

This notebook provides a step-by-step process for cleaning and analyzing a dataset related to sentencing. We will explore the data, handle missing and erroneous values, and visualize the cleaned dataset.

```
In [ ]: # Create a virtual environment using Anaconda/Miniconda
        # conda create -n sentence python=3.10 -y
        # conda activate sentence

        # Install required packages
        # !pip install pandas numpy seaborn matplotlib ydata-profiling
        # !pip install -r requirements.txt
```

```
In [ ]: # Dataset can be found at:
        # https://datacatalog.cookcountyil.gov/browse?tags=state%27s+attorney+case-level
```

```
In [1]: !conda create -n msds python=3.10.14 -y
        !conda activate msds
        !pip install --upgrade pycaret
```

Channels:

- defaults

Platform: win-64

Collecting package metadata (repodata.json): ...working... done

Solving environment: ...working... done

## Package Plan ##

environment location: C:\Users\pghat\anaconda3\envs\msds

added / updated specs:

- python=3.10.14

The following NEW packages will be INSTALLED:

bzip2	pkgs/main/win-64::bzip2-1.0.8-h2bbff1b_6
ca-certificates	pkgs/main/win-64::ca-certificates-2024.9.24-haa95532_0
libffi	pkgs/main/win-64::libffi-3.4.4-hd77b12b_1
openssl	pkgs/main/win-64::openssl-3.0.15-h827c3e9_0
pip	pkgs/main/win-64::pip-24.2-py310haa95532_0
python	pkgs/main/win-64::python-3.10.14-he1021f5_1
setuptools	pkgs/main/win-64::setuptools-75.1.0-py310haa95532_0
sqlite	pkgs/main/win-64::sqlite-3.45.3-h2bbff1b_0
tk	pkgs/main/win-64::tk-8.6.14-h0416ee5_0
tzdata	pkgs/main/noarch::tzdata-2024b-h04d1e81_0
vc	pkgs/main/win-64::vc-14.40-h2eaa2aa_1
vs2015_runtime	pkgs/main/win-64::vs2015_runtime-14.40.33807-h98bb1dd_1
wheel	pkgs/main/win-64::wheel-0.44.0-py310haa95532_0
xz	pkgs/main/win-64::xz-5.4.6-h8cc25b3_1
zlib	pkgs/main/win-64::zlib-1.2.13-h8cc25b3_1

Downloading and Extracting Packages: ...working... done

Preparing transaction: ...working... done

Verifying transaction: ...working... done

Executing transaction: ...working... done

#

# To activate this environment, use

#

#     \$ conda activate msds

#

# To deactivate an active environment, use

#

#     \$ conda deactivate

Requirement already satisfied: pycaret in c:\users\pghat\appdata\roaming\python\python312\site-packages (3.3.2)

Requirement already satisfied: ipython>=5.5.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (8.27.0)

Requirement already satisfied: ipywidgets>=7.6.5 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (7.8.1)

Requirement already satisfied: tqdm>=4.62.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (4.66.5)

Requirement already satisfied: numpy<1.27,>=1.21 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (1.26.4)

Requirement already satisfied: pandas<2.2.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (2.1.4)

Requirement already satisfied: jinja2>=3 in c:\users\pghat\anaconda3\lib\site-pac

