

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

```
In [2]: import numpy as np
```

```
In [3]: df=pd.read_csv(r"C:\Users\SSD\Downloads\movies.csv")
```

```
In [4]: df
```

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	S1
0	Blood Red Sky	-2021	\nAction, Horror, Thriller	6.1	\nA woman with a mysterious illness is forced ...	\n Director:\nPeter Thorwarth\n \n Star.
1	Masters of the Universe: Revelation	(2021–)	\nAnimation, Action, Adventure	5.0	\nThe war for Eternia begins again in what may...	\n \n Stars:\nChris W\n\nSara...
2	The Walking Dead	(2010– 2022)	\nDrama, Horror, Thriller	8.2	\nSheriff Deputy Rick Grimes wakes up from a c...	\n \n Stars:\nAndrew Lincoln, \n...
3	Rick and Morty	(2013–)	\nAnimation, Adventure, Comedy	9.2	\nAn animated series that follows the exploits...	\n \n Stars:\nJustin Roiland, \n...
4	Army of Thieves	-2021	\nAction, Crime, Horror	NaN	\nA prequel, set before the events of Army of ...	\n Director:\nMatthia Schweighöfer\n \n ..
...
9994	NaN	NaN	NaN	NaN	NaN	NaN
9995	NaN	NaN	NaN	NaN	NaN	NaN
9996	NaN	NaN	NaN	NaN	NaN	NaN
9997	NaN	NaN	NaN	NaN	NaN	NaN
9998	NaN	NaN	NaN	NaN	NaN	NaN

9999 rows x 9 columns

In [5]: `df.head()`

	MOVIES	YEAR	GENRE	RATING	ONE-LINE	STARS
0	Blood Red Sky	-2021	\nAction, Horror, Thriller	6.1	\nA woman with a mysterious illness is forced ...	\n Director:\nPeter Thorwarth\n \n Star..
1	Masters of the Universe: Revelation	(2021–)	\nAnimation, Action, Adventure	5.0	\nThe war for Eternia begins again in what may...	\n \n Stars:\nChris W\n\nSara...
2	The Walking Dead	(2010– 2022)	\nDrama, Horror, Thriller	8.2	\nSheriff Deputy Rick Grimes wakes up from a c...	\n \n Stars:\nAndrew Lincoln, \n...
3	Rick and Morty	(2013–)	\nAnimation, Adventure, Comedy	9.2	\nAn animated series that follows the exploits...	\n \n Stars:\nJustin Roiland, \n...
4	Army of Thieves	-2021	\nAction, Crime, Horror	NaN	\nA prequel, set before the events of Army of ...	\n Director:\nMatthias Schweighöfer\n \n ...

In [6]: `df.shape`

(9999, 9)

In [7]: `df.columns`

```
Index(['MOVIES', 'YEAR', 'GENRE', 'RATING', 'ONE-LINE', 'STARS', 'VOTES',
      'RunTime', 'Gross'],
      dtype='object')
```

In [8]: `df.columns=['Films','Year','Genre','Rating','Online','Stars','Votes','Runtime',`

```
In [9]: df.head()
```

	Films	Year	Genre	Rating		Online
0	Blood Red Sky	-2021	\nAction, Horror, Thriller	6.1	\nA woman with a mysterious illness is forced ...	\n Director:\nPeter Thorwarth\n \n Star.
1	Masters of the Universe: Revelation	(2021–)	\nAnimation, Action, Adventure	5.0	\nThe war for Eternia begins again in what may...	\n \n Stars:\nChris W\nSara...
2	The Walking Dead	(2010–2022)	\nDrama, Horror, Thriller	8.2	\nSheriff Deputy Rick Grimes wakes up from a c...	\n \n Stars:\nAndrew Lincoln, \n...
3	Rick and Morty	(2013–)	\nAnimation, Adventure, Comedy	9.2	\nAn animated series that follows the exploits...	\n \n Stars:\nJustin F\n...
4	Army of Thieves	-2021	\nAction, Crime, Horror	NaN	\nA prequel, set before the events of Army of ...	\n Director:\nMatthia: Schweighöfer\n \n ..

```
In [10]: # Drop the last 2 rows by specifying their index  
df = df.drop(df.index[-9499:])  
print(df)
```

