In [1]:
```
# importing major libraries
import pandas as pd
import numpy as np
```

In [2]:
```
# dataframes --> tables
l1=[101,90,10]
l2=[98,70,5]
l3=[104,95,22]
pd.DataFrame([l1,l2,l3],columns=['iq','Marks','PKG'],index=['Gaurav','Anshu','
```

Out[2]:

	iq	Marks	PKG
Gaurav	101	90	10
Anshu	98	70	5
Vipul	104	95	22

In [3]:
```
# dictionaries --> dataframe
dict1={
    'Name':['Gaurav','Shariq','Vipul'],
    'Course':['Data Analytics','Data Science','Data Science'],
    'PKG':[10,20,30],
    'Trainer':['Anshum','Sanchit','Aseem']
}
pd.DataFrame(dict1)
```

Out[3]:

	Name	Course	PKG	Trainer
0	Gaurav	Data Analytics	10	Anshum
1	Shariq	Data Science	20	Sanchit
2	Vipul	Data Science	30	Aseem

In [4]:
```
# numpy arrays --> dataframe
pd.DataFrame(np.random.randint(0,101,600).reshape(100,6),columns=['A','B','C',
```

Out[4]:

	A	B	C	D	E	F
0	24	31	8	65	17	67
1	14	40	20	37	84	8
2	50	44	69	49	100	59
3	19	22	27	43	51	10
4	31	58	5	27	37	62
...	...	...	...	...	...	...
95	22	15	36	67	94	9
96	11	0	90	94	22	81
97	21	24	5	14	11	58
98	8	56	66	99	89	74
99	24	38	7	33	34	27

100 rows × 6 columns

In [5]:
```
df=pd.read_csv('movies.csv')
```

In [6]:
```
df
#manual assessing
```

Out[6]:

	title_x	imdb_id	poster_path	wiki_link	
0	Uri: The Surgical Strike	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Uri:_The_Surgica...	
1	Battalion 609	tt9472208	NaN	https://en.wikipedia.org/wiki/Battalion_609	E
2	The Accidental Prime Minister (film)	tt6986710	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/The_Accidental_P...	Ac
3	Why Cheat India	tt8108208	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Why_Cheat_India	
4	Evening Shadows	tt6028796	NaN	https://en.wikipedia.org/wiki/Evening_Shadows	S
...	...	...	...	...	
1624	Tera Mera Saath Rahen	tt0301250	https://upload.wikimedia.org/wikipedia/en/2/2b...	https://en.wikipedia.org/wiki/Tera_Mera_Saath_...	Te
1625	Yeh Zindagi Ka Safar	tt0298607	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Yeh_Zindagi_Ka_S...	
1626	Sabse Bada Sukh	tt0069204	NaN	https://en.wikipedia.org/wiki/Sabse_Bada_Sukh	
1627	Daaka	tt10833860	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Daaka	
1628	Humsafar	tt2403201	https://upload.wikimedia.org/	https://en.wikipedia.org/	H

	title_x	imdb_id	poster_path	wiki_link
			wikipedia/en/thum...	wiki/Humsafar

1629 rows × 18 columns

In [7]:
```
#Attributes
#shape
df.shape
```

Out[7]:
```
(1629, 18)
```

In [8]:
```
#ndim
df.ndim
```

Out[8]:
```
2
```

In [9]:
```
#columns
df.columns
```

Out[9]:
```
Index(['title_x', 'imdb_id', 'poster_path', 'wiki_link', 'title_y',
       'original_title', 'is_adult', 'year_of_release', 'runtime', 'genres',
       'imdb_rating', 'imdb_votes', 'story', 'summary', 'tagline', 'actors',
       'wins_nominations', 'release_date'],
      dtype='object')
```