kages (from pycaret) (3.1.4)  
Requirement already satisfied: scipy<=1.11.4,>=1.6.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (1.11.4)  
Requirement already satisfied: joblib<1.4,>=1.2.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (1.3.2)  
Requirement already satisfied: scikit-learn>1.4.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (1.4.2)  
Requirement already satisfied: pyod>=1.1.3 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (2.0.2)  
Requirement already satisfied: imbalanced-learn>=0.12.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (0.12.3)  
Requirement already satisfied: category-encoders>=2.4.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (2.6.4)  
Requirement already satisfied: lightgbm>=3.0.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (4.5.0)  
Requirement already satisfied: numba>=0.55.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (0.60.0)  
Requirement already satisfied: requests>=2.27.1 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (2.32.3)  
Requirement already satisfied: psutil>=5.9.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (6.0.0)  
Requirement already satisfied: markupsafe>=2.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (2.1.3)  
Requirement already satisfied: importlib-metadata>=4.12.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (7.0.1)  
Requirement already satisfied: nbformat>=4.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (5.10.4)  
Requirement already satisfied: cloudpickle in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (3.0.0)  
Requirement already satisfied: deprecation>=2.1.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (2.1.0)  
Requirement already satisfied: xxhash in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (3.5.0)  
Requirement already satisfied: matplotlib<3.8.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (3.7.5)  
Requirement already satisfied: scikit-plot>=0.3.7 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (0.3.7)  
Requirement already satisfied: yellowbrick>=1.4 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (1.5)  
Requirement already satisfied: plotly>=5.14.0 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (5.24.1)  
Requirement already satisfied: kaleido>=0.2.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (0.2.1)  
Requirement already satisfied: schemdraw==0.15 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (0.15)  
Requirement already satisfied: plotly-resampler>=0.8.3.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (0.10.0)  
Requirement already satisfied: statsmodels>=0.12.1 in c:\users\pghat\anaconda3\lib\site-packages (from pycaret) (0.14.2)  
Requirement already satisfied: sktime==0.26.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (0.26.0)  
Requirement already satisfied: tbats>=1.1.3 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (1.1.3)  
Requirement already satisfied: pmdarima>=2.0.4 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pycaret) (2.0.4)  
Requirement already satisfied: packaging in c:\users\pghat\appdata\roaming\python\python312\site-packages (from sktime==0.26.0->pycaret) (24.1)  
Requirement already satisfied: scikit-base<0.8.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from sktime==0.26.0->pycaret) (0.7.8)  
Requirement already satisfied: patsy>=0.5.1 in c:\users\pghat\anaconda3\lib\site-

packages (from category-encoders>=2.4.0->pycaret) (0.5.6)  
 Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from imbalanced-learn>=0.12.0->pycaret) (3.5.0)  
 Requirement already satisfied: zipp>=0.5 in c:\users\pghat\anaconda3\lib\site-packages (from importlib-metadata>=4.12.0->pycaret) (3.17.0)  
 Requirement already satisfied: decorator in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (5.1.1)  
 Requirement already satisfied: jedi>=0.16 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (0.19.1)  
 Requirement already satisfied: matplotlib-inline in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (0.1.7)  
 Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (3.0.47)  
 Requirement already satisfied: pygments>=2.4.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (2.18.0)  
 Requirement already satisfied: stack-data in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (0.6.3)  
 Requirement already satisfied: traitlets>=5.13.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (5.14.3)  
 Requirement already satisfied: colorama in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipython>=5.5.0->pycaret) (0.4.6)  
 Requirement already satisfied: comm>=0.1.3 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipywidgets>=7.6.5->pycaret) (0.2.2)  
 Requirement already satisfied: ipython-genutils<=0.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from ipywidgets>=7.6.5->pycaret) (0.2.0)  
 Requirement already satisfied: widgetsnbextension<=3.6.6 in c:\users\pghat\anaconda3\lib\site-packages (from ipywidgets>=7.6.5->pycaret) (3.6.6)  
 Requirement already satisfied: jupyterlab-widgets<3,>=1.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from ipywidgets>=7.6.5->pycaret) (1.0.0)  
 Requirement already satisfied: contourpy>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.8.0->pycaret) (1.2.0)  
 Requirement already satisfied: cycycler>=0.10 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.8.0->pycaret) (0.11.0)  
 Requirement already satisfied: fonttools>=4.22.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.8.0->pycaret) (4.51.0)  
 Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.8.0->pycaret) (1.4.4)  
 Requirement already satisfied: pillow>=6.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.8.0->pycaret) (10.4.0)  
 Requirement already satisfied: pyparsing>=2.3.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.8.0->pycaret) (3.1.2)  
 Requirement already satisfied: python-dateutil>=2.7 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from matplotlib<3.8.0->pycaret) (2.9.0.post0)  
 Requirement already satisfied: fastjsonschema>=2.15 in c:\users\pghat\anaconda3\lib\site-packages (from nbformat>=4.2.0->pycaret) (2.16.2)  
 Requirement already satisfied: jsonschema>=2.6 in c:\users\pghat\anaconda3\lib\site-packages (from nbformat>=4.2.0->pycaret) (4.23.0)  
 Requirement already satisfied: jupyter-core!=5.0.\*,>=4.12 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from nbformat>=4.2.0->pycaret) (5.7.2)  
 Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in c:\users\pghat\anaconda3\lib\site-packages (from numba>=0.55.0->pycaret) (0.43.0)  
 Requirement already satisfied: pytz>=2020.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas<2.2.0->pycaret) (2024.1)  
 Requirement already satisfied: tzdata>=2022.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas<2.2.0->pycaret) (2023.3)  
 Requirement already satisfied: tenacity>=6.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from plotly>=5.14.0->pycaret) (8.2.3)  
 Requirement already satisfied: dash>=2.9.0 in c:\users\pghat\appdata\roaming\python

