# Capstone Project -

# Netflix Movies And TV Shows Clustering

By

Jijabai Dhanwate

# CONTENT

Defining problem statement

EDA and feature engineering

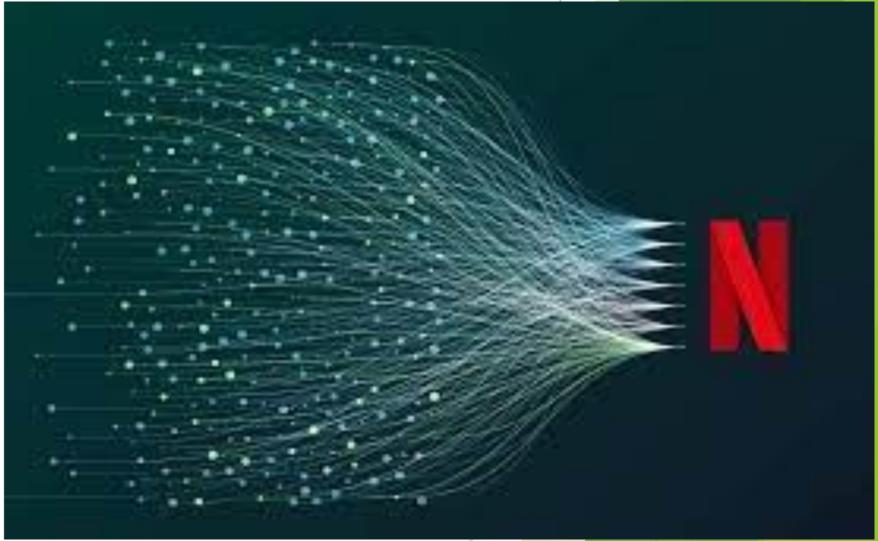
Dimensionality reduction

Clustering

Evaluating Clusters

Interpretation

7 Visualization



# OVERVIEW

The objective of this project is to create a predictive model for Netflix movies and TV shows that can estimate their user ratings based on various factors such as title, genre, cast, director, release year, runtime, and production budget. The project is broken down into several parts starting with data collection, where we gather relevant data on Netflix movies and TV shows. The collected data will then be cleaned and transformed into a format that is suitable for analysis.

In the next step, feature selection, we will determine which factors have the most significant impact on the ratings of Netflix movies and TV shows. We will analyze the data to identify which features are correlated with higher or lower ratings and use them as inputs for the predictive model. Additionally, we will use domain knowledge and feature engineering techniques to create new features that can improve the predictive power of the model.

- After selecting the features, we will prepare the dataset for modeling by splitting it into training, validation, and test sets and perform necessary data preprocessing or normalization steps. Then, we will train machine learning algorithms, such as regression or classification models, on the training set to develop a predictive model that can accurately predict Netflix movie or TV show ratings.

- We will evaluate the performance of the model on the validation set and select the best-performing model based on metrics like mean squared error or accuracy. Finally, we will test the final model on the test set to ensure its accuracy and reliability. With this predictive model, Netflix can use the movie and TV show specifications to estimate the expected user ratings and make strategic decisions on content acquisition and production. Additionally, users can use the model to discover new titles that match their preferences or to anticipate the ratings of upcoming releases.

# Data Pipeline

The first part of the Netflix movies and TV shows clustering project involved removing unnecessary features that contained null values.

In the second part, we encoded categorical features and changed datetime columns to make the data easier to work with and convert categorical data into numerical data.

The third part of the project involved exploratory data analysis (EDA) to identify trends and patterns in the selected features that could be used for clustering.

In the final part of the project, we created a clustering model using algorithms such as K-means or Hierarchical Clustering.

The objective of the clustering model was to group similar movies and TV shows together and separate them from dissimilar ones.

Clustering can provide valuable insights for both users and the company and enhance the streaming experience for everyone.

The clustering model can help Netflix in content acquisition and production decisions.

Overall, clustering Netflix movies and TV shows can help better understand their characteristics and recommend similar content to users.

# Data Summary

show_id : Unique ID for every Movie / Tv Show

type : Identifier - A Movie or TV Show

title : Title of the Movie / Tv Show

director : Director of the Movie

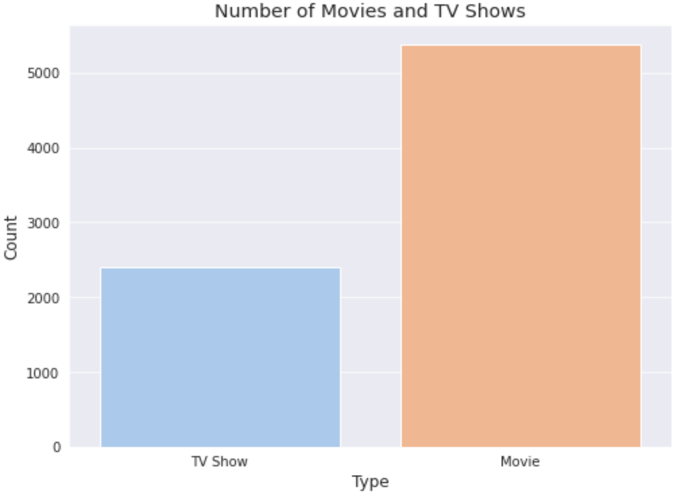
cast : Actors involved in the movie / show

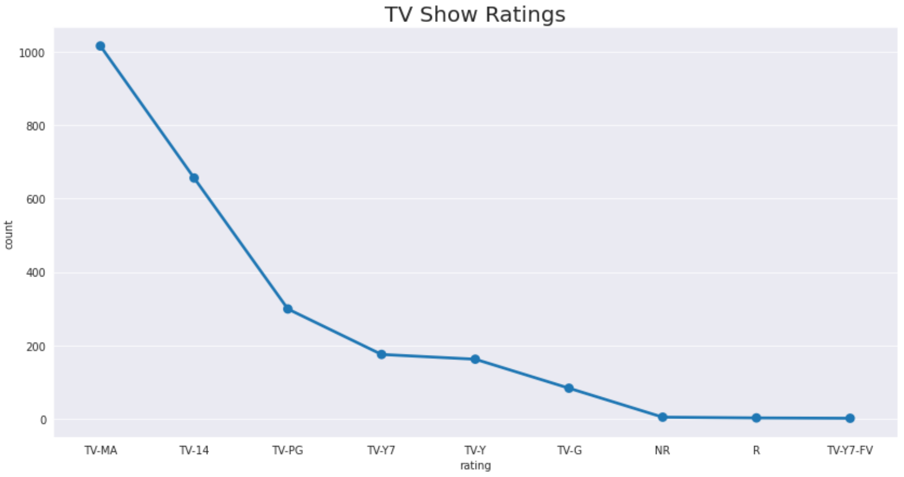
country : Country where the movie / show was produced

# Data Summary

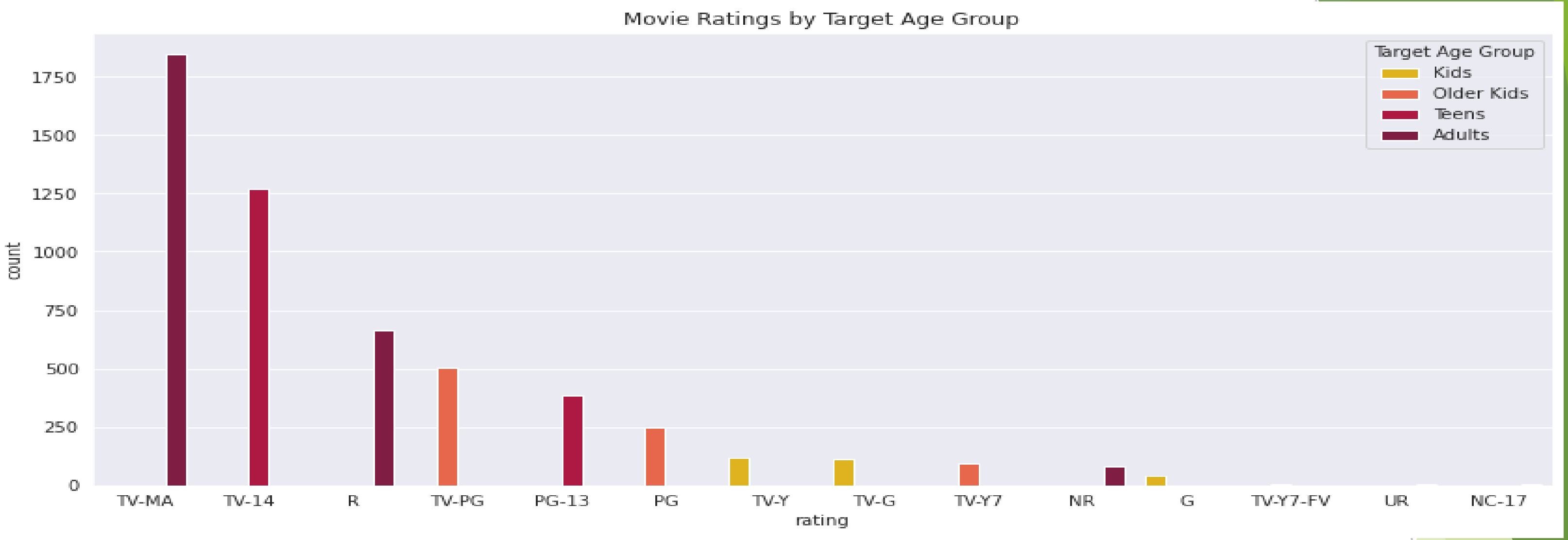
- date_added : Date it was added on Netflix
- release_year : Actual Releaseyear of the movie / show
- rating : TV Rating of the movie / show
- duration : Total Duration - in minutes or number of seasons
- listed_in : Genere
- description: The Summary description

# EDA

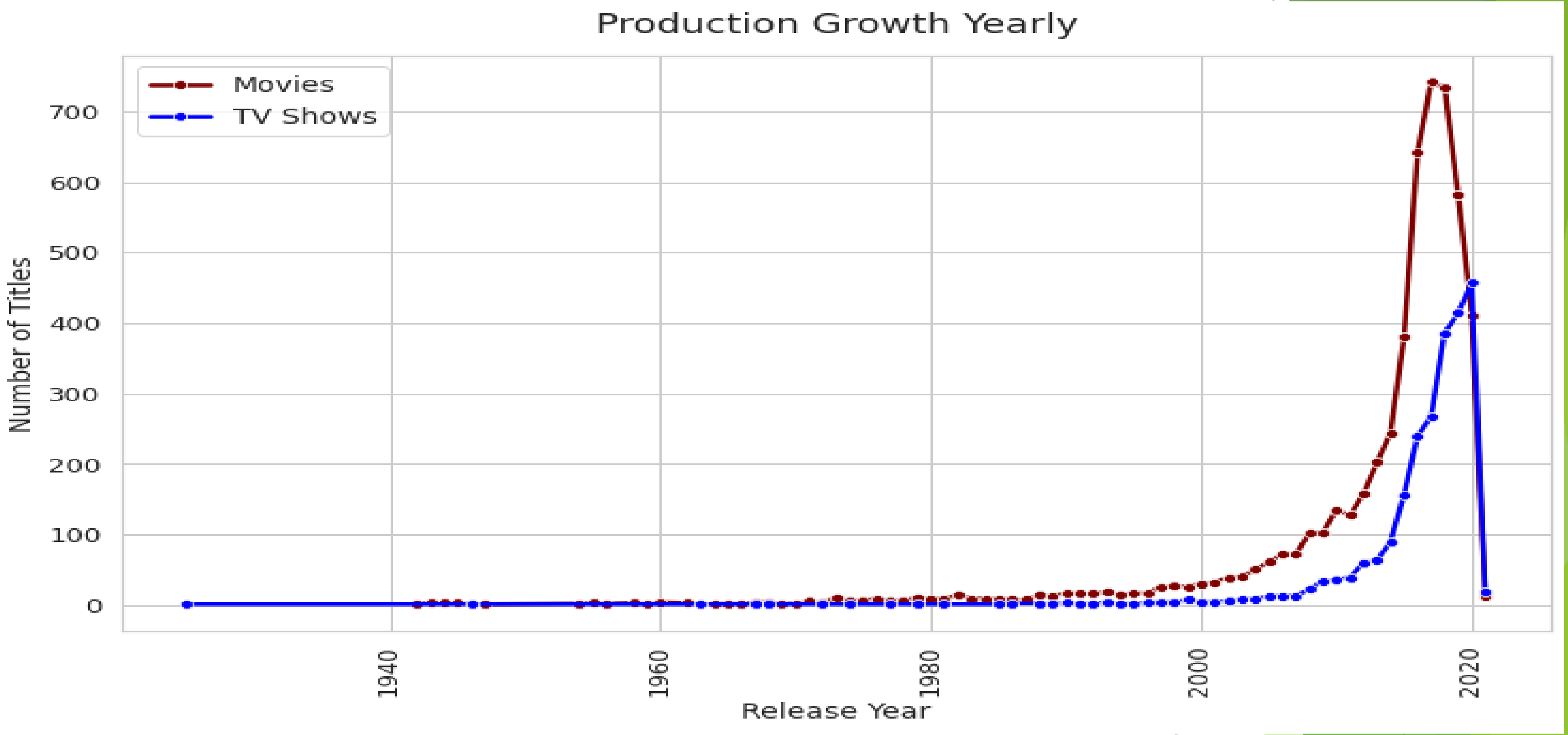


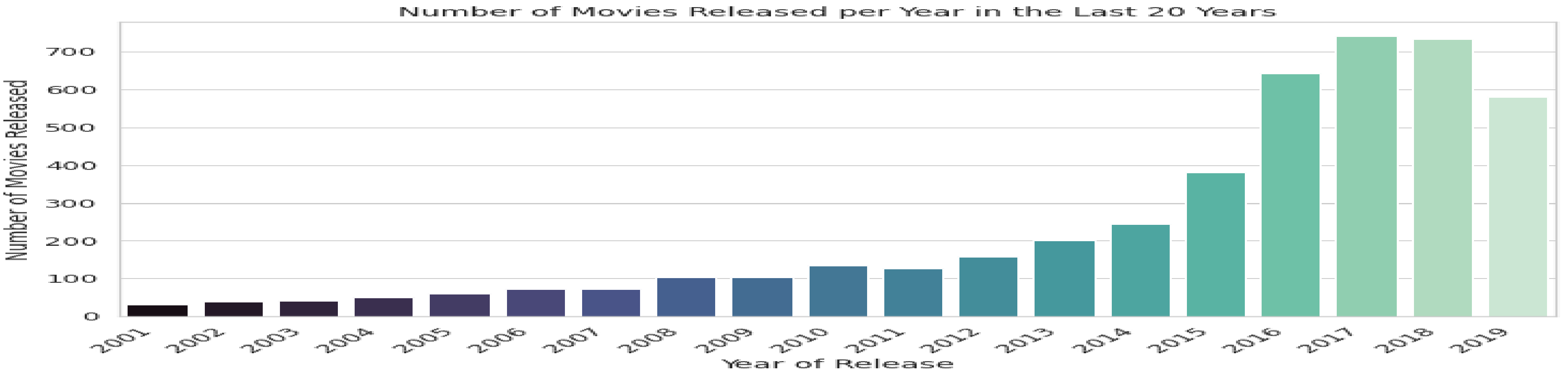


1.The number of movies on Netflix is greater than the number of TV shows, with 5372 movies and 2398 TV shows currently available on the platform.



