the mode or a placeholder like "Unknown", "Not Rated", "Director\_info\_missing" and "Cast\_info\_missing".
- 6.Standardized capitalization in categorical columns 'type' and 'rating' for consistency.
- 7. Checked for and remove any duplicate entries and outliers.

# APPROACH:

1. **Data Preparation:**
  - 1.1 Load the Netflix dataset and clean the dataset by handling missing values and ensuring data consistency.
  - 1.2 Filter the dataset to include only TV shows and movies.
  - 1.3 Subset the dataset to include data relevant to the United States, the United Kingdom, and Turkey.
2. **Data Exploration:**
  - 2.1 Top Countries on Netflix Content.
  - 2.2 Perform a cross-tabulation to analyze the distribution of content types across the three countries.
  - 2.3 Trend of Netflix content additions over the years.
  - 2.4 Investigate the distribution of content ratings for each country.
  - 2.5 Examine the most common genres in each country's content.
3. **Data Analysis:**
  4. 3.1.Counting the number of movies for each country
  5. 3.2 Compare the total duration of Movies & TV Shows from each country.
  6. 3.3 Determine the most popular genres for TV shows and movies in each country.
  7. 3.4 Investigate the most common directors for movies and TV shows in each country.
  8. 3.5 Investigate the ratings of the most common directors for movies and TV shows in each country
  9. 3.6 Investigate the most common casts for movies and TV shows in each country.
10. **Conclusion**

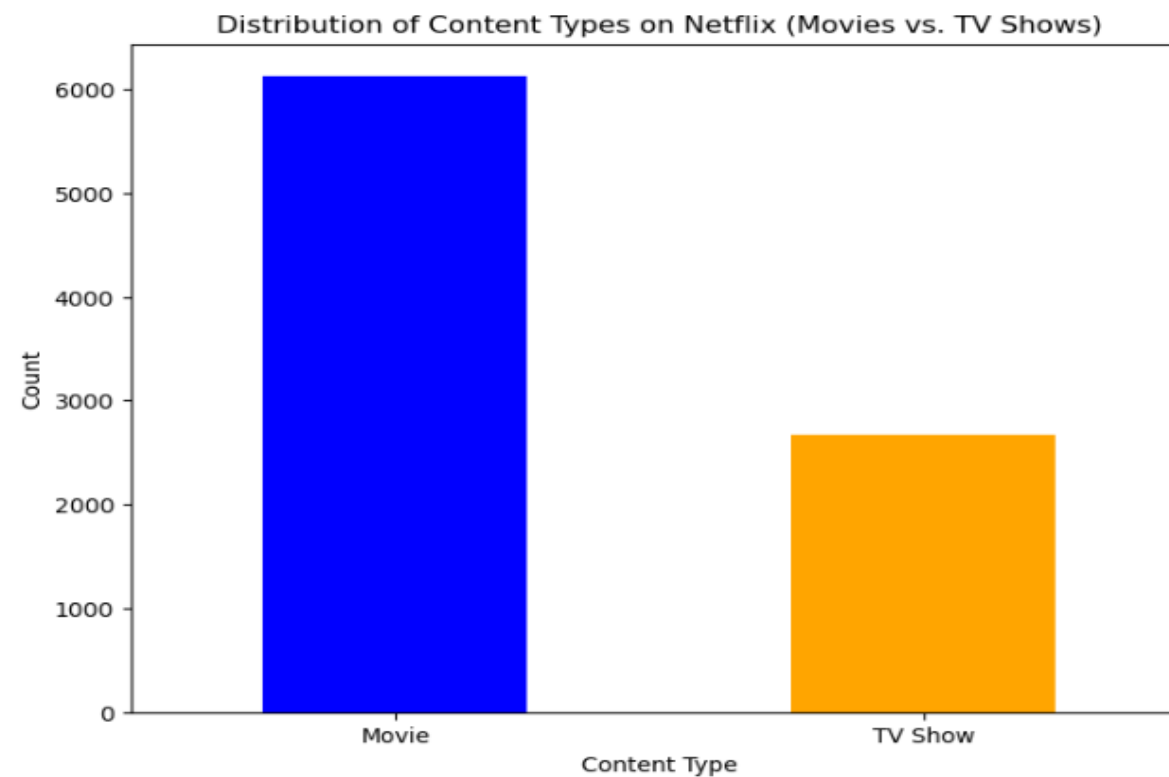
## 2.1 TOP COUNTRIES ON NETFLIX CONTENT.

- US ranked as the 1<sup>st</sup>,
- UK ranked as the 3<sup>rd</sup>,
- Turkey ranked as the 11<sup>th</sup>.



## 2.2 DISTRIBUTION OF CONTENT TYPES (MOVIES VS. TV SHOWS)

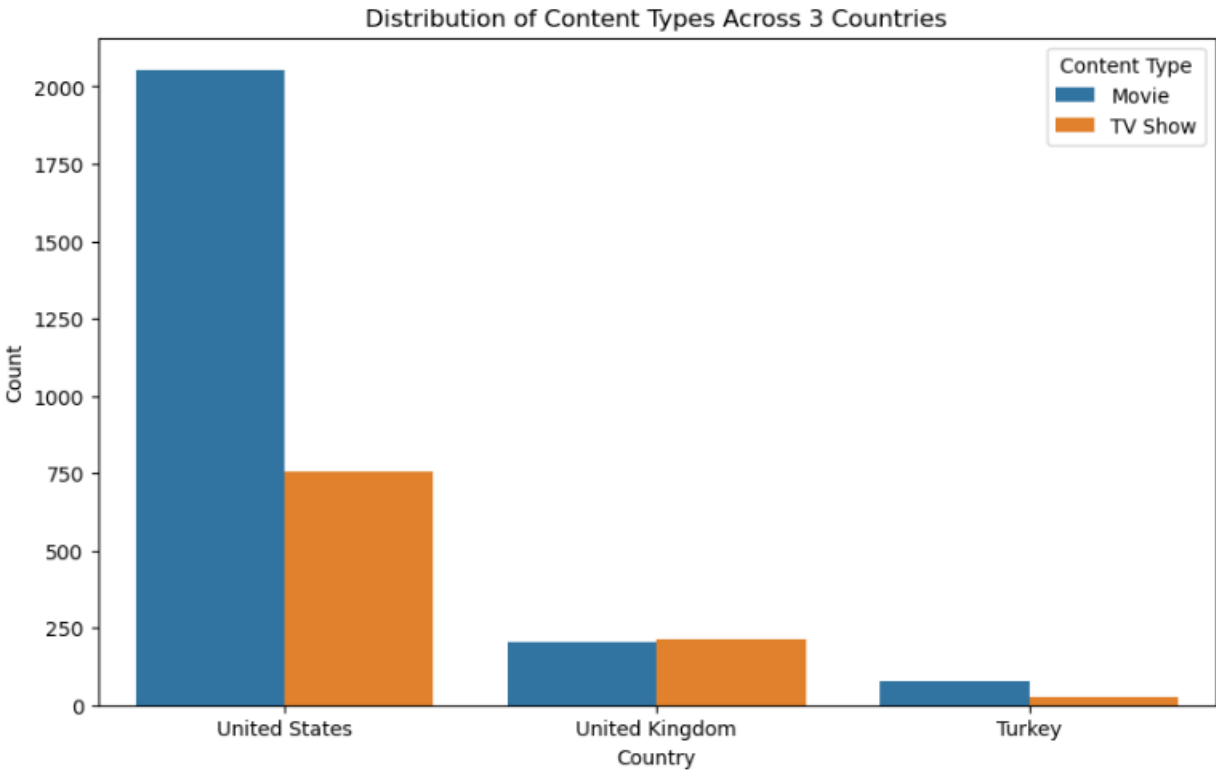
Number of Movies: 6128  
Number of TV Shows: 2666



# 2.2 DISTRIBUTION OF CONTENT TYPES (MOVIES VS.TV SHOWS)

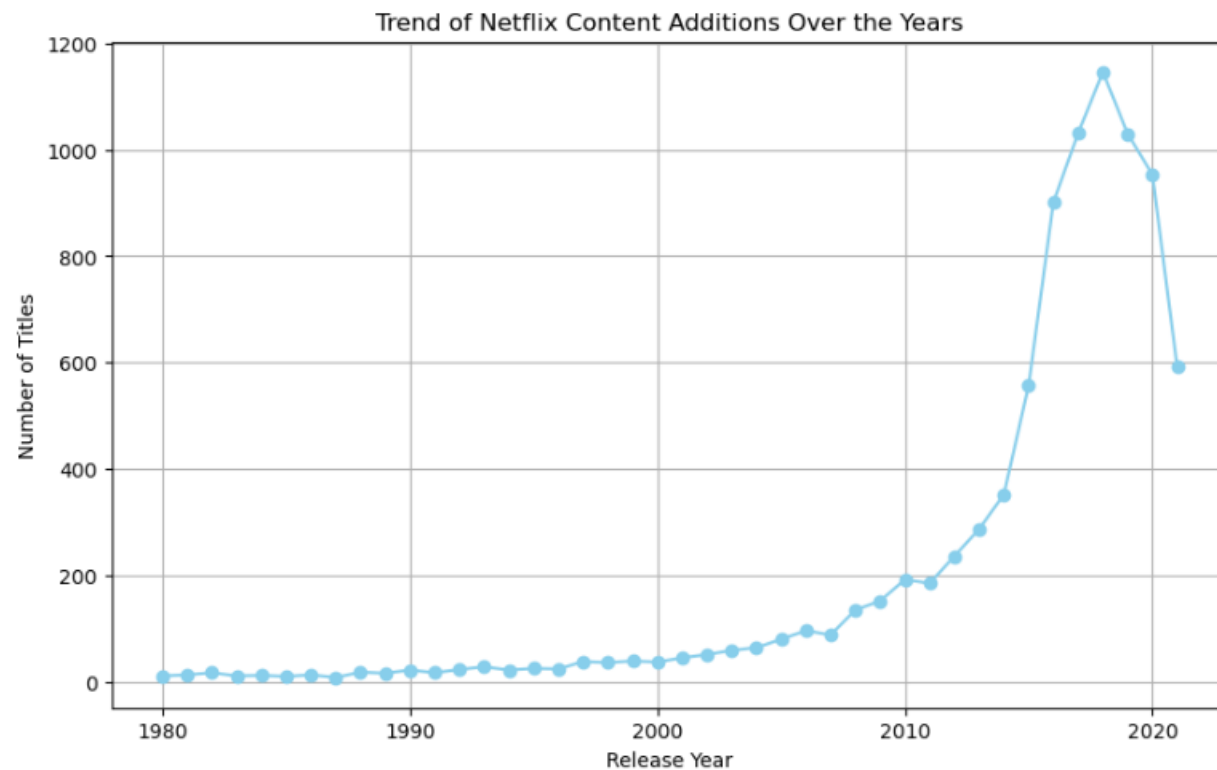
Distribution of Content Types Across Countries:

type	Movie	TV Show
country		
Turkey	76	29
United Kingdom	206	212
United States	2055	754



