```
import pandas as pd

df = pd.read_csv("/content/Music.csv")

#print all records of dataset
print(df)
#print artist_name
print(df['artist_name'])
#print name & release Date
print(df[['track_name','release_date']])
```

```
      15227   29  0.001880  0.001880    0.001880  ...  0.162384  0.001880
      15228   40  0.106823  0.001645    0.189822  ...  0.320371  0.001645
      15229   57  0.001012  0.053893    0.053253  ...  0.437774  0.001012
      15230   39  0.002392  0.002392    0.485467  ...       NaN       NaN

            danceability  loudness  acousticness  instrumentalness   valence  \
      0          0.357739  0.454119      0.997992          0.901822  0.339448
      1          0.331745  0.647540      0.954819          0.000002  0.325021
      2          0.456298  0.585288      0.840361          0.000000  0.351814
      3          0.686992  0.744404      0.083935          0.199393  0.775350
      4          0.291671  0.646489      0.975904          0.000246  0.597073
      ...             ...       ...           ...               ...       ...
      15226      0.636088  0.677512      0.741968          0.000000  0.967024
      15227      0.331745  0.427198      0.356425          0.008957  0.305441
      15228      0.534279  0.725098      0.353413          0.000000  0.793899
      15229      0.463880  0.657129      0.664658          0.000419  0.158079
      15230           NaN       NaN           NaN               NaN       NaN

              energy       topic       age
      0      0.137110     sadness  1.000000
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
      4      0.394375    romantic  1.000000
      ...         ...         ...       ...
      15226  0.691682    violence  0.342857
      15227  0.199174    romantic  0.342857
      15228  0.475459     sadness  0.342857
      15229  0.241218     sadness  0.342857
      15230       NaN         NaN       NaN

      [15231 rows x 31 columns]
      0                              mukesh
      1                       frankie laine
      2                         johnnie ray
      3                         pérez prado
      4                  giorgos papadopoulos
                             ...
      15226                  aretha franklin
      15227    the brian jonestown massacre
      15228                       taj mahal
      15229                       beth hart
      15230                       oblivians
      Name: artist_name, Length: 15231, dtype: object
                      track_name  release_date
      0       mohabbat bhi jhoothi          1950
      1                  i believe          1950
      2                        cry          1950
      3                   patricia          1950
      4          apopse eida oneiro          1950
      ...                       ...           ...
      15226  can't you just see me          1996
      15227              true love          1996
      15228          lonely avenue          1996
      15229         blame the moon          1996
      15230                trouble          1996

      [15231 rows x 2 columns]
```

```
df1 = df[['track_name','release_date']]
print(df1)
```

```
                      track_name  release_date
      0       mohabbat bhi jhoothi          1950
```

```
1                 i believe          1950
2                       cry          1950
3                   patricia          1950
4          apopse eida oneiro          1950
...                       ...           ...
15226  can't you just see me          1996
15227              true love          1996
15228          lonely avenue          1996
15229         blame the moon          1996
15230                trouble          1996

[15231 rows x 2 columns]
```

```
print(df.groupby('release_date').mean())
```

```
                Unnamed: 0         len    dating  violence  world/life  \
release_date
1950          17913.321429   60.357143  0.014076  0.065540    0.132519
1951          18292.562500   67.531250  0.007259  0.092994    0.139605
1952          13389.031250   60.968750  0.008117  0.066209    0.098522
1953          18262.913043   54.217391  0.025343  0.063537    0.174455
1954          15294.862069   51.344828  0.018001  0.149862    0.132404
...                    ...         ...       ...       ...         ...
2015          32115.629870   97.266234  0.025182  0.120196    0.074959
2016          26267.771144  108.155000  0.021388  0.102792    0.075707
2017          27301.448864  111.170455  0.025992  0.077854    0.092731
2018          26875.000000  124.324176  0.017965  0.081737    0.066533
2019          29395.006623  109.569536  0.018422  0.072054    0.089110

              night/time  shake the audience  family/gospel  romantic  \
release_date
1950            0.092779            0.004044       0.012586  0.129017
1951            0.065741            0.003518       0.010707  0.057245
1952            0.037098            0.006505       0.032021  0.095474
1953            0.040227            0.015705       0.013060  0.085135
1954            0.023392            0.011467       0.022358  0.063821
...                  ...                 ...            ...       ...
2015            0.067965            0.025395       0.018309  0.032331
2016            0.049069            0.032029       0.013897  0.033610
Automatic saving failed. This file was updated remotely or in another tab.  Show diff
