

Nama : Taufik Ramdan  
Nim : 20210040154  
Kelas : TI21F

## Tugas Dasar Pemrograman Sesi 15 Case-Data Preprocessing

### Soal

Lakukan Proses Persiapan Data

1. Download data dengan ekstensi csv, atau cari url dari file tersebut (Bebas) - Rekomendasi (Kaggle.com)
2. Load Data tersebut ke dalam python jupyter Notebook
3. Buatlah beberapa proses yaitu wawasan apa yang bisa diperoleh dari data tersebut / Hal apa yang dianalisa ?

Ini link Github nya : <https://github.com/taufik-01/Tugas-Dasar-Pemrograman-Sesi-15-Case-Data-Preprocessing.git>

### Jawaban

Mohon pak taufik ngerjain nya pake google colab karna gak punya vs code sama python.

Google Colab interface showing a Jupyter Notebook for data analysis.

**Code Cell [2]:**

```
import pandas as pd
df = pd.read_csv('netflix_titles.csv')
df.head()
```

**Data Preview:**

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	Amma Oamata, Khosi Ngema, Gail Makhelane, Thab...	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...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoos, 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	Jaibirds 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 K...	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...

**Next steps:** [Generate code with AI](#) [View recommended plots](#) [New interactive sheet](#)

**Analisis Data Netflix - Exploratory Data Analysis**

Notebook ini digunakan untuk menganalisis dataset Netflix Movies and TV Shows.

Data diambil dari Kaggle dan berisi informasi seperti:

- Judul film / acara TV
- Tipe konten (Movie atau TV Show)
- Nama sutradara dan pemeran
- Negara asal, tanggal ditambahkan, tahun rilis
- Rating usia, durasi, dan kategori genre
- Deskripsi singkat konten

**Tujuan Analisis**

1. Mengetahui distribusi konten berdasarkan tahun rilis
2. Mengetahui negara produksi terbanyak
3. Melihat genre/kategori yang paling umum
4. Mengetahui insight dari konten TV dan Movie di Netflix

Double-click (or enter) to edit

```
[3] df['type'].value_counts()
```

type	count
Movie	6131
TV Show	2676

dtype: int64

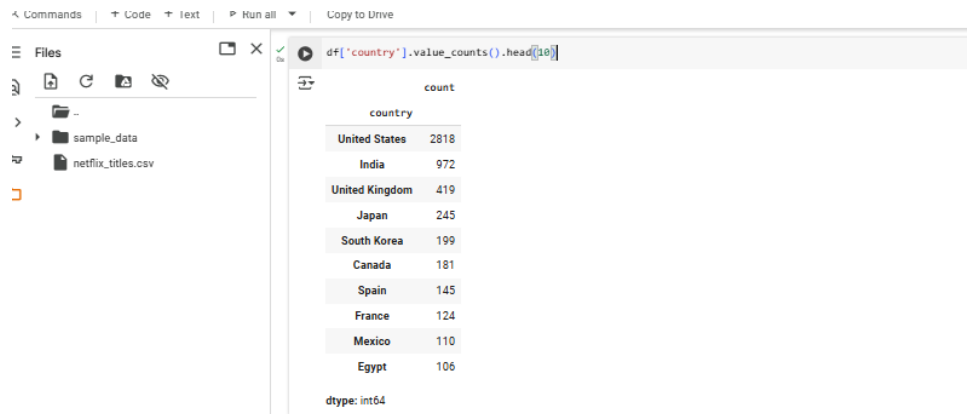
**Persiapan & Pembersihan Data**

Sebelum melakukan analisis lebih lanjut, kita perlu memahami struktur data dan membersihkan data yang tidak lengkap atau tidak relevan.

Langkah-langkah yang dilakukan:

- Menampilkan informasi struktur dataset
- Mengetahui jumlah nilai yang hilang (missing value)
- Menghapus duplikat data jika ada

Langkah ini penting agar data yang akan dianalisis bersih dan siap digunakan.