In [10]:
```
#size
df.size
```

Out[10]:
```
29322
```

In [11]:
```
#1629*18=29322(r*c)
```

In [12]:
```
#values
#index
df.index
df.values[0]
```

Out[12]:
```
array(['Uri: The Surgical Strike', 'tt8291224',
       'https://upload.wikimedia.org/wikipedia/en/thumb/3/3b/URI_-_New_poste
r.jpg/220px-URI_-_New_poster.jpg',
       'https://en.wikipedia.org/wiki/Uri:_The_Surgical_Strike',
       'Uri: The Surgical Strike', 'Uri: The Surgical Strike', 0, 2019,
       '138', 'Action|Drama|War', 8.4, 35112,
       'Divided over five chapters  the film chronicles the events of the sur
gical strike conducted by the Indian military against suspected militants in
Pakistan occupied Kashmir. It tells the story of the 11 tumultuous events ove
r which the operation was carried out. Indian army special forces carry out a
covert operation to avenge the killing of fellow army men at their base by a
terrorist group.',
       'Indian army special forces execute a covert operation  avenging the k
illing of fellow army men at their base by a terrorist group.',
       nan,
       'Vicky Kaushal|Paresh Rawal|Mohit Raina|Yami Gautam|Kirti Kulhari|Raji
t Kapoor|Ivan Rodrigues|Manasi Parekh|Swaroop Sampat|Riva Arora|Yogesh Soma
n|Fareed Ahmed|Akashdeep Arora|Kallol Banerjee|',
       '4 wins', '11 January 2019 (USA)'], dtype=object)
```

In [13]:
```
#dtypes
df.dtypes
```

Out[13]:
```
title_x              object
imdb_id              object
poster_path          object
wiki_link            object
title_y              object
original_title       object
is_adult              int64
year_of_release       int64
runtime              object
genres               object
imdb_rating         float64
imdb_votes            int64
story                object
summary              object
tagline              object
actors               object
wins_nominations     object
release_date         object
dtype: object
```

In [14]:
```
#index related attributes
#iloc
#loc
```

In [15]:
```
#iloc
df.iloc[0:5,0:5]
```

Out[15]:

	title_x	imdb_id	poster_path	wiki_link	title
0	Uri: The Surgical Strike	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Uri:_The_Surgica...	Uri: T Surgi Str
1	Battalion 609	tt9472208	NaN	https://en.wikipedia.org/wiki/Battalion_609	Battal 6
2	The Accidental Prime Minister (film)	tt6986710	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/The_Accidental_P...	T Acciden Pri Minis
3	Why Cheat India	tt8108208	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Why_Cheat_India	W Ch In
4	Evening Shadows	tt6028796	NaN	https://en.wikipedia.org/wiki/Evening_Shadows	Even Shado

In [16]:
```
#fancy indexing
df.iloc[0:5,[0,2,1,7,9]]
```

Out[16]:

	title_x	poster_path	imdb_id	year_of_release	gen
0	Uri: The Surgical Strike	https://upload.wikimedia.org/wikipedia/en/thum...	tt8291224	2019	Action|Drama|
1	Battalion 609	NaN	tt9472208	2019	
2	The Accidental Prime Minister (film)	https://upload.wikimedia.org/wikipedia/en/thum...	tt6986710	2019	Biography|Dra
3	Why Cheat India	https://upload.wikimedia.org/wikipedia/en/thum...	tt8108208	2019	Crime|Dra
4	Evening Shadows	NaN	tt6028796	2018	Dra

In [17]:
```
#loc
#fancy indexes
df.loc[0:5,['title_x','runtime','genres']]
df.loc[0:5,'title_x':'genres']
```

Out[17]:

	title_x	imdb_id	poster_path	wiki_link	title
0	Uri: The Surgical Strike	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Uri:_The_Surgica...	Uri: T Surgi Str
1	Battalion 609	tt9472208	NaN	https://en.wikipedia.org/wiki/Battalion_609	Battal 6
2	The Accidental Prime Minister (film)	tt6986710	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/The_Accidental_P...	T Acciden Pri Minis
3	Why Cheat India	tt8108208	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Why_Cheat_India	W Ch In
4	Evening Shadows	tt6028796	NaN	https://en.wikipedia.org/wiki/Evening_Shadows	Even Shado
5	Soni (film)	tt6078866	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Soni_(film)	S