According to the dataset, TV-MA is the most common rating for TV shows, with the highest number of occurrences in the 'rating' column. This indicates that a significant portion of the TV shows available on Netflix are intended for adult audiences.According to the dataset, TV-MA is the most common rating for both movies and TV shows. This indicates that a significant portion of the content available on Netflix is intended for adult audiences. Specifically, TV-MA has the highest number of occurrences in the 'rating' column for TV shows, while for movies it is also the most common rating. This suggests that Netflix's content caters to a primarily adult demographic, with a focus on mature and potentially controversial themes.


Production Growth Yearly
Movies
TV Shows
Number of Titles
0
100
200
300
400
500
600
700
1940
1960
1980
2000
2020
Release Year



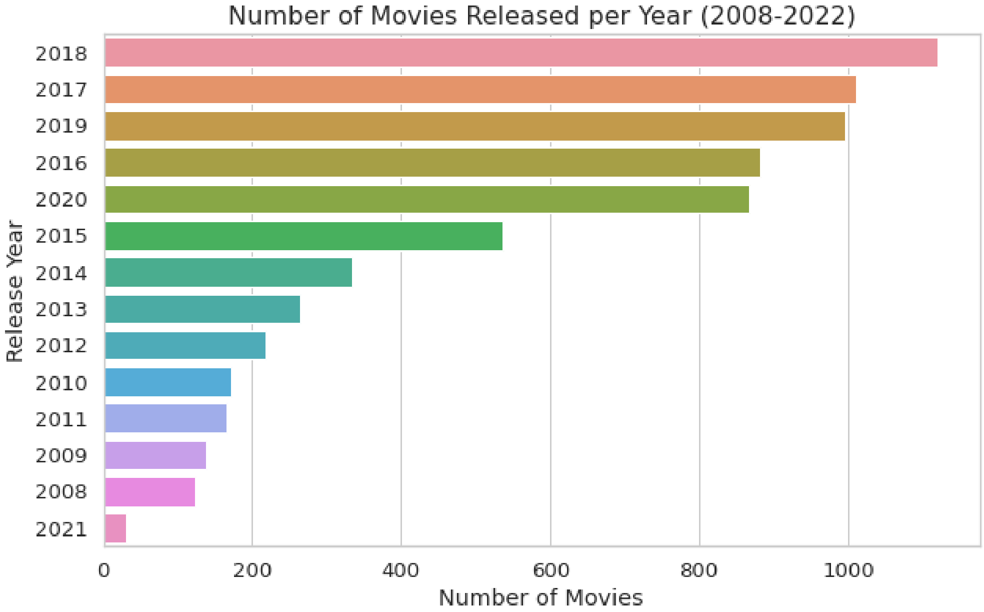
The years 2017 and 2018 had the highest number of movie releases, while 2020 had the highest number of TV show releases.

The growth rate of movie releases on Netflix is significantly faster than that of TV shows.

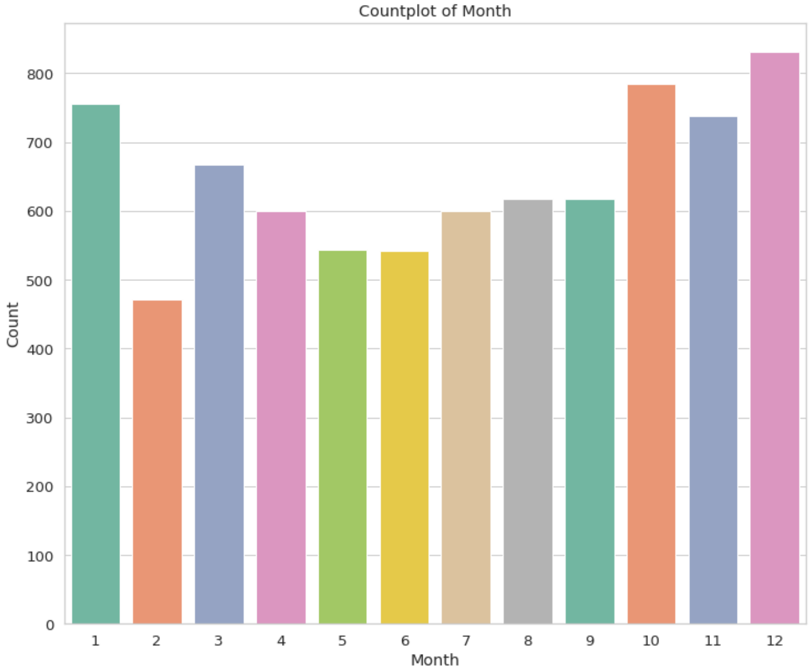
Since 2015, there has been a substantial increase in the number of movies and TV show episodes available on Netflix.

However, there has been a notable drop in the number of movies and TV show episodes produced after 2020.

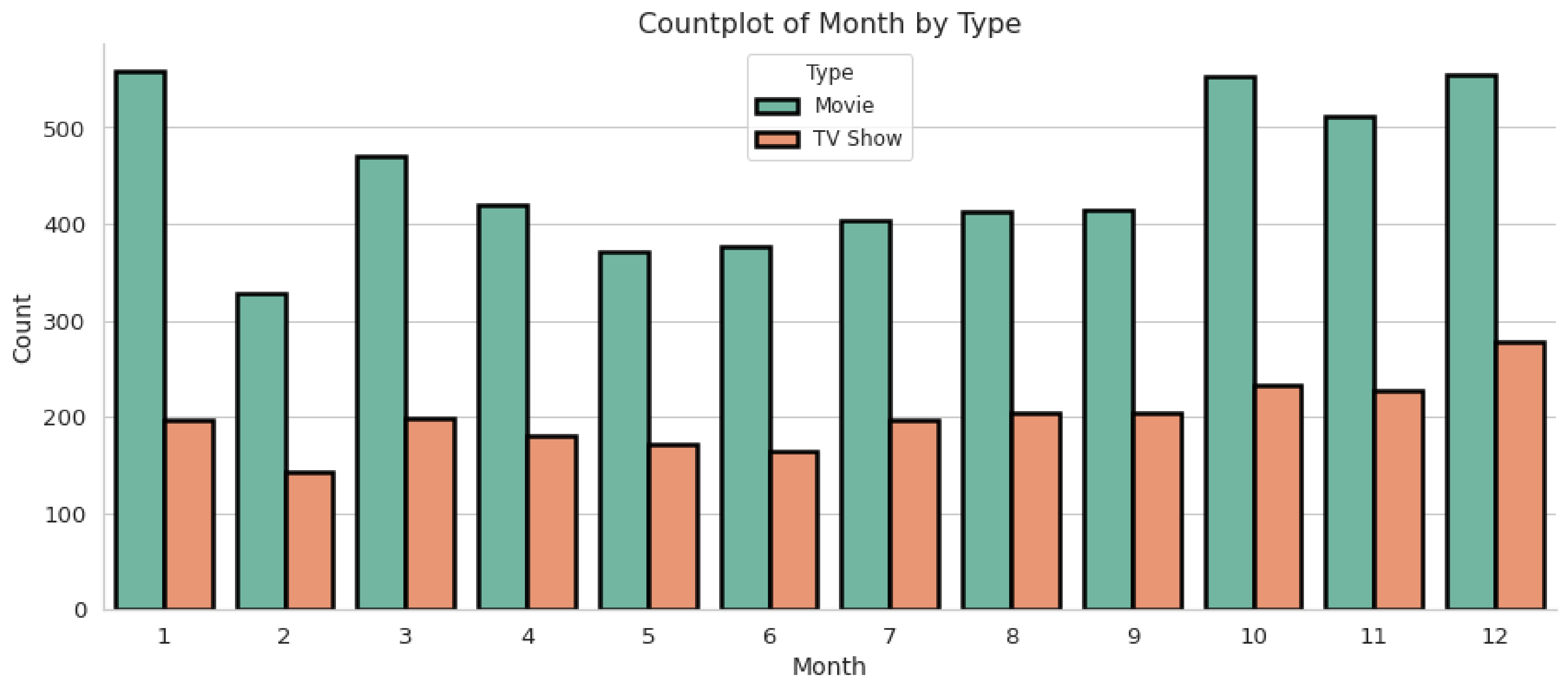
It appears that Netflix has given more attention to increasing its movie content rather than TV shows, as the growth rate of movies has been much more significant than that of TV shows.



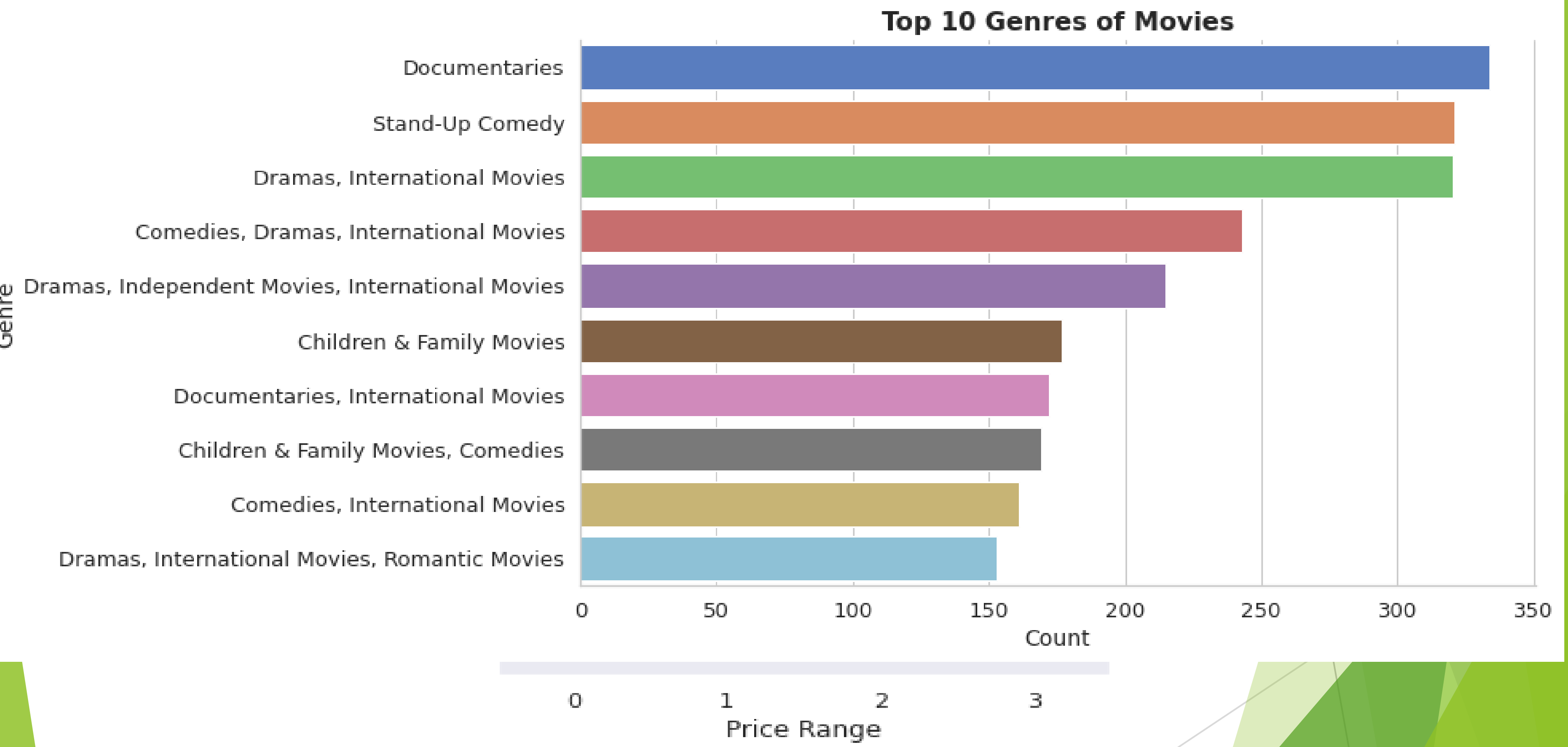
Number of Movies Released per Year (2008-2022)