2019            0.054273            0.044300       0.013502  0.018751

              communication  ...  like/girls   sadness  feelings  \
release_date                 ...
1950               0.090467  ...    0.020841  0.110400  0.031841
1951               0.052060  ...    0.019713  0.120134  0.040735
1952               0.064258  ...    0.012299  0.176331  0.043857
1953               0.051146  ...    0.004223  0.138667  0.024428
1954               0.068160  ...    0.027902  0.145519  0.017944
...                     ...  ...         ...       ...       ...
2015               0.078012  ...    0.034025  0.118397  0.031751
2016               0.085859  ...    0.037809  0.104324  0.037228
2017               0.084059  ...    0.047693  0.116999  0.035793
2018               0.084757  ...    0.033412  0.095892  0.024625
2019               0.073929  ...    0.039571  0.094867  0.028095

              danceability  loudness  acousticness  instrumentalness  \
release_date
1950              0.547431  0.611537      0.817326          0.039675
1951              0.574454  0.601727      0.778865          0.161342
1952              0.529405  0.616984      0.802491          0.075751
1953              0.471649  0.628288      0.848001          0.079213
1954              0.431350  0.592374      0.800575          0.063011
...                    ...       ...           ...               ...
2015              0.565502  0.748231      0.180479          0.039839
2016              0.602778  0.744599      0.198113          0.011562
2017              0.620839  0.743620      0.237573          0.003213
2018              0.646615  0.747603      0.227130          0.007363
2019              0.614434  0.755920      0.208698          0.003588

               valence    energy       age
```

```
print(df.corr())
print(df.cov())
```

```
                           Unnamed: 0  release_date       len    dating  \
Unnamed: 0                   1.000000      0.291681 -0.051844  0.036871
release_date                 0.291681      1.000000  0.333133 -0.051569
len                         -0.051844      0.333133  1.000000 -0.011538
dating                       0.036871     -0.051569 -0.011538  1.000000
violence                     0.067391      0.109022  0.046035 -0.102117
```

```
world/life                -0.048644     -0.054048 -0.137096 -0.078919
night/time                 0.020140     -0.003262 -0.040822  0.025812
shake the audience         0.007485      0.104351  0.131677  0.032142
family/gospel              0.057182     -0.046287  0.008915  0.006841
romantic                  -0.045452     -0.173085 -0.168207  0.027631
communication             -0.030890      0.013569 -0.058879 -0.072537
obscene                    0.019570      0.206512  0.450726 -0.003274
music                      0.003697     -0.131733 -0.076703 -0.012770
movement/places            0.071268      0.057359  0.134678 -0.093232
light/visual perceptions  -0.020461     -0.057971 -0.053956 -0.096784
family/spiritual           0.010635     -0.055985 -0.028776 -0.062268
like/girls                -0.008648      0.040975  0.066501 -0.000941
sadness                   -0.062892     -0.053769 -0.179975 -0.029493
feelings                   0.029148     -0.015054 -0.018812  0.023050
danceability              -0.034514      0.140548  0.271371  0.047658
loudness                   0.032012      0.461189  0.257638  0.022045
acousticness              -0.032182     -0.442971 -0.258219  0.008038
instrumentalness           0.084338     -0.003933 -0.109056 -0.006470
valence                   -0.024886     -0.146823  0.102141  0.087700
energy                     0.038508      0.374374  0.264541  0.031786
age                       -0.291559     -1.000000 -0.333133  0.051569

                          violence  world/life  night/time  \
Unnamed: 0                0.067391   -0.048644    0.020140
release_date              0.109022   -0.054048   -0.003262
len                       0.046035   -0.137096   -0.040822
dating                   -0.102117   -0.078919    0.025812
violence                  1.000000   -0.183558   -0.120229
world/life               -0.183558    1.000000   -0.120976
night/time               -0.120229   -0.120976    1.000000
shake the audience       -0.031470   -0.064548   -0.007171
family/gospel            -0.027276   -0.044637   -0.015298
romantic                 -0.143581   -0.066631   -0.051945
communication            -0.068365   -0.067059   -0.049986
obscene                  -0.136870   -0.196480   -0.097528
music                    -0.141586   -0.139582   -0.079191
movement/places           0.011136   -0.074647   -0.041154
light/visual perceptions  0.006481   -0.010994   -0.049722
family/spiritual          0.064049    0.010122   -0.066377
like/girls                0.043334    0.041549    0.019757
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
danceability             -0.105658   -0.079498    0.022060
loudness                  0.082795   -0.057194   -0.003901
acousticness             -0.134398    0.049036   -0.024578
instrumentalness          0.117272   -0.005982    0.010749
valence                  -0.054426   -0.096562    0.040185
energy                    0.149999   -0.085741    0.022845
age                      -0.109022    0.054048    0.003262

                          shake the audience  family/gospel  romantic  ...  \
```

```
# To drop duplicate rows :
df.drop_duplicates()
```

