

#1 Identify and handle missing values using .isnull() in Python

```
import pandas as pd
df = pd.read_csv("/content/netflix.csv")
print(df.shape)
print(df.head())
```

```
(8807, 12)
  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 \
0  September 25, 2021  2020.0  PG-13  90 min
1  September 24, 2021  2021.0  TV-MA  2 Seasons
2  September 24, 2021  2021.0  TV-MA  1 Season
3  September 24, 2021  NaN  TV-MA  1 Season
4  September 24, 2021  2021.0  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, 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...
```

```
print(df.isnull())
print(df.isnull().sum())
print(df[df['rating'].isnull()])
```

```
(8806, 12)
  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 \
0  September 25, 2021  2020.0  PG-13  90 min
1  September 24, 2021  2021.0  TV-MA  2 Seasons
2  September 24, 2021  2021.0  TV-MA  1 Season
3  September 24, 2021  NaN  TV-MA  1 Season
4  September 24, 2021  2021.0  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, 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...

[8807 rows x 12 columns]
show_id      0
type         0
title        0
director    2634
cast        825
country     831
```

```

5989      NaN      Oprah Winfrey, Ava DuVernay
6827      NaN      Kaito Ishikawa, Hisako Kanemoto, Ai Kayano, Ka...
7312      NaN      Flynn Curry, Olivia Deeble, Madison Lu, Oisín ...
7537  Alessandro Pepe      Leone Frisa, Paolo Vaccarino, Francesco Miglio...

      country      date_added      release_year      rating      duration \
5989      NaN      January 26, 2017      2017.0      NaN      37 min
6827      Japan      December 1, 2016      2013.0      NaN      1 Season
7312  Australia      February 1, 2018      2015.0      NaN      1 Season
7537      Italy      March 1, 2017      2015.0      NaN      115 min

      listed_in \
5989      Movies
6827  Anime Series, International TV Shows
7312      Kids' TV, TV Comedies
7537      Dramas

      description
5989  Oprah Winfrey sits down with director Ava DuVe...
6827  After falling through a wormhole, a space-dwel...
7312  Adopting a child's perspective, this show take...
7537  Amid the chaos and horror of World War II, a c...

```

```

df=df.dropna(subset=['director','cast'])
df['country'] = df['country'].fillna('Unknown')
print(df.isnull().sum())
print(df.head())

```

```

director      0
cast          0
country       0
date_added    0
release_year  0
rating        1
duration      3
listed_in     0
description   0
dtype: int64
show_id  type      title \
2      s3  TV Show      Ganglands
5      s6  TV Show      Midnight Mass
6      s7  Movie    My Little Pony: A New Generation
7      s8  Movie      Sankofa
8      s9  TV Show    The Great British Baking Show

      director \
2      Julien Leclercq
5      Mike Flanagan
6  Robert Cullen, José Luis Ucha
7      Haile Gerima
8      Andy Devonshire

      cast \
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
5  Kate Siegel, Zach Gilford, Hamish Linklater, H...
6  Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7  Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
8  Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...

      country      date_added \
2      Unknown      September 24, 2021
5      Unknown      September 24, 2021
6      Unknown      September 24, 2021
7  United States, Ghana, Burkina Faso, United Kin...      September 24, 2021
8      United Kingdom      September 24, 2021

      release_year      rating      duration \
2      2021.0      TV-MA      1 Season
5      2021.0      TV-MA      1 Season
6      2021.0      PG      91 min
7      1993.0      TV-MA      125 min
8      2021.0      TV-14      9 Seasons

      listed_in \
2  Crime TV Shows, International TV Shows, TV Act...
5      TV Dramas, TV Horror, TV Mysteries
6      Children & Family Movies

```

-----

```

description
2 To protect his family from a powerful drug lor...
5 The arrival of a charismatic young priest brin...
6 Equestria's divided. But a bright-eyed hero be...
7 On a photo shoot in Ghana, an American model s...
8 A talented batch of amateur bakers face off in...

```

```

mode_val = df['rating'].mode()[0]
df['rating'] = df['rating'].fillna(mode_val)
df=df.dropna()
print(df.isnull().sum())
print(df.shape)

```

```

↗ 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
(5697, 12)

```

```
#2 Remove duplicate rows using .drop_duplicates()
```

```

duplicates = df.duplicated().sum()
print("Number of duplicate rows:", duplicates)

```

```
df = df.drop_duplicates()
```

```
↗ Number of duplicate rows: 0
```

```

ignore = 'show_id'
check = [col for col in df.columns if col != ignore]

```

```

count = df.duplicated(subset = check).sum()
df = df.drop_duplicates(subset= check)

```

```
print(f"duplicate rows (excluding 'show_id'): {count}")
```

```
↗ duplicate rows (excluding 'show_id'): 2
```

```
#3 Standardize text values
```

```

df['type'] = df['type'].str.strip().str.title().replace('Tv Show', 'TV Show')
df['country'] = df['country'].str.strip().str.title()
df['rating'] = df['rating'].str.strip().str.upper()
df['duration'] = df['duration'].str.strip().str.title()
df['listed_in'] = df['listed_in'].str.lower().str.strip()

