



Commands + Code + Text

✓ RAM  
Disk

```
[17] import pandas as pd
import numpy as np
```

```
[18] from google.colab import files
uploaded = files.upload()

netflix = pd.read_csv('netflix_titles.csv')
```

Choose Files netflix\_titles.csv  
netflix\_titles.csv(text/csv) - 3399671 bytes, last modified: n/a - 100% done  
Saving netflix\_titles.csv to netflix\_titles (1).csv

```
print("Netflix Head")
print(netflix.head())
print()

#data types and missing values
print("Netflix Info")
print(netflix.info())
print()

#count missing values in column
print("Netflix Missing Values")
print(netflix.isnull().sum())
print()

print(len(netflix))
```

Netflix Head

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	NaN	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	NaN	
4	s5	TV Show	Kota Factory	NaN	

  

	cast	country	\
0	NaN	United States	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	
3	NaN	NaN	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	

date\_added release\_year rating duration \



Q Commands + Code + Text

✓ RAM  
Disk

```
0   date_added  release_year rating  duration \
1   September 25, 2021      2020  PG-13    90 min
2   September 24, 2021      2021  TV-MA    2 Seasons
3   September 24, 2021      2021  TV-MA    1 Season
4   September 24, 2021      2021  TV-MA    2 Seasons
```

```
0   listed_in \
1   Documentaries
2   International TV Shows, TV Dramas, TV Mysteries
3   Crime TV Shows, International TV Shows, TV Act...
4   Docuseries, Reality TV
5   International TV Shows, Romantic TV Shows, TV ...
```

```
0   description
1   As her father nears the end of his life, filmm...
2   After crossing paths at a party, a Cape Town t...
3   To protect his family from a powerful drug lor...
4   Feuds, flirtations and toilet talk go down amo...
5   In a city of coaching centers known to train I...
```

Netflix Info

&lt;class 'pandas.core.frame.DataFrame'&gt;

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

None

Netflix Missing Values

show_id	0
type	0
title	0
director	2634
cast	825
country	831



Q Commands + Code + Text

✓ RAM  Disk  ^

Ds

```
# Fill missing 'director' and 'cast' with 'No Data'
netflix['director'] = netflix['director'].fillna('No Data')
netflix['cast'] = netflix['cast'].fillna('No Data')

# Fill missing 'country' with the most common country (mode)
most_common_country = netflix['country'].mode()[0]
netflix['country'] = netflix['country'].fillna(most_common_country)

# Drop rows with missing 'date_added' or 'rating'
netflix = netflix.dropna(subset=['date_added', 'rating'])
```

```
[21] #removing duplicate rows, if any
netflix.drop_duplicates(inplace=True)
```

```
[22] # Standardize column names
netflix.columns = netflix.columns.str.lower().str.replace(' ', '_')

# Convert 'date_added' to datetime format
# Strip leading/trailing whitespace before conversion and use errors='coerce'
netflix['date_added'] = pd.to_datetime(netflix['date_added'].str.strip(), errors='coerce')
#the earlier code I tried had an error where one of the data was not in the correct format for Pnada to detect it and covert
#to standart date and time format. and so google colab suggested this change in code.

# Ensure 'release_year' is integer
netflix['release_year'] = netflix['release_year'].astype(int)

# (Optional) Check and standardize 'type' and 'rating' capitalization
netflix['type'] = netflix['type'].str.title()
netflix['rating'] = netflix['rating'].str.upper()
```

```
[23] #confirming the changes made
print(netflix.isnull().sum())

print(netflix.dtypes)

print(netflix.head())

print(len(netflix))
```

```
title      0
director   0
cast       0
country    0
```



Commands + Code + Text

✓ RAM Disk



```
title      0
director   0
cast       0
country    0
date_added 0
release_year 0
rating     0
duration   3
listed_in  0
description 0
dtype: int64

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

show_id  type  title  director \
0      s1  Movie  Dick Johnson Is Dead  Kirsten Johnson
1      s2  Tv Show  Blood & Water  No Data
2      s3  Tv Show  Ganglands  Julien Leclercq
3      s4  Tv Show  Jailbirds New Orleans  No Data
4      s5  Tv Show  Kota Factory  No Data

cast  country \
0      No Data  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...  United States
3      No Data  United States
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...  India

date_added  release_year  rating  duration \
0  2021-09-25      2020  PG-13    90 min
1  2021-09-24      2021  TV-MA    2 Seasons
2  2021-09-24      2021  TV-MA    1 Season
3  2021-09-24      2021  TV-MA    1 Season
4  2021-09-24      2021  TV-MA    2 Seasons

listed_in \
0      Documentaries
1  International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3      Docuseries, Reality TV
4  International TV Shows, TV Dramas, TV Mysteries
```



Q Commands

+ Code + Text

✓ RAM  
Disk

```
1 International TV Shows, TV Dramas, TV Mysteries
2 Crime TV Shows, International TV Shows, TV Act...
3 Docuseries, Reality TV
4 International TV Shows, Romantic TV Shows, TV ...

description
0 As her father nears the end of his life, filmm...
1 After crossing paths at a party, a Cape Town t...
2 To protect his family from a powerful drug lor...
3 Feuds, flirtations and toilet talk go down amo...
4 In a city of coaching centers known to train I...
8793
```

✓  
0s

```
from google.colab import files
```

```
#save a CSV file
```

```
netflix.to_csv('netflix_cleaned.csv', index=False)
```

```
# Download the file
```

```
files.download('netflix_cleaned.csv')
```

