```
In [1]:    import pandas as pd
           import numpy as np
```

Load Excel file

```
In [2]:    df=pd.read_csv("C:/Users/nacha/Downloads/dataset_netflix1.csv")
           df
```

Out[2]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
...	...	...	...	...	...	...	...	...	...	...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	3 Seasons	Kids' TV
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	TV-Y	1 Season	Kids' TV
8789	s8786	TV Show	YOM	Not Given	Pakistan	6/7/2018	2016	TV-Y7	1 Season	Kids' TV

8790 rows × 10 columns

Convert date in proper format

```
In [3]:    df['date_added']=pd.to_datetime(df['date_added'],format='%m/%d/%Y')
```

```
In [4]:    df['date_added']=pd.to_datetime(df['date_added'],format='%d-%m-%Y')
```

```
In [5]:    df['date_added']=df['date_added'].dt.strftime('%d-%m-%Y')
           df
```

Out[5]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	25-09-2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	24-09-2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	24-09-2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	22-09-2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	24-09-2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
...	...	...	...	...	...	...	...	...	...	...
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	17-01-2017	2016	TV-PG	2 Seasons	International TV Shows, TV Dramas
8786	s8798	TV Show	Zak Storm	Not Given	United States	13-09-2018	2016	TV-Y7	3 Seasons	Kids' TV
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	15-12-2016	2012	TV-PG	1 Season	International TV Shows, Romantic TV Shows, TV ...
8788	s8784	TV Show	Yoko	Not Given	Pakistan	23-06-2018	2016	TV-Y	1 Season	Kids' TV
8789	s8786	TV Show	YOM	Not Given	Pakistan	07-06-2018	2016	TV-Y7	1 Season	Kids' TV

8790 rows × 10 columns

Download dataset in Excel format

```
In [9]:    df.to_excel("C:\\Users\\nacha\\Downloads\\analysis1.xlsx")
```

Check Null values

```
In [10]:   df.isna().sum()
```

```
Out[10]:   show_id        0
           type           0
           title          0
           director       0
           country        0
           date_added     0
           release_year   0
           rating         0
           duration       0
           listed_in      0
           dtype: int64
```

```
In [11]:   df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

Summary

```
In [12]:   df.describe()
```

Out[12]:

	release_year
count	8790.000000
mean	2014.183163
std	8.825466
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

Replace 'Not Given' with NaN in the 'director' and 'country' columns

```
In [13]:   df['director'].replace('Not Given', pd.NA, inplace=True)
           df['country'].replace('Not Given', pd.NA, inplace=True)
```

Remove rows where 'director' or 'country' is NaN

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

```
In [14]:   df
```

Out[14]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	25-09-2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	24-09-2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	24-09-2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	22-09-2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	24-09-2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
...	...	...	...	...	...	...	...	...	...	...
6565	s3351	Movie	True: Grabbleapple Harvest	Mark Thornton, Todd Kauffman	Canada	01-11-2019	2019	TV-Y	24 min	Children & Family Movies
6566	s3509	Movie	True: Tricky Treat Day	Mark Thornton, Todd Kauffman	Canada	20-09-2019	2019	TV-Y	24 min	Children & Family Movies
6567	s3554	TV Show	True and the Rainbow Kingdom	Mark Thornton, Todd Kauffman	United States	30-08-2019	2019	TV-Y	3 Seasons	Kids' TV
6568	s4241	Movie	Sam Kinison: The Scream Continues	Walter C. Miller	United States	31-12-2018	2016	TV-MA	29 min	Documentaries
6569	s1315	TV Show	Chappelle's Show	Stan Lathan	United States	12-02-2021	2005	TV-MA	2 Seasons	TV Comedies

5915 rows × 10 columns

Download file

```
In [15]:   df.to_excel("C:\\Users\\nacha\\Downloads\\clean_data.xlsx")
```