```

```
print(df[['type', 'country', 'rating', 'duration', 'listed_in']].head(10))
```

```

↗
   type      country rating \
2  TV Show      Unknown  TV-MA
5  TV Show      Unknown  TV-MA
6  Movie      Unknown    PG
7  Movie  United States, Ghana, Burkina Faso, United Kin...  TV-MA
8  TV Show      United Kingdom  TV-14
9  Movie      United States  PG-13
11 TV Show      Unknown  TV-MA
12 Movie      Germany, Czech Republic  TV-MA
13 Movie      Unknown  TV-PG
18 Movie      Unknown  TV-14

   duration      listed_in
2  1 Season  crime tv shows, international tv shows, tv act...
5  1 Season          tv dramas, tv horror, tv mysteries
6    91 Min      children & family movies
7   125 Min  dramas, independent movies, international movies
8   9 Seasons      british tv shows, reality tv
9   104 Min          comedies, dramas
11  1 Season  crime tv shows, international tv shows, tv act...

```

```

12      127 Min      dramas, international movies
13      91 Min      children & family movies, comedies
18      94 Min      thrillers

```

#4 Convert date formats to a consistent type (e.g., dd-mm-yyyy).

```

df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
print(df['date_added'].dt.strftime('%d-%m-%Y'))
print()
print(df['date_added'].dtype)

```

```

2      24-09-2021
5      24-09-2021
6      24-09-2021
7      24-09-2021
8      24-09-2021
...
8801   09-03-2016
8802   20-11-2019
8804   01-11-2019
8805   11-01-2020
8806   02-03-2019
Name: date_added, Length: 5695, dtype: object

datetime64[ns]

```

```

df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
df['date_added'] = df['date_added'].dt.strftime('%d-%m-%Y')
print(df['date_added'])
print()
print(df['date_added'].dtype)

```

```

2      24-09-2021
5      24-09-2021
6      24-09-2021
7      24-09-2021
8      24-09-2021
...
8801   09-03-2016
8802   20-11-2019
8804   01-11-2019
8805   11-01-2020
8806   02-03-2019
Name: date_added, Length: 5695, dtype: object

object

```

#5 Rename column headers to be clean and uniform

```

renamed = df.columns.str.strip().str.lower().str.replace(' ', '_')
count = sum(df.columns != renamed)
print(f"No. of column names to be renamed: {count}")

```

df.columns = renamed #updated to df

```

No. of column names to be renamed: 0

```

#6 Check and fix data types

```
print(df.dtypes)
```

```

show_id      object
type         object
title        object
director      object
cast          object
country       object
date_added   object
release_year  float64
rating        object
duration      object
listed_in    object
description   object
dtype: object

```

```
df = df.convert_dtypes()
print(df.dtypes)
```

```
show_id      string[python]
type         string[python]
title        string[python]
director     string[python]
cast         string[python]
country      string[python]
date_added   string[python]
release_year  Int64
rating       string[python]
duration     string[python]
listed_in    string[python]
description   string[python]
dtype: object
```

```
# Display final o/p
```

```
print(df.head())
print(df.shape)
```

```
show_id  type  title \
2      s3  TV Show  Ganglands
5      s6  TV Show  Midnight Mass
6      s7  Movie  My Little Pony: A New Generation
7      s8  Movie  Sankofa
8      s9  TV Show  The Great British Baking Show

director \
2      Julien Leclercq
5      Mike Flanagan
6  Robert Cullen, José Luis Ucha
7      Haile Gerima
8      Andy Devonshire

cast \
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
5  Kate Siegel, Zach Gilford, Hamish Linklater, H...
6  Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7  Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
8  Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...

country  date_added \
2      Unknown  24-09-2021
5      Unknown  24-09-2021
6      Unknown  24-09-2021
7  United States, Ghana, Burkina Faso, United Kin...  24-09-2021
8      United Kingdom  24-09-2021

release_year  rating  duration \
2      2021  TV-MA  1 Season
5      2021  TV-MA  1 Season
6      2021  PG  91 Min
7      1993  TV-MA  125 Min
8      2021  TV-14  9 Seasons

listed_in \
2  crime tv shows, international tv shows, tv act...
5      tv dramas, tv horror, tv mysteries
6      children & family movies
7  dramas, independent movies, international movies
8      british tv shows, reality tv

description
2  To protect his family from a powerful drug lor...
5  The arrival of a charismatic young priest brin...
6  Equestria's divided. But a bright-eyed hero be...
7  On a photo shoot in Ghana, an American model s...
8  A talented batch of amateur bakers face off in...
(5695, 12)
```

```
# saving altered csv
df.to_csv('altered_file.csv', index=False)
```

```
from google.colab import files
files.download('altered_file.csv')
```



