

dsbda-practical-1

February 9, 2024

```
[8]: pip install pandas
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pandas in
c:\users\rohan\appdata\roaming\python\python312\site-packages (2.2.0)
Requirement already satisfied: numpy<2,>=1.26.0 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from pandas)
(1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from pandas)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from pandas)
(2024.1)
Requirement already satisfied: tzdata>=2022.7 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from pandas)
(2023.3)
Requirement already satisfied: six>=1.5 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from python-
dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
[8]: import pandas as pd
```

```
C:\Users\Rohan\AppData\Local\Temp\ipykernel_16964\4080736814.py:1:
DeprecationWarning:
Pyyarrow will become a required dependency of pandas in the next major release of
pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better
interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466
```

```
import pandas as pd
```

```
[9]: pip install seaborn
```

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: seaborn in
c:\users\rohan\appdata\roaming\python\python312\site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from seaborn)
(1.26.4)
Requirement already satisfied: pandas>=1.2 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from seaborn)
(2.2.0)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from seaborn)
(3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (4.48.1)
Requirement already satisfied: kiwisolver>=1.3.1 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from
pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in
c:\users\rohan\appdata\roaming\python\python312\site-packages (from python-
dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

```
[10]: import seaborn as sns
```

```
[11]: ds_name=sns.get_dataset_names()
```

```
[12]: print(ds_name)
```

```
['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes',  
'diamonds', 'dots', 'dowjones', 'exercise', 'flights', 'fmri', 'geyser', 'glue',  
'healthexp', 'iris', 'mpg', 'penguins', 'planets', 'seaice', 'taxis', 'tips',  
'titanic']
```

```
[13]: df = sns.load_dataset("flights")  
df
```

```
[13]:
```

	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121
..
139	1960	Aug	606
140	1960	Sep	508
141	1960	Oct	461
142	1960	Nov	390
143	1960	Dec	432

[144 rows x 3 columns]

```
[14]: df.head(n=5)
```

```
[14]:
```

	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121

```
[19]: df.head(n=50)
```

```
[19]:
```

	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121
5	1949	Jun	135

6	1949	Jul	148
7	1949	Aug	148
8	1949	Sep	136
9	1949	Oct	119
10	1949	Nov	104
11	1949	Dec	118
12	1950	Jan	115
13	1950	Feb	126
14	1950	Mar	141
15	1950	Apr	135
16	1950	May	125
17	1950	Jun	149
18	1950	Jul	170
19	1950	Aug	170
20	1950	Sep	158
21	1950	Oct	133
22	1950	Nov	114
23	1950	Dec	140
24	1951	Jan	145
25	1951	Feb	150
26	1951	Mar	178
27	1951	Apr	163
28	1951	May	172
29	1951	Jun	178
30	1951	Jul	199
31	1951	Aug	199
32	1951	Sep	184
33	1951	Oct	162
34	1951	Nov	146
35	1951	Dec	166
36	1952	Jan	171
37	1952	Feb	180
38	1952	Mar	193
39	1952	Apr	181
40	1952	May	183
41	1952	Jun	218
42	1952	Jul	230
43	1952	Aug	242
44	1952	Sep	209
45	1952	Oct	191
46	1952	Nov	172
47	1952	Dec	194
48	1953	Jan	196
49	1953	Feb	196

```
[15]: df.tail(n=2)
```

```
[15]:      year month  passengers
      142  1960   Nov         390
      143  1960   Dec         432
```

```
[16]: df.index
```

```
[16]: RangeIndex(start=0, stop=144, step=1)
```

```
[17]: df.shape
```

```
[17]: (144, 3)
```

```
[18]: df.dtypes
```

```
[18]: year          int64
      month        category
      passengers    int64
      dtype: object
```

```
[19]: df.columns.values
```

```
[19]: array(['year', 'month', 'passengers'], dtype=object)
```

```
[25]: df.describe(include='all')
```

```
[25]:      year month  passengers
count    144.000000    144    144.000000
unique         NaN     12         NaN
top         NaN    Jan         NaN
freq         NaN     12         NaN
mean    1954.500000    NaN    280.298611
std         3.464102    NaN    119.966317
min    1949.000000    NaN    104.000000
25%    1951.750000    NaN    180.000000
50%    1954.500000    NaN    265.500000
75%    1957.250000    NaN    360.500000
max    1960.000000    NaN    622.000000
```

```
[20]: df.sort_index(axis=1, ascending=False)
```

```
[20]:      year  passengers month
0    1949         112    Jan
1    1949         118    Feb
2    1949         132    Mar
3    1949         129    Apr
4    1949         121    May
..     ...         ...    ...
```

139	1960	606	Aug
140	1960	508	Sep
141	1960	461	Oct
142	1960	390	Nov
143	1960	432	Dec

[144 rows x 3 columns]

```
[21]: df.iloc[5]
```

```
[21]: year          1949
      month         Jun
      passengers    135
      Name: 5, dtype: object
```

```
[22]: df[0:3]
```

```
[22]:   year month  passengers
0  1949   Jan          112
1  1949   Feb          118
2  1949   Mar          132
```

```
[23]: df.iloc[:5, :]
```

```
[23]:   year month  passengers
0  1949   Jan          112
1  1949   Feb          118
2  1949   Mar          132
3  1949   Apr          129
4  1949   May          121
```

```
[24]: df.iloc[:3, :7]
```

```
[24]:   year month  passengers
0  1949   Jan          112
1  1949   Feb          118
2  1949   Mar          132
```

```
[25]: df.isnull()
```

```
[25]:   year  month  passengers
0   False  False         False
1   False  False         False
2   False  False         False
3   False  False         False
4   False  False         False
..     ...   ...           ...
```

139	False	False	False
140	False	False	False
141	False	False	False
142	False	False	False
143	False	False	False

[144 rows x 3 columns]

```
[26]: df.iloc[[1, 2, 4], [0, 2]]
```

```
[26]:   year  passengers
1  1949          118
2  1949          132
4  1949          121
```

```
[27]: df.iloc[1, 1]
```

```
[27]: 'Feb'
```

```
[28]: cols_2_4=df.columns[2:4]
df[cols_2_4]
```

```
[28]:   passengers
0          112
1          118
2          132
3          129
4          121
..         ...
139         606
140         508
141         461
142         390
143         432
```

[144 rows x 1 columns]

```
[29]: df[df.columns[2:4]].iloc[5:10]
```

```
[29]:   passengers
5          135
6          148
7          148
8          136
9          119
```

```
[30]: df.isnull().any()
```

```
[30]: year          False
      month          False
      passengers     False
      dtype: bool
```

```
[31]: df.isnull().sum().sum()
```

```
[31]: 0
```

```
[32]: df.isnull().sum(axis = 1)
```

```
[32]: 0      0
      1      0
      2      0
      3      0
      4      0
      ..
     139     0
     140     0
     141     0
     142     0
     143     0
      Length: 144, dtype: int64
```

```
[33]: df.isnull().sum()
```

```
[33]: year          0
      month          0
      passengers     0
      dtype: int64
```

```
[34]: df.isna().sum()
```

```
[34]: year          0
      month          0
      passengers     0
      dtype: int64
```

```
[37]: df.month.isnull().sum()
```

```
[37]: 0
```

```
[38]: df.groupby(['year'])['month'].apply(lambda x:x.isnull().sum())
```

```
[38]: year
     1949     0
     1950     0
```



```
1951    0
1952    0
1953    0
1954    0
1955    0
1956    0
1957    0
1958    0
1959    0
1960    0
Name: month, dtype: int64
```

```
[ ]:
```