

```
import pandas as pd
import numpy as np
import seaborn as sns
```

```
df = pd.read_csv('train.csv')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  891 non-null    int64
1   Survived     891 non-null    int64
2   Pclass       891 non-null    int64
3   Name         891 non-null    object
4   Sex          891 non-null    object
5   Age          714 non-null    float64
6   SibSp        891 non-null    int64
7   Parch        891 non-null    int64
8   Ticket       891 non-null    object
9   Fare         891 non-null    float64
10  Cabin        204 non-null    object
11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
df.shape
```

```
(891, 12)
```

```
df.size
```

```
10692
```

```
df.head(5)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Next steps:


[Generate code with df](#)

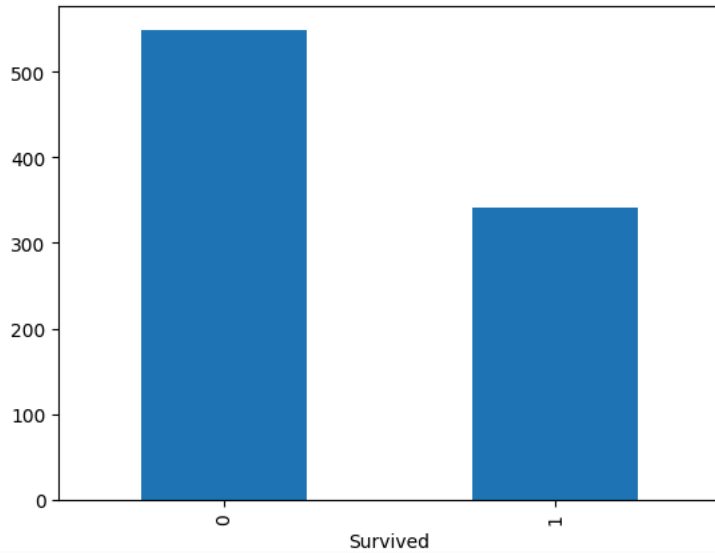
[View recommended plots](#)

[New interactive sheet](#)


```
import matplotlib.pyplot as plt
```

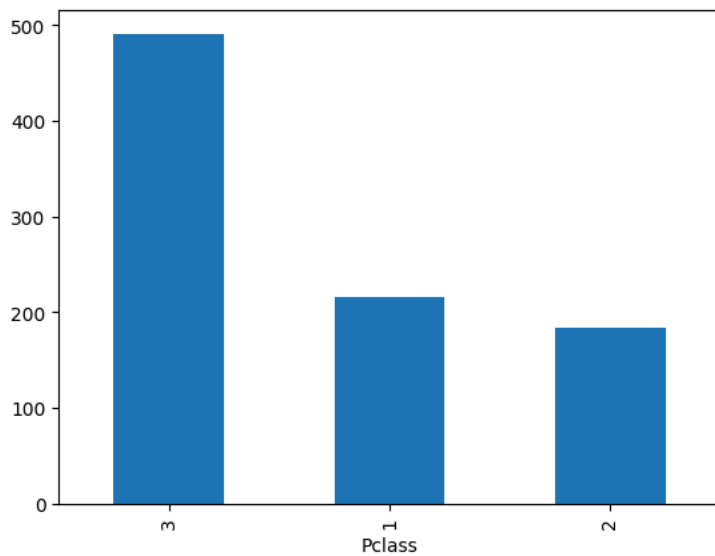
```
#sns.countplot(df['Survived'])
df['Survived'].value_counts().plot(kind='bar')
```

 <Axes: xlabel='Survived'>