	Films	Year \
0	Blood Red Sky	-2021
1	Masters of the Universe: Revelation	(2021-)
2	The Walking Dead	(2010-2022)
3	Rick and Morty	(2013-)
4	Army of Thieves	-2021
..
495	La catedral del mar	-2018
496	Beckett	-2021
497	Triple Frontier	-2019
498	Fondeados	-2021
499	Sweet Home	(2020-)

	Genre	Rating \
0	\nAction, Horror, Thriller	6.1
1	\nAnimation, Action, Adventure	5.0
2	\nDrama, Horror, Thriller	8.2
3	\nAnimation, Adventure, Comedy	9.2
4	\nAction, Crime, Horror	NaN
..
495	\nDrama, History	7.5
496	\nAction, Crime, Drama	NaN
497	\nAction, Adventure, Crime	6.4
498	\nComedy	4.2
499	\nDrama, Fantasy, Horror	7.4

	Online \
0	\nA woman with a mysterious illness is forced ...
1	\nThe war for Eternia begins again in what may...
2	\nSheriff Deputy Rick Grimes wakes up from a c...
3	\nAn animated series that follows the exploits...
4	\nA prequel, set before the events of Army of ...
..	...
495	\nBased upon Ildefonso Falcones novel.
496	\nFollowing a tragic car accident in Greece, B...
497	\nLoyalties are tested when five friends and f...
498	\nTwo directionless millennial bros get high a...
499	\nHyun, a loner high school student who lost h...

	Stars	Votes	Runtime	Gross
0	\n Director:\nPeter Thorwarth\n \n Star...	21,062	121.0	NaN
1	\n Stars:\nChris Wood, \nSara...	17,870	25.0	NaN
2	\n Stars:\nAndrew Lincoln, \n...	885,805	44.0	NaN
3	\n Stars:\nJustin Roiland, \n...	414,849	23.0	NaN
4	\n Director:\nMatthias Schweighöfer\n \n ...	NaN	NaN	NaN
..
495	\n Stars:\nJordi Aguilar, \nA...	3,574	55.0	NaN
496	\n Director:\nFerdinando Cito Filomarino\n ...	NaN	108.0	NaN
497	\n Director:\nJ.C. Chandor\n \n Stars:...	115,059	125.0	NaN
498	\n Director:\nMarcos Bucay\n \n Stars:...	161	97.0	NaN
499	\n Stars:\nSong Kang, \nLee J...	12,428	52.0	NaN

[500 rows x 9 columns]

In [11]:

```
# Drop column 'B' in-place  
df.drop('Gross', axis=1, inplace=True)  
print(df)
```

	Films	Year \
0	Blood Red Sky	-2021
1	Masters of the Universe: Revelation	(2021-)
2	The Walking Dead	(2010-2022)
3	Rick and Morty	(2013-)
4	Army of Thieves	-2021
..
495	La catedral del mar	-2018
496	Beckett	-2021
497	Triple Frontier	-2019
498	Fondeados	-2021
499	Sweet Home	(2020-)

	Genre	Rating \
0	\nAction, Horror, Thriller	6.1
1	\nAnimation, Action, Adventure	5.0
2	\nDrama, Horror, Thriller	8.2
3	\nAnimation, Adventure, Comedy	9.2
4	\nAction, Crime, Horror	NaN
..
495	\nDrama, History	7.5
496	\nAction, Crime, Drama	NaN
497	\nAction, Adventure, Crime	6.4
498	\nComedy	4.2
499	\nDrama, Fantasy, Horror	7.4

	Online \
0	\nA woman with a mysterious illness is forced ...
1	\nThe war for Eternia begins again in what may...
2	\nSheriff Deputy Rick Grimes wakes up from a c...
3	\nAn animated series that follows the exploits...
4	\nA prequel, set before the events of Army of ...
..	...
495	\nBased upon Ildefonso Falcones novel.
496	\nFollowing a tragic car accident in Greece, B...
497	\nLoyalties are tested when five friends and f...
498	\nTwo directionless millennial bros get high a...
499	\nHyun, a loner high school student who lost h...

