

preprocessing-and-eda-books-data

August 22, 2024

```
[1]: !pip install ydata-profiling
!pip install pydantic-settings

import pandas as pd
from ydata_profiling import ProfileReport

# Load your data into a DataFrame
df = pd.read_csv('books_data.csv')

# Generate the profile report
profile = ProfileReport(df, title="Books Data Profiling Report",
    ↪explorative=True)

# Save the report as an HTML file
profile.to_file("books_data_report.pdf")

# Display the report directly in a Jupyter notebook
profile.to_notebook_iframe()
```

```
Requirement already satisfied: ydata-profiling in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (4.9.0)
Requirement already satisfied: scipy<1.14,>=1.4.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (1.11.4)
Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (2.1.4)
Requirement already satisfied: matplotlib<3.10,>=3.5 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (3.7.4)
Requirement already satisfied: pydantic>=2 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (2.8.2)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (6.0.1)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
```

ydata-profiling) (3.1.3)
Requirement already satisfied: visions[type_image_path]<0.7.7,>=0.7.5 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (0.7.6)
Requirement already satisfied: numpy<2,>=1.16.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (1.24.4)
Requirement already satisfied: htmlmin==0.1.12 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (0.1.12)
Requirement already satisfied: phik<0.13,>=0.11.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (0.12.4)
Requirement already satisfied: requests<3,>=2.24.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (2.31.0)
Requirement already satisfied: tqdm<5,>=4.48.2 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (4.66.1)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (0.13.1)
Requirement already satisfied: multimethod<2,>=1.4 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (1.12)
Requirement already satisfied: statsmodels<1,>=0.13.2 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (0.14.1)
Requirement already satisfied: typeguard<5,>=3 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (4.3.0)
Requirement already satisfied: imagehash==4.3.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (4.3.1)
Requirement already satisfied: wordcloud>=1.9.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (1.9.3)
Requirement already satisfied: dacite>=1.8 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (1.8.1)
Requirement already satisfied: numba<1,>=0.56.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
ydata-profiling) (0.60.0)
Requirement already satisfied: PyWavelets in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
imagehash==4.3.1->ydata-profiling) (1.7.0)
Requirement already satisfied: pillow in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from

imagehash==4.3.1->ydata-profiling) (10.2.0)
 Requirement already satisfied: MarkupSafe>=2.0 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 jinja2<3.2,>=2.11.1->ydata-profiling) (2.1.3)
 Requirement already satisfied: contourpy>=1.0.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (1.2.0)
 Requirement already satisfied: cycler>=0.10 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (0.12.1)
 Requirement already satisfied: fonttools>=4.22.0 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (4.47.2)
 Requirement already satisfied: kiwisolver>=1.0.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (1.4.5)
 Requirement already satisfied: packaging>=20.0 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (23.2)
 Requirement already satisfied: pyparsing>=2.3.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (3.1.1)
 Requirement already satisfied: python-dateutil>=2.7 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 matplotlib<3.10,>=3.5->ydata-profiling) (2.8.2)
 Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 numba<1,>=0.56.0->ydata-profiling) (0.43.0)
 Requirement already satisfied: pytz>=2020.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 pandas!=1.4.0,<3,>1.1->ydata-profiling) (2023.3.post1)
 Requirement already satisfied: tzdata>=2022.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 pandas!=1.4.0,<3,>1.1->ydata-profiling) (2023.4)
 Requirement already satisfied: joblib>=0.14.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 phik<0.13,>=0.11.1->ydata-profiling) (1.1.1)
 Requirement already satisfied: annotated-types>=0.4.0 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 pydantic>=2->ydata-profiling) (0.7.0)
 Requirement already satisfied: pydantic-core==2.20.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 pydantic>=2->ydata-profiling) (2.20.1)
 Requirement already satisfied: typing-extensions>=4.6.1 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
 pydantic>=2->ydata-profiling) (4.12.2)
 Requirement already satisfied: charset-normalizer<4,>=2 in
 c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from

```

requests<3,>=2.24.0->ydata-profiling) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
requests<3,>=2.24.0->ydata-profiling) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
requests<3,>=2.24.0->ydata-profiling) (2.1.0)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
requests<3,>=2.24.0->ydata-profiling) (2023.11.17)
Requirement already satisfied: patsy>=0.5.4 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
statsmodels<1,>=0.13.2->ydata-profiling) (0.5.6)
Requirement already satisfied: colorama in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
tqdm<5,>=4.48.2->ydata-profiling) (0.4.6)
Requirement already satisfied: attrs>=19.3.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
visions[type_image_path]<0.7.7,>=0.7.5->ydata-profiling) (23.2.0)
Requirement already satisfied: networkx>=2.4 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
visions[type_image_path]<0.7.7,>=0.7.5->ydata-profiling) (3.3)
Requirement already satisfied: six in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
patsy>=0.5.4->statsmodels<1,>=0.13.2->ydata-profiling) (1.16.0)

```

WARNING: Skipping

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'

WARNING: Skipping

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'

WARNING: Skipping

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'

WARNING: Skipping

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'

WARNING: Skipping

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'

WARNING: Skipping

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'

[notice] A new release of pip is available: 23.2.1 -> 24.2

[notice] To update, run: python.exe -m pip install --upgrade pip

WARNING: Skipping

```

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'
WARNING: Skipping
C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'
WARNING: Skipping
C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'
WARNING: Skipping
C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'
WARNING: Skipping
C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'
WARNING: Skipping
C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\numpy-1.26.3.dist-info due to invalid metadata entry 'name'
WARNING: Skipping

```

```

[notice] A new release of pip is available: 23.2.1 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip

```

```

Requirement already satisfied: pydantic-settings in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (2.4.0)
Requirement already satisfied: pydantic>=2.7.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
pydantic-settings) (2.8.2)
Requirement already satisfied: python-dotenv>=0.21.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
pydantic-settings) (1.0.1)
Requirement already satisfied: annotated-types>=0.4.0 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
pydantic>=2.7.0->pydantic-settings) (0.7.0)
Requirement already satisfied: pydantic-core==2.20.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
pydantic>=2.7.0->pydantic-settings) (2.20.1)
Requirement already satisfied: typing-extensions>=4.6.1 in
c:\users\admin\appdata\local\programs\python\python311\lib\site-packages (from
pydantic>=2.7.0->pydantic-settings) (4.12.2)

```

```

Summarize dataset: 0%|          | 0/5 [00:00<?, ?it/s]

```

```

Generate report structure: 0%|          | 0/1 [00:00<?, ?it/s]

```

```

Render HTML: 0%|          | 0/1 [00:00<?, ?it/s]

```

```

C:\Users\Admin\AppData\Local\Programs\Python\Python311\Lib\site-
packages\ydata_profiling\profile_report.py:384: UserWarning: Extension .pdf not
supported. For now we assume .html was intended. To remove this warning, please
use .html or .json.

```

```

    warnings.warn(

```

Export report to file: 0%| | 0/1 [00:00<?, ?it/s]

<IPython.core.display.HTML object>

[]: