

Nama : Kharisma Kusuma Dewi

NIM : 1900018299

Kelas : C

Judul : Dashboard Netflix dan TV Show Analysis

1. Data Preparation

Pada tahap ini data yang digunakan merupakan data yang diambil dari situs Kaggle, yaitu data Netflix Movies and TV Shows. Dari link berikut :

<https://www.kaggle.com/datasets/shivamb/netflix-shows>. Kemudian Dataset tersebut diolah dengan menggunakan library pandas pada Python untuk mengecek, cleaning (menghapus data yang memiliki missing value), dan kemudian hasilnya di export menjadi file csv lain untuk dilakukan proses selanjutnya. Berikut merupakan langkah-langkah dan hasil dari proses data preparation :

```
In [1]: 1 # import library yang akan digunakan
        2 import numpy as np
        3 import pandas as pd
```

```
In [2]: 1 # Load dataset menjadi data frame
        2 df=pd.read_csv("netflix_titles.csv")
```

```
In [3]: 1 # melihat data
        2 df.head(5)
```

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---------|---------|-----------------------|-----------------|---|---------------|--------------------|--------------|--------|-----------|---|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmm... |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata, Khosi Ngema, Gail Mababane, Thaban... | South Africa | September 24, 2021 | 2021 | TV-MA | 2 Seasons | International TV Shows, TV Dramas, TV Mysteries | After crossing paths at a party, a Cape Town l... |
| 2 | s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi... | NaN | September 24, 2021 | 2021 | TV-MA | 1 Season | Crime TV Shows, International TV Shows, TV Act... | To protect his family from a powerful drug lor... |
| 3 | s4 | TV Show | Jailbirds New Orleans | NaN | NaN | NaN | September 24, 2021 | 2021 | TV-MA | 1 Season | Docuseries, Reality TV | Feuds, flirtations and toilet talk go down amo... |
| 4 | s5 | TV Show | Kota Factory | NaN | Mayur More, Jitendra Kumar, Ranjan Raj, Alam | India | September 24, 2021 | 2021 | TV-MA | 2 Seasons | International TV Shows, Romantic TV Shows, TV | In a city of coaching centers known to train l... |

```
In [4]: 1 # melihat informasi dari dataset
        2 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   show_id         8807 non-null   object
1   type            8807 non-null   object
2   title           8807 non-null   object
3   director        6173 non-null   object
4   cast            7982 non-null   object
5   country         7976 non-null   object
6   date_added      8797 non-null   object
7   release_year    8807 non-null   int64
8   rating          8803 non-null   object
9   duration        8804 non-null   object
10  listed_in       8807 non-null   object
11  description      8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

Poses Cleaning Data

```
In [5]: 1 # mengecek missing value
        2 df.isnull().sum()
```

```
Out[5]: show_id      0
        type         0
        title        0
        director    2634
        cast        825
        country     831
        date_added   10
        release_year  0
        rating       4
        duration     3
        listed_in    0
        description  0
        dtype: int64
```

```
In [6]: 1 # menghapus data dengan missing value
        2 df.dropna(inplace=True)
```

```
In [7]: 1 df.isnull().sum()
```

```
Out[7]: show_id      0
        type         0
        title        0
        director     0
        cast         0
        country      0
        date_added   0
        release_year  0
        rating       0
        duration     0
        listed_in    0
        description  0
        dtype: int64
```

```
In [9]: 1 # mengubah data kategori pada kolom rating
        2 new_categories = {
        3     'TV-PG': 'Parental Guidance',
        4     'TV-MA': 'Mature Audience',
        5     'TV-Y7-FV': 'Teens',
        6     'TV-Y7': 'Teens',
        7     'TV-14': 'Teens',
        8     'R': 'Mature Audience',
        9     'TV-Y': 'General Audience',
       10     'NR': 'Mature Audience',
       11     'PG-13': 'Teens',
       12     'TV-G': 'General Audience',
       13     'PG': 'Teens',
       14     'G': 'General Audience',
       15     'UR': 'Mature Audience',
       16     'NC-17': 'Mature Audience'
```

