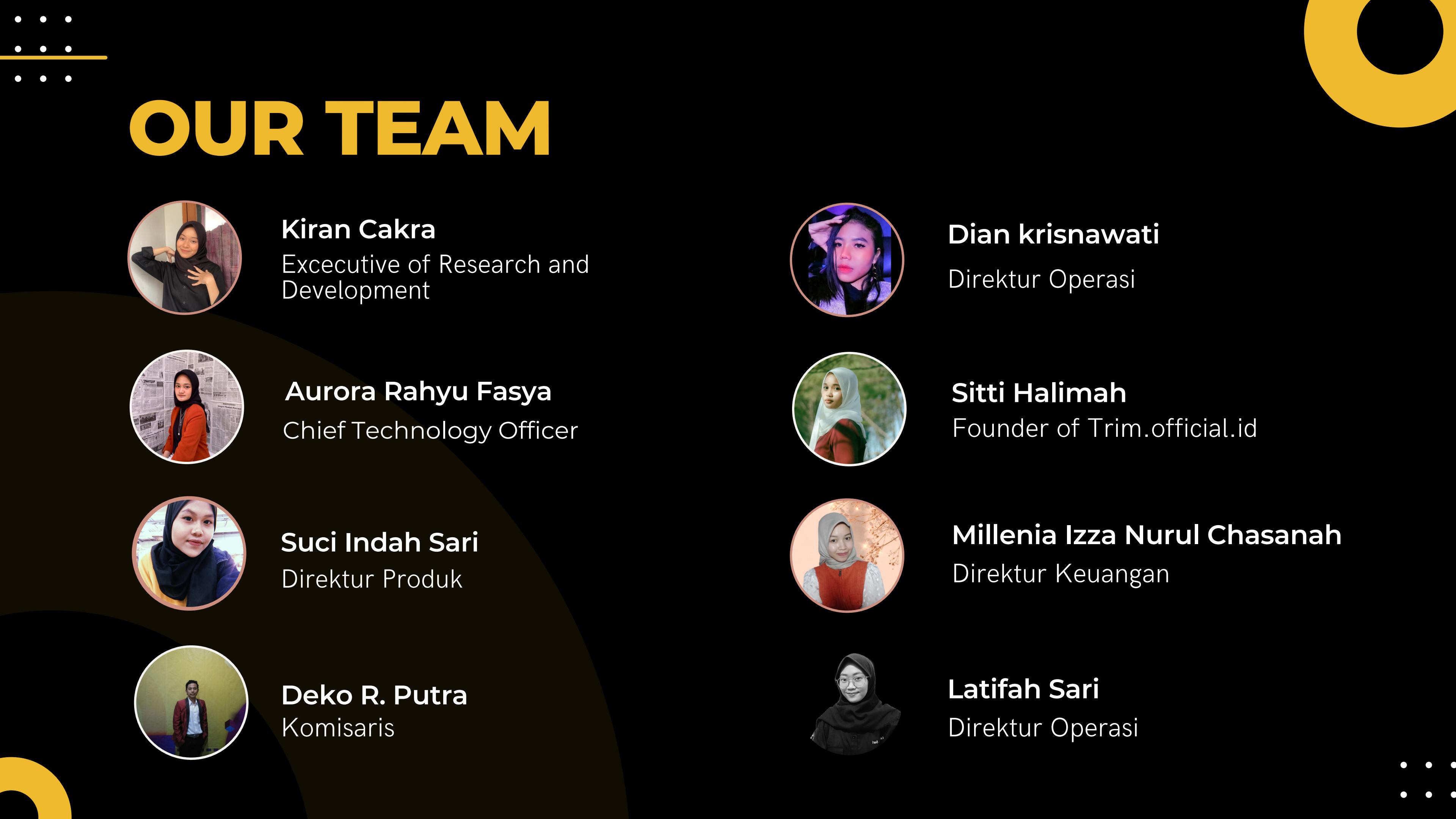


CONTENT ANALYST WITH EDA

NETFLIX





OUR TEAM



Kiran Cakra
Executive of Research and Development



Dian Krisnawati
Direktur Operasi



Aurora Rahyu Fasya
Chief Technology Officer



Sitti Halimah
Founder of Trim.official.id



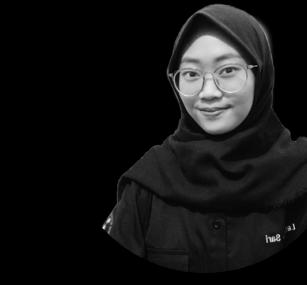
Suci Indah Sari
Direktur Produk



Millenia Izza Nurul Chasanah
Direktur Keuangan



Deko R. Putra
Komisaris



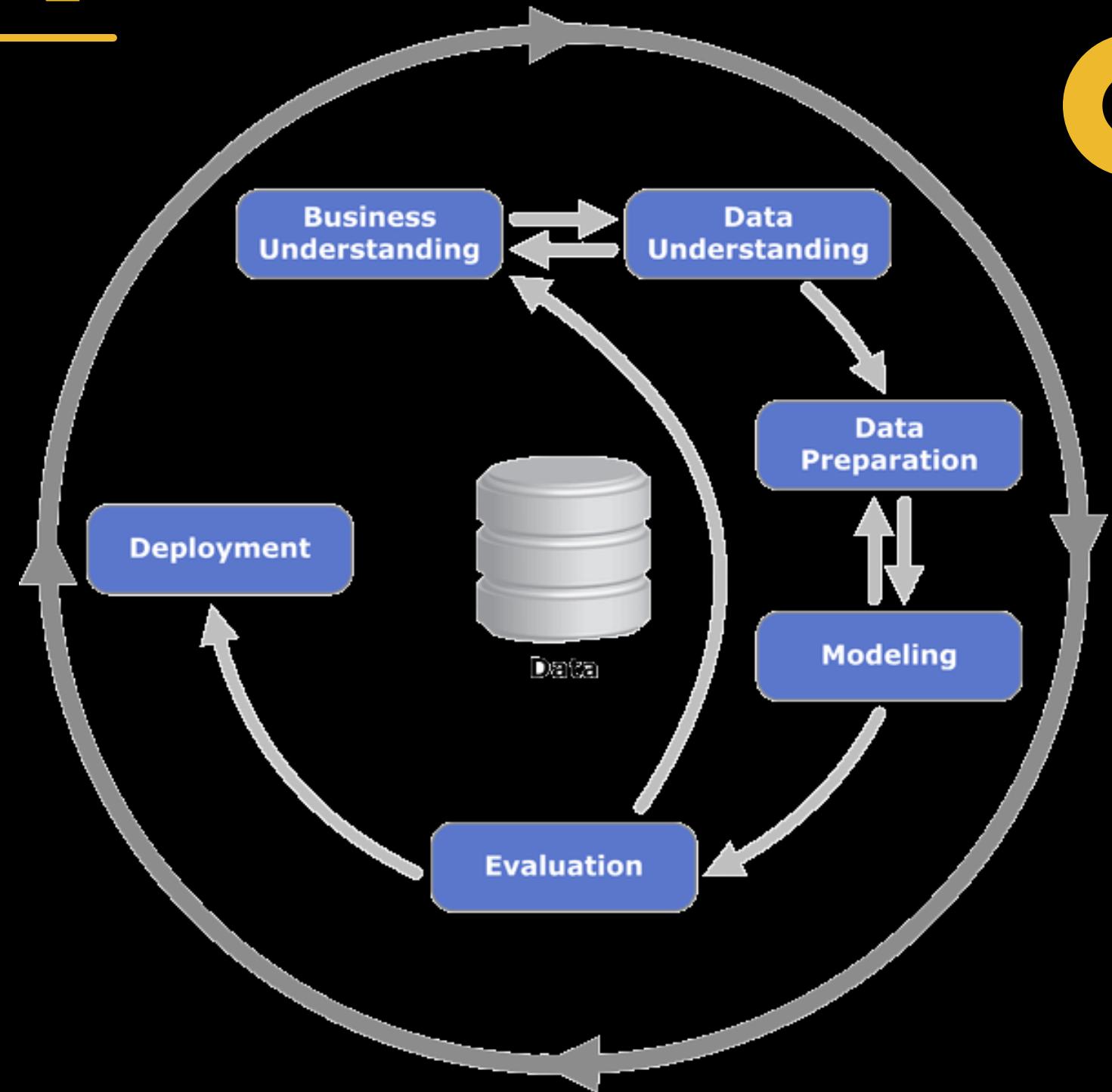
Latifah Sari
Direktur Operasi

INTRODUCTION

- Project Mini Clustering -

Netflix is an application that keeps growing bigger and faster with its popularity, shows and content. This is an EDA or a story telling through its data along with a content-based recommendation system and a wide range of different graphs and visuals.

This activity is carried out to find out the grouping of content that Netflix, TV Show or Movie users are interested in using the Cross-Industry Standard Process for Data Mining or CRISP-DM.



BUSINESS UNDERSTANDING



Netflix is a digital streaming media service provider that offers a variety of entertainment content through devices connected to the Internet.



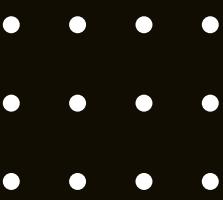
In this project, our team will conduct a Content Analyst on Netflix using Exploratory Data Analysis (EDA).

- • •
- • •
- • • •
- • • •



EDA is an analytical method that is used to obtain data quickly using graphs as a form of data visualization, without using statistical models, or formulating hypotheses.

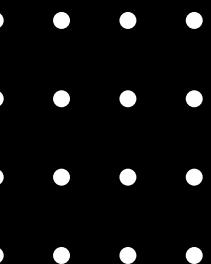
• • •
• • •
• • •
• • •



TUJUAN

Kegiatan Content Analyst dengan menggunakan EDA ini bertujuan untuk mengetahui kualitas tayangan dan mengembangkan tayangan yang diminati di aplikasi Netflix. dengan memaksimalkan wawasan analis ke dalam kumpulan data dan ke dalam struktur yang mendasari kumpulan data, sambil menyediakan semua item spesifik yang analis ingin ekstrak dari kumpulan data.