	Stars	Votes	Runtime
0	\n Director:\nPeter Thorwarth\n \n Star...	21,062	121.0
1	\n Stars:\nChris Wood, \nSara...	17,870	25.0
2	\n Stars:\nAndrew Lincoln, \n...	885,805	44.0
3	\n Stars:\nJustin Roiland, \n...	414,849	23.0
4	\n Director:\nMatthias Schweighöfer\n \n ...	NaN	NaN
..
495	\n Stars:\nJordi Aguilar, \nA...	3,574	55.0
496	\n Director:\nFerdinando Cito Filomarino\n ...	NaN	108.0
497	\n Director:\nJ.C. Chandor\n \n Stars:...	115,059	125.0
498	\n Director:\nMarcos Bucay\n \n Stars:...	161	97.0
499	\n Stars:\nSong Kang, \nLee J...	12,428	52.0

[500 rows x 8 columns]

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

	Films	Year	Genre	Rating		Online
0	Blood Red Sky	-2021	\nAction, Horror, Thriller	6.1	\nA woman with a mysterious illness is forced ...	\n Director:\n Thorwarth\n
1	Masters of the Universe: Revelation	(2021–)	\nAnimation, Action, Adventure	5.0	\nThe war for Eternia begins again in what may...	\n \n Stars:\n \nSara...
2	The Walking Dead	(2010– 2022)	\nDrama, Horror, Thriller	8.2	\nSheriff Deputy Rick Grimes wakes up from a c...	\n \n Stars:\n \n...
3	Rick and Morty	(2013–)	\nAnimation, Adventure, Comedy	9.2	\nAn animated series that follows the exploits...	\n \n Stars:\n \n...
4	Army of Thieves	-2021	\nAction, Crime, Horror	NaN	\nA prequel, set before the events of Army of ...	\n Director:\n Schweighöfe

In [13]: `df.shape`

(500, 8)

In [14]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Films       500 non-null   object
1   Year        500 non-null   object
2   Genre       500 non-null   object
3   Rating      490 non-null   float64
4   Online      500 non-null   object
5   Stars       500 non-null   object
6   Votes       490 non-null   object
7   Runtime     478 non-null   float64
dtypes: float64(2), object(6)
memory usage: 31.4+ KB
```



```
In [15]: df.isnull().sum()
```

```
Films      0
Year        0
Genre       0
Rating     10
Online      0
Stars       0
Votes      10
Runtime     22
dtype: int64
```

```
In [16]: df['Films'] = df['Films'].str.replace(r'\W', '')
```

C:\Users\SSD\AppData\Local\Temp\ipykernel_8388\3190433430.py:1: FutureWarning: The default value of `regex` is deprecated and will be changed to `False` in a future version.

```
df['Films'] = df['Films'].str.replace(r'\W', '')
```

```
In [17]: df['Films']
```

```
0          BloodRedSky
1  MastersoftheUniverseRevelation
2          TheWalkingDead
3          RickandMorty
4          ArmyofThieves
...
495          Lacatedraldelmar
496          Beckett
497          TripleFrontier
498          Fondeados
499          SweetHome
Name: Films, Length: 500, dtype: object
```

```
In [18]: df['Year'] = df['Year'].str.replace(r'\W', '')
```

C:\Users\SSD\AppData\Local\Temp\ipykernel_8388\3522469387.py:1: FutureWarning: The default value of `regex` is deprecated and will be changed to `False` in a future version.