```
17 }
18 df["rating"] = df['rating'].replace(new_categories)
19 df.head()
```

```
Out[9]:
```

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|----|---------|---------|-------------------------------|---------------------|---|---|--------------------|--------------|-----------------|-----------|--|---|
| 7 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | United States, Ghana, Burkina Faso, United Kin... | September 24, 2021 | 1993 | Mature Audience | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... |
| 8 | s9 | TV Show | The Great British Baking Show | Andy Devonshire | Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho... | United Kingdom | September 24, 2021 | 2021 | Teens | 9 Seasons | British TV Shows, Reality TV | A talented batch of amateur bakers face off in... |
| 9 | s10 | Movie | The Starling | Theodore Melfi | Melissa McCarthy, Chris O'Dowd, Kevin Kline, T... | United States | September 24, 2021 | 2021 | Teens | 104 min | Comedies, Dramas | A woman adjusting to life after a loss contend... |
| 12 | s13 | Movie | Je Suis Karl | Christian Schwöchow | Luna Wedler, Jannis Niewöhner, Milan Peschel, ... | Germany, Czech Republic | September 23, 2021 | 2021 | Mature Audience | 127 min | Dramas, International Movies | After most of her family is murdered in a terr... |
| 24 | s25 | Movie | Jeans | S. Shankar | Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi... | India | September 21, 2021 | 1998 | Teens | 166 min | Comedies, International Movies, Romantic Movies | When the father of the man she loves insists t... |

```
In [14]: 1 # menyimpan file hasil preparation data ke file csv
        2 df.to_csv("Netflix_TV_Show.csv", index=False)
```

2. Analisis Data Eksplorasi

Pada tahap EDA ini, dilakukan proses seperti load dataset hasil preparation, kemudian mengecek ukuran data dan hasilnya dataset ini memiliki 5332 baris dengan 13 kolom. Selanjutnya yaitu mengecek tipe data tiap kolom dan mengecek apakah masih ada missing value. Selanjutnya dilakukan pengecekan untuk melihat jumlah data berdasarkan kolom tertentu, melihat summary statistic dan mengubah tipe data pada kolom date_time menjadi type datetime. Berikut merupakan proses EDA dengan Python pada jupyter notebook :

```
In [1]: 1 #import library yang akan digunakan
        2 import numpy as np
        3 import pandas as pd
```

```
In [2]: 1 #load dataset hasil preparation
        2 df=pd.read_csv("Netflix_&TV_Show.csv")
        3 df.head()
```

```
Out[2]:
```

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description | data_time |
|---|---------|---------|-------------------------------|---------------------|---|---|--------------------|--------------|-----------------|-----------|--|---|------------|
| 0 | s8 | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D... | United States, Ghana, Burkina Faso, United Kin... | September 24, 2021 | 1993 | Mature Audience | 125 min | Dramas, Independent Movies, International Movies | On a photo shoot in Ghana, an American model s... | 2021-09-24 |
| 1 | s9 | TV Show | The Great British Baking Show | Andy Devonshire | Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho... | United Kingdom | September 24, 2021 | 2021 | Teens | 9 Seasons | British TV Shows, Reality TV | A talented batch of amateur bakers face off in... | 2021-09-24 |
| 2 | s10 | Movie | The Starling | Theodore Melfi | Melissa McCarthy, Chris O'Dowd, Kevin Kline, T... | United States | September 24, 2021 | 2021 | Teens | 104 min | Comedies, Dramas | A woman adjusting to life after a loss contend... | 2021-09-24 |
| 3 | s13 | Movie | Je Suis Karl | Christian Schwochow | Luna Wedler, Jannis Niewöhner, Milan Peschel, ... | Germany, Czech Republic | September 23, 2021 | 2021 | Mature Audience | 127 min | Dramas, International Movies | After most of her family is murdered in a terr... | 2021-09-23 |
| 4 | s25 | Movie | Jeans | S. Shankar | Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi... | India | September 21, 2021 | 1998 | Teens | 166 min | Comedies, International Movies, Romantic Movies | When the father of the man she loves insists t... | 2021-09-21 |

```
In [3]: 1 #mengecek ukuran data
        2 df.shape
```

```
Out[3]: (5332, 13)
```

```
In [4]: 1 #menampilkan tipe data
        2 df.dtypes
```

```
Out[4]: show_id      object
        type        object
        title        object
        director      object
        cast          object
        country       object
        date_added    object
        release_year  int64
        rating        object
        duration      object
        listed_in     object
        description   object
        data_time     object
        dtype: object
```