```
df['Pclass'].value_counts().plot(kind='bar')
```

 <Axes: xlabel='Pclass'>




```
sns.countplot(df['Pclass'])
```

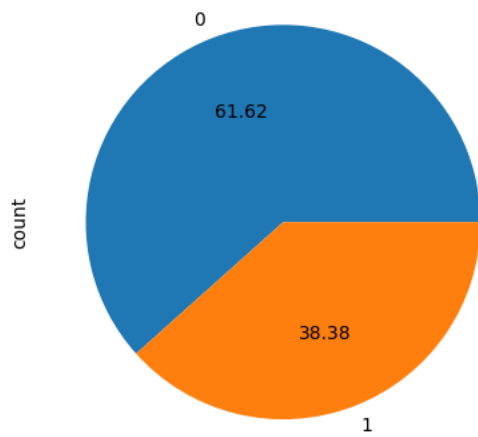
 **NameError** Traceback (most recent call last)
 <ipython-input-9-1dd2dd8e780a> in <cell line: 0>()
 ----> 1 sns.countplot(df['Pclass'])

NameError: name 'sns' is not defined


Next steps: [Explain error](#)

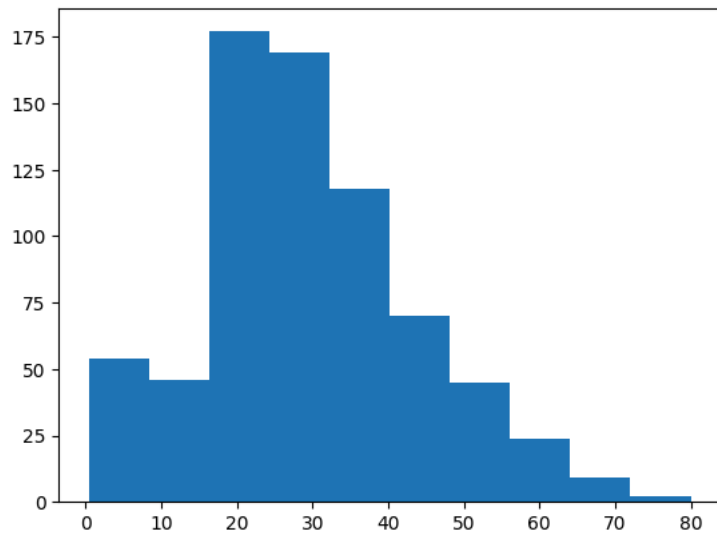
```
df['Survived'].value_counts().plot(kind='pie', autopct='%0.2f')
```

 <Axes: ylabel='count'>



```
#histogram for numerical data  
plt.hist(df['Age'], bins= 10)
```

 (array([54., 46., 177., 169., 118., 70., 45., 24., 9., 2.]),
array([0.42 , 8.378, 16.336, 24.294, 32.252, 40.21 , 48.168, 56.126,
64.084, 72.042, 80.]),
<BarContainer object of 10 artists>)



```
sns.distplot(df['Age'])
```

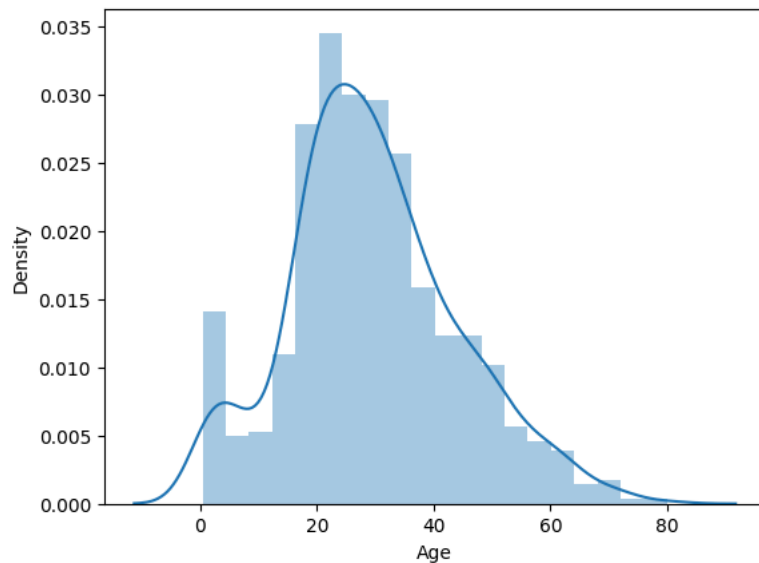
 <ipython-input-15-0fafe04ea3f6>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.


Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

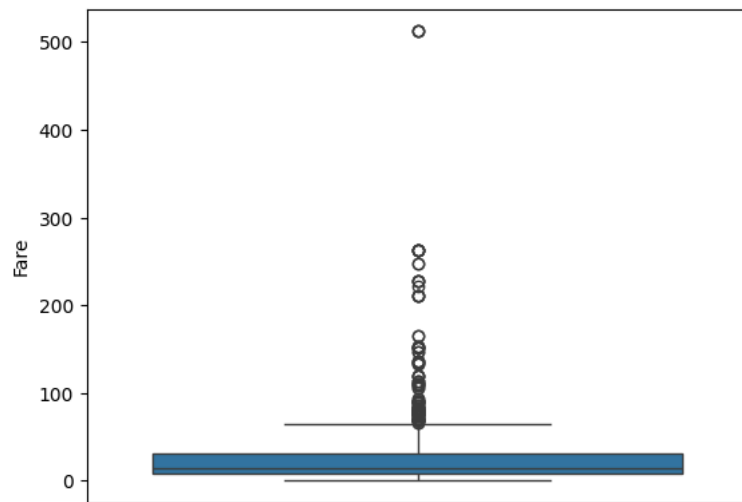
For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['Age'])  
<Axes: xlabel='Age', ylabel='Density'>
```