on\python312\site-packages (from plotly-resampler>=0.8.3.1->pycaret) (2.18.1)  
Requirement already satisfied: orjson<4.0.0,>=3.8.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from plotly-resampler>=0.8.3.1->pycaret) (3.10.7)  
Requirement already satisfied: tsdownsample>=0.1.3 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from plotly-resampler>=0.8.3.1->pycaret) (0.1.3)  
Requirement already satisfied: Cython!=0.29.18,!0.29.31,>=0.29 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pmdarima>=2.0.4->pycaret) (3.0.11)  
Requirement already satisfied: urllib3 in c:\users\pghat\anaconda3\lib\site-packages (from pmdarima>=2.0.4->pycaret) (2.2.3)  
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in c:\users\pghat\anaconda3\lib\site-packages (from pmdarima>=2.0.4->pycaret) (75.1.0)  
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\pghat\anaconda3\lib\site-packages (from requests>=2.27.1->pycaret) (3.3.2)  
Requirement already satisfied: idna<4,>=2.5 in c:\users\pghat\anaconda3\lib\site-packages (from requests>=2.27.1->pycaret) (3.7)  
Requirement already satisfied: certifi>=2017.4.17 in c:\users\pghat\anaconda3\lib\site-packages (from requests>=2.27.1->pycaret) (2024.8.30)  
Requirement already satisfied: Flask<3.1,>=1.0.4 in c:\users\pghat\anaconda3\lib\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (3.0.3)  
Requirement already satisfied: Werkzeug<3.1 in c:\users\pghat\anaconda3\lib\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (3.0.3)  
Requirement already satisfied: dash-html-components==2.0.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.0)  
Requirement already satisfied: dash-core-components==2.0.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.0.0)  
Requirement already satisfied: dash-table==5.0.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (5.0.0)  
Requirement already satisfied: typing-extensions>=4.1.1 in c:\users\pghat\anaconda3\lib\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (4.11.0)  
Requirement already satisfied: retrying in c:\users\pghat\appdata\roaming\python\python312\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.3.4)  
Requirement already satisfied: nest-asyncio in c:\users\pghat\appdata\roaming\python\python312\site-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.6.0)  
Requirement already satisfied: parso<0.9.0,>=0.8.3 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jedi>=0.16->ipython>=5.5.0->pycaret) (0.8.4)  
Requirement already satisfied: attrs>=22.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (23.1.0)  
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\pghat\anaconda3\lib\site-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (2023.7.1)  
Requirement already satisfied: referencing>=0.28.4 in c:\users\pghat\anaconda3\lib\site-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (0.30.2)  
Requirement already satisfied: rpds-py>=0.7.1 in c:\users\pghat\anaconda3\lib\site-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (0.10.6)  
Requirement already satisfied: platformdirs>=2.5 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jupyter-core!=5.0.\*,>=4.12->nbformat>=4.2.0->pycaret) (4.2.2)  
Requirement already satisfied: pywin32>=300 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jupyter-core!=5.0.\*,>=4.12->nbformat>=4.2.0->pycaret) (306)