```
In [5]: 1 #mengecek missing value
        2 df.isnull().sum()
```

```
Out[5]: show_id      0
        type        0
        title        0
        director      0
        cast          0
        country       0
        date_added    0
        release_year  0
        rating        0
        duration      0
        listed_in     0
        description   0
```

```

data_time      0
dtype: int64

In [6]: 1 #mengecek jumlah data berdasarkan kolom type
        2 df.type.value_counts()

Out[6]: Movie      5185
        TV Show    147
        Name: type, dtype: int64

In [7]: 1 #mengecek jumlah data berdasarkan kolom country
        2 df.country.value_counts()

Out[7]: United States      1846
        India              875
        United Kingdom     183
        Canada             107
        Spain              91
        ...
        Canada, Luxembourg      1
        China, United States, United Kingdom      1
        Pakistan, Norway, United States      1
        United States, Canada, Belgium, United Kingdom      1
        United States, France, Italy, United Kingdom      1
        Name: country, Length: 604, dtype: int64

In [8]: 1 # summary statistic
        2 df.describe().T

Out[8]:
           count      mean      std      min      25%      50%      75%      max
release_year  5332.0  2012.742123  9.625831  1942.0  2011.0  2016.0  2018.0  2021.0

In [9]: 1 # mengubah tipe data
        2 df["data_time"] = pd.to_datetime(df["date_added"])