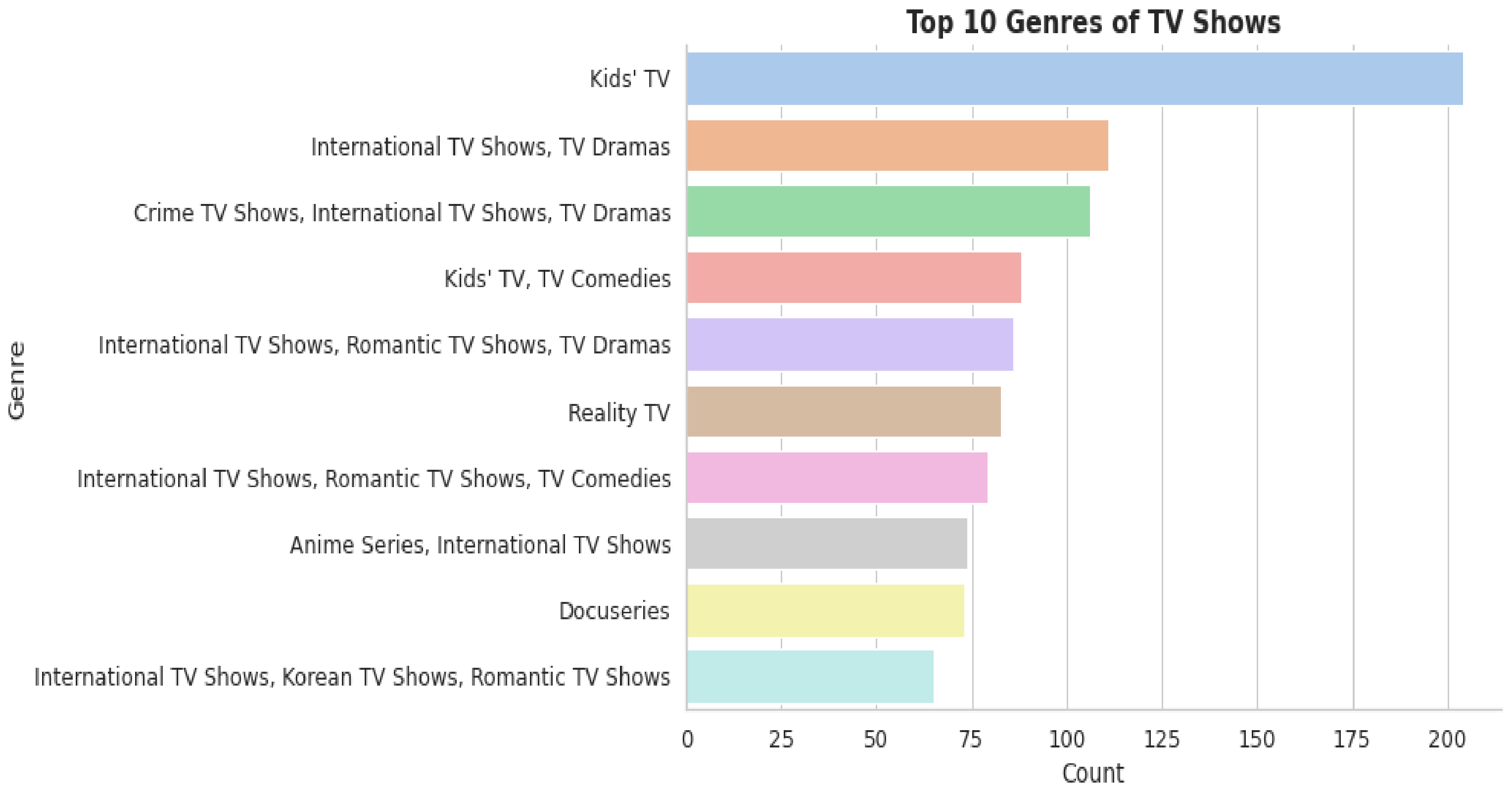
Countplot of Month



According to the countplot, it appears that Netflix adds the highest number of movies and TV shows during the period between October and January. This period seems to be the busiest time of year for Netflix in terms of adding new content to its platform.


Top 10 Genres of Movies
Genre
Documentaries
Stand-Up Comedy
Dramas, International Movies
Comedies, Dramas, International Movies
Dramas, Independent Movies, International Movies
Children & Family Movies
Documentaries, International Movies
Children & Family Movies, Comedies
Comedies, International Movies
Dramas, International Movies, Romantic Movies
0
50
100
150
200
250
300
350
Count
0
1
2
3
Price Range

# EDA (continued)



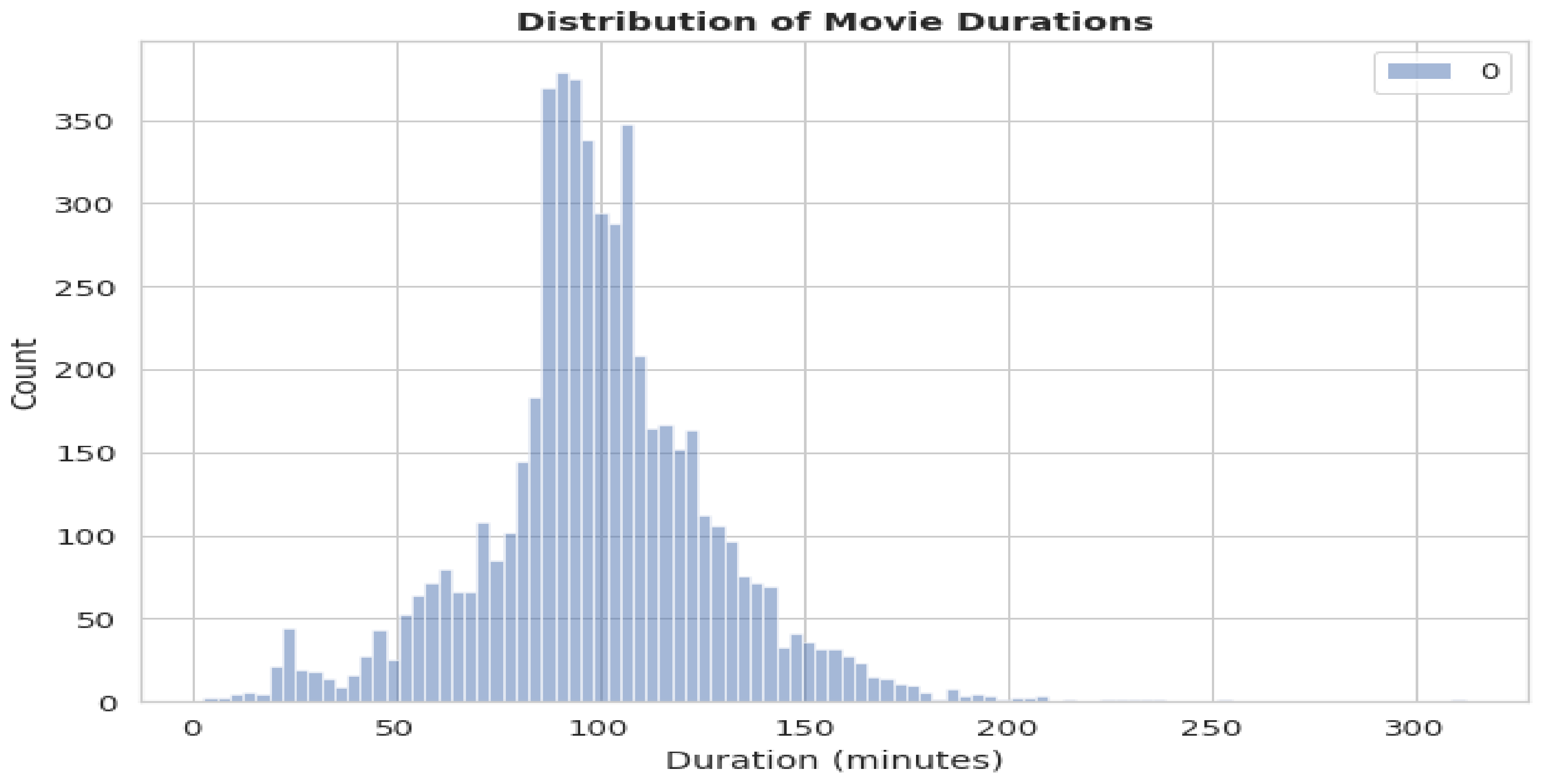
Netflix offers a diverse range of TV show genres, each with its own unique flavor and appeal. However, one genre that stands out as a perennial favorite among viewers of all ages is kids TV.

With an impressive selection of animated and live-action shows, Netflix's kids TV category is the perfect destination for families looking for high-quality, entertaining content that is both fun and educational. From beloved classics like SpongeBob SquarePants and Power Rangers to exciting new series like Carmen Sandiego and The Dragon Prince, Netflix's kids TV library has something for every young viewer.

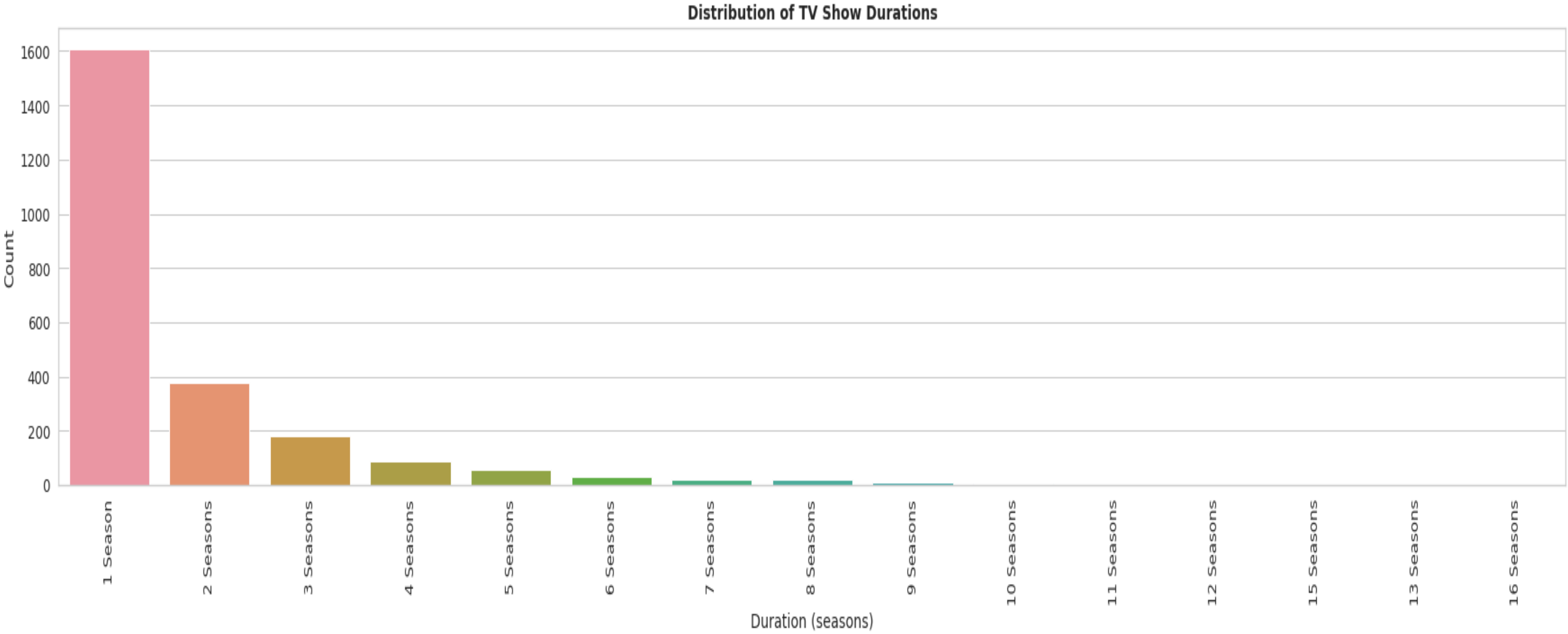
Moreover, Netflix's kids TV category is designed with parents in mind, offering a safe and secure viewing environment that allows them to have peace of mind while their kids enjoy their favorite shows. The parental controls feature allows parents to set age-appropriate content filters, monitor viewing history, and restrict access to certain shows or movies.