```
df['Year'] = df['Year'].str.replace(r'\W', '')
```

```
In [19]: df['Year']
```

```
0      2021
1      2021
2    20102022
3      2013
4      2021
...
495    2018
496    2021
497    2019
498    2021
499    2020
Name: Year, Length: 500, dtype: object
```

```
In [20]: df['Genre'] = df['Genre'].str.replace(r'\W', '')
```

```
C:\Users\SSD\AppData\Local\Temp\ipykernel_8388\686410405.py:1: FutureWarning: The default value of regex will be changed to True in a future version.
  df['Genre'] = df['Genre'].str.replace(r'\W', '')
```

```
In [21]: df['Genre']
```

```
0      ActionHorrorThriller
1    AnimationActionAdventure
2      DramaHorrorThriller
3    AnimationAdventureComedy
4      ActionCrimeHorror
...
495      DramaHistory
496      ActionCrimeDrama
497    ActionAdventureCrime
498              Comedy
499      DramaFantasyHorror
Name: Genre, Length: 500, dtype: object
```

```
In [22]: df['Online'] = df['Online'].str.replace(r'\W', '')
```

```
C:\Users\SSD\AppData\Local\Temp\ipykernel_8388\3095981650.py:1: FutureWarning: The default value of regex will be changed to True in a future version.
  df['Online'] = df['Online'].str.replace(r'\W', '')
```

```
In [23]: df['Online']
```

```
0      Awomanwithamysteriousillnessisforcedintoaction...
1      ThewarforEterniabeginsagaininwhatmaybethefinal...
2      SheriffDeputyRickGrimeswakesupfromacomatolearn...
3      Ananimatedseriesthatfollowstheexploitsofasuper...
4      AprequelsetbeforetheeventsofArmyoftheDeadwhich...

...

495      BaseduponIldefonsoFalconesnovel
496      FollowingatragiccaraccidentinGreeceBeckettanAm...
497      Loyaltiesarettestedwhenfivefriendsandformerspec...
498      Twodirectionlessmillennialbrosethhighandpitcha...
499      Hyunalonerhighschoolstudentwholosthisentirefam...
Name: Online, Length: 500, dtype: object
```

```
In [24]: df['Stars'] = df['Stars'].str.replace(r'\W', '')
```

```
C:\Users\SSD\AppData\Local\Temp\ipykernel_8388\1295855449.py:1: FutureWarning: The default value of re
n a future version.
df['Stars'] = df['Stars'].str.replace(r'\W', '')
```

```
In [25]: df['Stars']
```

```
0      DirectorPeterThorwarthStarsPeriBaumeisterCarlaA...
1      StarsChrisWoodSarahMichelleGellarLenaHeadeyMar...
2      StarsAndrewLincolnNormanReedusMelissaMcBrideLa...
3      StarsJustinRoilandChrisParnellSpencerGrammerSa...
4      DirectorMatthiasSchweighöferStarsMatthiasSchwe...

...

495      StarsJordiAguilarAitorLunaPabloDerquiGinésGarc...
496      DirectorFerdinandoCitoFilomarinoStarsJohnDavid...
497      DirectorJCChandorStarsBenAffleckOscarIsaacChar...
498      DirectorMarcosBucayStarsAldoEscalanteRicardoPo...
499      StarsSongKangLeeJinWookLeeSiyoungLeeDoHyun
Name: Stars, Length: 500, dtype: object
```

```
In [26]: df['Votes'] = df['Votes'].str.replace(r'\W', '')
```

```
C:\Users\SSD\AppData\Local\Temp\ipykernel_8388\728037681.py:1: FutureWarning: The default value of reg
a future version.
df['Votes'] = df['Votes'].str.replace(r'\W', '')
```

```
In [27]: df['Votes']

0      21062
1      17870
2      885805
3      414849
4         NaN
...
495     3574
496         NaN
497    115059
498      161
499    12428
Name: Votes, Length: 500, dtype: object
```