	Unnamed: 0	artist_name	track_name	release_date	genre	lyrics	len	dating	violence	worl
0	0	mukesh	mohabbat bhi jhoothi	1950	pop	hold time feel break feel untrue convince spea...	95.0	0.000598	0.063746	0
1	4	frankie laine	i believe	1950	pop	believe drop rain fall grow believe darkest ni...	51.0	0.035537	0.096777	0
2	6	johnnie ray	cry	1950	pop	sweetheart send letter goodbye secret feel bet...	24.0	0.002770	0.002770	0
3	10	pérez prado	patricia	1950	pop	kiss lips want stroll charm mambo chacha merin...	54.0	0.048249	0.001548	0
4	12	giorgos papadopoulos	apopse eida oneiro	1950	pop	till darling till matter know till dream live ...	48.0	0.001350	0.001350	0
...	...	...	...	...	...	...	...	...	...	
						mama tell young come en ely ne ...	39.0	0.175480	0.001350	0
16703	53223	the wood brothers	raindrop	2015	blues	break heart cheap thrill time hear cryin	32.0	0.002193	0.002193	0

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
import pandas as pd


# Aggregation
#Compute the sum of values in each column
df.count()

#Compute the maximum value in each column
df.max()

#Compute the minimum value in each column
df.min()
```

```
<ipython-input-17-5c73911eefe7>:9: FutureWarning: The default value of numeric_only in DataFrame.max is deprecated. In a future version
  df.max()
<ipython-input-17-5c73911eefe7>:12: FutureWarning: The default value of numeric_only in DataFrame.min is deprecated. In a future versio
  df.min()
Unnamed: 0                                                          0
artist_name                                        "weird al" yankovic
track_name                                       "b" movie box car blues
release_date                                                        1950
genre                                                          blues
lyrics                    aaaa come home swing sweet come home swing swe...
len                                                               2.0
dating                                                       0.000292
violence                                                     0.000292
world/life                                                   0.000292
night/time                                                   0.000289
shake the audience                                           0.000306
family/gospel                                                0.000289
romantic                                                     0.000292
```

```
communication                                 0.000292
obscene                                       0.000289
music                                         0.000289
movement/places                               0.000289
light/visual perceptions                      0.000289
family/spiritual                              0.000289
like/girls                                    0.000289
sadness                                       0.000292
feelings                                      0.000289
danceability                                  0.021878
loudness                                       0.15568
acousticness                                       0.0
instrumentalness                                   0.0
valence                                       0.010305
energy                                        0.007056
age                                           0.014286
dtype: object
```

```
# To convert a column to a specific data type
df['len'] = df['len'].astype('float')
```

```
# To drop rows or columns with missing values
df.dropna() # Drops rows with any missing value
df.dropna(axis=1) #Drop columns with any missing value
```

	Unnamed: 0	artist_name	track_name	release_date	genre	lyrics
0	0	mukesh	mohabbat bhi jhoothi	1950	pop	hold time feel break feel untrue convince spea...
1	4	frankie laine	i believe	1950	pop	believe drop rain fall grow believe darkest ni...
2	9	andy russell	sweetheart send letter goodbye	1950	pop	sweetheart send letter goodbye secret feel bet...
3	10	pérez prado	patricia	1950	pop	kiss lips want stroll charm mambo chacha merin...
4	12	giorgos papadopoulos	apopse eida oneiro	1950	pop	till darling till matter know till dream live ...
...	...	...	...	...	...	...
16702	53222	gov't mule	simple man	2015	blues	mama tell young come listen closely help time ...
16703	53223	the wood brothers	raindrop	2015	blues	break heart cheap thrill time hear cryin muddy...