Hasil dari penelitian ini kemudian akan dianalisis untuk di rekomendasikan kepada user dan menjadi bahan pertimbangan user dalam memilih tayangan.



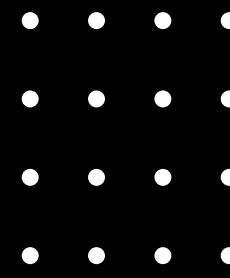
MANFAAT

Manfaat dari kegiatan Content Analyst dengan menggunakan EDA digarapkan dapat meningkatkan jumlah pelanggan yg berlangganan Netflix.

Dengan mengetahui tayangan apa saja yang paling dinikmati dan untuk dapat memberikan gambaran mengenai informasi, pandangan yang berhubungan dengan layanan streaming yang sedang berkembang khususnya video streaming.

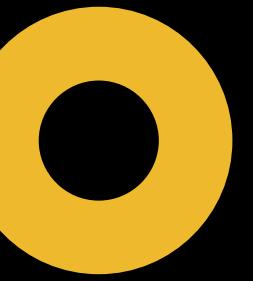
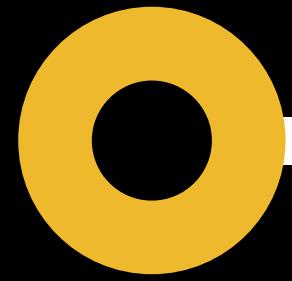


• • •
• • •
• • •



WORKFLOW

Cross-Industry Standard Process for
Data Mining atau CRISP-DM



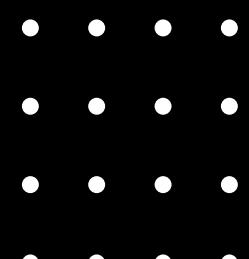
**DATA
UNDERSTANDING**

**DATA
PREPARATION**

MODELLING

EVALUATION

DEPLOYMENT



DATA UNDERSTANDING

Show Id = Season

Type = Tipe Content

Title Director = Sutradara

Cast = Pemeran

Country = Negara Asal

Date Added = Waktu ditambahkan

Release Year = Tahun Rilis

Rating = Pengelompokan Usia

Duration = Durasi

Listed In = Genre

Description = Deskripsi

DATA UNDERSTANDING

• • • •

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Dramas, International Movies
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Horror Movies, International Movies
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min	Dramas
...
7782	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmark...	October 19, 2020	2005	TV-MA	99 min	Dramas, International Movies
7783	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals
7784	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	2019	TV-MA	44 min	Documentaries, International Movies, Music & M...
7785	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	2019	TV-PG	1 Season	International TV Shows, Reality TV
7786	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	2019	TV-MA	90 min	Documentaries, Music & Musicals

7787 rows × 12 columns

• • • •
• • • •

DATA PREPARATION



Data yang diperlukan dalam melakukan analisis ini yaitu Type, Country, Release Year dan Rating dalam Tahap *Data Preparation* karena secara umum *Analysis of Movies vs TV Shows* berfokus pada Type Content.

Sedangkan *Movie Ratings Analysis* berfokus pada Rating dan Type Content. *Highest Genre Analysis* content berfokus pada Genre dan Rating. Terakhir *Year Wise Analysis*, berfokus pada Rating dan Release Year dari Data yang sudah disiapkan.

DATA PREPARATION

Type = Tipe Content

Release Year = Tahun Rilis

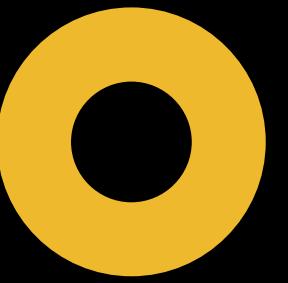
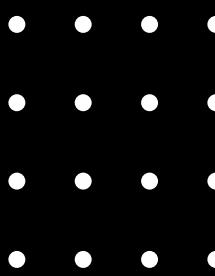
Rating = Pengelompokan Usia

Listed In = Genre

```
type(df[["type", "release_year", "rating", "listed_in"]])  
df[["type", "release_year", "rating", "listed_in"]]
```

	type	release_year	rating	listed_in
0	TV Show	2020	TV-MA	International TV Shows, TV Dramas, TV Sci-Fi &...
1	Movie	2016	TV-MA	Dramas, International Movies
2	Movie	2011	R	Horror Movies, International Movies
3	Movie	2009	PG-13	Action & Adventure, Independent Movies, Sci-Fi...
4	Movie	2008	PG-13	Dramas
...
7782	Movie	2005	TV-MA	Dramas, International Movies
7783	Movie	2015	TV-14	Dramas, International Movies, Music & Musicals
7784	Movie	2019	TV-MA	Documentaries, International Movies, Music & M...
7785	TV Show	2019	TV-PG	International TV Shows, Reality TV
7786	Movie	2019	TV-MA	Documentaries, Music & Musicals