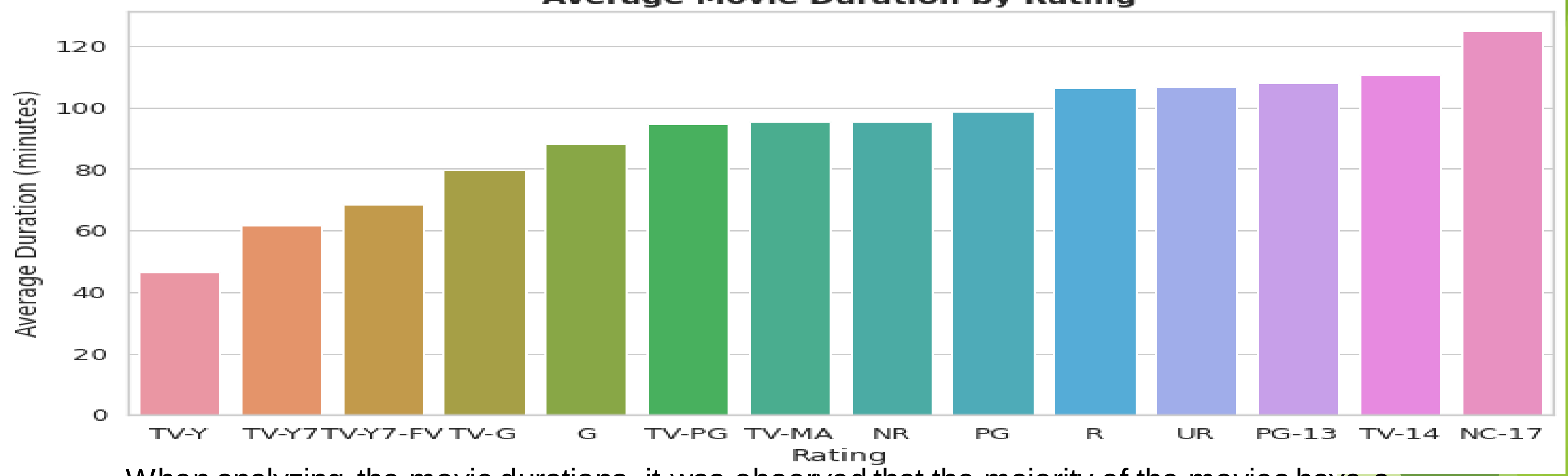
So, whether you're looking for a way to keep your little ones entertained on a rainy day, or just want to bond with your family over a great TV show, Netflix's kids TV category is the perfect place to start. With its vast selection of entertaining and educational content, it's no wonder that kids TV remains one of the top genres on the platform.

# EDA (continued)



# EDA (continued)

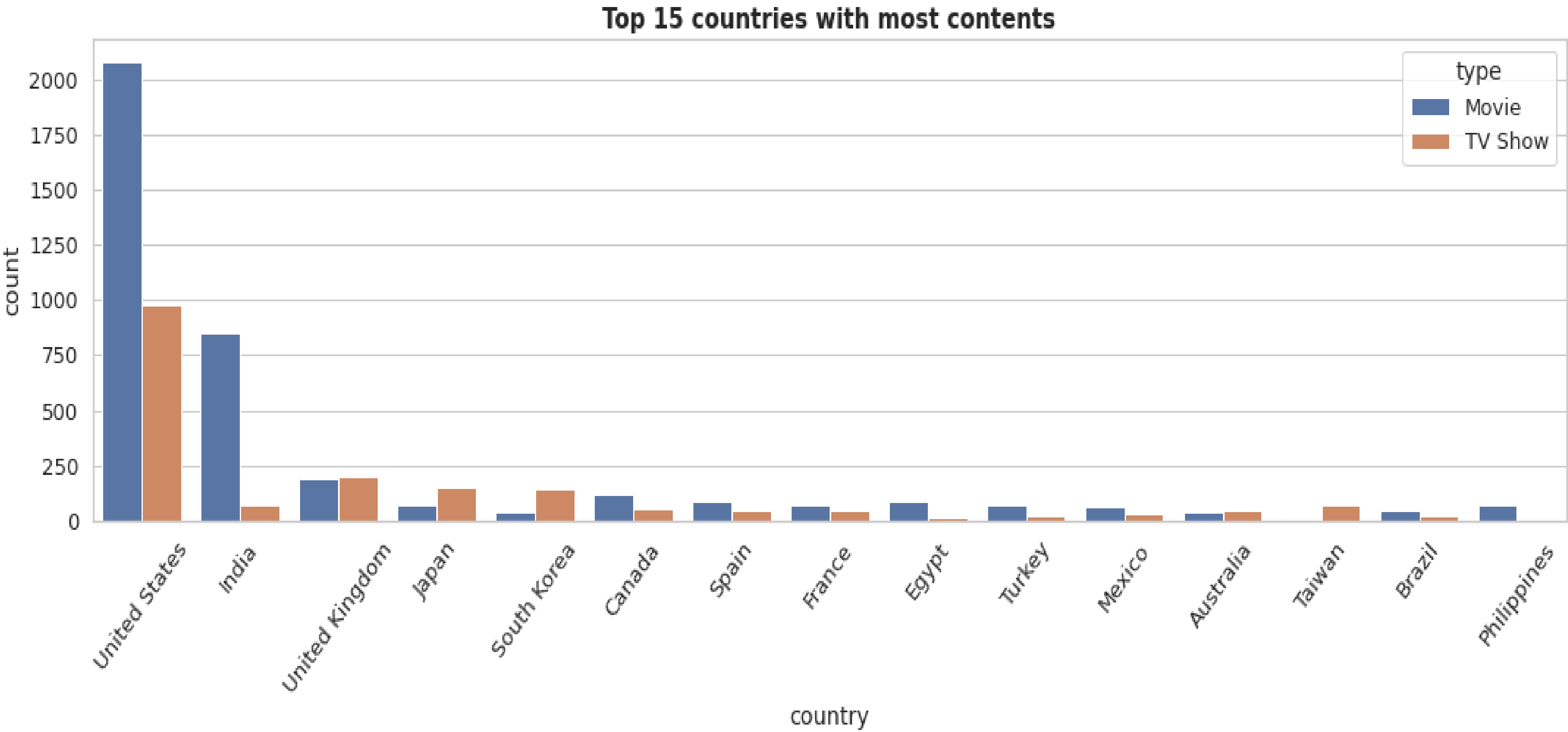


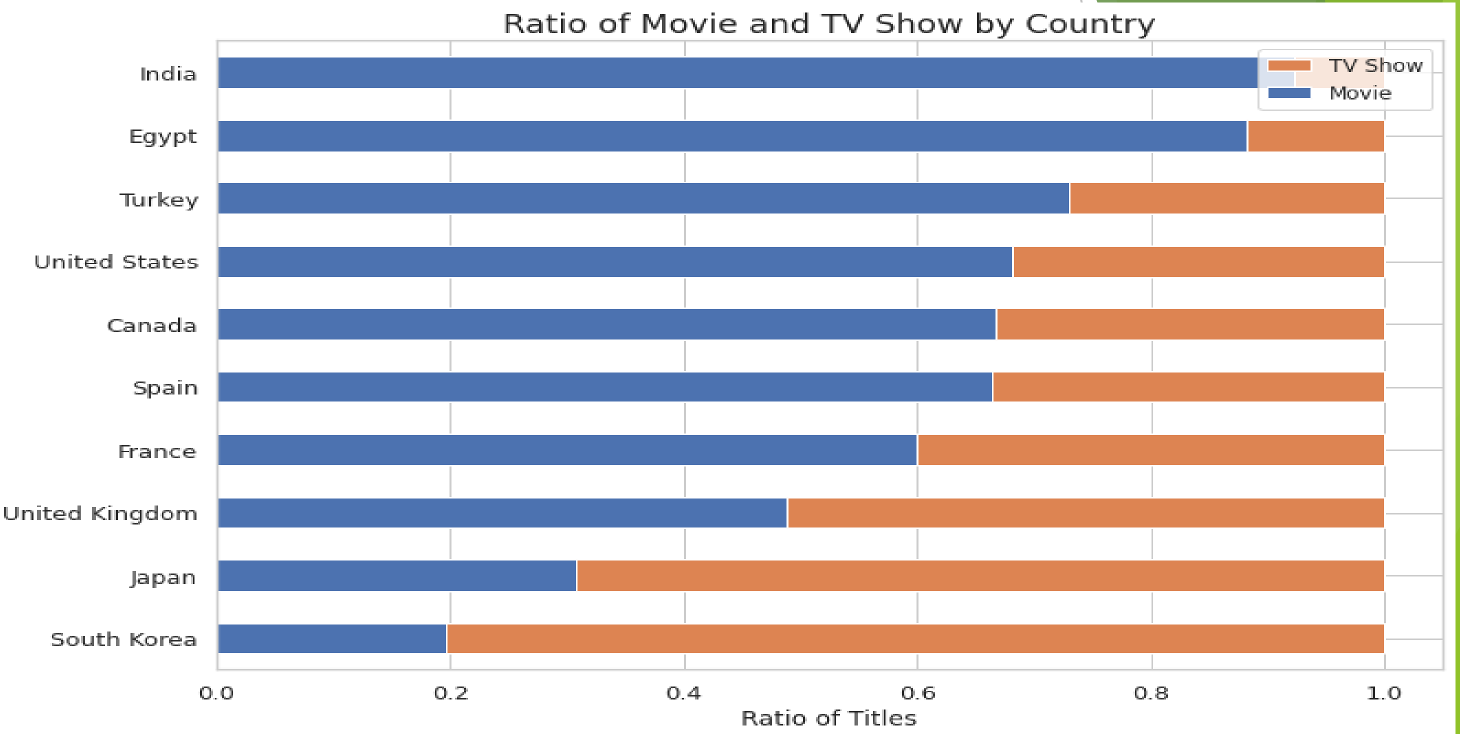


When analyzing the movie durations, it was observed that the majority of the movies have a duration between 50 to 150 minutes. On the other hand, the TV shows have a large number of single-season shows, which indicates that most of the TV shows on Netflix are relatively new.

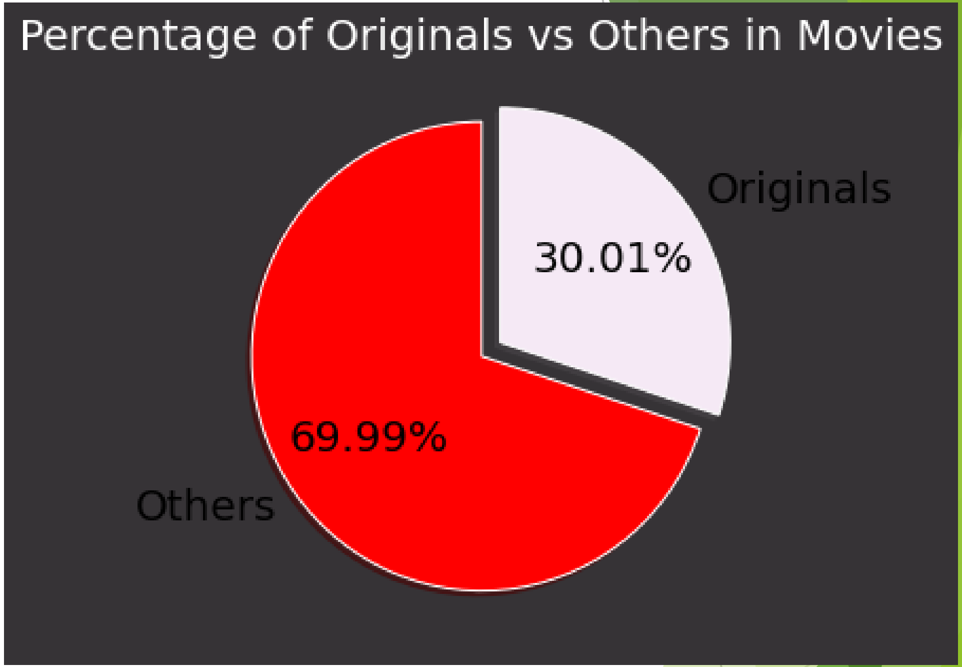
Furthermore, the analysis showed that movies with a rating of NC-17 have the longest average duration. This might be because the movies with such a rating can explore more mature themes and include more explicit content, which requires a longer runtime to tell a compelling story.

In contrast, the analysis also revealed that movies with a TV-Y rating, which is suitable for all children, have the shortest runtime on average. This suggests that the movies with this rating tend to be shorter and may have simpler plots and themes that are suitable for younger audiences.


Top 15 countries with most contents
count
2000
1750
1500
1250
1000
750
500
250
0
type
Movie
TV Show
United States
India
United Kingdom
Japan
South Korea
Canada
Spain
France
Egypt
Turkey
Mexico
Australia
Taiwan
Brazil
Philippines
country

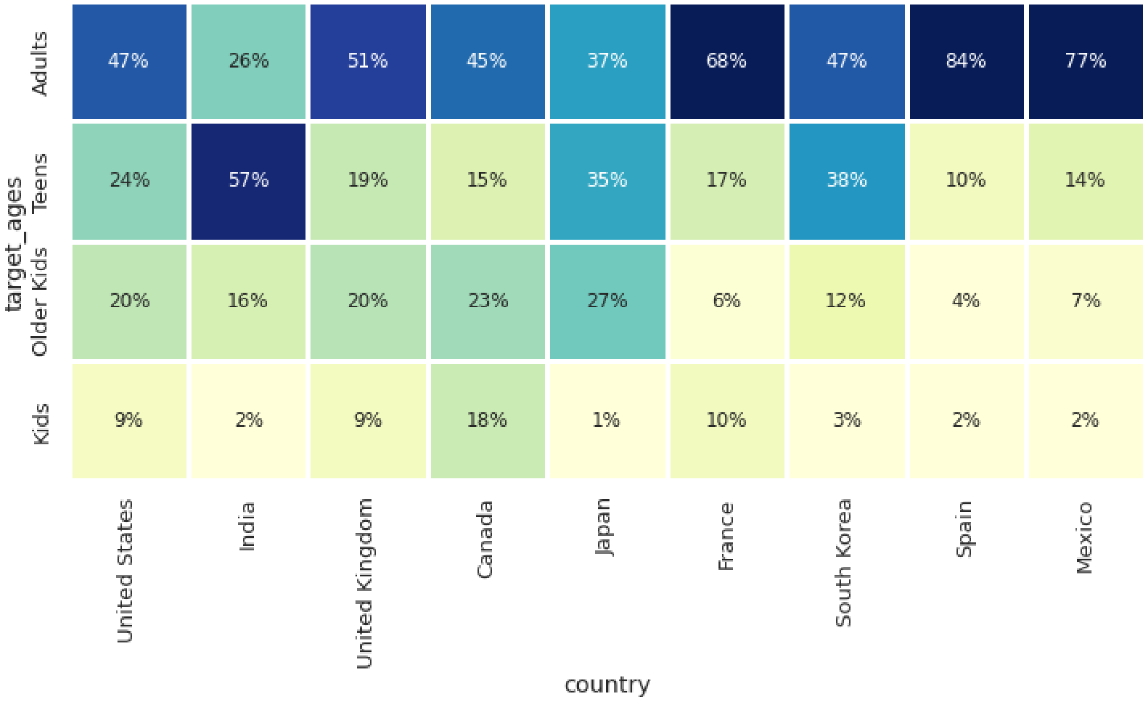


Netflix has the highest number of content in the United States, followed by India. India has the highest number of movies on Netflix.