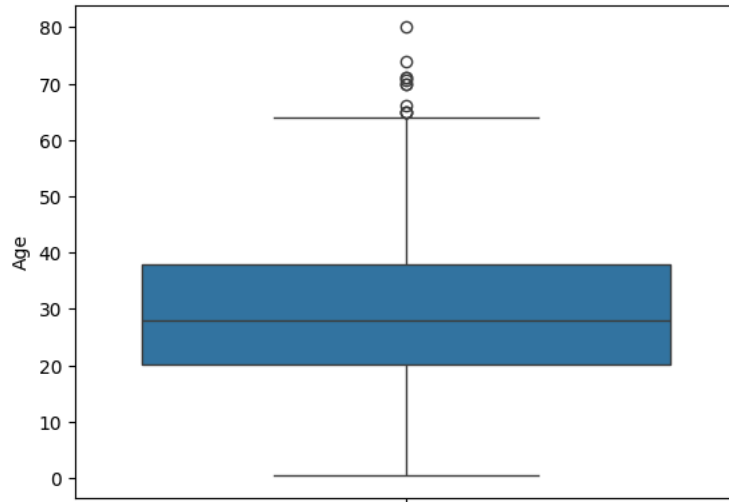
```
sns.boxplot(df['Fare'])
```

 <Axes: ylabel='Fare'>




```
sns.boxplot(df['Age'])
```


 <Axes: ylabel='Age'>



```
df['Age'].min()
```

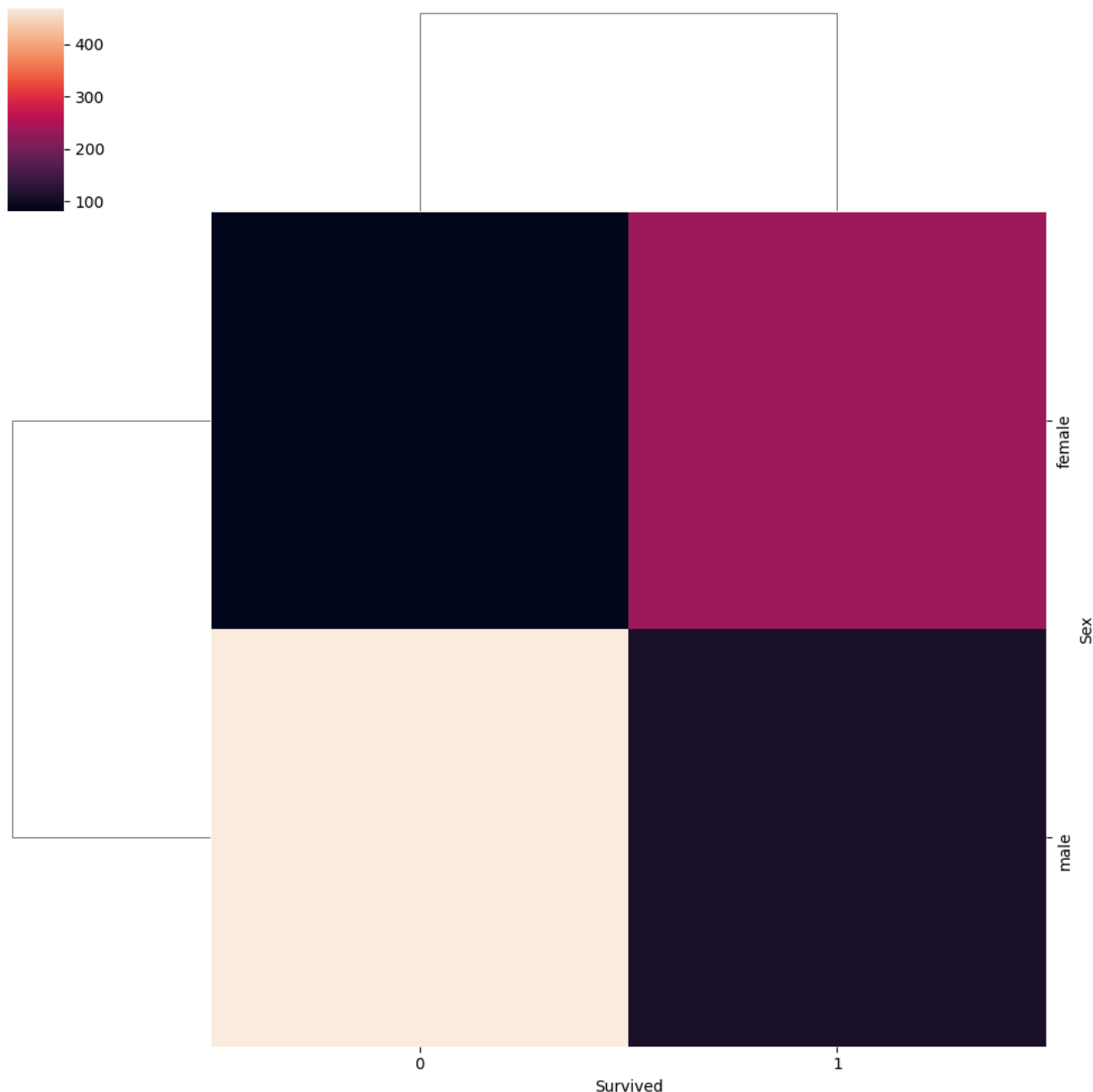
 0.42

```
df['Age'].skew()
```

 0.38910778230082704

```
#multivariant EDA  
sns.clustermap(pd.crosstab(df['Sex'], df['Survived']))
```

 <seaborn.matrix.ClusterGrid at 0x7a2b0a190990>




!pip install pandas.profilin

 Requirement already satisfied: pandas.profilin in /usr/local/lib/python3.11/dist-packages (3.2.0)
Requirement already satisfied: joblib~=1.1.0 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (1.1.1)
Requirement already satisfied: scipy>=1.4.1 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (1.15.2)
Requirement already satisfied: pandas!=1.0.0,!1.0.1,!1.0.2,!1.1.0,>=0.25.3 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (2.2.3)
Requirement already satisfied: matplotlib>=3.2.0 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (3.10.0)
Requirement already satisfied: pydantic>=1.8.1 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (2.11.4)
Requirement already satisfied: PyYAML>=5.0.0 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (6.0.2)
Requirement already satisfied: jinja2>=2.11.1 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (3.1.6)
Requirement already satisfied: markupsafe~=2.1.1 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (2.1.5)
Requirement already satisfied: visions==0.7.4 in /usr/local/lib/python3.11/dist-packages (from visions[type_image_path]==0.7.4->pandas.profilin) (0.7.4)
Requirement already satisfied: numpy>=1.16.0 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (1.26.4)
Requirement already satisfied: htmlmin>=0.1.12 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (0.1.12)