## 2.3 TREND OF NETFLIX CONTENT ADDITIONS OVER THE YEARS.

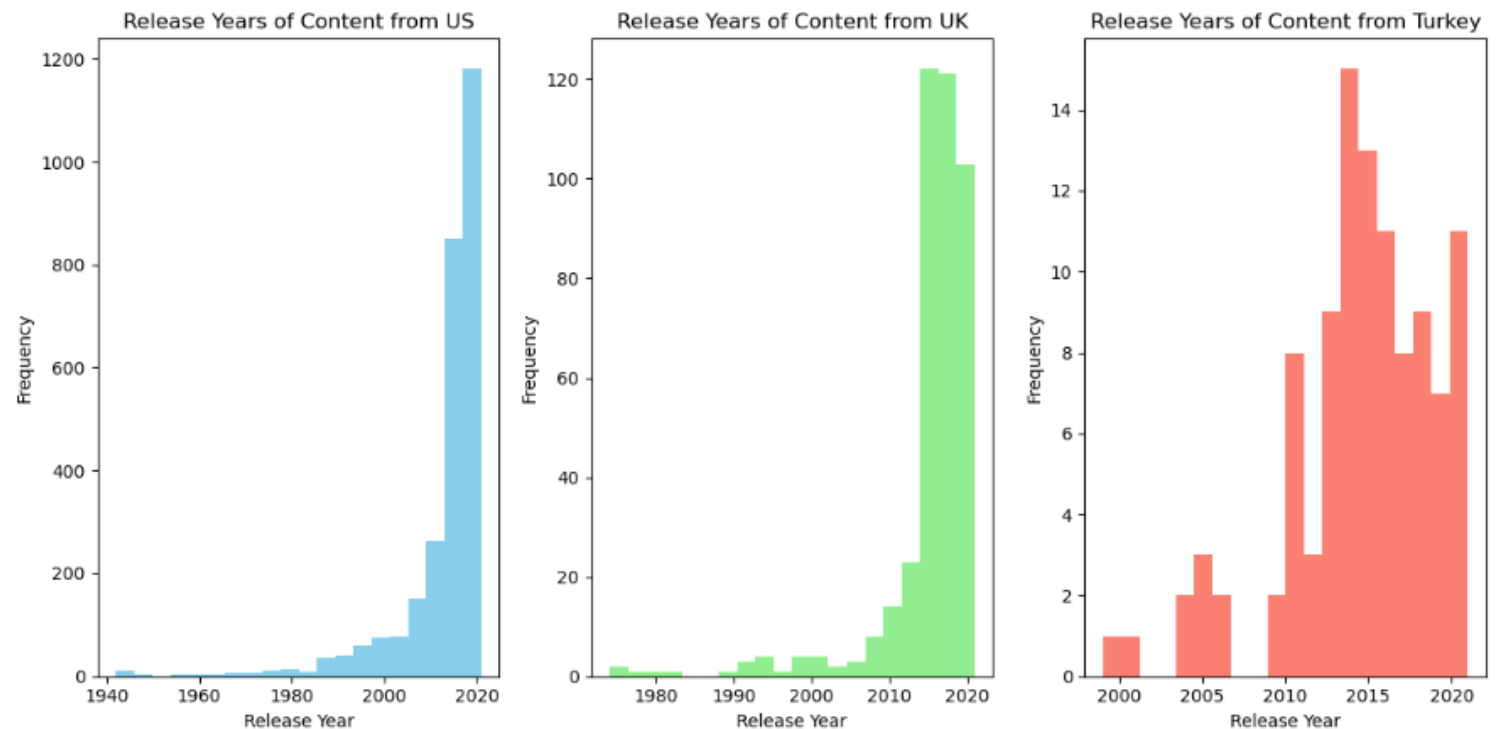
■ Note: Dataset dated by 2019(early 2020).



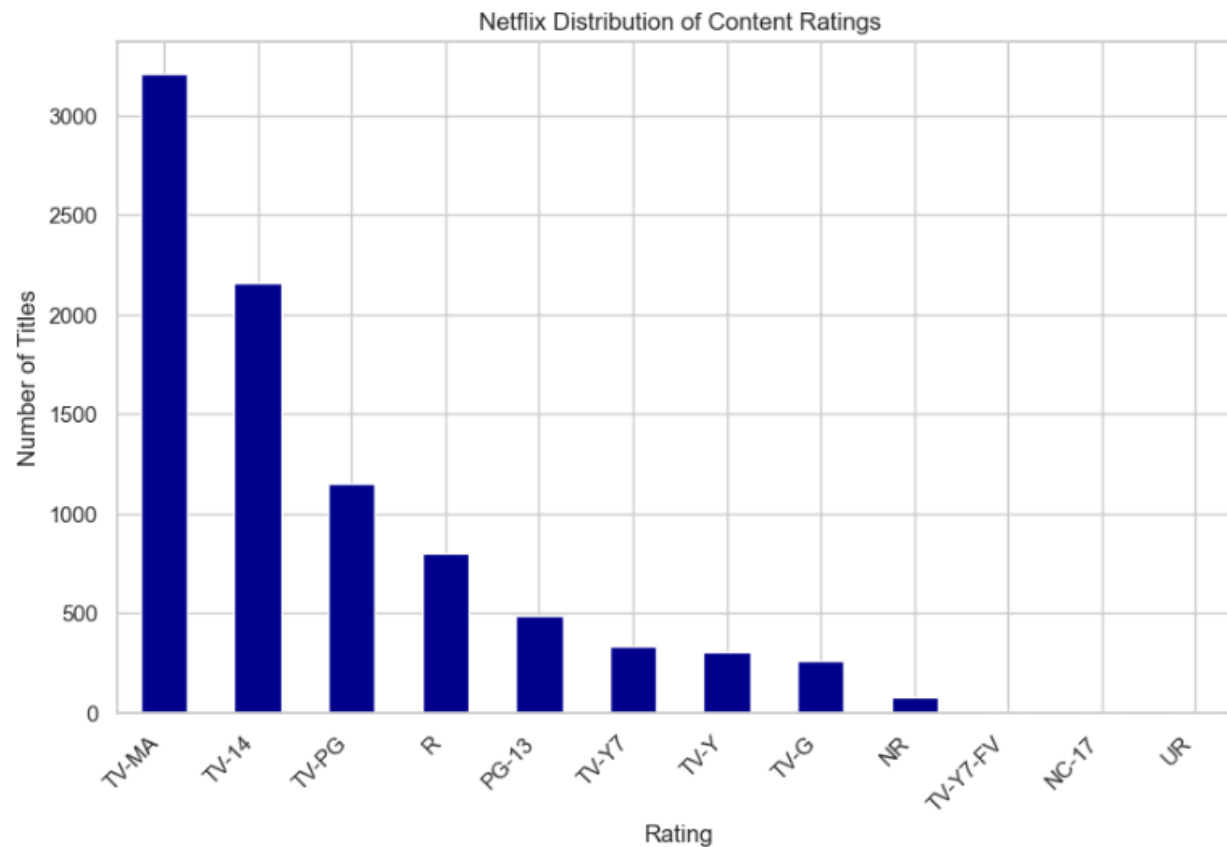


# TREND OF NETFLIX CONTENT ADDITIONS OVER THE YEARS

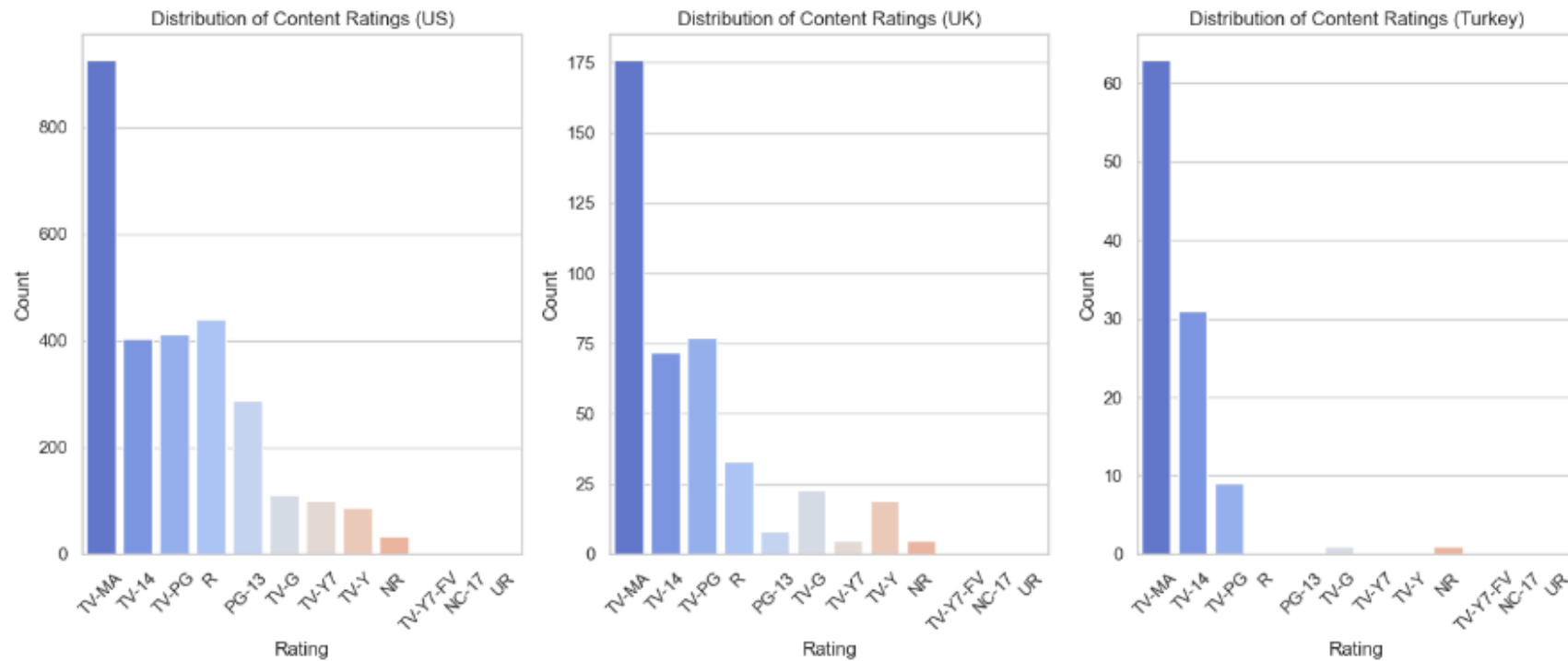
- The US entries are dated earlier than the UK and Turkey.