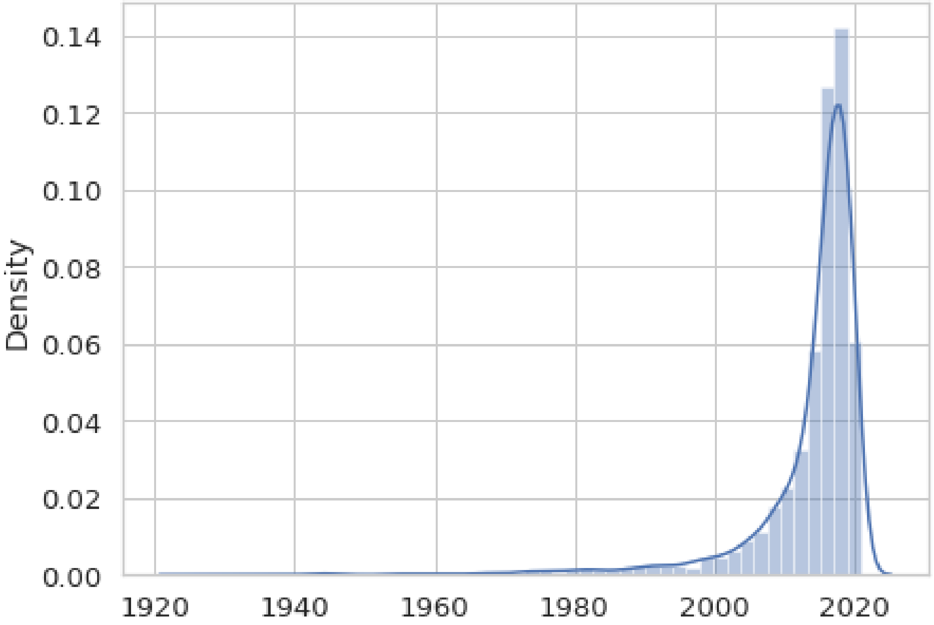


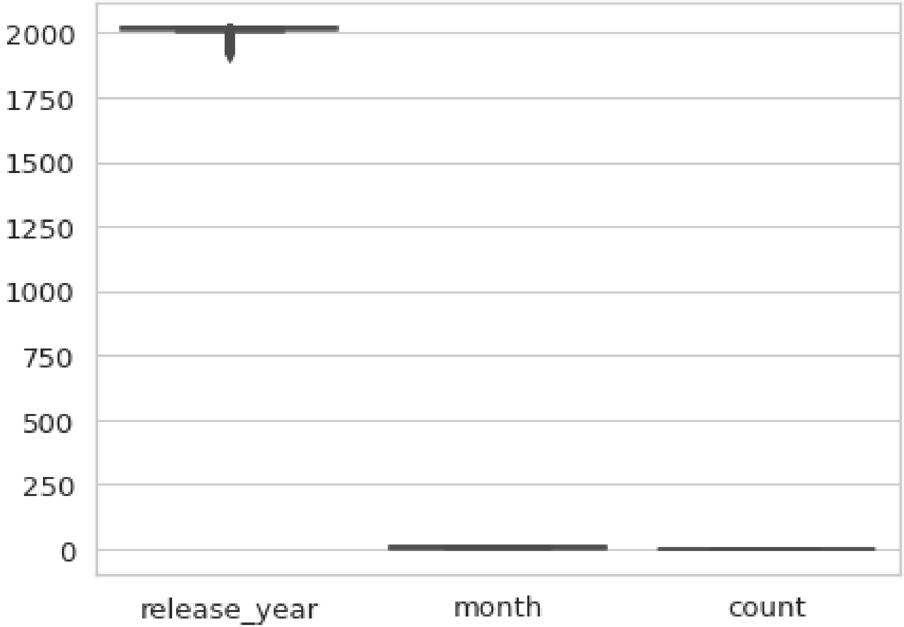
Netflix is known for producing original content, it is interesting to note that only 30% of the movies available on the platform were actually released by Netflix themselves. The remaining 70% of movies were added to Netflix after being released by different modes, such as theaters or other streaming platforms.

# Correlation Heatmap



# Outliers Handling

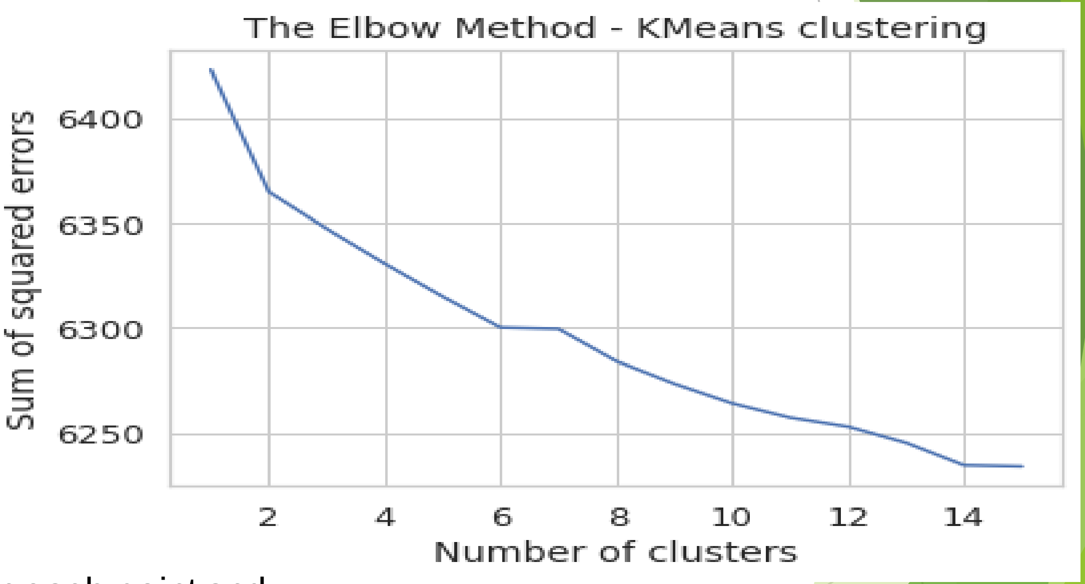




1.Except for the release year, almost all of the data are presented in text format.
2.The textual format contains the data we need to build a cluster/building model.
Therefore, there is no need to handle outliers.

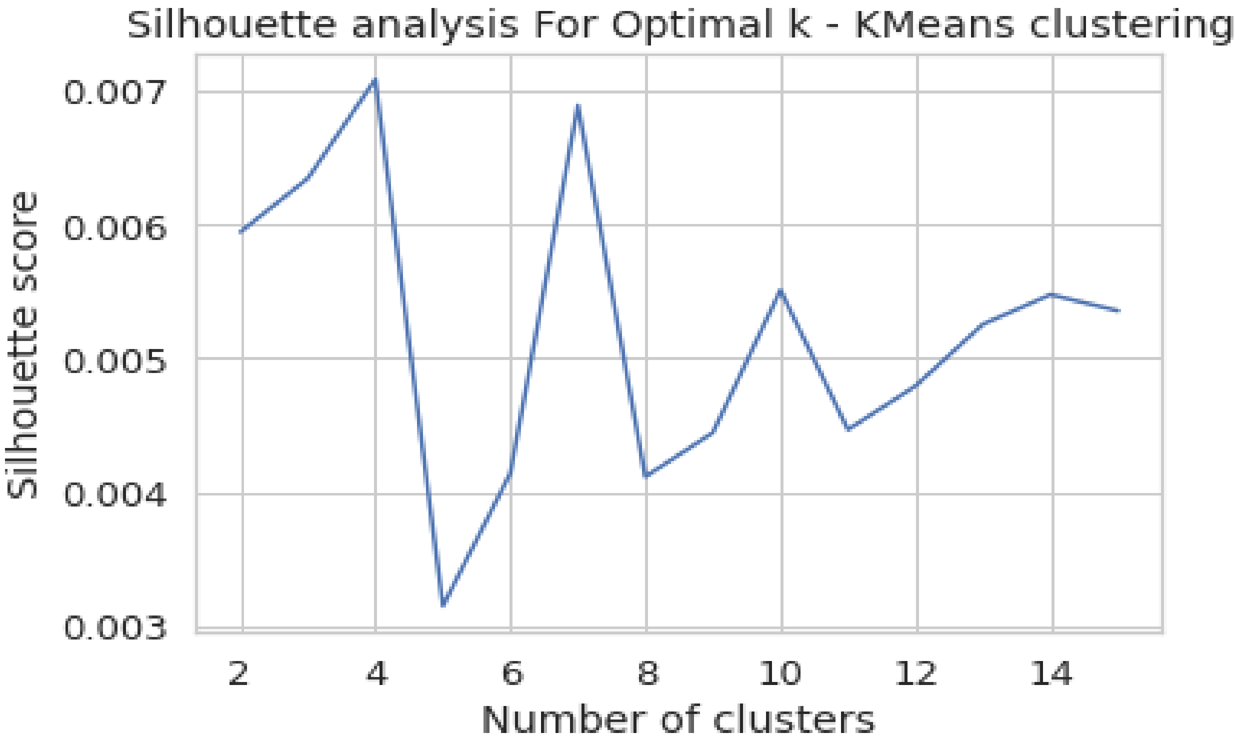
# Cluster Implementation

## K-Means Clustering

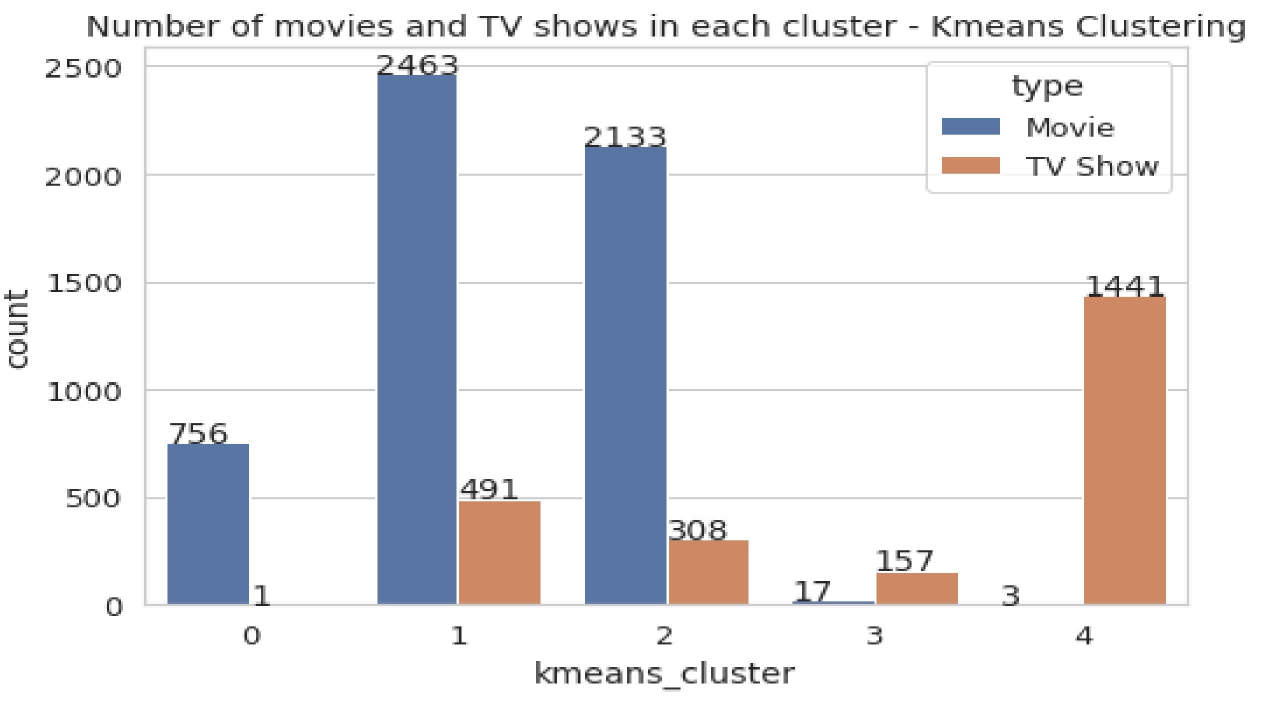


The sum of squared distance between each point and the centroid in a cluster decreases with the increase in the number of clusters.