Requirement already satisfied: missingno>=0.4.2 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (0.5.2)
Requirement already satisfied: phik>=0.11.1 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (0.12.4)
Requirement already satisfied: tangled-up-in-unicode==0.2.0 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (0.2.0)
Requirement already satisfied: requests>=2.24.0 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (2.32.3)
Requirement already satisfied: tqdm>=4.48.2 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (4.67.1)
Requirement already satisfied: seaborn>=0.10.1 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (0.13.2)
Requirement already satisfied: multimethod>=1.4 in /usr/local/lib/python3.11/dist-packages (from pandas.profilin) (2.0)
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.11/dist-packages (from visions==0.7.4->visions[type_image_path]==0.7.4->pandas.profilin) (25.1.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.11/dist-packages (from visions==0.7.4->visions[type_image_path]==0.7.4->pandas.profilin) (3.4.2)
Requirement already satisfied: imagehash in /usr/local/lib/python3.11/dist-packages (from visions[type_image_path]==0.7.4->pandas.profilin) (4.3.1)
Requirement already satisfied: Pillow in /usr/local/lib/python3.11/dist-packages (from visions[type_image_path]==0.7.4->pandas.profilin) (10.4.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profilin) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profilin) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profilin) (4.53.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profilin) (1.4.7)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profilin) (25.0)

Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profilin
 Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.2.0->pandas.profi
 Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas!=1.0.0,!1.0.1,!1.0.2,!1.1.0,>
 Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas!=1.0.0,!1.0.1,!1.0.2,!1.1.0,>
 Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic>=1.8.1->pandas.profi
 Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.11/dist-packages (from pydantic>=1.8.1->pandas.profil
 Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.11/dist-packages (from pydantic>=1.8.1->pandas.pr
 Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic>=1.8.1->pandas.prc
 Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests>=2.24.0->pandas.pr
 Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests>=2.24.0->pandas.profilin) (3
 Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests>=2.24.0->pandas.profilir
 Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.24.0->pandas.profilir
 Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib>=3.2.0->pa
 Requirement already satisfied: PyWavelets in /usr/local/lib/python3.11/dist-packages (from imagehash->visions[type_image_path]=0.7

```
from pandas_profiling import ProfileReport
prof = ProfileReport(df)
prof.to_file(output_file='output.html')
```

 /usr/local/lib/python3.11/dist-packages/numba/core/decorators.py:262: NumbaDeprecationWarning: numba.generated_jit is deprecated. Pl
 warnings.warn(msg, NumbaDeprecationWarning)
 /usr/local/lib/python3.11/dist-packages/visions/backends/shared/nan_handling.py:50: NumbaDeprecationWarning: The 'nopython' keyword
 @nb.jit

```
PydanticImportError                                Traceback (most recent call last)
<ipython-input-20-452a8c0cbe53> in <cell line: 0>()
----> 1 from pandas_profiling import ProfileReport
      2 prof = ProfileReport(df)
      3 prof.to_file(output_file='output.html')
```

5 frames

```
/usr/local/lib/python3.11/dist-packages/pydantic/_migration.py in wrapper(name)
    294     return import_string(REDIRECT_TO_V1[import_path])
    295     if import_path == 'pydantic:BaseSettings':
--> 296         raise PydanticImportError(
    297             '`BaseSettings` has been moved to the `pydantic-settings` package. '
    298             f'See https://docs.pydantic.dev/{version\_short\(\)}/migration/#basesettings-has-moved-to-pydantic-settings '

```

PydanticImportError: `BaseSettings` has been moved to the `pydantic-settings` package. See
<https://docs.pydantic.dev/2.11/migration/#basesettings-has-moved-to-pydantic-settings> for more details.

For further information visit <https://errors.pydantic.dev/2.11/u/import-error>


NOTE: If your import is failing due to a missing package, you can
 manually install dependencies using either `!pip` or `!apt`.

To view examples of installing some common dependencies, click the
 "Open Examples" button below.