In [11]: 1 df.dtypes

Out[11]: show_id      object
         type        object
         title      object
         director  object
         cast       object
         country   object
         date_added object
         release_year  int64
         rating      object
         duration    object
         listed_in   object
         description  object
         data_time    datetime64[ns]
         dtype: object

```

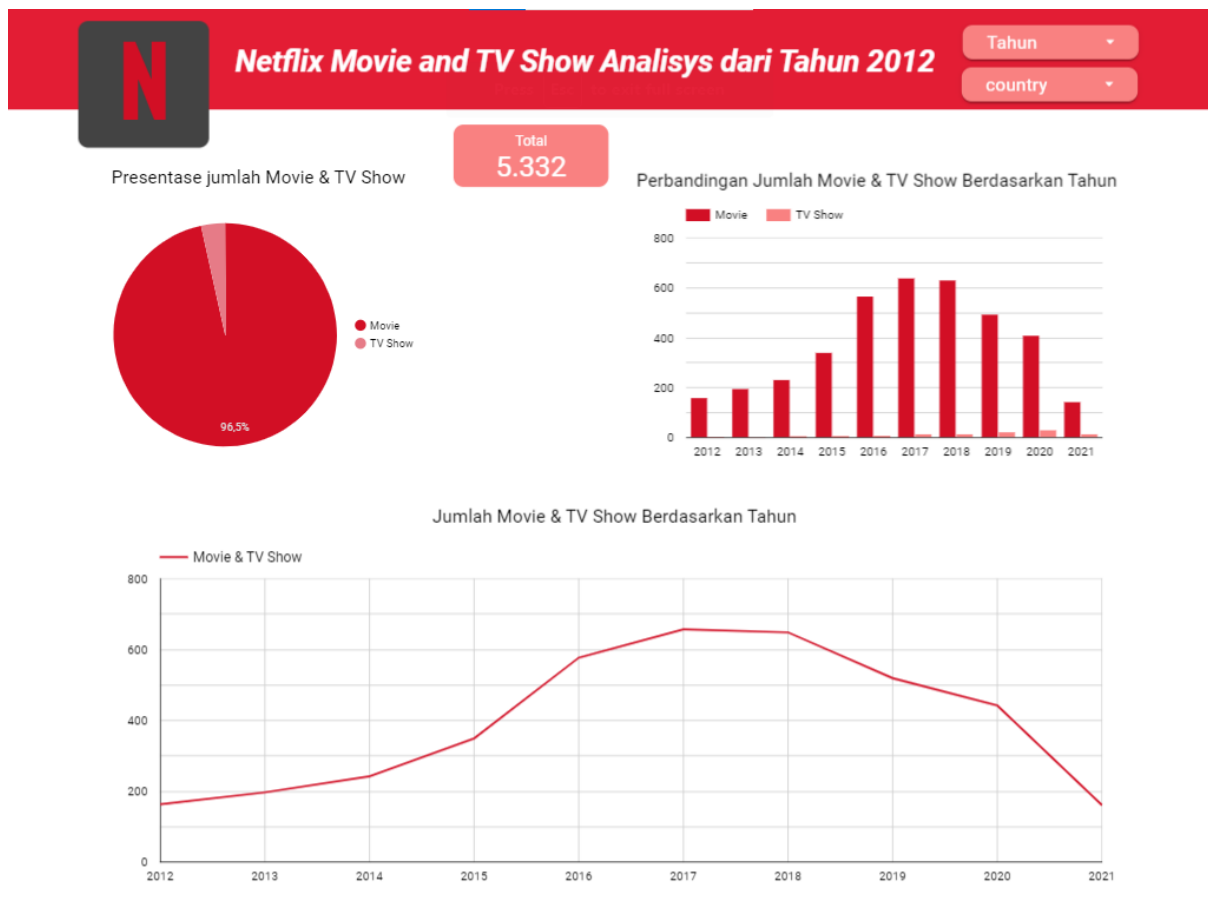
3. Storytelling dan teknologi dan langkah yang digunakan

Netflix merupakan platform streaming media dan video terkenal. Mereka memiliki lebih dari 8000 film atau serial televisi yang tersedia di situs mereka, dan pada pertengahan 2021, mereka memiliki lebih dari 200 juta pelanggan di seluruh dunia. Tujuan dari visualisasi data netflix ini yaitu untuk mengetahui persentase perbandingan jumlah jenis siaran pada netflix yaitu movie dan tv show yang digambarkan dengan **pie chart**. Kemudian digunakan juga **line chart** untuk mengetahui perkembangan jumlah siaran yang ditayangkan tiap tahunnya dari tahun 2012 sampai 2021. Selain itu, untuk lebih jelasnya digunakan **bar chart** untuk mengetahui perbandingan jumlah jenis siaran berupa movie dan tv show di tiap tahunnya dari tahun 2012 sampai 2021. Pada halaman kedua, visualisasi yang ditampilkan memiliki tujuan untuk mengetahui jumlah siaran berdasarkan rating (penggolongan umur yang diperbolehkan menonton) dengan **bar chart**. Untuk lebih jelasnya ditampilkan juga **column chart** untuk menunjukkan perbandingan jumlah rating berdasarkan jenis siarannya (movie dan tv show). Visualisasi yang terakhir yaitu **map chart**, untuk menunjukkan negara mana saja yang memproduksi siaran di netflix beserta banyaknya siaran yang diproduksi. Dashboard web interaktif ini juga terdapat kontrol pada bagian kanan pojok untuk menyaring data berdasarkan tahun dan atau negara produksi siaran.

Teknologi yang digunakan untuk membangun dashboard interaktif ini yaitu dengan Google Data studio dan Google Spreadsheet. Yang mana datanya sudah dilakukan preparation dan EDA dengan jupyter notebook menggunakan Python. Proses preparation dan EDA sudah dijelaskan sebelumnya. Hasil dari EDA, datanya akan disimpan dan file nya di upload ke Google Spreadsheet. Kemudian pada GDS, untuk datanya di import melalui data pada Google Spreadsheet tadi. Dan barulah dimulai pemilihan chart yang akan digunakan beserta pengaturan datanya. Hasil dari visualisasi ini dapat dilihat pada link berikut :

<https://datastudio.google.com/reporting/94f4512f-51c1-456f-b3d4-a743af96d726>

4. Hasil





Netflix Movie and TV Show Analisis dari Tahun 2012

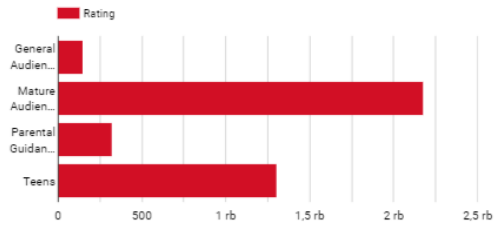
Tahun



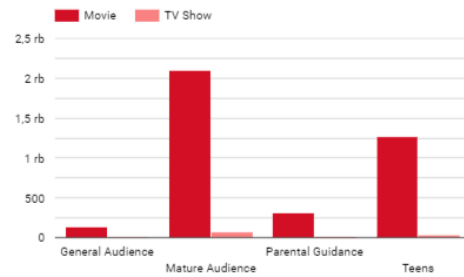
country



Jumlah Rating



Perbandingan Jumlah Rating Berdasarkan Tipe Movie & TV Show



Persebaran Jumlah Movie & TV Show Berdasarkan Negara