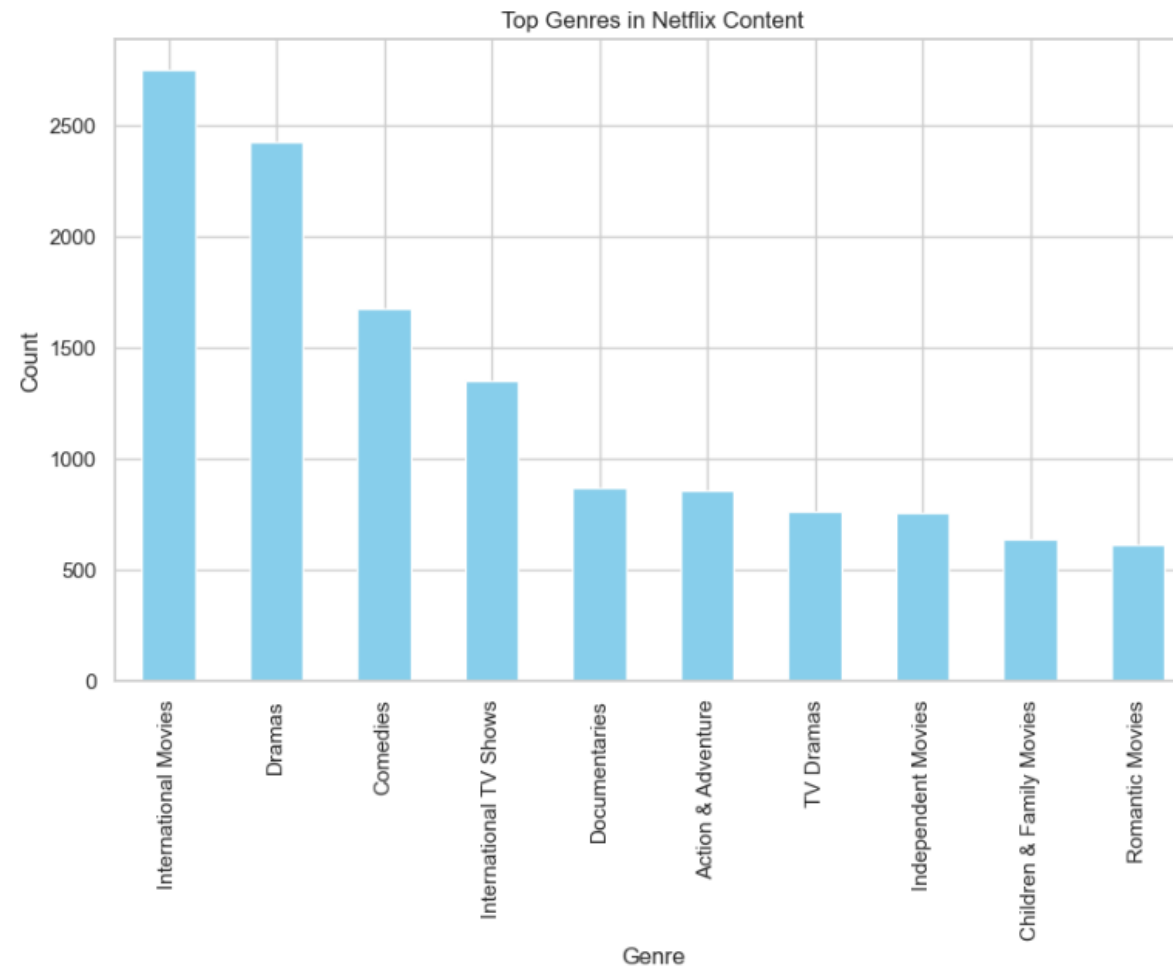
## 2.4 INVESTIGATE THE DISTRIBUTION OF CONTENT RATINGS.



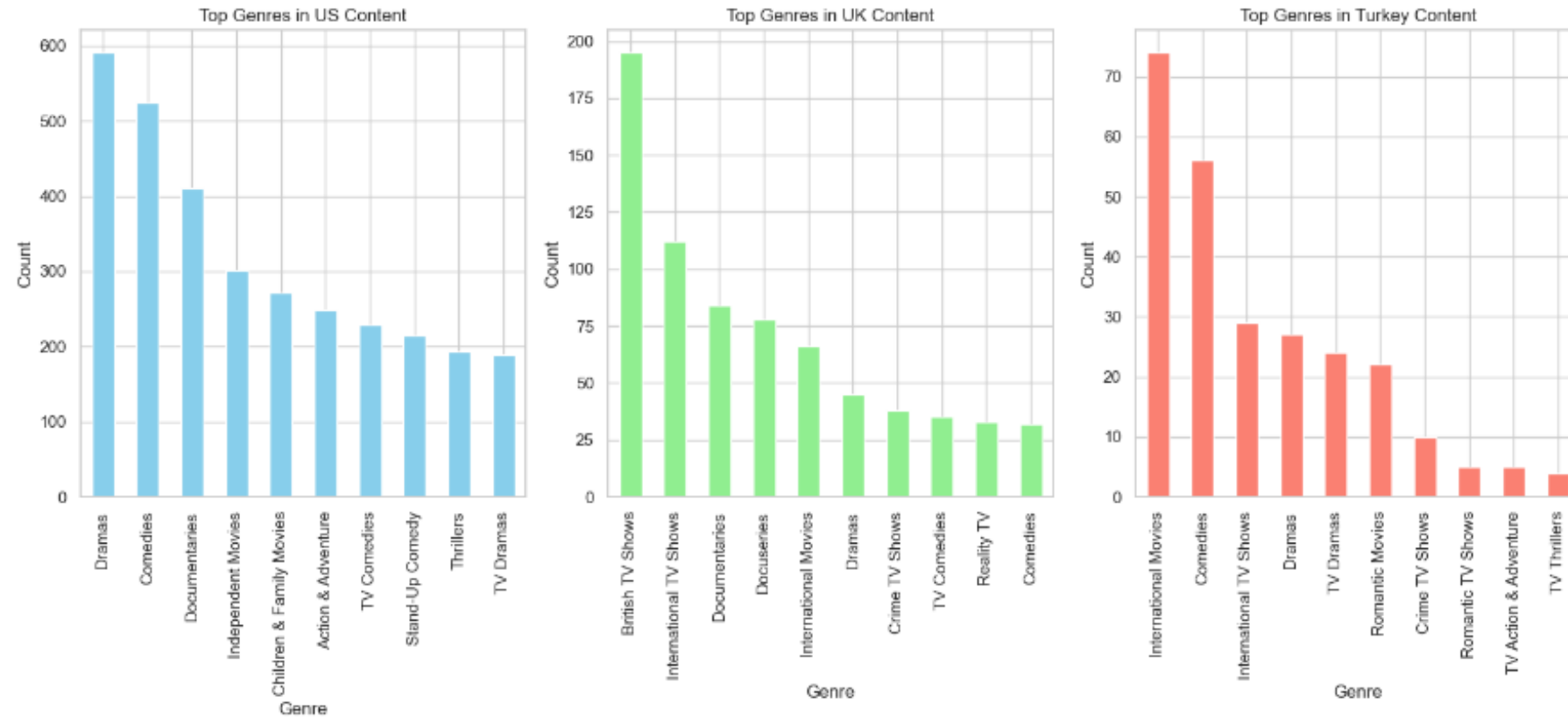
## 2.4 INVESTIGATE THE DISTRIBUTION OF CONTENT RATINGS.



