

```
In [116... import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [117... df=pd.read_csv(r"D:\Stranger Things\New folder\you_ex_data science\pactice\data_
```

```
In [118... df.head()
```

```
Out[118...      name      author      narrator  time  releasedate  la
```

0	Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobely	2 hrs and 20 mins	04-08-08
---	----------------------------	---------------------------	-----------------------	-------------------	----------

1	The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	01-05-18
---	------------------	-----------------------	--------------------------	-------------------	----------

2	The Deep End	Writtenby:JeffKinney	Narratedby:DanRussell	2 hrs and 3 mins	06-11-20
---	--------------	----------------------	-----------------------	------------------	----------

3	Daughter of the Deep	Writtenby:RickRiordan	Narratedby:SoneelaNankani	11 hrs and 16 mins	05-10-21
---	----------------------	-----------------------	---------------------------	--------------------	----------

4	The Lightning Thief: Percy Jackson, Book 1	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs	13-01-10
---	--	-----------------------	---------------------------	--------	----------



```
In [119... df.columns
```

```
Out[119... Index(['name', 'author', 'narrator', 'time', 'releasedate', 'language',
      'stars', 'price'],
      dtype='object')
```

```
In [120... df.sample(5)
```

Out[120...

	name	author	narrator	time	release
<b>52397</b>	Glitterschnitter	Writtenby:SvenRegener	Narratedby:SvenRegener	10 hrs and 29 mins	09-(
<b>63207</b>	Saving the News	Writtenby:MarthaMinow	Narratedby:ElizaFoss	6 hrs and 17 mins	06-(
<b>85796</b>	Die Hexe aus dem Brackermoor	Writtenby:RhiannonWilliams	Narratedby:FundaVanroy	8 hrs and 18 mins	29-(
<b>65160</b>	Capital City	Writtenby:SamuelStein	Narratedby:EmilyBeresford	5 hrs and 28 mins	25-(
<b>20977</b>	Joseph McCarthy	Writtenby:ArthurHerman	Narratedby:SeanPratt	16 hrs and 15 mins	30-(

In [121...

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 87489 entries, 0 to 87488
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   name        87489 non-null  object
1   author      87489 non-null  object
2   narrator    87489 non-null  object
3   time        87489 non-null  object
4   releasedate 87489 non-null  object
5   language    87489 non-null  object
6   stars       87489 non-null  object
7   price       87489 non-null  object
dtypes: object(8)
memory usage: 5.3+ MB
```

In [122...

```
df.isnull().sum()
```

```
Out[122...  name          0
          author    0
          narrator   0
          time       0
          releasedate 0
          language   0
          stars      0
          price      0
          dtype: int64
```

```
In [123... df2=df.copy()
```

```
In [124... df2
```

	name	author	narrator	time	released
0	Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobely	2 hrs and 20 mins	04-08-
1	The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	01-05-
2	The Deep End	Writtenby:JeffKinney	Narratedby:DanRussell	2 hrs and 3 mins	06-11-
3	Daughter of the Deep	Writtenby:RickRiordan	Narratedby:SoneelaNankani	11 hrs and 16 mins	05-10-
4	The Lightning Thief: Percy Jackson, Book 1	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs	13-01-
...	...	...	...	...	...
87484	Last Days of the Bus Club	Writtenby:ChrisStewart	Narratedby:ChrisStewart	7 hrs and 34 mins	09-03-
87485	The Alps	Writtenby:StephenO'Shea	Narratedby:RobertFass	10 hrs and 7 mins	21-02-
87486	The Innocents Abroad	Writtenby:MarkTwain	Narratedby:FloGibson	19 hrs and 4 mins	30-12-
87487	A Sentimental Journey	Writtenby:LaurenceSterne	Narratedby:AntonLesser	4 hrs and 8 mins	23-02-
87488	Havana	Writtenby:MarkKurlansky	Narratedby:FleetCooper	6 hrs and	07-03-

name	author	narrator	time	released:
			1	min

87489 rows × 8 columns

```
In [125... df2['author'].sample(10)
```

```
Out[125... 51338          Writtenby:SidneyGardner
11004          Writtenby:div.
84935          Writtenby:JenniferJenkins
64523          Writtenby:GeorgeOrwell
43422          Writtenby:DamianBarr
23052          Writtenby:AnanyoBhattacharya
17622          Writtenby:GaryBarlow
21813  Writtenby:DengThiakAdut,BenMckelvey
73464          Writtenby:PaulFéval
27072          Writtenby:ShaunaM.Ahern
Name: author, dtype: object
```