In [18]:
```
#head
#tail
#sample
```

In [19]:
```
df.head(10)
```

Out[19]:

	title_x	imdb_id	poster_path	wiki_link	
0	Uri: The Surgical Strike	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Uri:_The_Surgica...	
1	Battalion 609	tt9472208	NaN	https://en.wikipedia.org/wiki/Battalion_609	B
2	The Accidental Prime Minister (film)	tt6986710	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/The_Accidental_P...	Acc
3	Why Cheat India	tt8108208	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Why_Cheat_India	Wh
4	Evening Shadows	tt6028796	NaN	https://en.wikipedia.org/wiki/Evening_Shadows	S
5	Soni (film)	tt6078866	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Soni_(film)	
6	Fraud Saiyaan	tt5013008	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Fraud_Saiyaan	
7	Bombairiya	tt4971258	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Bombairiya	Bom
8	Manikarnika: The Queen of Jhansi	tt6903440	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Manikarnika:_The...	Manil The o

	title_x	imdb_id	poster_path	wiki_link	
9	Thackeray (film)	tt7777196	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Thackeray_(film)	Tha

In [20]:
```
df.tail()
```

Out[20]:

	title_x	imdb_id	poster_path	wiki_link	
1624	Tera Mera Saath Rahen	tt0301250	https://upload.wikimedia.org/wikipedia/en/2/2b...	https://en.wikipedia.org/wiki/Tera_Mera_Saath_...	
1625	Yeh Zindagi Ka Safar	tt0298607	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Yeh_Zindagi_Ka_S...	K
1626	Sabse Bada Sukh	tt0069204	NaN	https://en.wikipedia.org/wiki/Sabse_Bada_Sukh	
1627	Daaka	tt10833860	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Daaka	
1628	Humsafar	tt2403201	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Humsafar	Hu

In [21]:
```
df.sample()
```

Out[21]:

	title_x	imdb_id	poster_path	wiki_link	title_y
721	Ajab Gazabb Love	tt2356959	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Ajab_Gazabb_Love	Ajab Gazabb Love

In [22]:
```
#rename
df.columns
df.rename(columns={'title_x':'title','imdb_id':'id','year_of_release':'release
df.rename(columns={'title_x':'title','imdb_id':'id','year_of_release':'release
df.columns
```

Out[22]:
```
Index(['title', 'id', 'poster_path', 'wiki_link', 'title_y', 'original_titl
e',
       'is_adult', 'release_year', 'runtime', 'genres', 'imdb_rating',
       'imdb_votes', 'story', 'summary', 'tagline', 'actors',
       'wins_nominations', 'release_date'],
      dtype='object')
```

In [23]:
```
#info
#describe()
#seeking info()
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1629 entries, 0 to 1628
Data columns (total 18 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   title             1629 non-null   object
 1   id                1629 non-null   object
 2   poster_path       1526 non-null   object
 3   wiki_link         1629 non-null   object
 4   title_y           1629 non-null   object
 5   original_title    1629 non-null   object
 6   is_adult          1629 non-null   int64
 7   release_year      1629 non-null   int64
 8   runtime           1629 non-null   object
 9   genres            1629 non-null   object
 10  imdb_rating       1629 non-null   float64
 11  imdb_votes        1629 non-null   int64
 12  story             1609 non-null   object
 13  summary           1629 non-null   object
 14  tagline           557 non-null    object
 15  actors            1624 non-null   object
 16  wins_nominations  707 non-null    object
 17  release_date      1522 non-null   object
dtypes: float64(1), int64(3), object(14)
memory usage: 229.2+ KB
```

Information

- our dataset consists of 1629 entries
- our dataset contains null values
- validity issue,runtime is in object type
- 4 numerical columns and 14 object type columns

In [24]:
```
#descibe
df.imdb_rating.mean()
df.imdb_rating.std()
df.imdb_rating.quantile(0.25)
df.imdb_rating.quantile(0.50)
df.imdb_rating.quantile(0.75)
df.describe()
```