```
In [28]: clean_data = df.copy()
```

```
In [29]: clean_data
```

	Films	Year	Genre	Rating	
0	BloodRedSky	2021	ActionHorrorThriller	6.1	Awomanwithamysterio
1	MastersoftheUniverseRevelation	2021	AnimationActionAdventure	5.0	ThewarforEterniabegir
2	TheWalkingDead	20102022	DramaHorrorThriller	8.2	SheriffDeputyRickGrim
3	RickandMorty	2013	AnimationAdventureComedy	9.2	Ananimatedseriesthatf
4	ArmyofThieves	2021	ActionCrimeHorror	NaN	Aprequelsetbeforethee
...
495	Lacatedraldelmar	2018	DramaHistory	7.5	BaseduponIldefonsoFz
496	Beckett	2021	ActionCrimeDrama	NaN	Followingatragiccaracc
497	TripleFrontier	2019	ActionAdventureCrime	6.4	Loyaltiesaretestedwhe
498	Fondeados	2021	Comedy	4.2	Twodirectionlessmillen
499	SweetHome	2020	DramaFantasyHorror	7.4	Hyunalonerhighschool

500 rows x 8 columns

```
In [30]: clean_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Films       500 non-null   object 
 1   Year        500 non-null   object 
 2   Genre       500 non-null   object 
 3   Rating      490 non-null   float64
 4   Online      500 non-null   object 
 5   Stars       500 non-null   object 
 6   Votes       490 non-null   object 
 7   Runtime     478 non-null   float64
dtypes: float64(2), object(6)
memory usage: 31.4+ KB
```

```
In [31]: clean_data['Rating'] = clean_data['Rating'].fillna(np.mean(pd.to_numeric( clean_data['Rating'], errors='coerce', dtype=float64)))
```

```
In [32]: clean_data['Rating']
```

```
0      6.100000
1      5.000000
2      8.200000
3      9.200000
4      7.408367
...
495     7.500000
496     7.408367
497     6.400000
498     4.200000
499     7.400000
Name: Rating, Length: 500, dtype: float64
```

```
In [33]: clean_data['Votes'] = clean_data['Votes'].fillna(np.mean(pd.to_numeric( clean_data['Votes'], errors='coerce', dtype=float64)))
```

```
In [34]: clean_data['Votes']
```

```
0          21062
1          17870
2         885805
3         414849
4    155778.14898
...
495          3574
496    155778.14898
497         115059
498           161
499         12428
Name: Votes, Length: 500, dtype: object
```

```
In [35]: clean_data['Runtime'] = clean_data['Runtime'].fillna(np.mean(pd.to_numeric( clea
```

```
In [36]: clean_data['Runtime']
```

```
0      121.000000
1       25.000000
2       44.000000
3       23.000000
4      80.167364
...
495      55.000000
496     108.000000
497     125.000000
498      97.000000
499      52.000000
Name: Runtime, Length: 500, dtype: float64
```

```
In [37]: clean_data.isnull().sum()
```

```
Films      0
Year       0
Genre      0
Rating     0
Online     0
Stars      0
Votes      0
Runtime    0
dtype: int64
```

```
In [38]: clean_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Films       500 non-null   object 
 1   Year        500 non-null   object 
 2   Genre       500 non-null   object 
 3   Rating      500 non-null   float64
 4   Online     500 non-null   object 
 5   Stars       500 non-null   object 
 6   Votes       500 non-null   object 
 7   Runtime     500 non-null   float64
dtypes: float64(2), object(6)
memory usage: 31.4+ KB
```

```
In [39]: clean_data.describe()
```

	Rating	Runtime
count	500.000000	500.000000
mean	7.408367	80.167364
std	1.053228	62.686813
min	2.700000	11.000000
25%	6.800000	43.000000
50%	7.600000	60.000000
75%	8.200000	109.250000
max	9.400000	572.000000