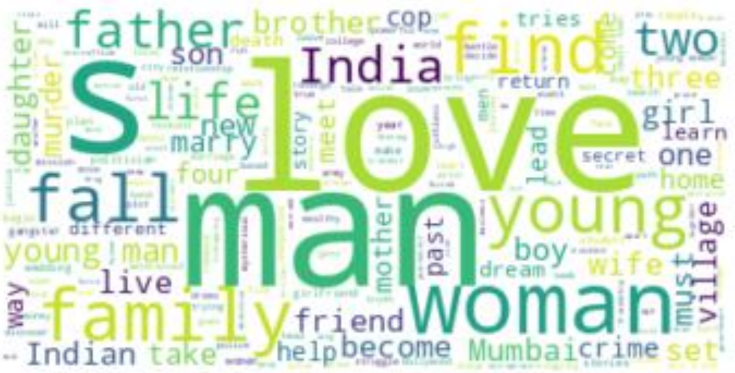
The highest Silhouette score is obtained for 5 clusters.

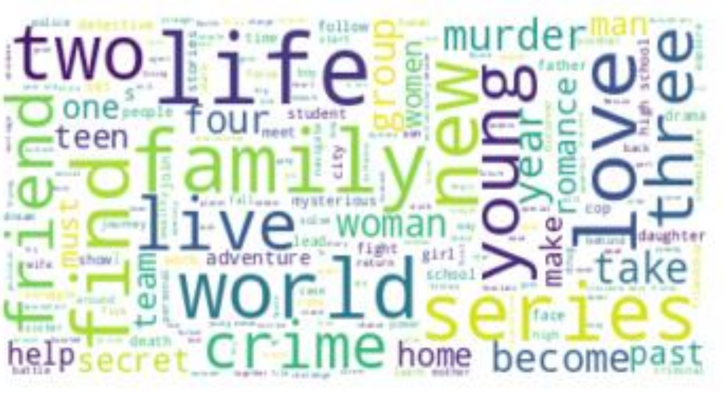


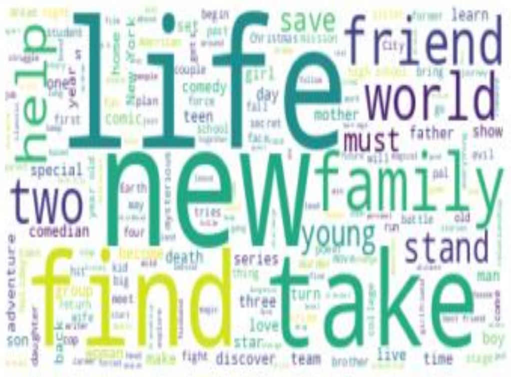
Successfully built 5 clusters using the k-means clustering algorithm.

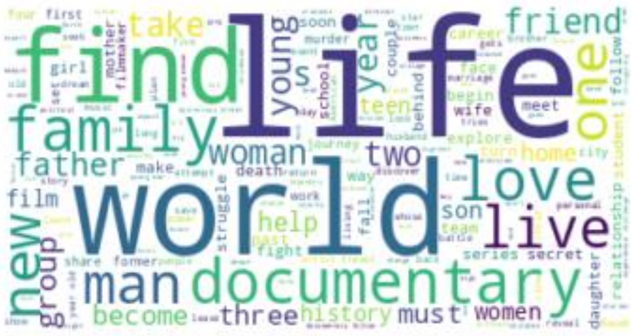


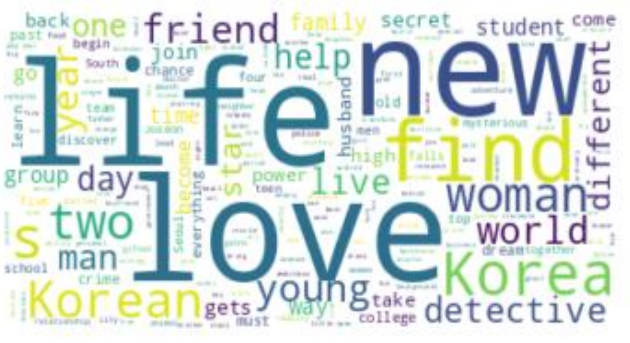
# Description



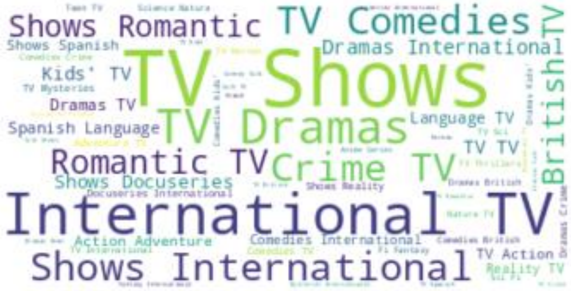


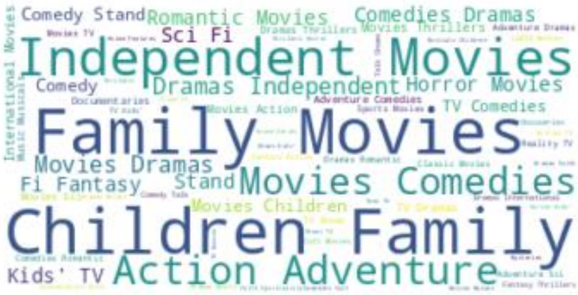


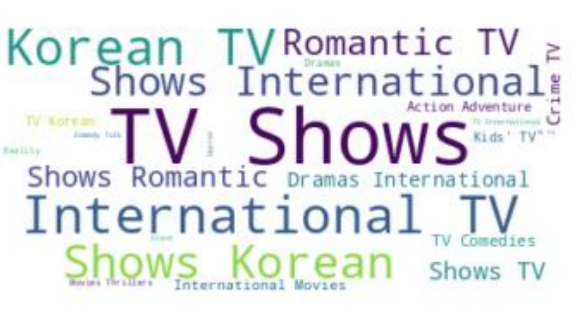


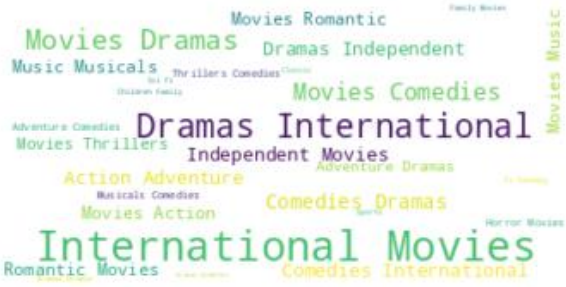


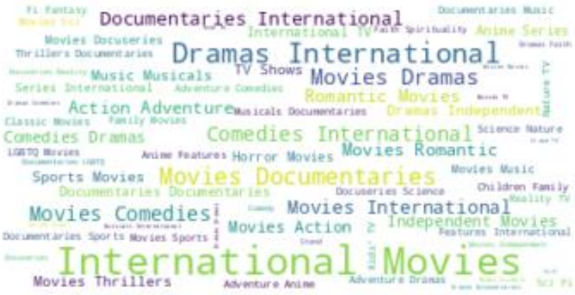
# Listed_in


Shows Romantic TV Comedies TV
Shows Spanish
Dramas International
Kids' TV
TV Shows
TV Mysteries
Dramas TV
British
Spanish Language TV Dramas Language TV
TV TV
Romantic TV Crime TV
Shows Docuseries
Docuseries International
International TV
Action Adventure
Comedies International
Shows International
TV Action
Reality TV
Dramas Crime

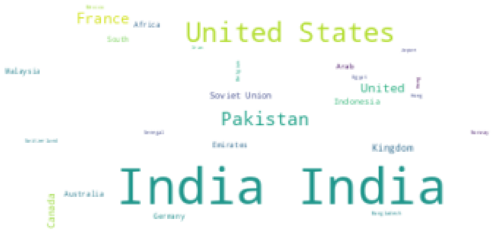

Comedy Stand Romantic Movies Comedies Dramas
Sci Fi
Independent Movies
Comedy
Dramas Independent Horror Movies
International Movies
Movies Action
TV Comedies
Family Movies
Movies Dramas
Fi Fantasy Stand Movies Comedies
Movies Children
Children Family
Kids' TV Action Adventure


Korean TV Romantic TV
Shows International
Action Adventure
Crime TV
TV Korean
TV Shows
Kids' TV
Shows Romantic Dramas International
International TV
TV Comedies
Shows Korean Shows TV
International Movies


Movies Romantic
Movies Dramas
Dramas Independent
Music Musicals
Thrillers Comedies
Movies Music
Movies Comedies
Adventure Comedies
Dramas International
Movies Thrillers
Independent Movies
Adventure Dramas
Action Adventure
Musicals Comedies
Comedies Dramas
Movies Action
Horror Movies
International Movies
Romantic Movies
Comedies International


Fi Fantasy
Documentaries International
International TV
Anime Series
Movies Docuseries
Dramas International
Thrillers Documentaries
Music Musicals
TV Shows
Movies Dramas
Series International
Romantic Movies
Action Adventure
Musicals Documentaries
Dramas Independent
Nature TV
Classic Movies
Family Movies
Comedies Dramas
Comedies International
Science Nature
LGBTQ Movies
Anime Features
Horror Movies
Movies Romantic
Sports Movies
Movies Documentaries
Movies Music
Documentaries Documentaries
Docuseries Science
Children Family
Movies Comedies
Movies International
Reality TV
Movies Action
Independent Movies
Features International
Documentaries Sports
Movies Sports
International Movies
Movies Thrillers
Adventure Anime
Adventure Dramas
Sci Fi

# Countries


France
Africa
United States
South
Malaysia
Arab
United
Soviet Union
Indonesia
Pakistan
Emirates
Kingdom
Australia
India India
Canada
Germany

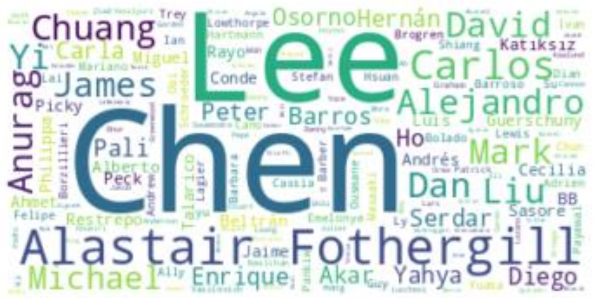
Korea South
China
Singapore
Kong
Malaysia
South Korea
Hong
United States

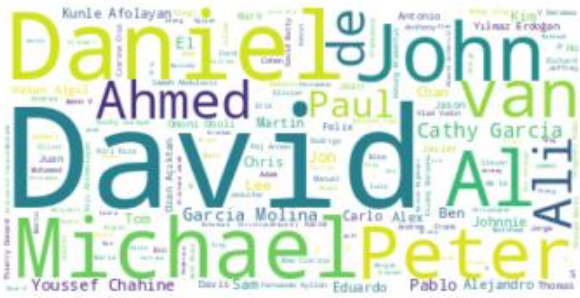
Qatar
Turkey Turkey
Australia
Kong China
South Korea
Taiwan
Singapore
Philippines
France
Thailand
Brazil
United Arab
Colombia
Belgium
United Kingdom
India India
Japan
Nigeria
Indonesia Indonesia
France Belgium
Japan Japan
China
Egypt
Hong Kong
Switzerland
South Africa
Turkey
Spain
Canada
Peru
Israel
Lebanon
Norway
United States
Argentina
States United
Indonesia
Arab Emirates
Malaysia
Denmark
Germany
Netherlands
Ireland
Italy
Mexico
Poland
Uruguay
New Zealand
India

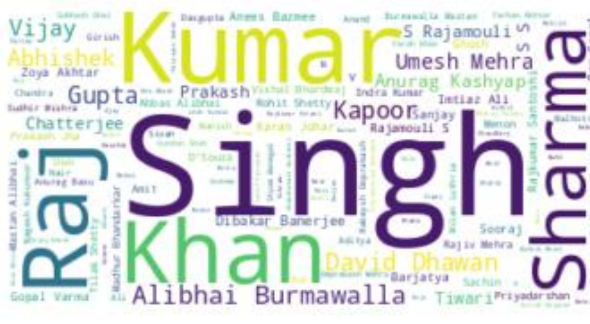
Mexico
France
Italy
Belgium
India
Australia
United Kingdom
Germany
New Zealand
Denmark
Ireland
States United
South Korea
States Canada
China
United States
Kingdom United
Canada United
Japan

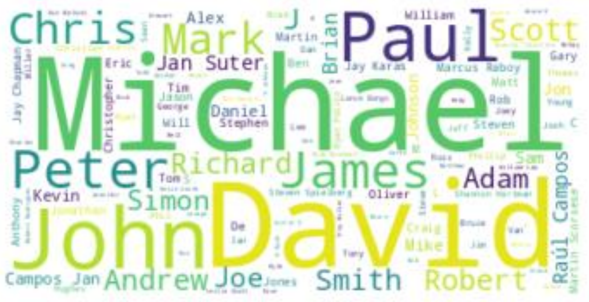
Germany
Japan
Nigeria
United States
Taiwan
India
South Africa
Kingdom United
Thailand
Spain
Colombia
China
Singapore
States United
Argentina
Lebanon
Brazil
South Korea
Turkey
France
Australia
United Kingdom
Mexico
Canada
Italy