Out[24]:

	is_adult	release_year	imdb_rating	imdb_votes
count	1629.0	1629.000000	1629.000000	1629.000000
mean	0.0	2010.263966	5.557459	5384.263352
std	0.0	5.381542	1.567609	14552.103231
min	0.0	2001.000000	0.000000	0.000000
25%	0.0	2005.000000	4.400000	233.000000
50%	0.0	2011.000000	5.600000	1000.000000
75%	0.0	2015.000000	6.800000	4287.000000
max	0.0	2019.000000	9.400000	310481.000000

Information

- variance threshold --> is adult,experts 0.0-0.05,drop column
- potential outliers in imdb_votes
- movies from year 2001-2019

In [25]:
```
# sort_values
# sort_index
# reset_index
```

In [26]:
```
df.sort_values(by='release_year')
```

Out[26]:

	title	id	poster_path	wiki_link	
1625	Yeh Zindagi Ka Safar	tt0298607	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Yeh_Zindagi_Ka_S...	
1624	Tera Mera Saath Rahen	tt0301250	https://upload.wikimedia.org/wikipedia/en/2/2b...	https://en.wikipedia.org/wiki/Tera_Mera_Saath_...	
1623	Zubeidaa	tt0255713	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Zubeidaa	
1622	Yeh Teraa Ghar Yeh Meraa Ghar	tt0298606	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Yeh_Teraa_Ghar_Y...	
1621	Tum Bin	tt0290326	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Tum_Bin	
...	...	...	...	...	
45	Chicken Curry Law	tt7189494	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Chicken_Curry_Law	
47	Jabariya Jodi	tt8785426	https://upload.wikimedia.org/wikipedia/en/9/97...	https://en.wikipedia.org/wiki/Jabariya_Jodi	
44	Judgementall Hai Kya	tt8108196	https://upload.wikimedia.org/wikipedia/en/c/c8...	https://en.wikipedia.org/wiki/Judgementall_Hai...	J
19	Risknamaa	tt9795264	NaN	https://en.wikipedia.org/wiki/Risknamaa	
0	Uri: The Surgical	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Uri:_The_Surgica...	

	title	id	poster_path	wiki_link
	Strike			

1629 rows × 18 columns

In [27]:
```
df.sort_index()
```

Out[27]:

	title	id	poster_path	wiki_link	
0	Uri: The Surgical Strike	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Uri:_The_Surgica...	
1	Battalion 609	tt9472208	NaN	https://en.wikipedia.org/wiki/Battalion_609	E
2	The Accidental Prime Minister (film)	tt6986710	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/The_Accidental_P...	Ac
3	Why Cheat India	tt8108208	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Why_Cheat_India	
4	Evening Shadows	tt6028796	NaN	https://en.wikipedia.org/wiki/Evening_Shadows	S
...	...	...	...	...	
1624	Tera Mera Saath Rahen	tt0301250	https://upload.wikimedia.org/wikipedia/en/2/2b...	https://en.wikipedia.org/wiki/Tera_Mera_Saath_...	Te
1625	Yeh Zindagi Ka Safar	tt0298607	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Yeh_Zindagi_Ka_S...	
1626	Sabse Bada Sukh	tt0069204	NaN	https://en.wikipedia.org/wiki/Sabse_Bada_Sukh	
1627	Daaka	tt10833860	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.org/wiki/Daaka	
1628	Humsafar	tt2403201	https://upload.wikimedia.org/	https://en.wikipedia.org/	H

	title	id	poster_path	wiki_link
			wikipedia/en/thum...	wiki/Humsafar

1629 rows × 18 columns

In [28]:
```
df.reset_index()
```

Out[28]:

	index	title	id	poster_path	wiki_
0	0	Uri: The Surgical Strike	tt8291224	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.wiki/Uri:_The_Surgi
1	1	Battalion 609	tt9472208	NaN	https://en.wikipedia.wiki/Battalion_
2	2	The Accidental Prime Minister (film)	tt6986710	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.wiki/The_Accidental
3	3	Why Cheat India	tt8108208	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.wiki/Why_Cheat_I
4	4	Evening Shadows	tt6028796	NaN	https://en.wikipedia.wiki/Evening_Shad
...	...	...	...	...	
1624	1624	Tera Mera Saath Rahen	tt0301250	https://upload.wikimedia.org/wikipedia/en/2/2b...	https://en.wikipedia.Tera_Mera_Saat
1625	1625	Yeh Zindagi Ka Safar	tt0298607	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.wiki/Yeh_Zindagi_Ka
1626	1626	Sabse Bada Sukh	tt0069204	NaN	https://en.wikipedia.wiki/Sabse_Bada_S
1627	1627	Daaka	tt10833860	https://upload.wikimedia.org/wikipedia/en/thum...	https://en.wikipedia.wiki/Da
1628	1628	Humsafar	tt2403201	https://upload.wikimedia.org/	https://en.wikipedia.

index	title	id	poster_path	wiki_
			wikipedia/en/thum...	wiki/Hums

1629 rows × 19 columns

In [29]:
```
df.genres.value_counts().reset_index()
```

Out[29]:

	genres	count
0	Drama	162
1	Comedy|Drama|Romance	101
2	Comedy|Drama	88
3	Drama|Romance	86
4	Action|Crime|Drama	86
...	...	...
200	Action|Musical|Romance	1
201	Documentary|War	1
202	Action|Crime|Horror	1
203	Comedy|Fantasy	1
204	Comedy|Musical|Mystery	1

205 rows × 2 columns

In [30]:
```
# checking null values
df.isnull().sum().sum()
df.isnull().sum()
df.isnull().mean()*100

#theory
#0-5% null values--> mean,median,mode
#5-10% null values --> cautions,mean,median,mode
#10-30% null values --> knn imputers'multivariate imputions --> machine learni
# >>>40 --> patterns,check if the column is important or not
#>>60 --> plenty of column is empty --> drop column
```

Out[30]:
```
title                0.000000
id                   0.000000
poster_path          6.322897
wiki_link            0.000000
title_y              0.000000
original_title       0.000000
is_adult             0.000000
release_year         0.000000
runtime              0.000000
genres               0.000000
imdb_rating          0.000000
imdb_votes           0.000000
story                1.227747
summary              0.000000
tagline             65.807244
actors               0.306937
wins_nominations    56.599141
release_date         6.568447
dtype: float64
```

Action:

- poster_path --> placeholder--> unknown
- story --> placeholder/no description/dropna
- tagline --> drop column
- actors --> dropna
- win_nomination --> drop column
- release_date--> dropna

In [31]:
```
df.release_date.str.split(expand=True).loc[:,[1]]
#pd.to_datetime()--> formula for date and time this is not fit in this data be
```

Out[31]:

	1
0	January
1	January
2	January
3	January
4	January
...	...
1624	November
1625	November
1626	NaN
1627	November
1628	Series

1629 rows × 1 columns

In [32]:

```
# implement
#filling
#dropna
#drop
df.poster_path.fillna('unknown',inplace=True)
df.story.fillna('no description',inplace=True)
df.drop('tagline',axis=1,inplace=True)
df.dropna(subset='actors')
df.drop(columns='wins_nominations',inplace=True)
df.dropna(subset='release_date',inplace=True)
#once you call simple df
#it give the output here it give give warning only no output for the upper coc
```

```
C:\Users\vipul\AppData\Local\Temp\ipykernel_19992\875845210.py:5: FutureWarnin
g: A value is trying to be set on a copy of a DataFrame or Series through chain
ed assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work bec
ause the intermediate object on which we are setting values always behaves as a
copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.me
thod({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, t
o perform the operation inplace on the original object.


  df.poster_path.fillna('unknown',inplace=True)
C:\Users\vipul\AppData\Local\Temp\ipykernel_19992\875845210.py:6: FutureWarnin
g: A value is trying to be set on a copy of a DataFrame or Series through chain
ed assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work bec
ause the intermediate object on which we are setting values always behaves as a
copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.me
thod({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, t
o perform the operation inplace on the original object.


  df.story.fillna('no description',inplace=True)
```