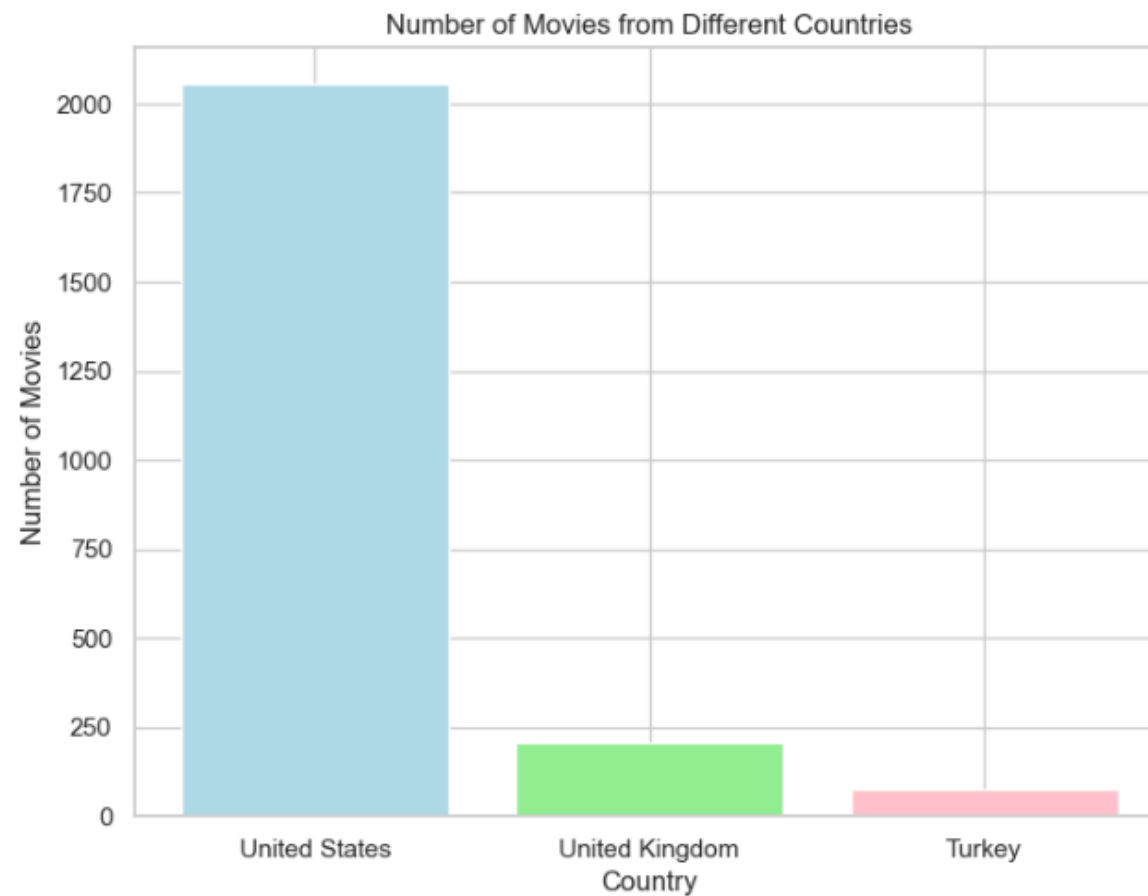
## 2.5 TOP GENRES IN NETFLIX CONTENTS



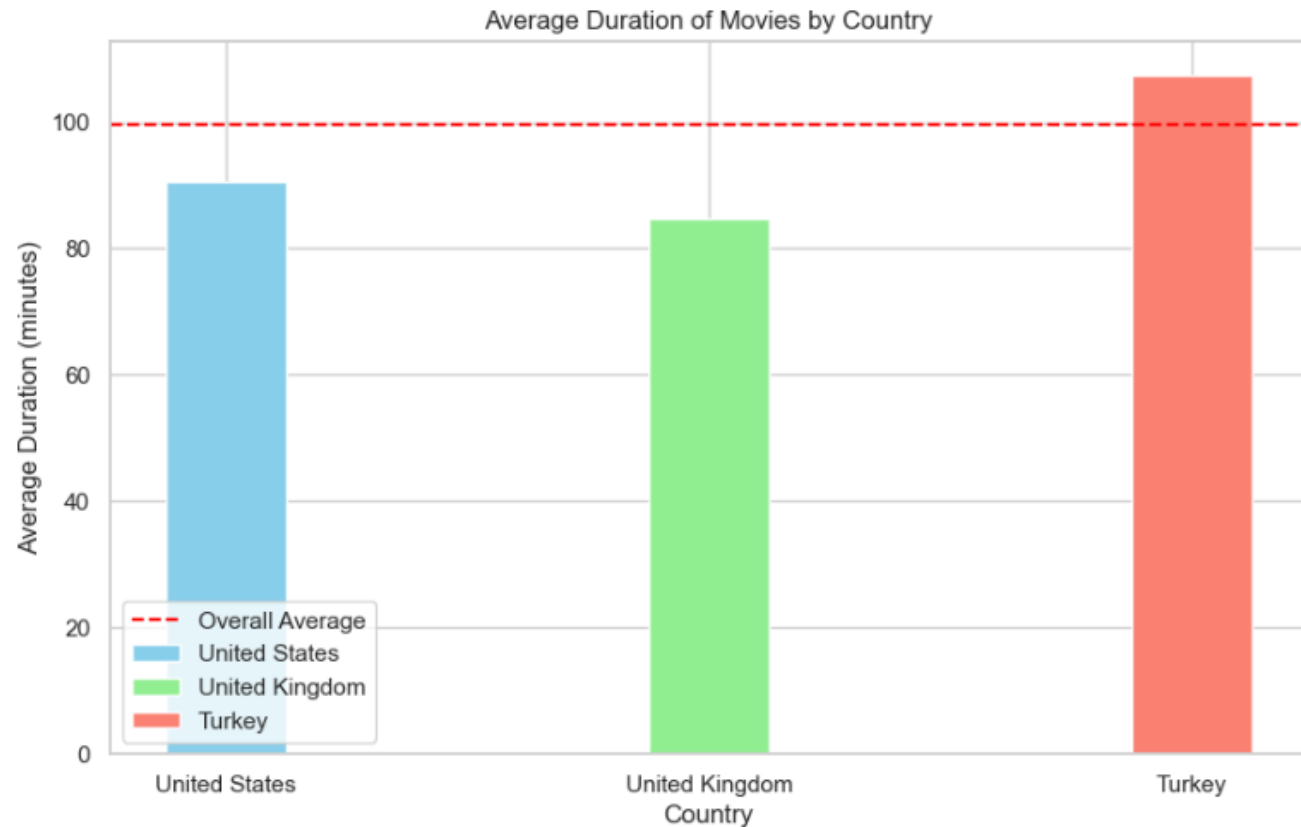
## 2.5 TOP GENRES IN NETFLIX CONTENTS



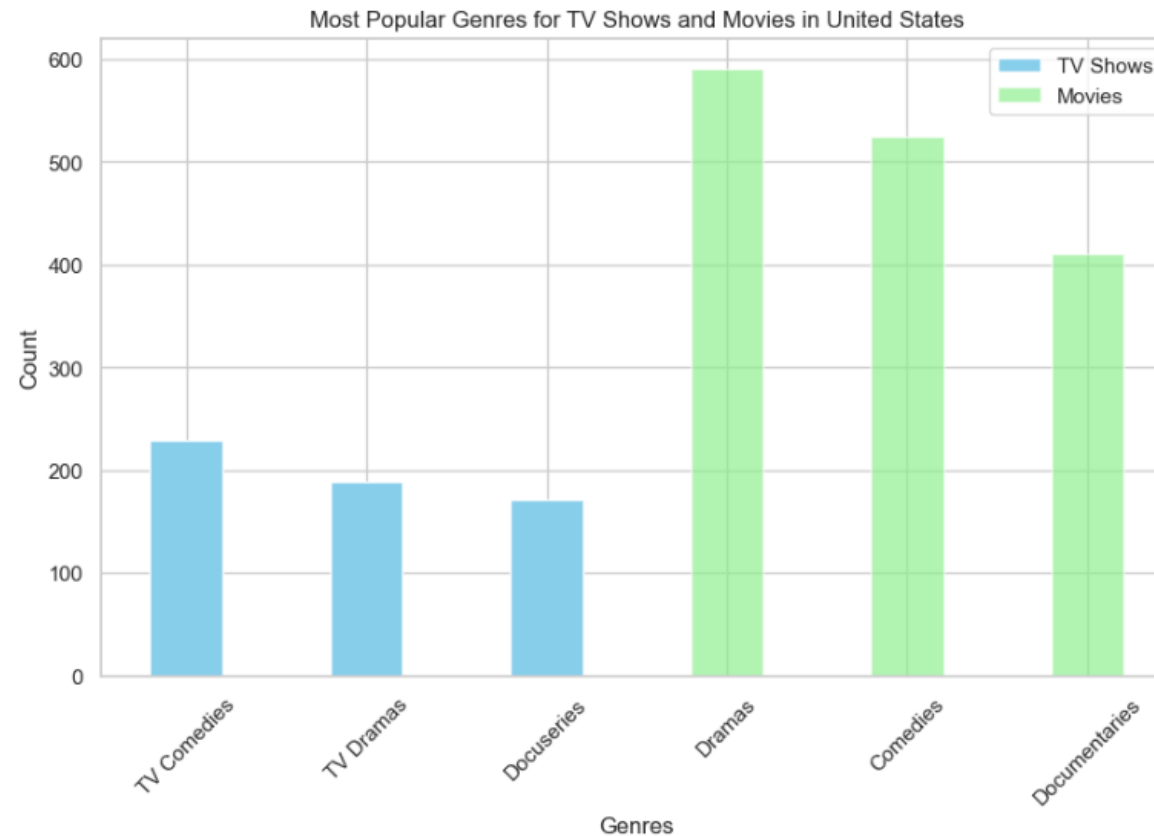
## 3.1 COUNTING THE NUMBER OF THE MOVIES FOR EACH COUNTRY:



## 3.2. CALCULATE THE AVERAGE DURATION OF MOVIES IN EACH COUNTRY.



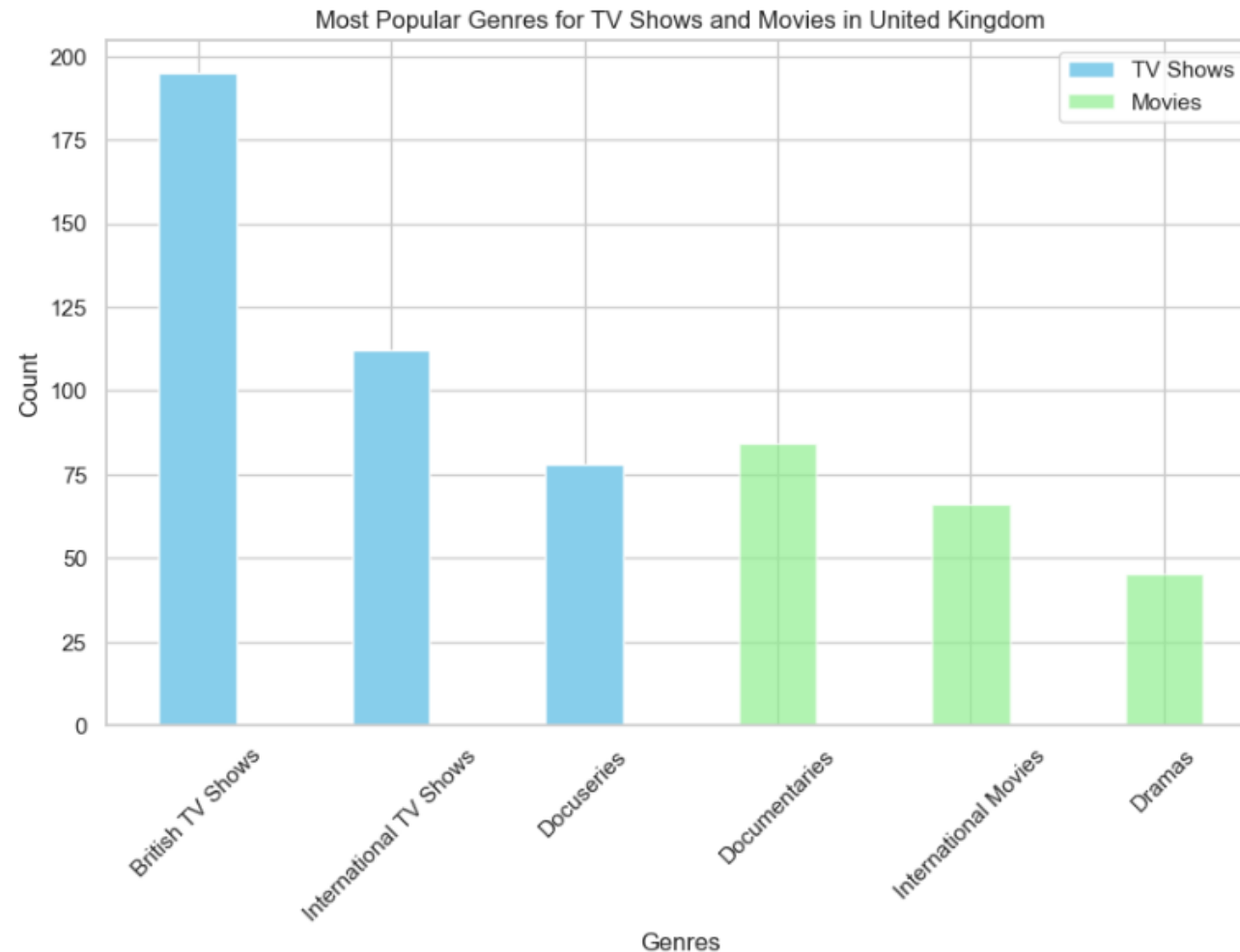
### 3.3 DETERMINE THE MOST POPULAR GENRES FOR TV SHOWS AND MOVIES IN THE US.