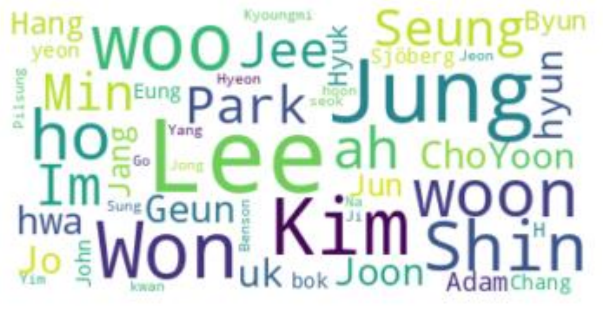
# Director


Chen
Lee
Alastair
Fothergill
Chuang
Carlos
Alejandro
David
Osorno
Hernán
James
Anurag
Yi
Carla
Miguel
Conde
Rayo
Peter
Barros
Mark
Dan
Liu
Serdar
Ho
Pali
Peck
Picky
Michael
Enrique
Akar
Yahya
Diego
Andrés
Cecilia
Sasore
Restrepo
Beltran
Jaime
Trey
Katıksız
Luis
Guerschuny
Alberto
Ahmet
Felipe
Hsuan
Stefan
Lowthorpe
Hartmann
Brogren
Shiang
Barroso
Talarico
Barber
Ly
BB
Adrian
Lewis
Cassia
Philippe
Borzillieri
Ally
Guy
Yuma


David
Daniel
John
Michael
Peter
Ahmed
van
Ali
Paul
de
Cathy Garcia
Garcia Molina
Kunle Afolayan
Youssef Chahine
Chris
Jon
Lee
Martin
Felix
Tom
Carlo
Alex
Ben
Johnnie
Antonio
Yılmaz Erdoğan
Kim
Jason
Chan
El
Pablo
Alejandro
Eduardo
Sam
Thomas
Omoni Oboli


Singh
Kumar
Sharma
Raj
Khan
Gupta
Kapoor
Vijay
Abhishek
Zoya Akhtar
Prakash
Umesh Mehra
S Rajamouli
Anurag Kashyap
Imtiaz Ali
Chatterjee
Rajamouli S
Sanjay
Dibakar Banerjee
David Dhawan
Alibhai
Burmawalla
Tiwari
Priyadarshan
Barjatya
Rajiv Mehra
Sooraj
Gopal Varma
Abbas Alibhai
Rohit Shetty
Karan Johar
Indra Kumar
Prakash Jha
Anees Bazmee
Amit


Michael
David
John
Paul
Peter
James
Chris
Mark
Richard
Simon
Adam
Scott
Brian
Jan Suter
Alex
Daniel
Stephen
Kevin
Andrew
Joe
Smith
Robert
Campos
Jan
Raúl Campos
Christopher
Jay Chapman
Martin Scorsese
William
Marcus Raboy
Jay Karas
Gary
Jon
Steven
Sam
Mike
Craig
Oliver
Tom
Tim
Jason
George
Will
Eric
Johnson
Jones
Martin
Anthony
Jonathan
Rob
Matt
Joey


Jung
Lee
Kim
Won
Shin
woo
woon
Jee
Seung
Park
Min
ho
ah
Cho
Yoon
Im
Geun
hwa
Jo
uk
bok
Joon
Adam
Chang
hyun
Hyuk
Jun
Jang
Eung
Hyeon
Hang
yeon
Byun
Sjöberg
Jeon
Kyoungmi
Pilsung
John
Benson
Yang
Jong
Go
Sung
Yim
kwan
Na
Ji
H
hoon
seok

# Title

Hai
Love
Pyaar
Aur
Ke
Mumbai
Singh
Ki
Dil
Ek
Baltiboy
Main
Krish
Trish
Life
Version
Baahubali
One
Na
Zindagi
Deewana
Story
Saath
Great
Hum
Beginning
Lucky
Raja
Tamil
Jatt
Kabhi
Golmaal
Time
Right
Jaan

Love
La
El
World
de
Secret
Show
Life
Girl
Serie
Stories
Day
Man
Family
Crime
Wild
Adventure
Murder
Good
Boy
Game
Time
Planet
Story
King
Drug
Blood
Black
House
Tale
Call
Queen
Big
Event
Mind
Earth
Killer
DreamWorks
Mexico
Legend
Criminal
Next
Club

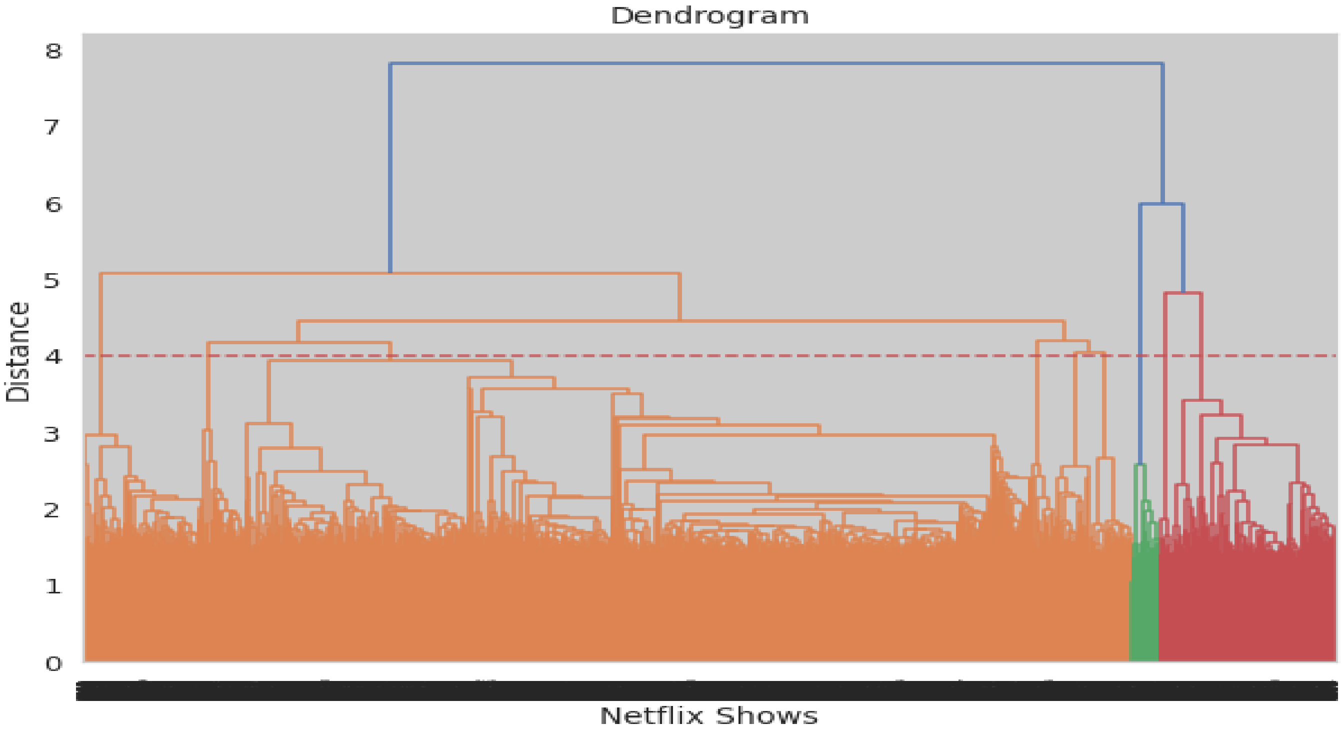
Christmas
Live
Power
Rangers
Time
Love
Last
Movie
Man
Monster
Special
Little
House
Girl
World
Boy
Day
Black
Holiday
Story
King
Tale
Good
Kill
American
Adventure
Night
True
Life
Secret
Legend
Mr
Halloween
Barbie
Baby
Super
First
Friend
City
Death
Brother
LEGO
Rescue
Bad
America
de

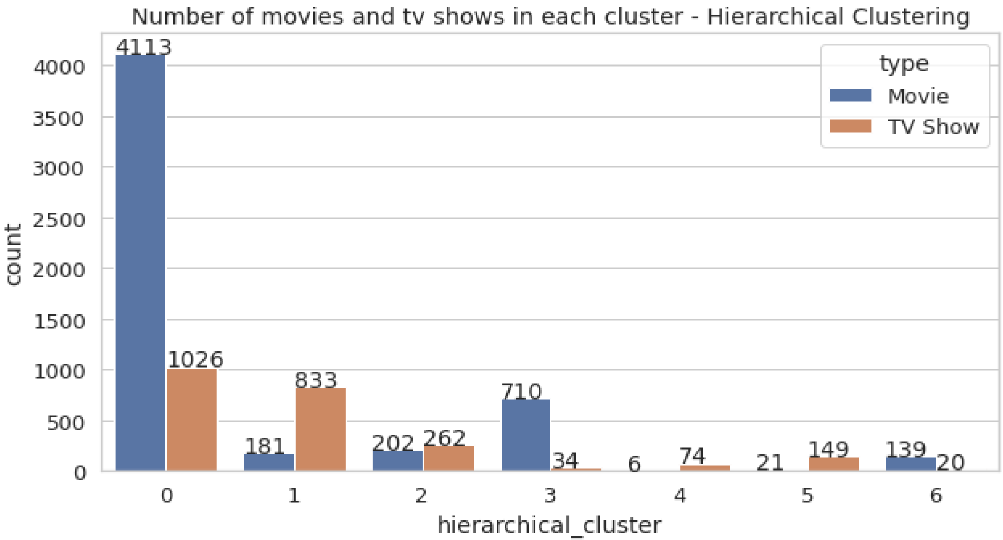
Love
Story
World
Movie
One
Life
U
Man
Girl
Secret
First
Last
El
NOVA
La
Game
Legend
Family
Film
Island
Day
City
American
Tale
War
Time
Sky
Dream
Kid
Fire
Son
de
Night
Heart
House
Boy
Black
Ghost
School

Love
Man
Heart
Rain
Reply
Friends
First
Girl
Live
Sol
Age
Little
Years
Miss
POP
Days
Youth
Wife
Bad
Lover
High
City
Guys
Ghost
Okay
Oh
Eat
Life
Memories
Night
One
Titipo
Miniforce
Secret
Romance
Prison
Time
Let
Mr

# Hierarchical clustering

At a distance of 4 units, 7 clusters can be built using the agglomerative clustering algorithm.





Successfully built 7 clusters using the Agglomerative (hierarchical) clustering algorithm.

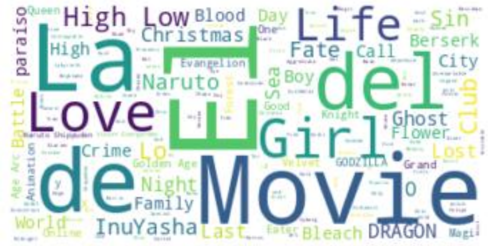
# Title

Time
World
Story
Man
Love
Movie
Girl
Little
Black
Tale
Live
Day
Christmas
Boy
Secret
War
Kid
American
Last
Life
One

Mission
Good
Crowds
Game
Life
Son
Station
Hajwala
Omar
Gospel
Black
Bolbol
Story
Aerials
Boushkash
Salma
Adam
Old
Egyptian
Right
El
Beirut
Player
Blessed
Alexandria
Land
Farm
Bye
Grandmother

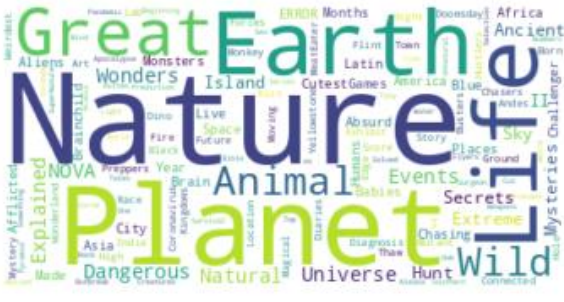
Ek
Baltiboy
Life
Time
Love
Ki
Zindagi
Main
Naa
Hain
Ke
Version
Krish
Trish
Singh
Aur
Hai
Dil
Story
One
Jatt
Girl
Mumbai
Pyaar
Lucky
Na
Deewana
India
Se

Murder
One
Family
Man
Killer
Dragon
Day
Love
Next
Life
Time
Marvel
Women
World
Girl
Game
Dark
Stories
Crime
Serie
Happy
King



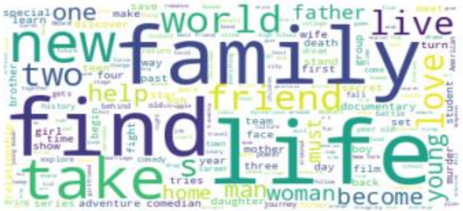
High
Low
Blood
Life
Sin
Christmas
Berserk
La
El
del
Fate
Call
City
Naruto
Evangelion
Boy
Love
Girl
Ghost
Flower
Club
de
Movie
Crime
Night
Family
InuYasha
Last
Bleach
DRAGON

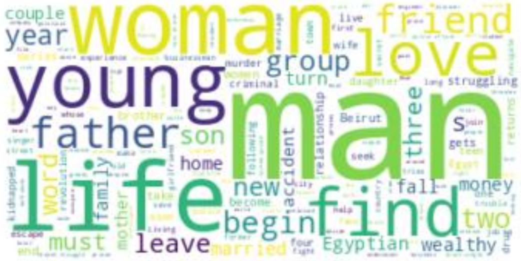
Rain
First
Time
Romance
High
Bad
Memories
Guys
Okay
Love
Oh
Age
Sol
Life
Days
Live
Eat
Years
Miss
Man
City
Island
Friends
Reply
Secret
Girl
Ghost
Wife
POP