Requirement already satisfied: six in c:\users\pghat\appdata\roaming\python\python312\site-packages (from patsy>=0.5.1->category-encoders>=2.4.0->pycaret) (1.16.0)

Requirement already satisfied: wcwidth in c:\users\pghat\appdata\roaming\python\python312\site-packages (from prompt-toolkit<3.1.0,>=3.0.41->ipython>=5.5.0->pycaret) (0.2.13)

Requirement already satisfied: notebook>=4.4.1 in c:\users\pghat\anaconda3\lib\site-packages (from widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (7.2.2)

Requirement already satisfied: executing>=1.2.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from stack-data->ipython>=5.5.0->pycaret) (2.1.0)

Requirement already satisfied: asttokens>=2.1.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from stack-data->ipython>=5.5.0->pycaret) (2.4.1)

Requirement already satisfied: pure-eval in c:\users\pghat\appdata\roaming\python\python312\site-packages (from stack-data->ipython>=5.5.0->pycaret) (0.2.3)

Requirement already satisfied: itsdangerous>=2.1.2 in c:\users\pghat\anaconda3\lib\site-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.2.0)

Requirement already satisfied: click>=8.1.3 in c:\users\pghat\anaconda3\lib\site-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (8.1.7)

Requirement already satisfied: blinker>=1.6.2 in c:\users\pghat\anaconda3\lib\site-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.6.2)

Requirement already satisfied: jupyter-server<3,>=2.4.0 in c:\users\pghat\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.14.1)

Requirement already satisfied: jupyterlab-server<3,>=2.27.1 in c:\users\pghat\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.27.3)

Requirement already satisfied: jupyterlab<4.3,>=4.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (4.2.5)

Requirement already satisfied: notebook-shim<0.3,>=0.2 in c:\users\pghat\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.2.3)

Requirement already satisfied: tornado>=6.2.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (6.4.1)

Requirement already satisfied: anyio>=3.1.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (4.2.0)

Requirement already satisfied: argon2-cffi>=21.1 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (21.3.0)

Requirement already satisfied: jupyter-client>=7.4.4 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (8.6.2)

Requirement already satisfied: jupyter-events>=0.9.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.10.0)

Requirement already satisfied: jupyter-server-terminals>=0.4.4 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.4.4)

Requirement already satisfied: nbconvert>=6.4.4 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (7.16.4)

Requirement already satisfied: overrides>=5.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=