```
In [40]: clean_data
```

	Films	Year	Genre	Rating	
0	BloodRedSky	2021	ActionHorrorThriller	6.100000	Awomanwithamysteri
1	MastersoftheUniverseRevelation	2021	AnimationActionAdventure	5.000000	ThewarforEterniabeg
2	TheWalkingDead	20102022	DramaHorrorThriller	8.200000	SheriffDeputyRickGri
3	RickandMorty	2013	AnimationAdventureComedy	9.200000	Ananimatedseriestha
4	ArmyofThieves	2021	ActionCrimeHorror	7.408367	Aprequelsetbeforethe
...
495	Lacatedraldelmar	2018	DramaHistory	7.500000	BaseduponIldefonso
496	Beckett	2021	ActionCrimeDrama	7.408367	Followingatragiccar
497	TripleFrontier	2019	ActionAdventureCrime	6.400000	Loyaltiesaretestedwh
498	Fondeados	2021	Comedy	4.200000	Twodirectionlessmille
499	SweetHome	2020	DramaFantasyHorror	7.400000	Hyunalonerhighschoc

500 rows x 8 columns

```
In [41]: clean_data['Votes'] = clean_data['Votes'].astype(int)
```

```
In [42]: clean_data['Votes']
```

```
0      21062
1      17870
2     885805
3     414849
4     155778
...
495      3574
496     155778
497     115059
498       161
499     12428
```

Name: Votes, Length: 500, dtype: int32

```
In [43]: clean_data['Year'] = clean_data['Year'].astype('category')
```



```
In [44]: clean_data['Year']
```

```
0      2021
1      2021
2    20102022
3      2013
4      2021
```

...

```
495    2018
496    2021
497    2019
498    2021
499    2020
```

Name: Year, Length: 500, dtype: category

Categories (145, object): ['19661969', '1968', '1971', '1975', ..., 'II2016', 'II2017', 'III2021', 'I\

```
In [45]: clean_data['Year'] = clean_data['Year'].str.extract('(\d+)')
```

```
In [46]: clean_data['Year']
```

```
0      2021
1      2021
2    20102022
3      2013
4      2021
```

...

```
495    2018
496    2021
497    2019
498    2021
499    2020
```

Name: Year, Length: 500, dtype: object

```
In [47]: clean_data['Year'] = clean_data['Year'].astype('int')
```

```
In [48]: clean_data['Year']
```

```
0      2021
1      2021
2    20102022
3      2013
4      2021
...
495    2018
496    2021
497    2019
498    2021
499    2020
Name: Year, Length: 500, dtype: int32
```

```
In [49]: clean_data.describe()
```

	Year	Rating	Votes	Runtime
count	5.000000e+02	500.000000	5.000000e+02	500.000000
mean	6.838356e+06	7.408367	1.557781e+05	80.167364
std	9.534464e+06	1.053228	2.254407e+05	62.686813
min	1.968000e+03	2.700000	2.700000e+01	11.000000
25%	2.017000e+03	6.800000	2.597575e+04	43.000000
50%	2.020000e+03	7.600000	7.750650e+04	60.000000
75%	2.009202e+07	8.200000	1.952758e+05	109.250000
max	2.020202e+07	9.400000	1.713028e+06	572.000000

```
In [50]: clean_data
```

	Films	Year	Genre	Rating	
0	BloodRedSky	2021	ActionHorrorThriller	6.100000	Awomanwithamysteri
1	MastersoftheUniverseRevelation	2021	AnimationActionAdventure	5.000000	ThewarforEterniabeg
2	TheWalkingDead	20102022	DramaHorrorThriller	8.200000	SheriffDeputyRickGri
3	RickandMorty	2013	AnimationAdventureComedy	9.200000	Ananimatedseriestha
4	ArmyofThieves	2021	ActionCrimeHorror	7.408367	Aprequelsetbeforethe
...
495	Lacatedraldelmar	2018	DramaHistory	7.500000	BaseduponIldefonso
496	Beckett	2021	ActionCrimeDrama	7.408367	Followingatragiccar
497	TripleFrontier	2019	ActionAdventureCrime	6.400000	Loyaltiesarettestedwh
498	Fondeados	2021	Comedy	4.200000	Twodirectionlessmille
499	SweetHome	2020	DramaFantasyHorror	7.400000	Hyunalonerhighschoc

500 rows x 8 columns