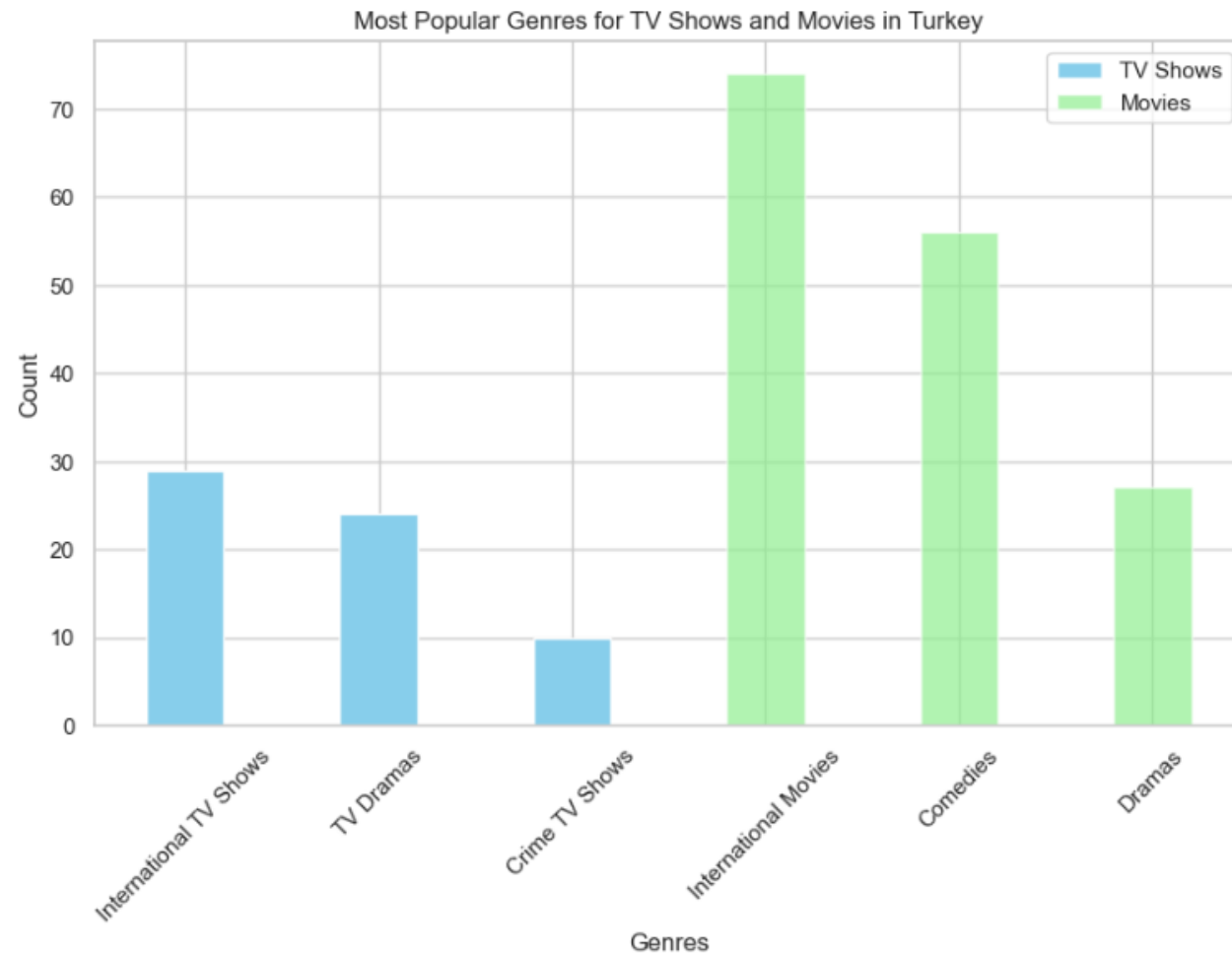


### 3.3 DETERMINE THE MOST POPULAR GENRES FOR TV SHOWS AND MOVIES IN THE UK.





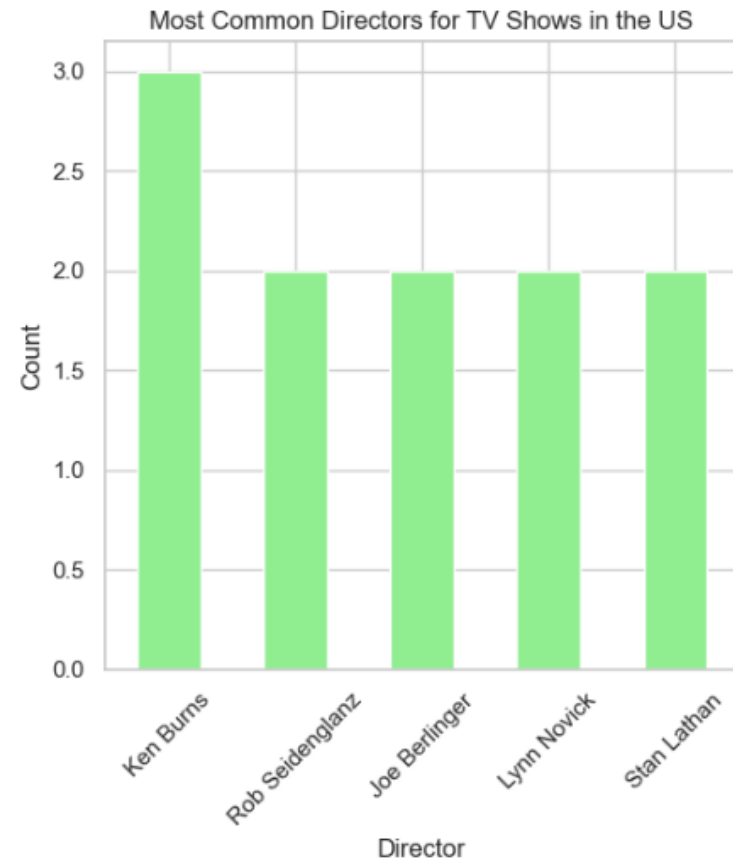
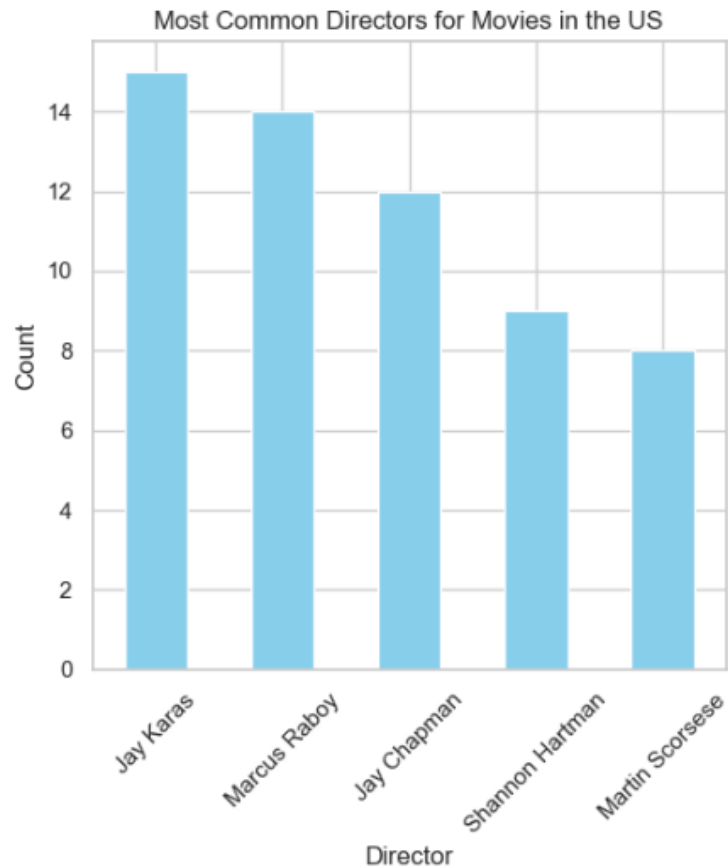
### 3.3 DETERMINE THE MOST POPULAR GENRES FOR TV SHOWS AND MOVIES IN THE TURKEY.