In [33]:
```
#fetching columns
df.title
```

Out[33]:
```
0                   Uri: The Surgical Strike
1                           Battalion 609
2       The Accidental Prime Minister (film)
3                           Why Cheat India
4                           Evening Shadows
                        ...
1623                                Zubeidaa
1624                  Tera Mera Saath Rahen
1625                   Yeh Zindagi Ka Safar
1627                                  Daaka
1628                                Humsafar
Name: title, Length: 1522, dtype: object
```

In [34]:
```
df['title']
```

Out[34]:
```
0                    Uri: The Surgical Strike
1                              Battalion 609
2       The Accidental Prime Minister (film)
3                            Why Cheat India
4                            Evening Shadows
                          ...
1623                                Zubeidaa
1624                   Tera Mera Saath Rahen
1625                    Yeh Zindagi Ka Safar
1627                                   Daaka
1628                                 Humsafar
Name: title, Length: 1522, dtype: object
```

In [35]:
```
#loc
df.loc[:,['title']]
```

Out[35]:

	title
0	Uri: The Surgical Strike
1	Battalion 609
2	The Accidental Prime Minister (film)
3	Why Cheat India
4	Evening Shadows
...	...
1623	Zubeidaa
1624	Tera Mera Saath Rahen
1625	Yeh Zindagi Ka Safar
1627	Daaka
1628	Humsafar

1522 rows × 1 columns

In [36]:
```
#iloc
df.iloc[:,[0]]
```

Out[36]:

	title
0	Uri: The Surgical Strike
1	Battalion 609
2	The Accidental Prime Minister (film)
3	Why Cheat India
4	Evening Shadows
...	...
1623	Zubeidaa
1624	Tera Mera Saath Rahen
1625	Yeh Zindagi Ka Safar
1627	Daaka
1628	Humsafar

1522 rows × 1 columns

In [37]:
```
# features selections
#manual,selectkbest,subset,optuna
# manual filteration
```

In [38]:
```
df.columns
```

Out[38]:
```
Index(['title', 'id', 'poster_path', 'wiki_link', 'title_y', 'original_titl
e',
       'is_adult', 'release_year', 'runtime', 'genres', 'imdb_rating',
       'imdb_votes', 'story', 'summary', 'actors', 'release_date'],
      dtype='object')
```

In [39]:
```
df.drop(columns=['poster_path','wiki_link','title_y','original_title',
                 'is_adult','story'],inplace=True)
```