```
In [51]: clean_data['Films'] = clean_data['Films'].astype('category')
```

```
In [52]: clean_data['Genre'] = clean_data['Genre'].astype('category')
```

```
In [53]: clean_data['Online'] = clean_data['Online'].astype('category')
```

```
In [54]: clean_data['Stars'] = clean_data['Stars'].astype('category')
```

```
In [55]: clean_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Films       500 non-null   category
1   Year        500 non-null   int32
2   Genre       500 non-null   category
3   Rating      500 non-null   float64
4   Online      500 non-null   category
5   Stars       500 non-null   category
6   Votes       500 non-null   int32
7   Runtime     500 non-null   float64
dtypes: category(4), float64(2), int32(2)
memory usage: 81.2 KB
```

```
In [56]: clean_data['Rating'] = clean_data['Rating'].astype('int')
```

```
In [57]: clean_data['Runtime'] = clean_data['Runtime'].astype('int')
```

```
In [58]: clean_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Films       500 non-null   category
 1   Year        500 non-null   int32   
 2   Genre       500 non-null   category
 3   Rating      500 non-null   int32   
 4   Online      500 non-null   category
 5   Stars       500 non-null   category
 6   Votes       500 non-null   int32   
 7   Runtime     500 non-null   int32   
dtypes: category(4), int32(4)
memory usage: 77.2 KB
```

```
In [59]: clean_data.isnull().sum()
```

```
Films      0
Year        0
Genre       0
Rating      0
Online      0
Stars       0
Votes       0
Runtime     0
dtype: int64
```

```
In [60]: clean_data.shape
```

```
(500, 8)
```

```
In [61]: # Drop column 'B' in-place
clean_data.drop('Stars', axis=1, inplace=True)
print(clean_data)
```

		Films	Year	Genre	\
0		BloodRedSky	2021	ActionHorrorThriller	
1	MastersoftheUniverseRevelation		2021	AnimationActionAdventure	
2	TheWalkingDead	20102022		DramaHorrorThriller	
3	RickandMorty	2013		AnimationAdventureComedy	
4	ArmyofThieves	2021		ActionCrimeHorror	
..	
495	Lacatedraldelmar	2018		DramaHistory	
496	Beckett	2021		ActionCrimeDrama	
497	TripleFrontier	2019		ActionAdventureCrime	
498	Fondeados	2021		Comedy	
499	SweetHome	2020		DramaFantasyHorror	

	Rating		Online	Votes	\
0	6	Awomanwithamysteriousillnessisforcedintoaction...		21062	
1	5	ThewarforEterniabeginsagaininwhatmaybethefinal...		17870	
2	8	SheriffDeputyRickGrimeswakesupfromacomatolearn...		885805	
3	9	Ananimatedseries that followstheexploitsofasuper...		414849	
4	7	AprequelsetbeforetheeventsofArmyoftheDeadwhich...		155778	
..	
495	7	BaseduponIldefonsoFalconesnovel		3574	
496	7	FollowingatragiccaraccidentinGreeceBeckettanAm...		155778	
497	6	Loyaltiesarettestedwhenfivefriendsandformerspec...		115059	
498	4	Twodirectionlessmillennialbrosgethighandpitcha...		161	
499	7	Hyunalonerhighschoolstudentwholostthisentirefam...		12428	

	Runtime
0	121
1	25
2	44
3	23
4	80
..	...
495	55
496	108
497	125
498	97
499	52

[500 rows x 7 columns]

```
In [62]: clean_data.shape
```

(500, 7)