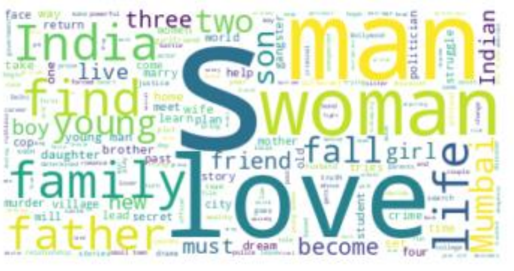


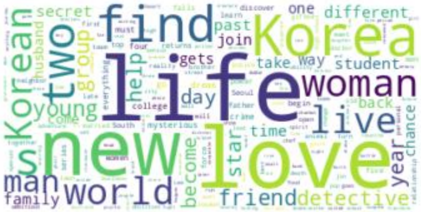
Great
Earth
Africa
Ancient
Monsters
Wonders
Island
Nature
Life
Animal
Events
Secrets
Explained
Planet
Dangerous
Natural
Universe
Hunt
Wild

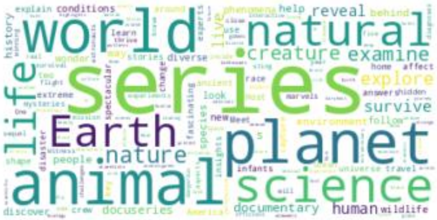
# Description

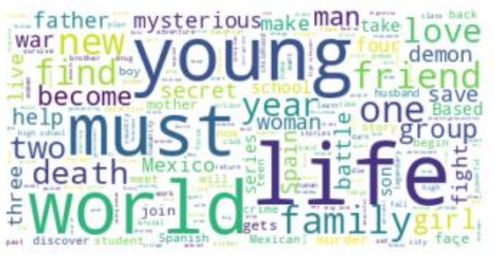


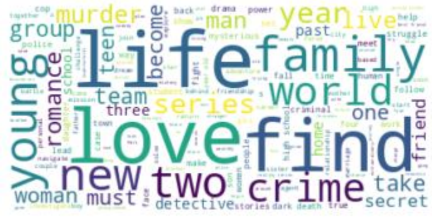








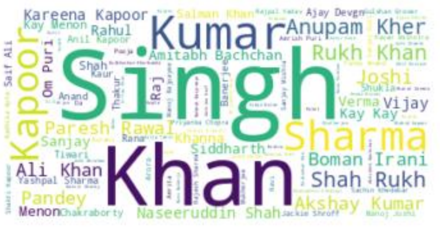




# Cast

cast
David Attenborough
Scott
Chris
Johnson
Robert
Sanders
Thompson
Clotworthy
Packham
Nasser
Ekulona
Mitch
Levin
Seidman
Zainab
Alie
Freida
Izquierdo
Bonnin
Ward
Liz
Gupta
Erik
Olivia
Lisa
Goritsas
Brisbane
Obeid
Silva
Felipe
Sahana
Cyril
Pinto
Chauquet
Bob
Steven
Sammy
Michael
Ryan
McGavin
Colman
Corey
Latif
Afi
Jason
Srinivasan
Demetri
Woody
Kelly
George

El
Al
Mohamed
Mohammed
Ahmad
Ahmed
Abdullah
Mahmoud
Khaled
Youssef
Abdel
Ali
Hassan
Ibrahim
Hussein
Saad
Zaki
Aziz
Salah
Lotfy
Omar
Saeed
Sherif
Dina
Hassan Hosny
Bayoumi Fouad
Yousra
Adel


Singh
Khan
Kumar
Sharma
Kapoor
Kareena Kapoor
Anupam Kher
Rukh Khan
Rahul
Amitabh Bachchan
Joshi
Vijay
Kay Kay
Paresh Rawal
Siddharth
Boman Irani
Shah Rukh
Ali Khan
Pandey
Akshay Kumar
Naseeruddin Shah
Menon
Chakraborty
Om Puri
Saif Ali
Anand
Thakur
Tiwari
Sanjay
Raj
Verma

Kim
Lee
Park
Jung
Jin
Min
hyun
young
Ji
Eun
Yoon
Jae
won
Choi
Shin
Soo
Kang
hee
Sung
Seo
Dong
Han
Joo
woo
Jeong
ho
Hye
seok
Seung
Jun
Moon
Sun
yeon
hwan
Song
Nam
Cho
Hong
Ye
yeong

cast
Michael
John
James
David
Paul
Richard
Peter
Robert
Tom
Lee
William
Mark
Adam
Scott
Kevin
Chris
Daniel
Martin
Andrew
Alex
Sarah
Kelly
Brian
Smith
Ryan
Jason
Taylor
Anthony
Jack
Christopher
Sam
Evan
Steve
Joe
Jones
Stephen
Ben
Aaron
Wilson

cast
Wang
Chen
David
Lee
De
Lin
Li
Yu
Yang
Michael
Zhang
John
James
Paul
Chang
Wu
Liu
Huang
Tom
Wei
Patrick
William
Taylor
Robert
Richard
Jonathan
Chris
Thomas
Smith
Jason
Ben
Peter
Scott

Luis
Carlo
María
Antonio
Daniel
Ana
Jorge
Juan
Juan Pablo
Alejandro
Diego
José
Pedro
Fernando
Gómez
Roberto
López
Sánchez
de la
de
Andrés
Javier
Yuki Kaji
Daisuke Ono
Castro
Kana Hanazawa
Eduardo
Takahiro Sakurai
Junichi Suwabe
Verónica
David
Ai Kayano

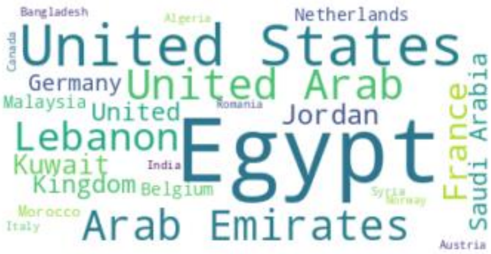
# Country

Australia
Spain
United Kingdom
States India
Turkey
France
India United
Brazil
Hong Kong
Nigeria
United States
States Canada
Canada United
Kingdom United
Philippines
Germany
States United
Japan

Singapore
China
Malaysia
Kong
Korea South
United States
Hong
France
South Korea
Kingdom United
Canada
Brazil
United States
Australia
Republic
States United
United Kingdom
Uruguay

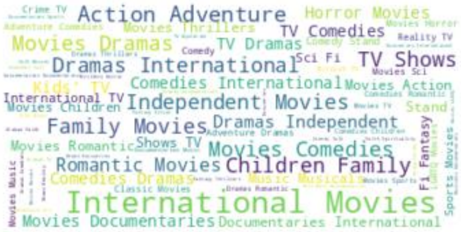
Pakistan
Germany
India India
Canada
Australia
United States
United Kingdom
France

Mexico
Chile
Argentina
Uruguay
Spain
Brazil
France
Colombia
Japan Japan
United States

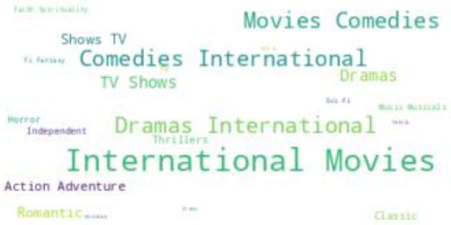

Bangladesh
Algeria
Netherlands
United States
Canada
Germany
United Arab
Malaysia
United
Romania
Jordan
Lebanon
Egypt
France
Saudi Arabia
Kuwait
India
Kingdom
Belgium
Syria
Norway
Morocco
Italy
Arab Emirates
Austria

United Kingdom
Taiwan
Norway
Spain
Hong Kong
India
Russia
Japan
Argentina
United States
Czech Republic
Thailand
Sweden
Canada
Germany
Brazil
Italy
Australia
Poland
France
Turkey
Belgium
New Zealand
Singapore
China
States United
Denmark
Israel
South Africa

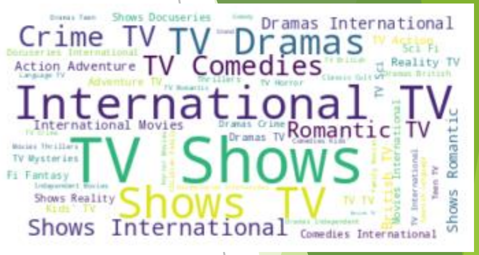
# Listed_in



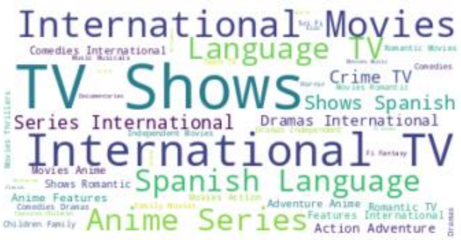
Action Adventure
Horror Movies
Movies Thrillers
TV Comedies
Movies Dramas
TV Dramas
Dramas International
Sci Fi
TV Shows
Comedies International
Movies Action
International TV
Independent Movies
Movies Children
Kids' TV
Stand
Family Movies
Dramas Independent
Movies Romantic
Shows TV
Movies Comedies
Romantic Movies
Children Family
Comedies Dramas
Music Musicals
International Movies
Movies Documentaries
Documentaries International
Fi Fantasy
Sports Movies



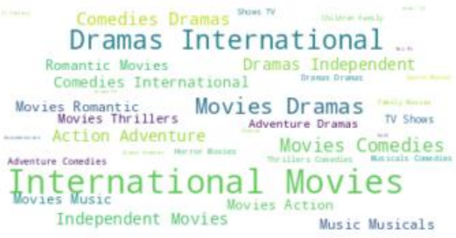
Movies Comedies
Shows TV
Comedies International
TV Shows
Dramas
Dramas International
Thrillers
Horror
Independent
International Movies
Action Adventure
Romantic
Classic



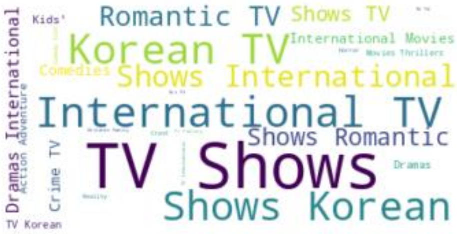
Shows Docuseries
Dramas International
Crime TV
TV Dramas
Action Adventure
TV Comedies
Reality TV
Sci Fi
International TV
International Movies
Romantic TV
TV Mysteries
TV Shows
Fi Fantasy
Shows Reality
Shows TV
Shows International
Comedies International
Shows Romantic



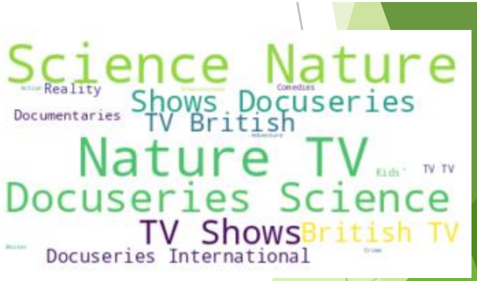
International Movies
Comedies International
Language TV
TV Shows
Crime TV
Shows Spanish
Series International
Dramas International
International TV
Movies Anime
Spanish Language
Shows Romantic
Anime Features
Anime Series
Adventure Anime
Romantic TV
Features International
Action Adventure
Children Family
Movies Thrillers



Comedies Dramas
Shows TV
Dramas International
Romantic Movies
Dramas Independent
Comedies International
Movies Romantic
Movies Thrillers
Movies Dramas
Adventure Dramas
TV Shows
Action Adventure
Movies Comedies
Adventure Comedies
Horror Movies
International Movies
Movies Music
Movies Action
Independent Movies
Music Musicals



Kids'
Romantic TV
Shows TV
Korean TV
International Movies
Comedies
Shows International
Dramas International
Action Adventure
International TV
Shows Romantic
Crime TV
TV Shows
Dramas
Shows Korean
TV Korean



Science Nature
Reality
Shows Docuseries
Documentaries
TV British
Nature TV
Kids'
TV TV
Docuseries Science
TV Shows
British TV
Docuseries International

# CONCLUSION

- Data set contains 7787 rows and 12 columns in that cast and director features contains large number of missing values so we can drop it.
- We have two types of content TV shows and Movies (30.9% contains TV shows and 69.1% contains Movies).
- Most films were released in the years 2018, 2019, and 2020 and united states have the maximum content on Netflix.
- The months of October, November, December and January had the largest number of films and Tv-shows released.
- The USA, India, the United Kingdom, Canada, and Egypt are the top five producer countries.
- For the clustering algorithm, we utilized type, director, nation, released year, genre, and year.
- LDA and LSA has sorted much more similar titles in a group of genre.
- Applied different clustering models Kmeans, hierarchical, Agglomerative clustering on data we got the best cluster arrangements.
- In Affinity Propagation, we had 13 clusters and a Silhouette Coefficient score of 0.244.
- We cut vertical lines with a horizontal line to obtain the number of clusters in Agglomerative Clustering. There were four clusters, with an average silhouette score of 0.17296314851287742.
- The final model we used was k-means clustering, which consisted of 2,3,4,5,6 clusters. 4 numbers of clusters gives us good fitting.
- After applying K - means optimal value of number of clusters is 5
- Silhouette score for a set of sample data points is used to measure how dense and well-separated the clusters are.

THANK

YOU!!