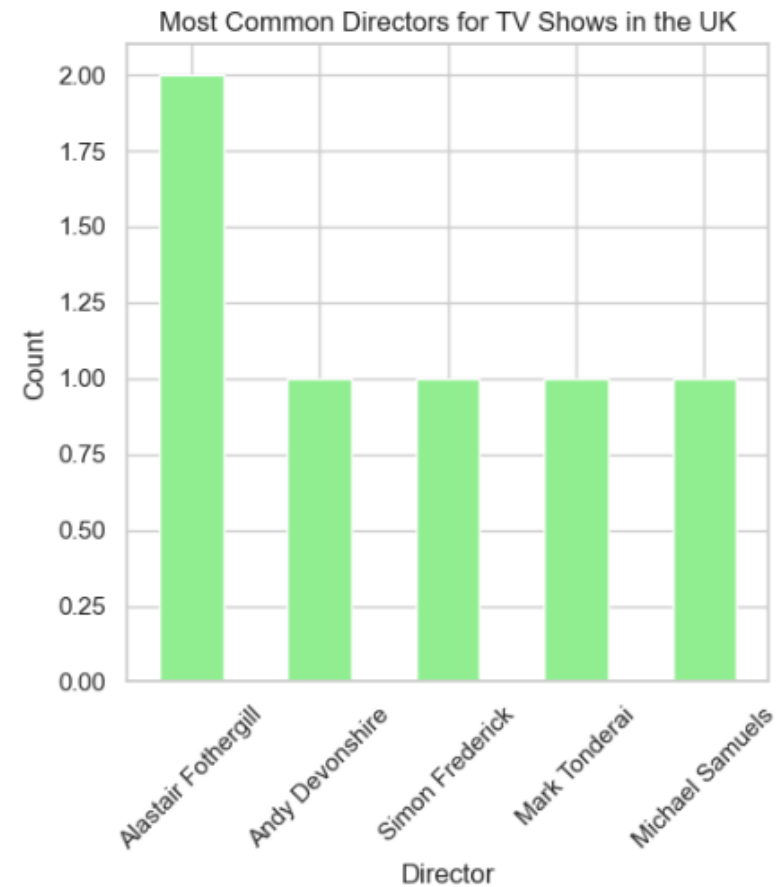
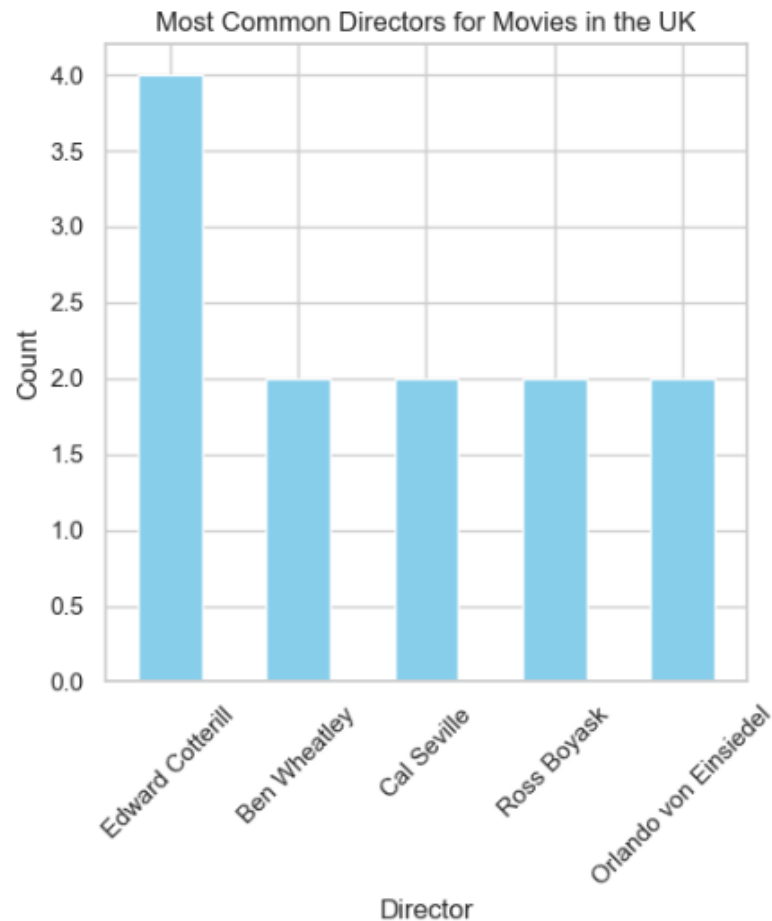




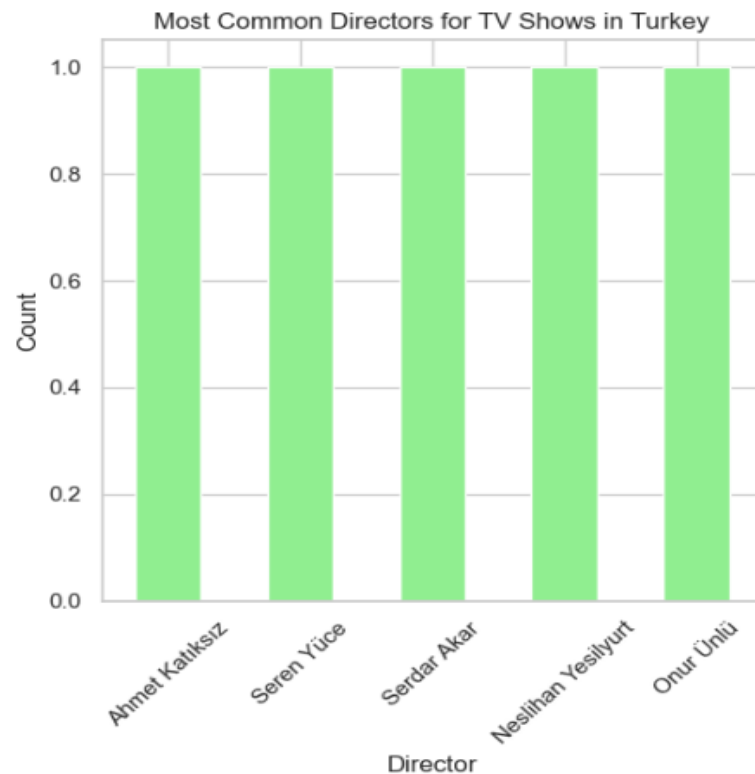
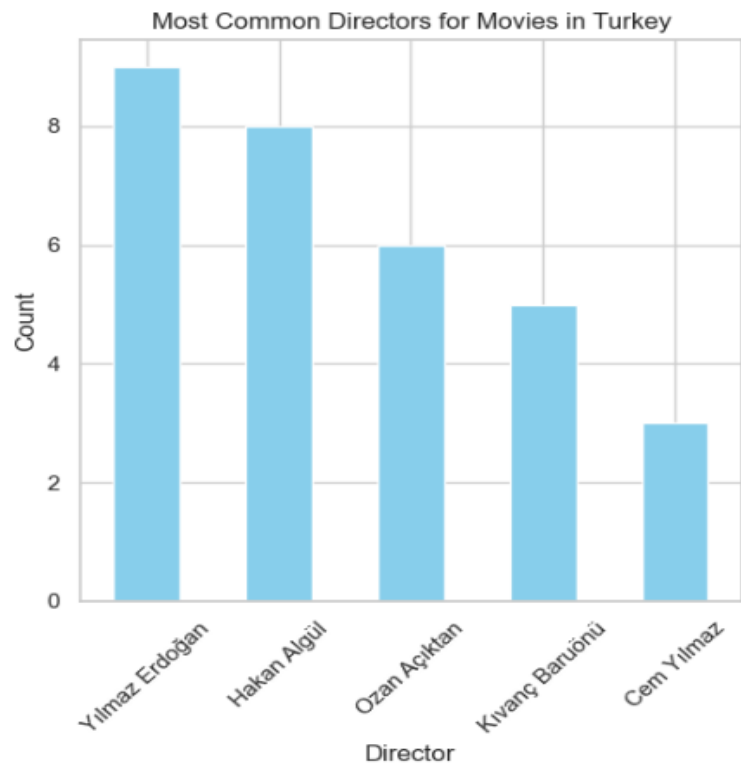
## 3.4 INVESTIGATE THE MOST COMMON DIRECTORS FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