```
# To fill missing values with a specific value or stratergy :
df.fillna(38)   #Fill missing values with the specific value
df.fillna(df.mean()) #fill missing values with the column mean
```

```
<ipython-input-15-89f70c0bc800>:3: FutureWarning: The default value of numeric_only in DataFrame.mean :
  df.fillna(df.mean()) #fill missing values with the column mean
```

	Unnamed: 0	artist_name	track_name	release_date	genre	lyrics	len	dating	violence
0	0	mukesh	mohabbat bhi jhoothi	1950	pop	hold time feel break feel untrue convince spea...	95.000000	0.000598	0.063746
1	4	frankie laine	i believe	1950	pop	believe drop rain fall grow believe darkest ni...	51.000000	0.035537	0.096777
2	6	johnnie ray	cry	1950	pop	sweetheart send letter goodbye secret feel bet...	24.000000	0.002770	0.002770
3	10	pérez prado	patricia	1950	pop	kiss lips want stroll charm mambo chacha merin...	54.000000	0.048249	0.001548
4	12	giorgos papadopoulos	apopse eida oneiro	1950	pop	till darling till matter know till dream live ...	48.000000	0.001350	0.001350
...	...	...	...	...	...	...	...	...	...
						mama tell ng ne en closely help time ...	39.000000	0.175480	0.001350
16703	53223	the wood brothers	raindrop	2015	blues	break heart cheap thrill time hear cryin muddy...	32.000000	0.002193	0.002193


```
# QUANTILES

# Compute the quantiles of a column
df['len'].quantile([0.25,0.5,0.75])

# Compute the quantiles of a multiple column
df[['len','release_date']].quantile([0.25,0.5,0.75])
```

	len	release_date
0.25	42.0	1972.0
0.50	61.0	1987.0
0.75	87.0	2003.0

```
mport pandas as pd

 Create a sample dataframe
ata={'artist_name':['mukesh','frankie laine','johnnie ray','pÃ©rez prado'],'track_name':['mohabbat bhi jhoothi','i believe','cry','patricia']
f=pd.DataFrame(data)

 Display the DataFrame
rint(df)

rint(df.describe())
```

```
     artist_name           track_name
0         mukesh  mohabbat bhi jhoothi
1  frankie laine             i believe
2    johnnie ray                   cry
3   pÃ©rez prado              patricia
       artist_name           track_name
count            4                    4
unique           4                    4
top         mukesh  mohabbat bhi jhoothi
freq             1                    1
```

```
# Data Cleaning
# To check for missing values in a DataFrame :
df.isnull()
```

	artist_name	track_name
0	False	False
1	False	False
2	False	False
3	False	False

```
# Data wragling
import pandas as pd
df1=pd.DataFrame({'key':['A','B','c','D'],
                  'value1':[1,2,3,4]})
df2=pd.DataFrame({'key':['B','D','E','F'],
                  'value2':[5,6,7,8]})
merged_df=pd.merge(df1,df2,on='key')

print(merged_df)
```

```
  key  value1  value2
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
import pandas as pd
data = {
    'artist_name':['mukesh','frankie laine','johnnie ray','pÃ©rez prado'],
    'track_name':['mohabbat bhi jhoothi','i believe','cry','patricia'],
    'len':[95,51,24,54]
}
df = pd.DataFrame(data)

# Using loc to select rows with index labels 1 & 2 , & columns 'artist_name' and 'track-name'
selected_loc = df.loc[[1,2],['artist_name','track_name']]
print("Using loc:")
print(selected_loc)
```

```
Using loc:
     artist_name track_name
1  frankie laine  i believe
3   pÃ©rez prado   patricia
```

```
# Data wragling
import pandas as pd
data = {
    'artist_name':['mukesh','frankie laine','johnnie ray','pÃ©rez prado'],
    'track_name':['mohabbat bhi jhoothi','i believe','cry','patricia'],
    'len':[95,51,24,54]
}
df = pd.DataFrame(data)

#Perform stack operation
stacked_df = df.set_index('artist_name').stack()
print(stacked_df)

# Perform unstack operation
Stacked_df = stacked_df.unstack()
print(stacked_df)
```

```
artist_name
mukesh         track_name    mohabbat bhi jhoothi
               len                             95
frankie laine  track_name                i believe
               len                             51
johnnie ray    track_name                      cry
               len                             24
pÃ©rez prado   track_name                 patricia
               len                             54
dtype: object
artist_name
mukesh         track_name    mohabbat bhi jhoothi
               len                             95
frankie laine  track_name                i believe
               len                             51
johnnie ray    track_name                      cry
               len                             24
pÃ©rez prado   track_name                 patricia
               len                             54
dtype: object
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

✓ 0s completed at 9:28 AM