3.6.6->ipywidgets>=7.6.5->pycaret) (7.4.0)  
Requirement already satisfied: prometheus-client>=0.9 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.14.1)  
Requirement already satisfied: pywinpty>=2.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.0.10)  
Requirement already satisfied: pyzmq>=24 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (26.2.0)  
Requirement already satisfied: send2trash>=1.8.2 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.8.2)  
Requirement already satisfied: terminado>=0.8.3 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.17.1)  
Requirement already satisfied: websocket-client>=1.7 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.8.0)  
Requirement already satisfied: async-lru>=1.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.0.4)  
Requirement already satisfied: httpx>=0.25.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.27.0)  
Requirement already satisfied: ipykernel>=6.5.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (6.29.5)  
Requirement already satisfied: jupyter-lsp>=2.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.2.0)  
Requirement already satisfied: babel>=2.10 in c:\users\pghat\anaconda3\lib\site-packages (from jupyterlab-server<3,>=2.27.1->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.11.0)  
Requirement already satisfied: json5>=0.9.0 in c:\users\pghat\anaconda3\lib\site-packages (from jupyterlab-server<3,>=2.27.1->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.9.6)  
Requirement already satisfied: sniffio>=1.1 in c:\users\pghat\anaconda3\lib\site-packages (from anyio>=3.1.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.3.0)  
Requirement already satisfied: argon2-cffi-bindings in c:\users\pghat\anaconda3\lib\site-packages (from argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (21.2.0)  
Requirement already satisfied: httpcore==1.\* in c:\users\pghat\anaconda3\lib\site-packages (from httpx>=0.25.0->jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.0.2)  
Requirement already satisfied: h11<0.15,>=0.13 in c:\users\pghat\anaconda3\lib\site-packages (from httpcore==1.\*->httpx>=0.25.0->jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.14.0)  
Requirement already satisfied: debugpy>=1.6.5 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ipykernel>=6.5.0->jupyterlab<4.3,>=4.2.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.8.5)  
Requirement already satisfied: python-json-logger>=2.0.4 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.0.7)  
Requirement already satisfied: pyyaml>=5.3 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (6.0.1)  
Requirement already satisfied: rfc3339-validator in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=

4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.1.4)  
Requirement already satisfied: rfc3986-validator>=0.1.1 in c:\users\pghat\anaconda3\lib\site-packages (from jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.1.1)  
Requirement already satisfied: beautifulsoup4 in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (4.12.3)  
Requirement already satisfied: bleach!=5.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (4.1.0)  
Requirement already satisfied: defusedxml in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.7.1)  
Requirement already satisfied: jupyterlab-pygments in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.1.2)  
Requirement already satisfied: mistune<4,>=2.0.3 in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.0.4)  
Requirement already satisfied: nbclient>=0.5.0 in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.8.0)  
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.5.0)  
Requirement already satisfied: tinycss2 in c:\users\pghat\anaconda3\lib\site-packages (from nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.2.1)  
Requirement already satisfied: webencodings in c:\users\pghat\anaconda3\lib\site-packages (from bleach!=5.0.0->nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (0.5.1)  
Requirement already satisfied: fqdn in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.5.1)  
Requirement already satisfied: isoduration in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (20.11.0)  
Requirement already satisfied: jsonpointer>1.13 in c:\users\pghat\anaconda3\lib\site-packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.1)  
Requirement already satisfied: uri-template in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.3.0)  
Requirement already satisfied: webcolors>=24.6.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (24.8.0)  
Requirement already satisfied: cffi>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.17.1)  
Requirement already satisfied: soupsieve>1.2 in c:\users\pghat\anaconda3\lib\site-packages (from beautifulsoup4->nbconvert>=6.4.4->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (2.5)  
Requirement already satisfied: pycparser in c:\users\pghat\anaconda3\lib\site-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret



t) (2.21)

Requirement already satisfied: arrow>=0.15.0 in c:\users\pghat\anaconda3\lib\site-packages (from isoduration->jsonschema[format-nongpl]>=4.18.0->jupyter-events>=0.9.0->jupyter-server<3,>=2.4.0->notebook>=4.4.1->widgetsnbextension~=3.6.6->ipywidgets>=7.6.5->pycaret) (1.2.3)