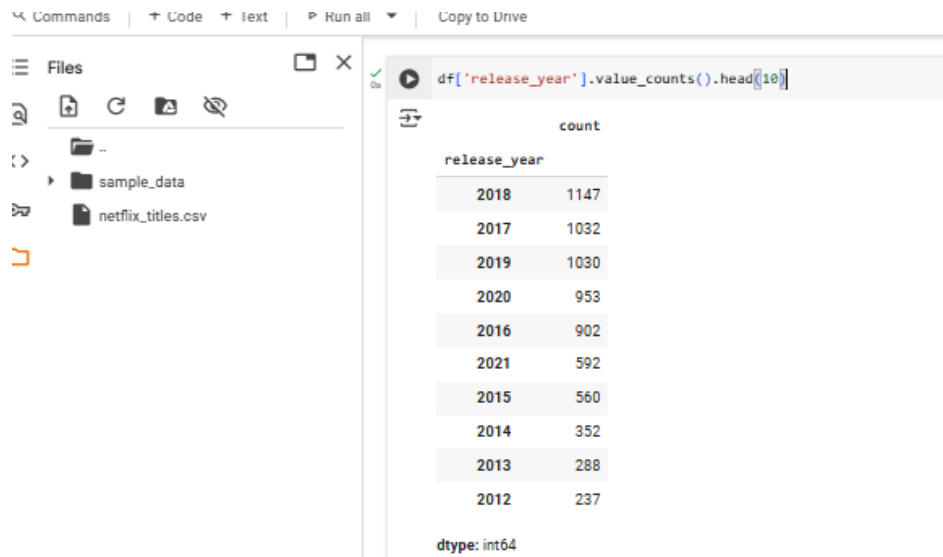
### 🔍 Analisis Awal & Wawasan

Setelah data bersih, kita melakukan analisis awal untuk melihat gambaran umum isi dataset.

Hal-hal yang dianalisis:

- Jumlah konten berdasarkan tipe (Movie/TV Show)
- Distribusi tahun rilis konten
- Negara dengan jumlah konten terbanyak
- Genre atau kategori yang paling sering muncul

Analisis ini dilakukan menggunakan fungsi `value_counts()`, `groupby()`, dan ditampilkan juga dalam bentuk grafik.



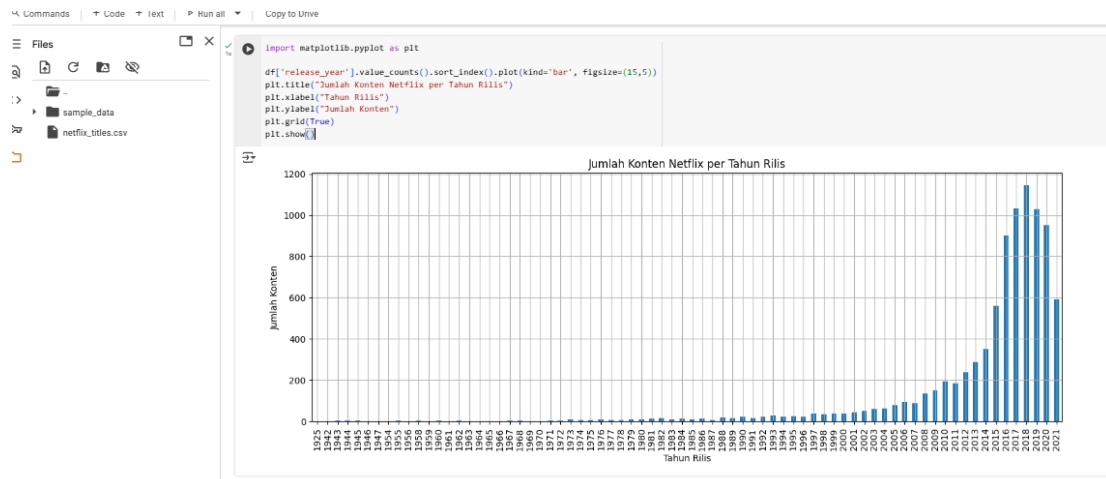
### 📊 Visualisasi Data

Visualisasi membantu memperjelas tren dan pola dalam data.

Grafik yang dibuat:

- Jumlah rilis konten per tahun
- Perbandingan jumlah Movie dan TV Show
- Genre terbanyak

Visualisasi dibuat menggunakan `matplotlib` dan `seaborn`.



File | Edit | View | Insert | Run | Tools | Help

Code | + Code | + Text | Run all | Copy to Drive

### Kesimpulan

Berdasarkan analisis dataset Netflix:

- Mayoritas konten di Netflix berupa **Movie**
- Tahun rilis konten terbanyak adalah sekitar **2018–2020**
- Negara penyumbang konten terbanyak adalah **United States**
- Genre/kategori yang paling sering muncul adalah **Drama, International TV Shows, dan Comedies**

Analisis ini memberikan gambaran umum tentang tren konten Netflix dalam beberapa tahun terakhir.

Commands | + Code | + Text | Run all | Copy to Drive

Files

- sample\_data
- netflix\_titles.csv

```
df.isnull().sum()
```

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

### Penutup

Notebook ini merupakan bagian dari tugas Sesi 15 - Data Preprocessing & Exploratory Data Analysis.

Dataset diambil dari Kaggle:

[Netflix Movies and TV Shows Dataset](#)

Analisis ini bersifat eksploratif dasar. Untuk pengembangan lanjutan, bisa dilakukan:

- Klasifikasi konten berdasarkan genre
- Sistem rekomendasi film
- Prediksi tren genre berdasarkan tahun

Terima kasih 🙏

Disk 70.21 GB available