In [40]:
```
df.shape
```

Out[40]:
```
(1522, 10)
```

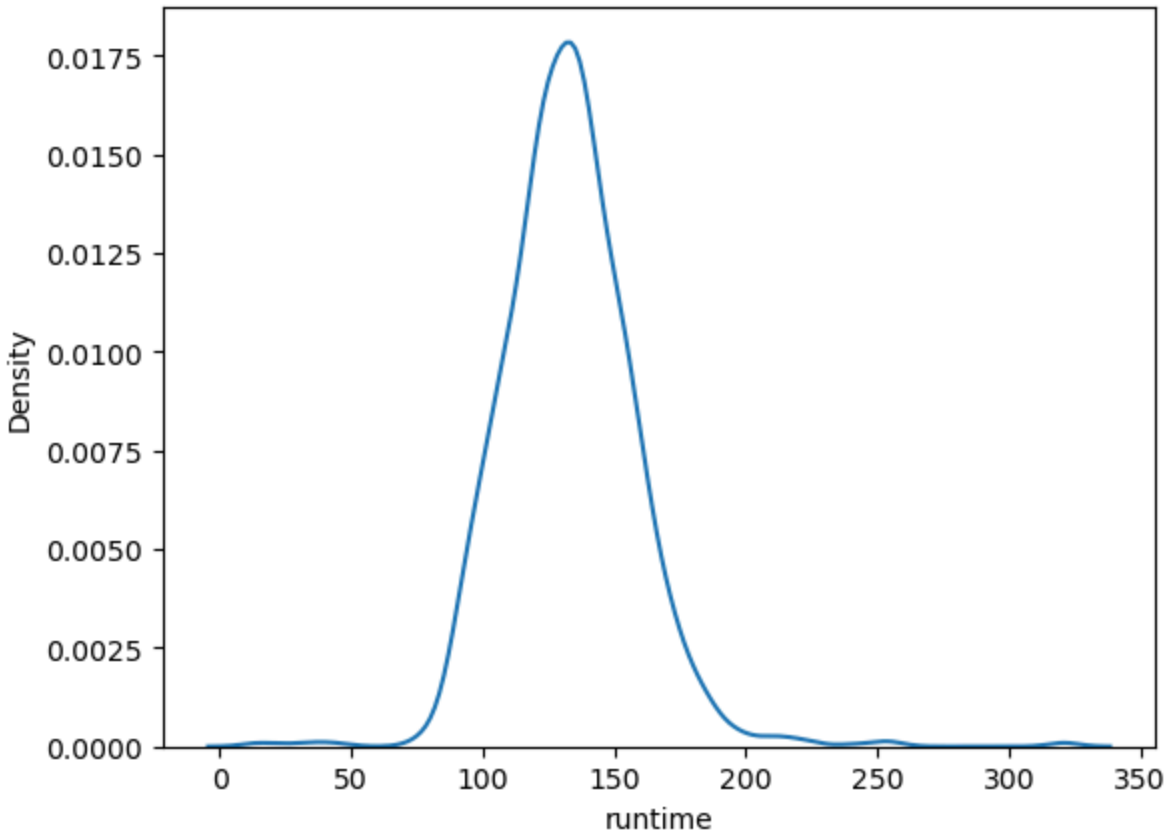
In [42]:
```
#reordering
df=df.iloc[:,[1,0,2,3,4,5,6,7,8,9]]
df.head()
```

Out[42]:

	id	title	release_year	runtime	genres	imdb_rating	im
0	tt8291224	Uri: The Surgical Strike	2019	138	Action|Drama|War	8.4	
1	tt9472208	Battalion 609	2019	131	War	4.1	
2	tt6986710	The Accidental Prime Minister (film)	2019	112	Biography|Drama	6.1	
3	tt8108208	Why Cheat India	2019	121	Crime|Drama	6.0	
4	tt6028796	Evening Shadows	2018	102	Drama	7.3	

In [45]:

```
#apply
#astype
#df.runtime.astype(int)
#gaussian or skewed
df.runtime[df.runtime.str.isnumeric()==False]#condition
#np.where--> np.where (condition,value,where)
df.runtime[df.runtime.str.isnumeric()==True].apply(lambda x: int(x)).median()#
import seaborn as sns
sns.kdeplot(df.runtime[df.runtime.str.isnumeric()==True].apply(lambda x: int(x
#median --> 132
df.runtime =np.where(df.runtime.str.isnumeric()==False,132,df.runtime)
df.runtime=df.runtime.astype(int)
```



In [46]:
```
df.runtime.dtype
```

Out[46]:
```
dtype('int64')
```

In [48]:
```
#kurtosis
df.runtime.kurtosis()
```

Out[48]:
```
np.float64(6.181511578056849)
```

Kurtosis

- >0 --> leptokurtic
- <0 --> platykurtic
- ~0 --> mesokurtic

In [49]:
```
#skewness
#left
#right
#gaussian
df.runtime.skew()
```

Out[49]:
```
np.float64(0.8240655151809133)
```

- ~0 --> gaussian
- >0 --> right skewed
- <0 --> left skewed

In [50]:
```
#corr()
df.corr(numeric_only=True)
```

Out[50]:

	release_year	runtime	imdb_rating	imdb_votes
release_year	1.000000	-0.242367	0.116570	0.043594
runtime	-0.242367	1.000000	0.022013	0.297058
imdb_rating	0.116570	0.022013	1.000000	0.350850
imdb_votes	0.043594	0.297058	0.350850	1.000000

In [51]:
```
#groupby --> aggregate functions
```

In [ ]:
```

```