7787 rows × 4 columns



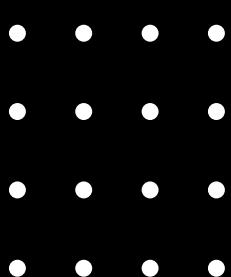
MODELING

o o o o

Kegiatan Analisis Konten untuk mengetahui
Analysis of Movies vs TV Shows, Movie Ratings Analysis, dan Year Wise Analysis,
menggunakan Algoritma TF-IDF



NETTEKIV



MODELING

TF-IDF(Term Frequency-Inverse Document Frequency (TF-IDF)) adalah frekuensi kemunculan kata dalam dokumen, dikurangi bobotnya berdasarkan jumlah dokumen tempat kemunculannya. Hal ini dilakukan untuk mengurangi pentingnya kata-kata yang sering muncul dalam tinjauan plot dan oleh karena itu, signifikansinya dalam menghitung skor kesamaan akhir.

```
✓ [1] from sklearn.feature_extraction.text import TfidfVectorizer

#removing stopwords
tfidf = TfidfVectorizer(stop_words='english')

#Replace NaN with an empty string
netflix_overall['description'] = netflix_overall['description'].fillna('')

#Construct the required TF-IDF matrix by fitting and transforming the data
tfidf_matrix = tfidf.fit_transform(netflix_overall['description'])

#output the shape of tfidf_matrix
tfidf_matrix.shape

[2] (7787, 17905)

✓ [17] # Import linear_kernel
from sklearn.metrics.pairwise import linear_kernel

# Compute the cosine similarity matrix
cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)

✓ [34] indices = pd.Series(netflix_overall.index, index=netflix_overall['title']).drop_duplicates()
```



o o o o

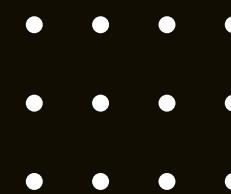
Bahasa pemrograman yang kami gunakan adalah Python
3.9 dengan Algoritma TF-IDF

```
✓ [27] import numpy as np # linear algebra
       import pandas as pd # data processing,
       import matplotlib.pyplot as plt
       import seaborn as sns
       import matplotlib.pyplot as plt

✓ [25] from sklearn.feature_extraction.text import TfidfVectorizer

✓ [26] # Import linear_kernel
       from sklearn.metrics.pairwise import linear_kernel
```