OPEN EXAMPLES

Next steps: [Explain error](#)

```
!pip install --force-reinstall "numba<0.59"
```

 Collecting numba<0.59
 Using cached numba-0.58.1-cp311-cp311-manylinux2014_x86_64.manylinux_2_17_x86_64.whl.metadata (2.7 kB)
 Collecting llvmlite<0.42,>=0.41.0dev0 (from numba<0.59)
 Using cached llvmlite-0.41.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.8 kB)
 Collecting numpy<1.27,>=1.22 (from numba<0.59)
 Using cached numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (61 kB)
 Using cached numba-0.58.1-cp311-cp311-manylinux2014_x86_64.manylinux_2_17_x86_64.whl (3.6 MB)
 Using cached llvmlite-0.41.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (43.6 MB)
 Using cached numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.3 MB)
 Installing collected packages: numpy, llvmlite, numba
 Attempting uninstall: numpy
 Found existing installation: numpy 1.26.4
 Uninstalling numpy-1.26.4:
 Successfully uninstalled numpy-1.26.4
 Attempting uninstall: llvmlite
 Found existing installation: llvmlite 0.41.1
 Uninstalling llvmlite-0.41.1:
 Successfully uninstalled llvmlite-0.41.1
 Attempting uninstall: numba
 Found existing installation: numba 0.58.1
 Uninstalling numba-0.58.1:
 Successfully uninstalled numba-0.58.1
 ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the sou
 numba-cuda 0.2.0 requires numba>=0.59.1, but you have numba 0.58.1 which is incompatible.
 cuml-cu12 25.2.1 requires numba<0.61.0a0,>=0.59.1, but you have numba 0.58.1 which is incompatible.
 distributed-ucxx-cu12 0.42.0 requires numba<0.61.0a0,>=0.59.1, but you have numba 0.58.1 which is incompatible.

```
scikit-learn 1.6.1 requires joblib>=1.2.0, but you have joblib 1.1.1 which is incompatible.
thinc 8.3.6 requires numpy<3.0.0,>=2.0.0, but you have numpy 1.26.4 which is incompatible.
dask-cuda 25.2.0 requires numba<0.61.0a0,>=0.59.1, but you have numba 0.58.1 which is incompatible.
cudf-cu12 25.2.1 requires numba<0.61.0a0,>=0.59.1, but you have numba 0.58.1 which is incompatible.
Successfully installed llvmlite-0.41.1 numba-0.58.1 numpy-1.26.4
```

```
from pandas_profiling import ProfileReport
prof = ProfileReport(df)
prof.to_file(output_file='output.html')
```

```
⚠ /usr/local/lib/python3.11/dist-packages/numba/core/decorators.py:262: NumbaDeprecationWarning: numba.generated_jit is deprecated. Please use warnings.warn(msg, NumbaDeprecationWarning)
/usr/local/lib/python3.11/dist-packages/visions/backends/shared/nan_handling.py:50: NumbaDeprecationWarning: The 'nopython' keyword @nb.jit
```

```
-----
PydanticImportError                                Traceback (most recent call last)
<ipython-input-2-452a8c0cbe53> in <cell line: 0>()
----> 1 from pandas_profiling import ProfileReport
      2 prof = ProfileReport(df)
      3 prof.to_file(output_file='output.html')
```

⏮ 5 frames ⏭

```
/usr/local/lib/python3.11/dist-packages/pydantic/_migration.py in wrapper(name)
    294     return import_string(REDIRECT_TO_V1[import_path])
    295     if import_path == 'pydantic:BaseSettings':
--> 296         raise PydanticImportError(
    297             '`BaseSettings` has been moved to the `pydantic-settings` package. '
    298             f'See https://docs.pydantic.dev/{version\_short\(\)}/migration/#basesettings-has-moved-to-pydantic-settings '

```

PydanticImportError: `BaseSettings` has been moved to the `pydantic-settings` package. See <https://docs.pydantic.dev/2.11/migration/#basesettings-has-moved-to-pydantic-settings> for more details.

For further information visit <https://errors.pydantic.dev/2.11/u/import-error>

NOTE: If your import is failing due to a missing package, you can manually install dependencies using either `!pip` or `!apt`.

To view examples of installing some common dependencies, click the "Open Examples" button below.

OPEN EXAMPLES

Next steps: [Explain error](#)

`!pip install ydata-profiling`

```
⚡ Collecting ydata-profiling
  Downloading ydata_profiling-4.16.1-py2.py3-none-any.whl.metadata (22 kB)
Requirement already satisfied: scipy<1.15.0,>=1.4.1 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (1.15.0)
```