## 3.4 INVESTIGATE THE MOST COMMON DIRECTORS FOR MOVIES AND TV SHOWS IN EACH COUNTRY.

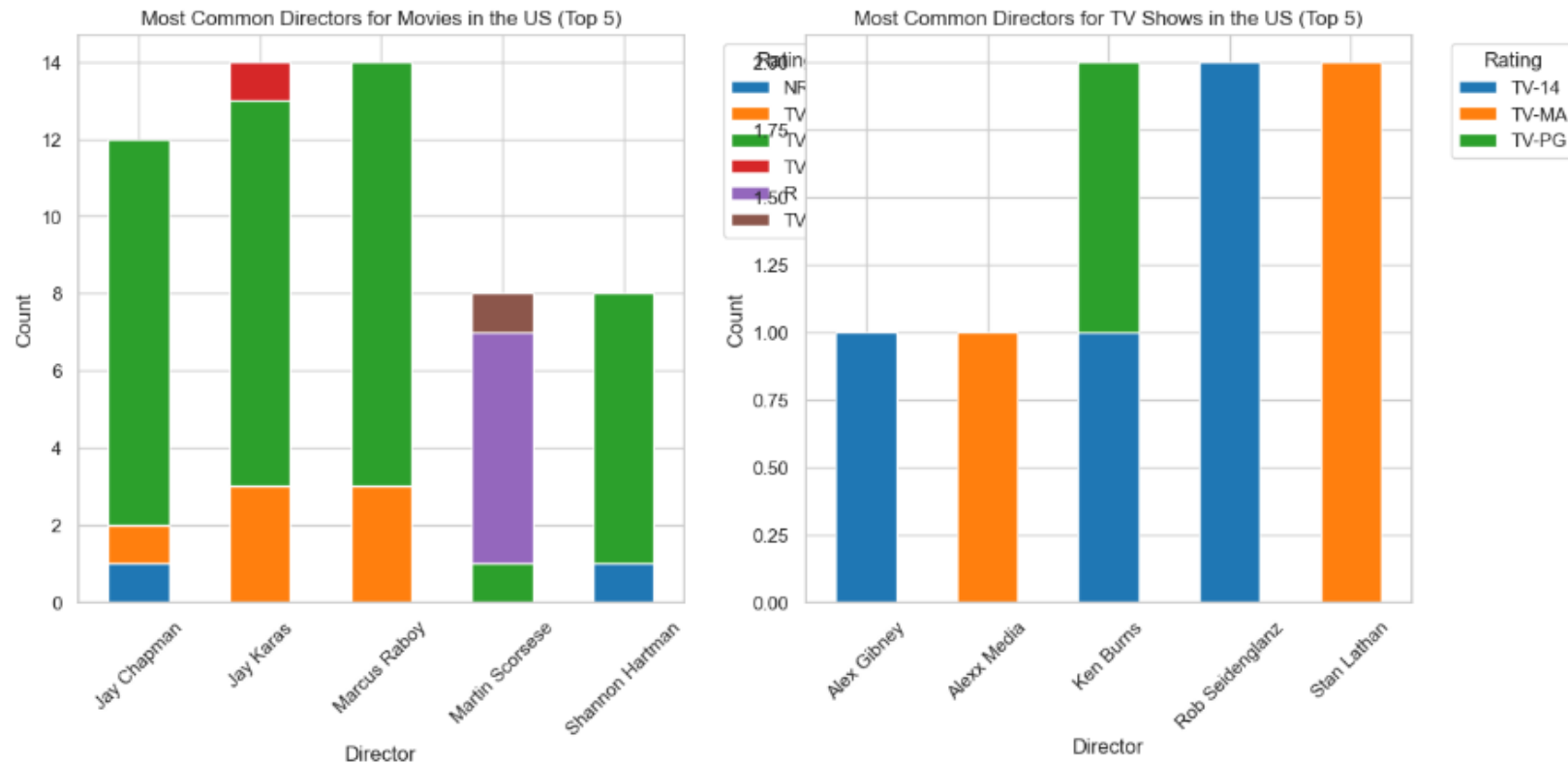


### 3.4 INVESTIGATE THE MOST COMMON DIRECTORS FOR MOVIES AND TV SHOWS IN EACH COUNTRY.

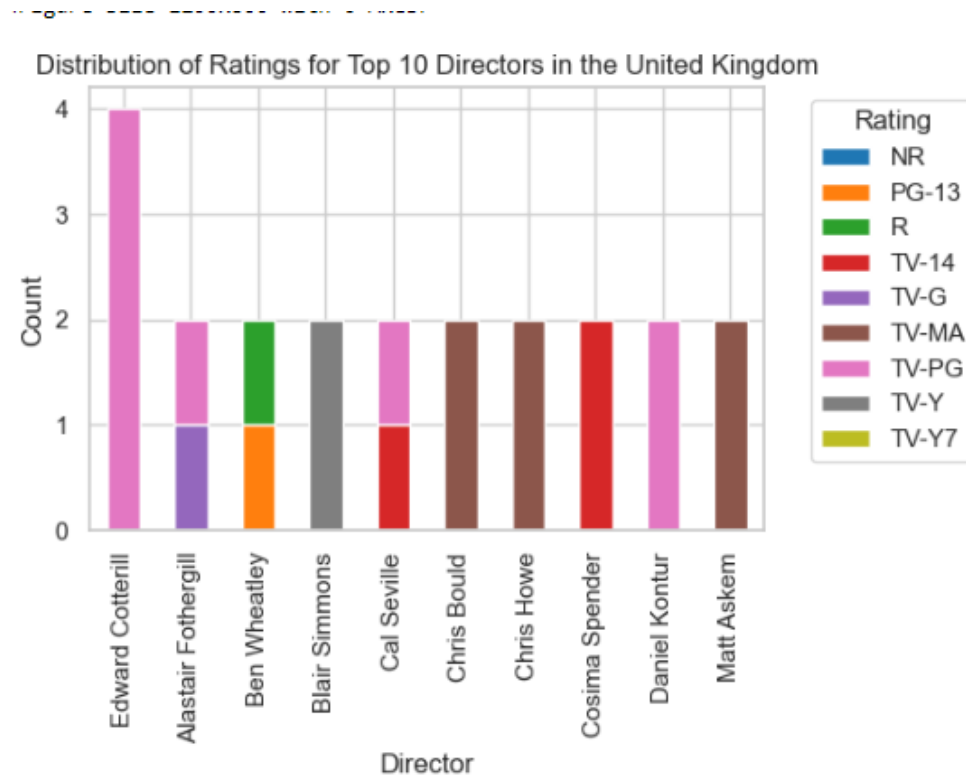




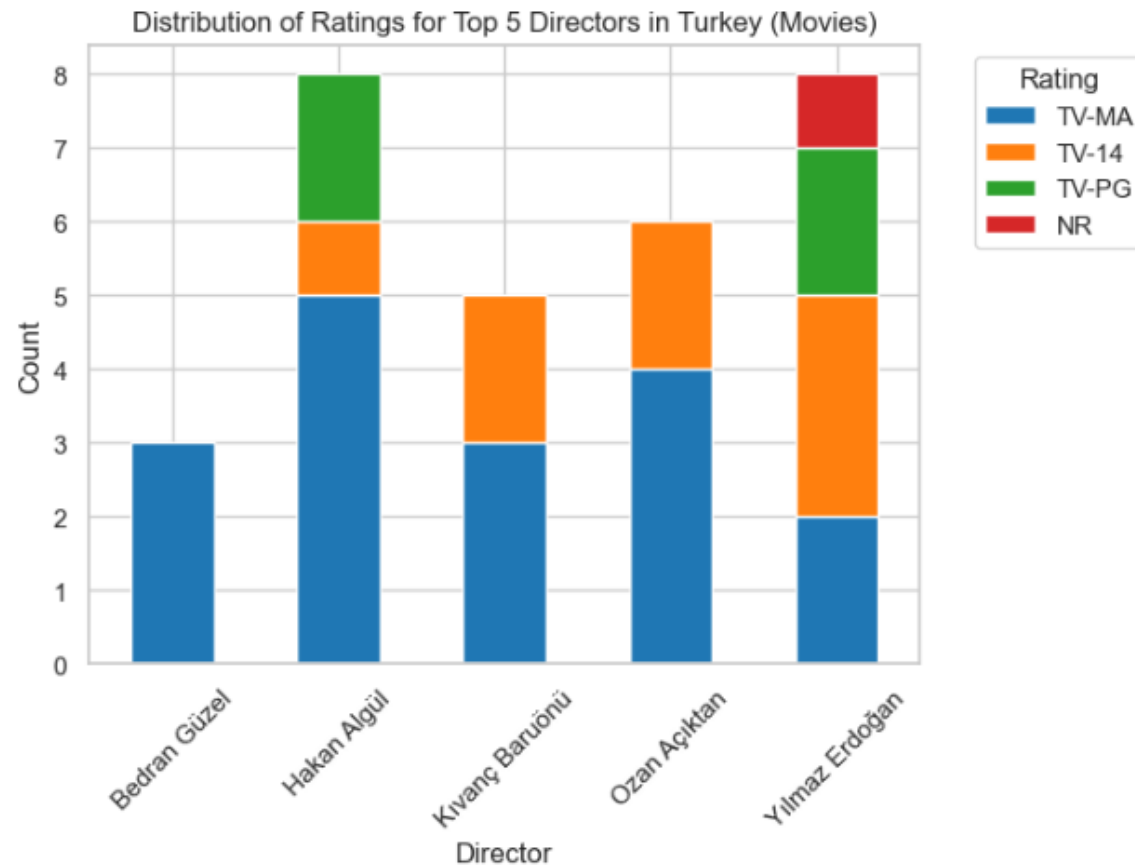
## 3.5 DETERMINE THE MOST COMMON DIRECTORS' RATINGS FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



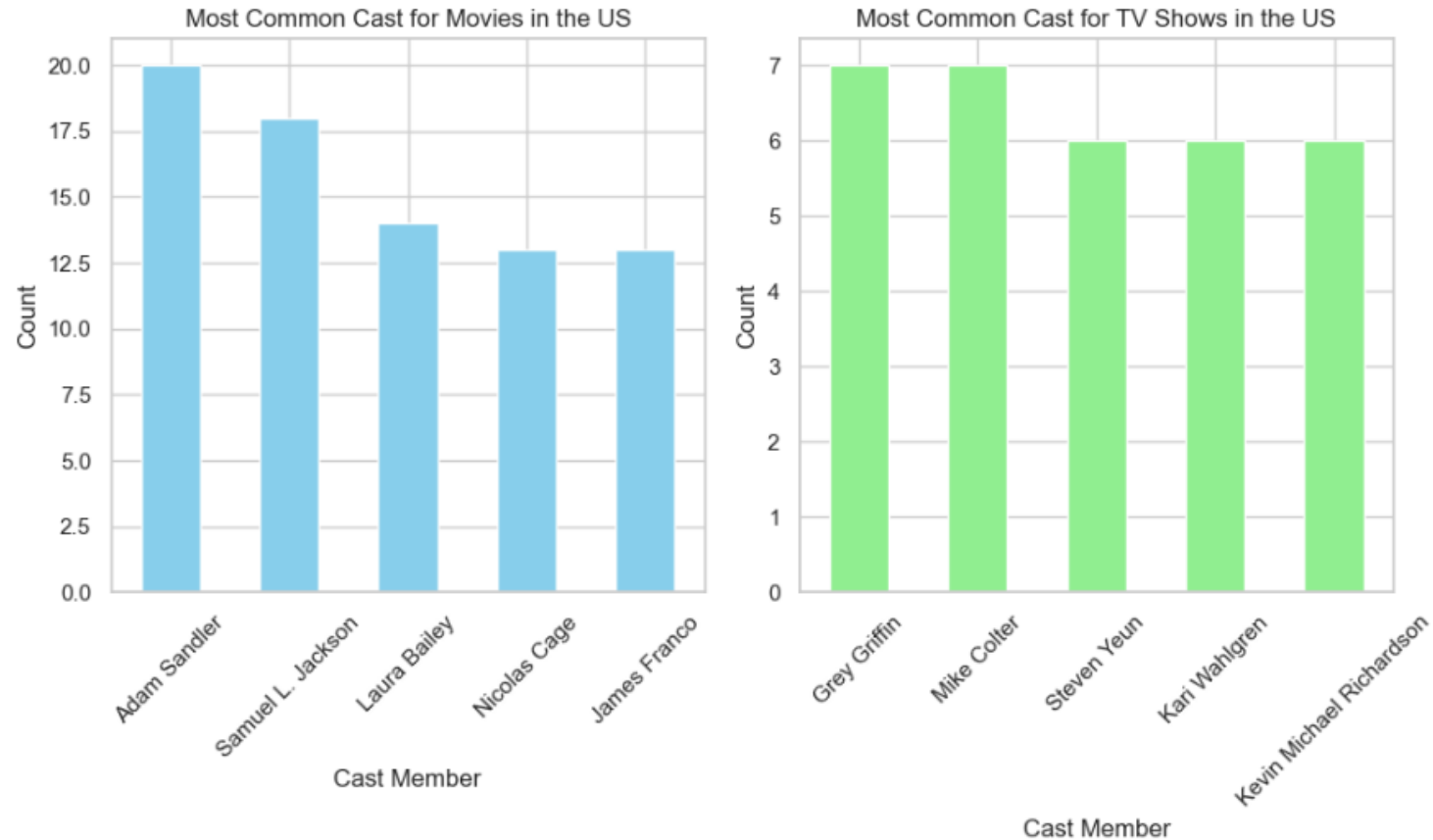
### 3.5 DETERMINE THE MOST COMMON DIRECTORS' RATINGS FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



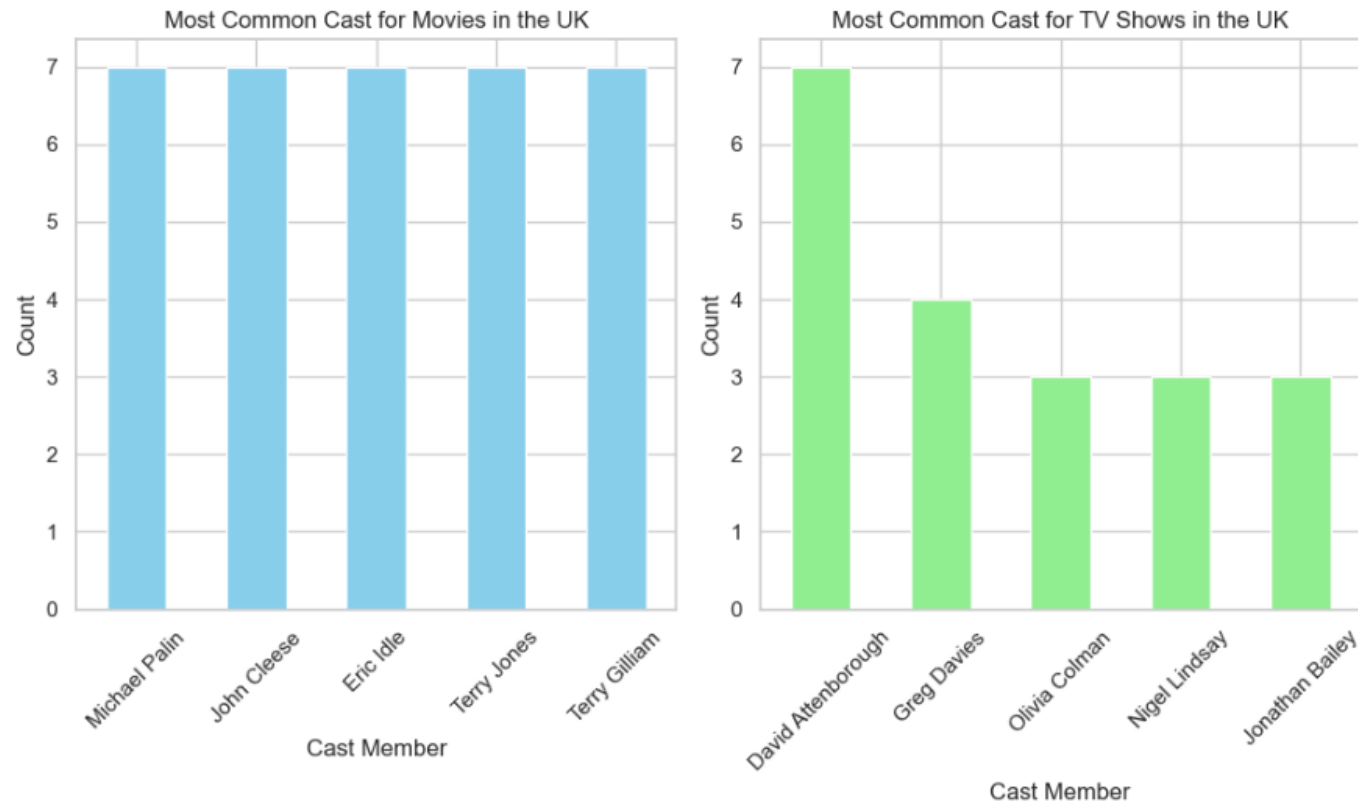
### 3.5 DETERMINE THE MOST COMMON DIRECTORS' RATINGS FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



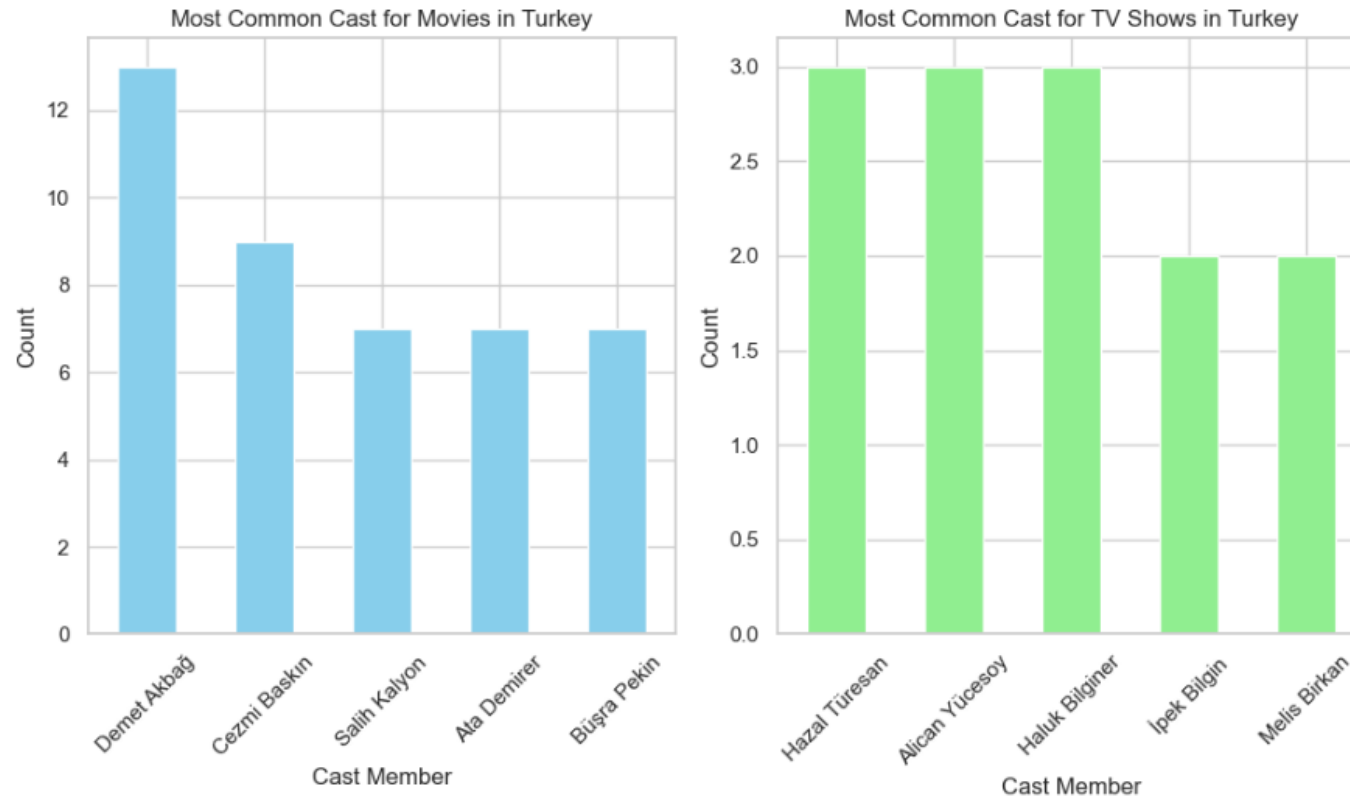
## 3.6 INVESTIGATE THE MOST COMMON CAST FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



## 3.6 INVESTIGATE THE MOST COMMON CAST FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



## 3.6 INVESTIGATE THE MOST COMMON CAST FOR MOVIES AND TV SHOWS IN EACH COUNTRY.



## 4.CONCLUSION:

- Upon completing the data analysis of Netflix content from the United States, the United Kingdom, and Turkey, several key insights have emerged. However, it's important to acknowledge that the analysis was impacted by the limited sample size of the dataset, which consisted of only 8807 entries. Despite this limitation, the data quality was deemed acceptable as there were no duplicates present.
- One notable aspect that significantly affected the analysis results was the presence of missing values. Missing values had a substantial impact on certain figures, particularly in variables such as country, cast, and director. As a result, the analysis of the number of movies for each country, as well as other metrics related to directors and casts, may be skewed due to incomplete data. The observed significant differences for the 1st, 4th, and 11th countries in the list should be interpreted with caution, considering the small sample size and the potential bias introduced by missing data.
- In conclusion, while the analysis provided valuable insights into Netflix content from different countries, the presence of missing values underscores the importance of rigorous data cleaning and preprocessing techniques to improve the accuracy and reliability of analyses in the future.

# C,NT.

Most common directors for movies in United States:

```
Director_info_missing  58
Jay Karas              15
Marcus Raboy          14
Name: director, dtype: int64
```

Most common directors for TV shows in United States:

```
Director_info_missing  706
Ken Burns              3
Rob Seidenglanz        2
Name: director, dtype: int64
```

Most common directors for movies in United Kingdom:

```
Director_info_missing  15
Edward Cotterill       4
Nick Broomfield        2
Name: director, dtype: int64
```

Most common directors for TV shows in United Kingdom:

```
Director_info_missing  191
Alastair Fothergill     2
Andy Devonshire         1
Name: director, dtype: int64
```

Most common directors for movies in Turkey:

```
Yılmaz Erdoğan        9
Hakan Algül            8
Ozan Açıktan           6
Name: director, dtype: int64
```

Most common directors for TV shows in Turkey:

```
Director_info_missing  24
Ahmet Katıksız         1
Seren Yüce             1
Name: director, dtype: int64
```

---