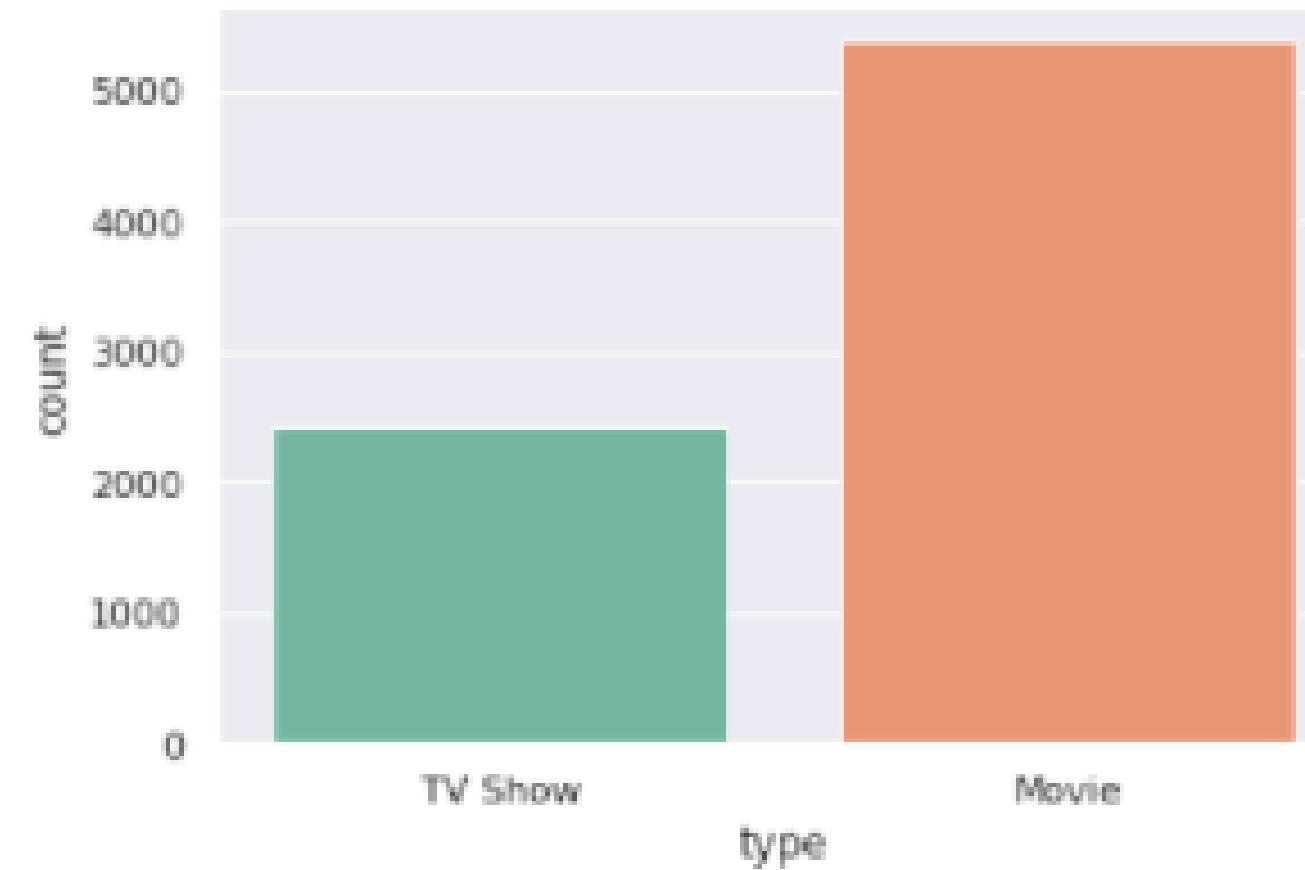
• • •
• • •
• • •
• • •

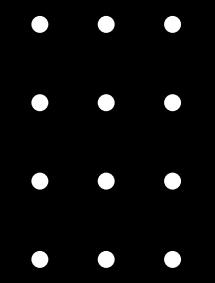


ANALYSIS OF MOVIES VS TV SHOWS

Analysis of Movies vs TV Shows.

```
[18] sns.set(style="darkgrid")
     ax = sns.countplot(x="type", data=netflix_overall, palette="Set2")
```

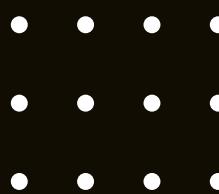




ANALYSIS OF MOVIES VS TV SHOWS

