

Step 1: Load the dataset

JupyterLab interface showing the loading of a dataset from a CSV file.

```
[4]: import pandas as pd
```

```
[6]: df = dfpd.read_csv("C:/Users/surek/OneDrive/Desktop/netflix.csv")
```

```
[7]: df
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	duration_minutes
0	s1	movie	Dick Johnson Is Dead	Kirsten Johnson	Unknown	United States	2021-09-25	2020	PG-13	90 min	documentaries	As her father nears the end of his life, filmm...	90.0
1	s2	tv show	Blood & Water	Unknown	Ama Qamata, Khosi Ngema, Gail Mablane, Thaban...	South Africa	2021-09-24	2021	TV-MA	2 Seasons	international tv shows, tv dramas, tv mysteries	After crossing paths at a party, a Cape Town L...	2.0
2	s3	tv show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown	2021-09-24	2021	TV-MA	1 Season	crime tv shows, international tv shows, tv act...	To protect his family from a powerful drug lor...	1.0
3	s4	tv show	Jailbirds New Orleans	Unknown	Unknown	Unknown	2021-09-24	2021	TV-MA	1 Season	docuseries, reality tv	Feuds, flirtations and toilet talk go down amo...	1.0
		tv		Mayur More, Jitendra					TV-	2	international tv	In a city of coaching	

600 / rows × 13 columns

[8]: df.head()

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	duration_minutes
0	s1	movie	Dick Johnson Is Dead	Kirsten Johnson	Unknown	United States	2021-09-25	2020	PG-13	90 min	documentaries	As her father nears the end of his life, filmm...	90.0
1	s2	tv show	Blood & Water	Unknown	Ama Qamata, Khosi Ngema, Gail Mababane, Thaban...	South Africa	2021-09-24	2021	TV-MA	2 Seasons	international tv shows, tv dramas, tv mysteries	After crossing paths at a party, a Cape Town t...	2.0
2	s3	tv show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Unknown	2021-09-24	2021	TV-MA	1 Season	crime tv shows, international tv shows, tv act...	To protect his family from a powerful drug lor...	1.0
3	s4	tv show	Jailbirds New Orleans	Unknown	Unknown	Unknown	2021-09-24	2021	TV-MA	1 Season	docuseries, reality tv	Feuds, flirtations and toilet talk go down amo...	1.0
4	s5	tv show	Kota Factory	Unknown	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	2021-09-24	2021	TV-MA	2 Seasons	international tv shows, romantic tv shows, tv ...	In a city of coaching centers known to train l...	2.0

Step 2: Check missing values

[16]: print(df.isnull().sum())

```
show_id      0
type         0
title        0
director     0
cast         0
country      0
date_added   10
release_year  0
rating       0
duration     0
listed_in    0
description   0
duration_minutes  3
dtype: int64
```

[17]: df['director'] = df['director'].fillna('Unknown')

Step3:Handle missing/null values and Remove duplicates

```
[19]: df['director'] = df['director'].fillna('Unknown')
      df['cast'] = df['cast'].fillna('Unknown')
```

```
[20]: df.dropna(subset=['country'], inplace=True)
```

```
[21]: print("\nNull Values After Cleaning:")
      print(df.isnull().sum())
```

Null Values After Cleaning:

show_id	0
type	0
title	0
director	0
cast	0
country	0
date_added	10
release_year	0
rating	0
duration	0
listed_in	0
description	0
duration_minutes	3
dtype: int64	

```
[22]: print(df.duplicated().sum())
```

0

Step 5: Standardize text/Clean column names

[25]:

```
# Extract numeric part from duration using a raw string for regex
df['duration_minutes'] = df['duration'].str.extract(r'(\d+)').astype('Int64')

# Check the output
print(df[['duration', 'duration_minutes']].head())
```

	duration	duration_minutes
0	90 min	90
1	2 Seasons	2
2	1 Season	1
3	1 Season	1
4	2 Seasons	2

Step 6: Verify cleaned dataset

```
[26]: print(df.info())  
      print(df.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 8807 entries, 0 to 8806  
Data columns (total 13 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              8807 non-null  object   
4   cast                  8807 non-null  object   
5   country               8807 non-null  object   
6   date_added            8797 non-null  datetime64[ns]  
7   release_year          8807 non-null  Int64    
8   rating                8807 non-null  object   
9   duration              8807 non-null  object   
10  listed_in             8807 non-null  object   
11  description            8807 non-null  object   
12  duration_minutes      8804 non-null  Int64    
dtypes: Int64(2), datetime64[ns](1), object(10)  
memory usage: 911.8+ KB  
None  
show_id      0  
type         0  
title        0  
director     0  
cast         0  
country      0  
date_added   10  
release_year  0  
rating       0  
duration     0  
listed_in    0  
description  0  
duration_minutes  3  
dtype: int64
```

Step 7: Save cleaned dataset

```
dtype: int64
[33]: df.to_csv("C:/Users/surek/OneDrive/Desktop/netflix.csv", index=False)

[34]: df
[34]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	duration_minutes
0	s1	movie	Dick Johnson Is Dead	Kirsten Johnson	Unknown	United States	2021-09-25	2020	PG-13	90 min	documentaries	As her father nears the end of his life, filmm...	90
1	s2	tv show	Blood & Water	Unknown	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	2021-09-24	2021	TV-MA	2 Seasons	international tv shows, tv dramas, tv mysteries	After crossing paths at a party, a Cape Town t...	2
2	s3	tv show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabl...	Unknown	2021-09-24	2021	TV-MA	1 Season	crime tv shows, international tv shows, tv act...	To protect his family from a powerful drug lor...	1
3	s4	tv show	Jailbirds New Orleans	Unknown	Unknown	Unknown	2021-09-24	2021	TV-MA	1 Season	docuseries, reality tv	Feuds, flirtations and toilet talk go down amo...	1
4	s5	tv show	Katastrophe	Unknown	Mayur More, Jitendra Kumar	India	2021-09-24	2021	TV-	2	international tv shows, romantic	In a city of coaching centers	2