```
In [126... df2['author'] = df2['author'].str.replace('Writtenby:', '', )
```

```
In [127... df2['author']
```

```
Out[127... 0      GeronimoStilton
1      RickRiordan
2      JeffKinney
3      RickRiordan
4      RickRiordan
...
87484   ChrisStewart
87485   StephenO'Shea
87486   MarkTwain
87487   LaurenceSterne
87488   MarkKurlansky
Name: author, Length: 87489, dtype: object
```

```
In [128... df2.sample(5)
```

Out[128...

	name	author	narrator	time
<b>32931</b>	The Mindful Vegan	LaniMuelrath,NealBarnard-foreword	Narratedby:LaniMuelrath	10 hrs and 53 mins
<b>40161</b>	Kremlin Winter	RobertService	Narratedby:LeightonPugh	14 hrs and 46 mins
<b>86796</b>	Warpath and Bivouac, or the Conquest of the Sioux	JohnFinerty	Narratedby:JohannZeiger	10 hrs and 36 mins
<b>53401</b>	The Poems of T. S. Eliot	T.S.Eliot	Narratedby:Jeremylrons,DameEileenAtkins	3 hrs and 41 mins
<b>54386</b>	For Once in My Life	ColleenColeman	Narratedby:RoisinRankin	7 hrs and 46 mins

In [129...

```
df2['narrator'].sample(10)
```

Out[129...

```
57189          Narratedby:NeilShah
20929          Narratedby:uncredited
19530          Narratedby:NancyBalbirer
30879          Narratedby:ToddBelcher
54306          Narratedby:ZacharyWebber
20559  Narratedby:LisaRauen,JuttaSeifert
79092          Narratedby:MarcViotor
27133          Narratedby:MiaEllis
35887          Narratedby:TancrediMilitano
57007          Narratedby:AlbertoLori
Name: narrator, dtype: object
```

In [130...

```
df2['narrator']=df2['narrator'].str.replace('Narratedby:', '')
```

In [131...

```
df2.sample(5)
```

Out[131...

	name	author	narrator	time	releaseda
<b>71111</b>	Closer	KylieScott	AndiArndt	2 hrs and 44 mins	05-11-
<b>75710</b>	Wolf's Claim	AnneMarsh	ErindeWard,NoahMichaelLevine	5 hrs and 44 mins	17-02-
<b>69975</b>	Tell it Slant	EugenePeterson	GroverGardner	9 hrs and 59 mins	10-02-(
<b>69371</b>	Simboli sacri degli indiani d'America	ChristopherDubois	PieroDiDomenico	2 hrs and 5 mins	13-12-
<b>21762</b>	Emiliano Zapata [Spanish Edition]	OnlineStudioProductions	uncredited	42 mins	25-03-

In [132...

```
df2['language'].sample(10)
```

Out[132...

```
8867      Hindi
62541    English
11944    spanish
34764    English
10822    English
6980     English
78560    English
34272    English
46394    English
37045    English
Name: language, dtype: object
```

In [133...

```
df2['language']=df2['language'].str.title()
```

In [134...

```
df2['language'].sample(10)
```

```
Out[134...] 46222 English
             61491 English
             3030  English
             31919 English
             54341 English
             4391  English
             61002 Spanish
             85818 English
             81897 English
             18277 English
Name: language, dtype: object
```

```
In [160...] df2['price'].sample(20)
```

```
Out[160...] 32717      74.0
             57924     668.0
             2684     268.0
             69646    1505.0
             68956     697.0
             29157     668.0
             64363    1063.0
             71251     469.0
             21869     754.0
             75703     938.0
             54950     586.0
             50620     379.0
             33037     376.0
             25056     586.0
             69798     586.0
             61437     501.0
             60096    1172.0
             22621    1005.0
             37278    1338.0
             44835     645.0
Name: price, dtype: float64
```