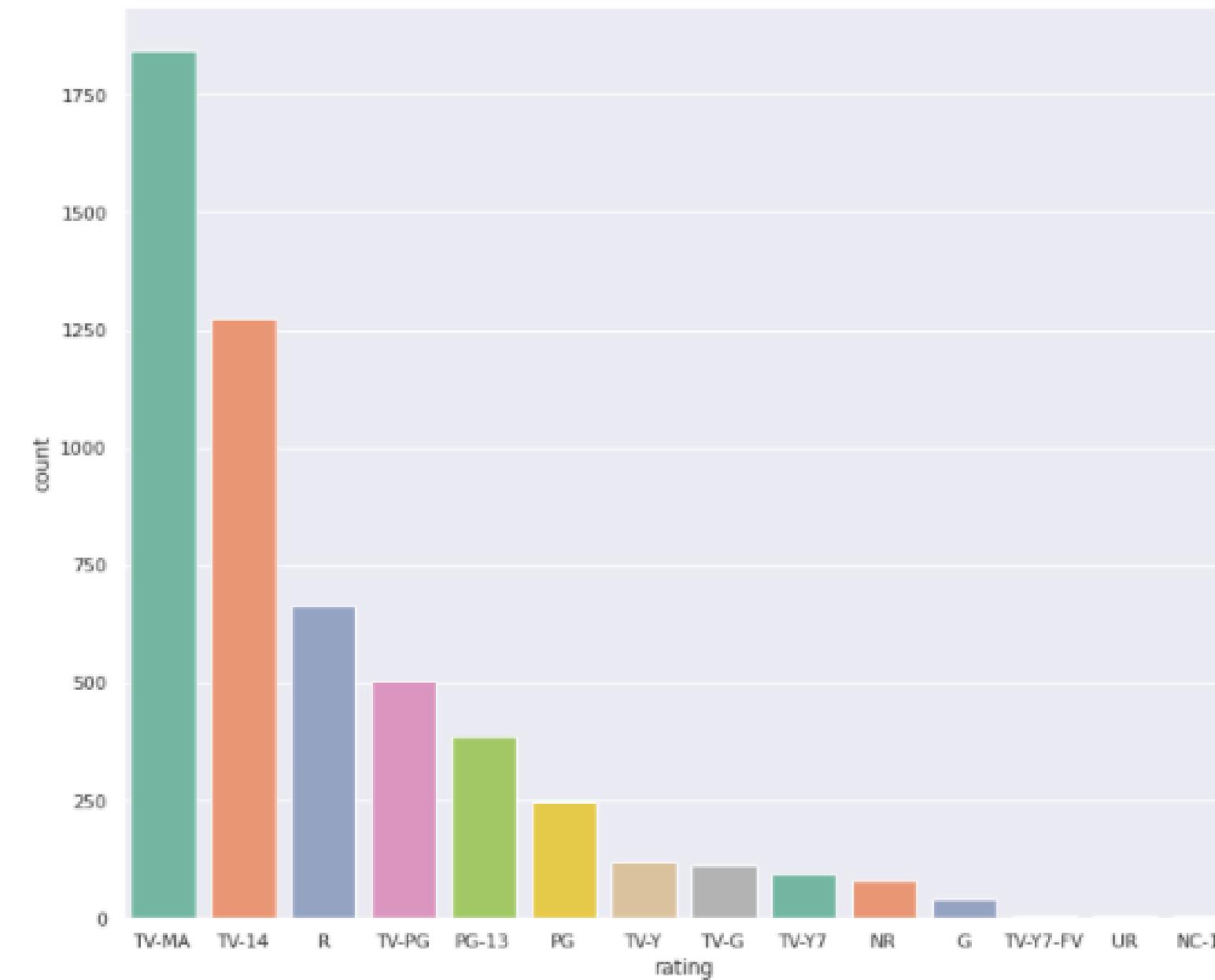
Berdasarkan proses analisis data yang sudah kita lakukan, secara umum dapat disimpulkan bahwa jumlah konten Movie dan TV Show di Netflix cenderung meningkat di setiap country nya. Terlihat dalam diagram analysis of Movie vs TV Show yaitu Netflix Movie countnya mencapai >5000 sedangkan Netflix TV Show hanya >2000 saja.