```
In [136...] df2[df2['price']=='Free']
```



Out[136...

	name	author	narrator	time	releasedate
983	Ringo & the Yogi	PrasunRoy	SiddhantaPinto	5 hrs and 18 mins	19-11-19
1414	The Velveteen Rabbit	MargeryWilliams	RichardArmitage	24 mins	14-08-20
1513	Billi Aur Moosarani [Cat and Musrani]	Shivani	SarahHashmi	26 mins	14-10-21
2932	Pariyon Ki Kahaniya [Fairy Tales]	JacobGrimm	AnuradhaChauhan	2 hrs and 15 mins	08-12-21
4547	101 Best Indian Fables for Children	StutiGupta	SunandhaRaghunathan	2 hrs and 26 mins	03-12-19
...	...	...	...	...	...
80021	FREE: The Undead	RRHaywood	DanMorgan	1 hr and 3 mins	24-02-15
81198	Supporters. Le Pilote	PierreNiboyet,AudreyPoiret	BorisRehlinger	24 mins	29-10-20
82644	Boots Belts Berets	TanushreePodder	RannvijaySinghSingha	6 hrs and 59 mins	27-08-19
86771	Canada Is Awesome	NeilPasricha	NeilPasricha	34 mins	09-10-18
86927	Osez l'aventure avec Mike Horn. Le pilote	MikeHorn,CharlesAudier	MikeHorn,CharlesAudier	24 mins	11-01-22

338 rows × 8 columns



In [137... (df2['price']=='Free').sum()

Out[137... 338

In [138... df2['price']=df2['price'].str.replace('Free','0')

In [139... (df2['price']=='Free').sum()

Out[139... 0

In [140... df2.sample(5)

Out[140... 

	name	author	narrator	tim
--	------	--------	----------	-----

58965	La Scienza e l'Arte della Ricchezza	JosephMurphy,WallaceD.Wattles,P.T.Barnum	StefanoTrillini,VitaWulff	h an 4 mir
-------	---	--	---------------------------	---------------------

72715	The Hex of Blackbriar Academy	OliviaAsh	AmandaDolan	h an 3 mir
-------	-------------------------------------	-----------	-------------	---------------------

28165	Project Management	MatthewBachelor	uncredited	h an 2 mir
-------	-----------------------	-----------------	------------	---------------------

8901	Love Made	QuinaAragon	QuinaAragon	mir
------	-----------	-------------	-------------	-----

43240	De Profundis (Portuguese Edition)	OscarWilde	RobertoGarcia	h an 3 mir
-------	---	------------	---------------	---------------------

◀  ▶

In [141... df2['time'].sample(20)

```
Out[141... 79754    13 hrs and 10 mins
64104     6 hrs and 31 mins
18042     8 hrs and 48 mins
59470     7 hrs and 16 mins
16450              33 mins
84104     3 hrs and 55 mins
25495     6 hrs and 50 mins
49503     5 hrs and 20 mins
30742     7 hrs and 10 mins
17662     4 hrs and 32 mins
16428              22 mins
85059     7 hrs and 26 mins
8574      3 hrs and 20 mins
48744              15 mins
45297     4 hrs and 9 mins
19150    26 hrs and 56 mins
57447    16 hrs and 34 mins
1305              53 mins
59676    10 hrs and 43 mins
86984              38 mins
Name: time, dtype: object
```

```
In [142... import re

def convert_to_minutes(time_str):
    match = re.match(r'(?:(\d+)\s*hrs)?\s*(?:and\s*)?(?:(\d+)\s*mins)?', time_

    if match:
        hours = int(match.group(1)) if match.group(1) else 0
        minutes = int(match.group(2)) if match.group(2) else 0
        total_minutes = hours * 60 + minutes
        return f"{total_minutes} min"
    else:
        return time_str

df2['time'] = df2['time'].apply(convert_to_minutes)
```

```
In [143... df2['time'].sample(20)
```

```
Out[143... 84043      466 min
37229      370 min
39233      673 min
65067     1056 min
22304     1212 min
44506     1071 min
53666       26 min
24141      356 min
34903       81 min
36357      342 min
20751      767 min
9808       131 min
56240      189 min
19912     2114 min
86883      806 min
63011      282 min
27418      667 min
6449        10 min
70631      467 min
18194      609 min
Name: time, dtype: object
```

```
In [144... df2.sample(5)
```

```
Out[144...
```

	name	author	narrator	time	releasedate	language	s
<b>34725</b>	The Coffee Lover's Diet	Dr.BobArnot	Dr.BobArnot	366 min	20-06-17	English	r
<b>4429</b>	How I Do My Homework	JenniferBoothroyd	Intuitive	1 min	22-01-20	English	r
<b>56404</b>	Blind Faith	RebeccaZanetti	KarenWhite	726 min	27-01-15	English	r
<b>21810</b>	Wallenstein	JakobWassermann	MarkusKrochmann	74 min	11-03-11	German	r
<b>69518</b>	Disunity in Christ	ChristenaCleveland	RandyKaye	388 min	24-10-17	English	r



```
In [145... df['stars'].sample(20)
```

```

Out[145... 27775      Not rated yet
           27005      Not rated yet
           18984      Not rated yet
           44590      4.5 out of 5 stars3 ratings
           52739      Not rated yet
           43575      Not rated yet
           9596       Not rated yet
           37443      Not rated yet
           10884      Not rated yet
           40094      Not rated yet
           22894      Not rated yet
           8350       Not rated yet
           32056      Not rated yet
           84865      Not rated yet
           83685      Not rated yet
           66969      Not rated yet
           44108      Not rated yet
           70252      Not rated yet
           24926      5 out of 5 stars1 rating
           33091      Not rated yet
Name: stars, dtype: object

```

```

In [146... df2['stars']=df2['stars'].str.replace('Not rated yet','NaN')

```

```

In [147... df2['stars'].sample(20)

```

```

Out[147... 4519      NaN
           36740      NaN
           61937      NaN
           72561      NaN
           76968      NaN
           79019      NaN
           55068      NaN
           75829      NaN
           5322       NaN
           502        NaN
           4751       NaN
           35906      NaN
           25112      NaN
           86033      NaN
           173        NaN
           41214      NaN
           24444      NaN
           67052      NaN
           69254      NaN
           57014      NaN
Name: stars, dtype: object

```

```

In [148... df2.sample(5)

```

Out[148...

	name	author	narrator	time	releasedate
<b>83105</b>	Tempesta	WilburSmith,ChrisWakling	FrancescoLeonardoFabbri	462 min	27-10-21
<b>4235</b>	Sir Cumference and the Sword in the Cone	CindyNeuschwander	AdamWeber	12 min	27-11-19
<b>15409</b>	War Machine: I mille volti degli Stati Uniti n...	FrancescoDeVito	MarcoCavicchioli	83 min	04-05-17
<b>19848</b>	Reagan	H.W.Brands	StephenHoye	1901 min	12-05-15
<b>59946</b>	Das doppelte Gesicht	HeidiRehn	MarkusBachmann	507 min	24-09-21

In [149...

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 87489 entries, 0 to 87488
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   name             87489 non-null  object
1   author           87489 non-null  object
2   narrator         87489 non-null  object
3   time             87489 non-null  object
4   releasedate      87489 non-null  object
5   language         87489 non-null  object
6   stars            87489 non-null  object
7   price            87489 non-null  object
dtypes: object(8)
memory usage: 5.3+ MB
```

In [155...

```
df2['price'] = df2['price'].str.replace(',', '').astype(float)
```

In [151...

```
df2['releasedate']=df['releasedate'].astype('datetime64[ns]')
```

C:\Users\arnak\AppData\Local\Temp\ipykernel\_15480\2816894412.py:1: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.

```
df2['releasedate']=df['releasedate'].astype('datetime64[ns]')
```

In [ ]:

In [156...

```
df2.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 87489 entries, 0 to 87488
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   name        87489 non-null  object
1   author      87489 non-null  object
2   narrator    87489 non-null  object
3   time        87489 non-null  object
4   releasedate 87489 non-null  datetime64[ns]
5   language    87489 non-null  object
6   price       87489 non-null  float64
dtypes: datetime64[ns](1), float64(1), object(5)
memory usage: 4.7+ MB

```

In [153... `df2.drop(columns=['stars'],inplace=True)`

In [159... `df2.sample(5)`

Out[159...

	name	author	narrator	time	releasedate	la
<b>41881</b>	Empire of Booze	HenryJeffreys	DavidThorpe	562 min	2018-01-09	
<b>34997</b>	Our Malady	TimothySnyder	TimothySnyder	181 min	2020-10-09	
<b>31924</b>	Le pouvoir de la volonté	RoyF.Baumeister,JohnTierney	JérômeCarrette	680 min	2020-07-30	
<b>38514</b>	Cäsar lässt grüßen - Die Geschichte der Römer 2	JoachimFernau	PhillippKreisselmeier	102 min	2009-04-09	
<b>26618</b>	Cry of Murder on Broadway	JulieMiller	TaviaGilbert	417 min	2020-10-15	