MOVIE RATINGS ANALYSIS

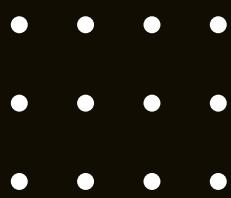
```
[19] plt.figure(figsize=(12,10))
sns.set(style="darkgrid")
ax = sns.countplot(x="rating", data=netflix_movies, palette="Set2", order=netflix_movies['rating'].value_counts().index[0:15])
```



MOVIE RATINGS ANALYSIS

Jumlah film terbesar dibuat dengan peringkat 'TV-MA'. "TV-MA" adalah peringkat yang diberikan oleh Pedoman Orang Tua untuk program televisi yang dirancang hanya untuk pemirsa dewasa, dan peringkat 'R' yang sangat populer. Film dengan peringkat R adalah film yang dinilai memiliki materi yang mungkin tidak cocok untuk anak-anak di bawah usia 17 tahun oleh Motion Picture Association of America; MPAA menulis "Di bawah 17 tahun membutuhkan pendamping orang tua atau wali dewasa".

NETFLIX



MOVIE RATINGS ANALYSIS & HIGHEST GENRE

```
from collections import Counter

genres=list(netflix_movies['listed_in'])
gen=[]

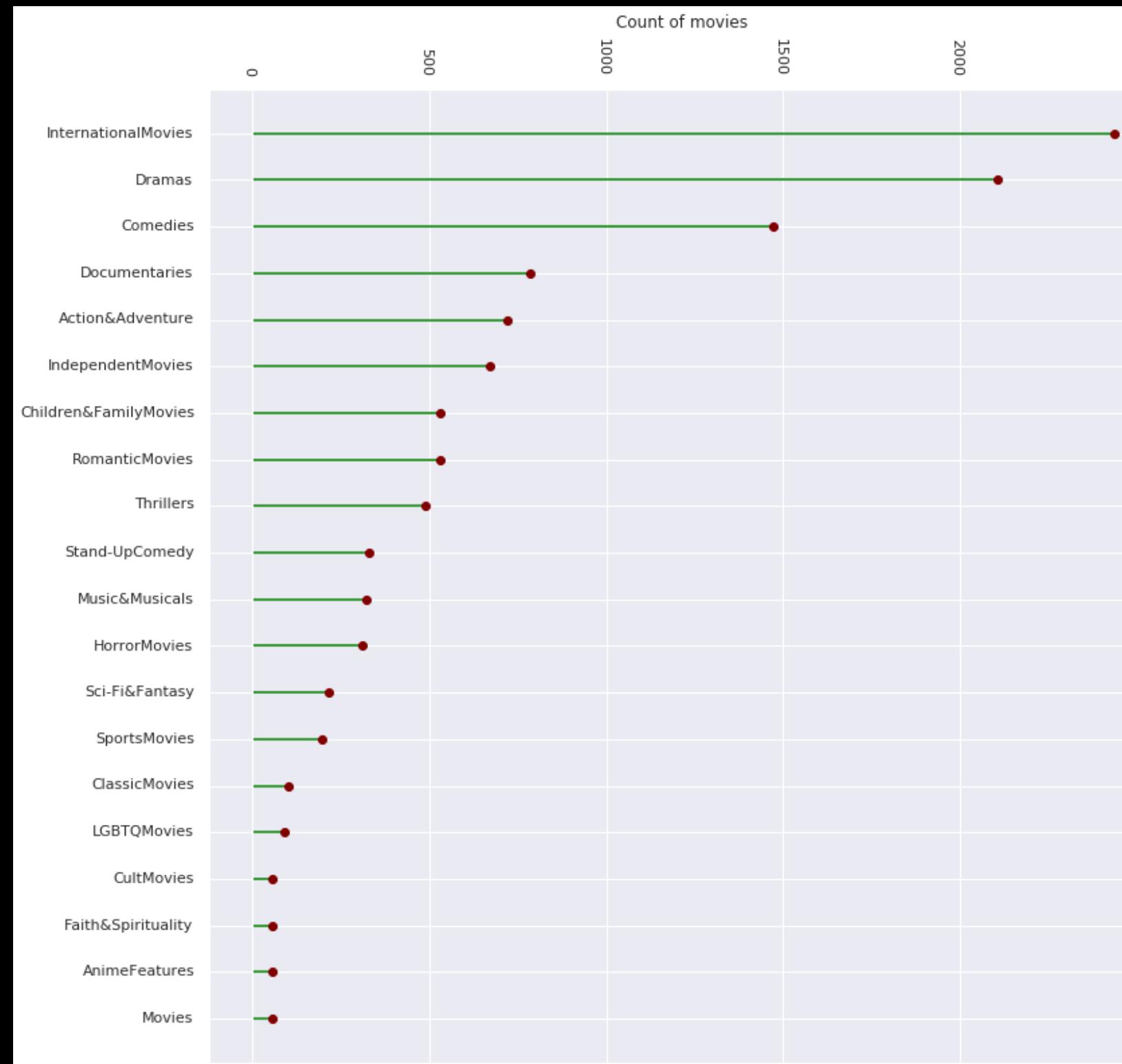
for i in genres:
    i=list(i.split(','))
    for j in i:
        gen.append(j.replace(' ', ""))
g=Counter(gen)

[34] g={k: v for k, v in sorted(g.items(), key=lambda item: item[1], reverse= True)}

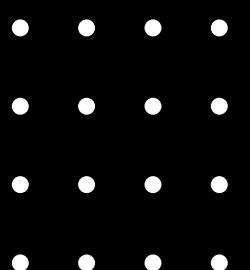
fig, ax = plt.subplots()

fig = plt.figure(figsize = (10, 10))
x=list(g.keys())
y=list(g.values())
ax.vlines(x, ymin=0, ymax=y, color='green')
ax.plot(x,y, "o", color='maroon')
ax.set_xticklabels(x, rotation = 90)
ax.set_ylabel("Count of movies")
# set a title
ax.set_title("Genres");
```

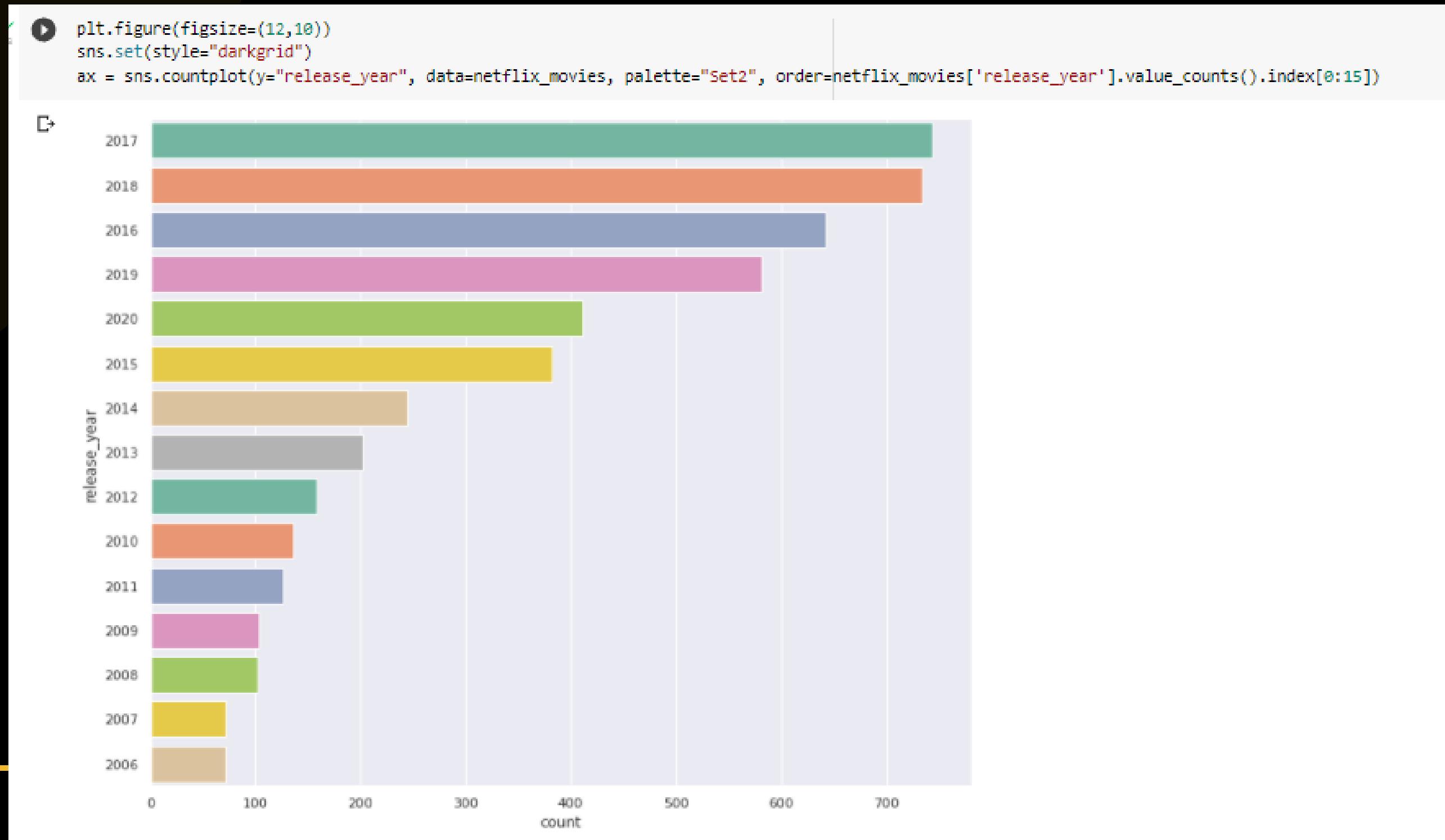
MOVIE RATINGS ANALYSIS & HIGHEST GENRE



International Movies, Comedy, dan Drama menjadi tiga Genre teratas yang memiliki jumlah konten tertinggi di Netflix



YEAR WISE ANALYSIS



YEAR WISE ANALYSIS

2017 adalah tahun ketika sebagian besar film dirilis

Pada tahun 2017, merilis sebuah laporan manarik yang menunjukkan bahwa jumlah film di Netflix meningkat sejak 2015, sementara jumlah acara TV Show di Netflix menurun pada tahun 2012.

Pada tahun 2017 adalah tahun ketika sebagian besar film pada aplikasi netflik dirilis, hingga mencapai sekitar 100 film yang di rilis, sehingga pada tahun 2017 lah menjadi year wise yang sempurna bagi perilisan film pada aplikasi netflik

REKOMENDASI HASIL EDA

Dasar pilihan Rekomendasi untuk Content Netflix dari tim kami adalah :

1. Rekomendasi harus berisikan Content TV Show dan Movies dengan Top Three Genre (International Movie, Comedy, dan Drama).
 2. Rekomendasi harus berisikan Content TV Show dan Movies dengan Top Three Ratings (TV-MA, TV-14, dan R).
 3. Rekomendasi harus berisikan Content TV Show dan Movies dengan periode rilis antara 2016-terbaru.
-

EVALUATION

Dalam kegiatan penelitian ini menghasilkan bahwa Movies lebih memiliki banyak peminat dibandingkan Content TV Shows. Maka dapat kita evaluasi untuk meningkatkan pengguna dalam menonton Content, Netflix harus membuat daftar rekomendasi yang sesuai dengan analisis lainnya, agar pengguna dapat meningkatkan daya minat didalam Content TV Shows.

DEPLOYMENT MOVIE



LIST GENRE RECOMENDATION

- List Rekomendasi Genre International Movies dengan Rating TV-MA.
- List Rekomendasi Genre Comedy - Drama dengan Rating TV-14 dan R.



GET OLD RECOMENDATION

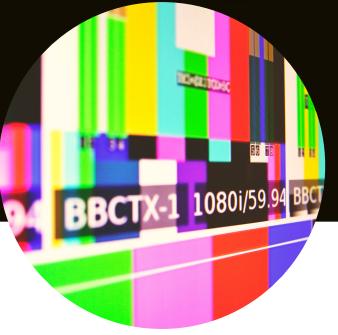
- List Rekomendasi Movies Lama, sebelum tahun 2015.
- List Rekomendasi Old Movie dengan Rating TV-MA, TV-14 dan R.



RATING RECOMENDATION

- List Rekomendasi Movies untuk Keluarga- Anak, Rating TV-MA & TV-14.
- List Rekomendasi Movie dengan Rating R untuk Dewasa.

DEPLOYMENT TV SHOW



LIST GENRE RECOMENDATION

- List Rekomendasi TV Show dengan Genre Comedy - Drama.
- List Rekomendasi Genre International Movies dengan Rating TV-MA dan R.



GET OLD RECOMENDATION

- List Rekomendasi TV Show sebelum tahun 2015.
- List Rekomendasi TV Show dengan Rating TV-MA, TV-14 dan R.



RATING RECOMENDATION

- List Rekomendasi TV Show untuk Keluarga- Anak, Rating TV-MA & TV-14.
- List Rekomendasi TV Show dengan Rating R untuk Dewasa.

CONLUSION

Berdasarkan pembuatan rekomendasi dari hasil analisis EDA, disimpulkan bahwa pembuatan rekomendasi mendapatkan hasil yang sesuai dengan tujuan

REFERENSI

o o o o

<https://www.kaggle.com/code/niharika41298/netflix-visualizations-recommendation-eda/notebook>
netflix-visualizations-recommendation-eda.ipynb – Colaboratory (google.com)

https://raw.githubusercontent.com/kirancakra/Machine-Learning-Netflix-/main/netflix_titles.csv