In [2]: !conda info

```

active environment : base
active env location : C:\Users\pghat\anaconda3
    shell level : 1
    user config file : C:\Users\pghat\.condarc
populated config files : C:\Users\pghat\anaconda3\.condarc
                        C:\Users\pghat\.condarc
    conda version : 24.9.2
conda-build version : 24.9.0
    python version : 3.12.7.final.0
        solver : libmamba (default)
virtual packages : __archspec=1=x86_64_v4
                  __conda=24.9.2=0
                  __win=0=0
base environment : C:\Users\pghat\anaconda3 (writable)
conda av data dir : C:\Users\pghat\anaconda3\etc\conda
conda av metadata url : None
    channel URLs : https://repo.anaconda.com/pkgs/main/win-64
                  https://repo.anaconda.com/pkgs/main/noarch
                  https://repo.anaconda.com/pkgs/r/win-64
                  https://repo.anaconda.com/pkgs/r/noarch
                  https://repo.anaconda.com/pkgs/msys2/win-64
                  https://repo.anaconda.com/pkgs/msys2/noarch
package cache : C:\Users\pghat\anaconda3\pkgs
                C:\Users\pghat\.conda\pkgs
                C:\Users\pghat\AppData\Local\conda\conda\pkgs
envs directories : C:\Users\pghat\anaconda3\envs
                  C:\Users\pghat\.conda\envs
                  C:\Users\pghat\AppData\Local\conda\conda\envs
    platform : win-64
    user-agent : conda/24.9.2 requests/2.32.3 CPython/3.12.7 Windows/11
Windows/10.0.22631 solver/libmamba conda-libmamba-solver/24.9.0 libmambapy/1.5.8
aau/0.4.4 c/. s/. e/.
    administrator : False
    netrc file : None
    offline mode : False

```

In [3]: !python -V

Python 3.12.7

In [4]: !pip install pandas  
!pip install numpy

Requirement already satisfied: pandas in c:\users\pghat\appdata\roaming\python\python312\site-packages (2.1.4)  
 Requirement already satisfied: numpy<2,>=1.26.0 in c:\users\pghat\anaconda3\lib\site-packages (from pandas) (1.26.4)  
 Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pandas) (2.9.0.post0)  
 Requirement already satisfied: pytz>=2020.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas) (2024.1)  
 Requirement already satisfied: tzdata>=2022.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas) (2023.3)  
 Requirement already satisfied: six>=1.5 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)  
 Requirement already satisfied: numpy in c:\users\pghat\anaconda3\lib\site-packages (1.26.4)

In [5]: !pip install seaborn

Requirement already satisfied: seaborn in c:\users\pghat\anaconda3\lib\site-packages (0.13.2)  
 Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\pghat\anaconda3\lib\site-packages (from seaborn) (1.26.4)  
 Requirement already satisfied: pandas>=1.2 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from seaborn) (2.1.4)  
 Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from seaborn) (3.7.5)  
 Requirement already satisfied: contourpy>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)  
 Requirement already satisfied: cycler>=0.10 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)  
 Requirement already satisfied: fonttools>=4.22.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)  
 Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)  
 Requirement already satisfied: packaging>=20.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.1)  
 Requirement already satisfied: pillow>=6.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.4.0)  
 Requirement already satisfied: pyparsing>=2.3.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.1.2)  
 Requirement already satisfied: python-dateutil>=2.7 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)  
 Requirement already satisfied: pytz>=2020.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas>=1.2->seaborn) (2024.1)  
 Requirement already satisfied: tzdata>=2022.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas>=1.2->seaborn) (2023.3)  
 Requirement already satisfied: six>=1.5 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)

In [6]: !pip install -r requirements.txt

Requirement already satisfied: pandas in c:\users\pghat\appdata\roaming\python\python312\site-packages (from -r requirements.txt (line 1)) (2.1.4)

Requirement already satisfied: seaborn in c:\users\pghat\anaconda3\lib\site-packages (from -r requirements.txt (line 2)) (0.13.2)

Requirement already satisfied: matplotlib in c:\users\pghat\appdata\roaming\python\python312\site-packages (from -r requirements.txt (line 3)) (3.7.5)

Requirement already satisfied: ydata-profiling in c:\users\pghat\anaconda3\lib\site-packages (from -r requirements.txt (line 4)) (4.12.0)

Requirement already satisfied: numpy<2,>=1.26.0 in c:\users\pghat\anaconda3\lib\site-packages (from pandas->-r requirements.txt (line 1)) (1.26.4)

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from pandas->-r requirements.txt (line 1)) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas->-r requirements.txt (line 1)) (2024.1)

Requirement already satisfied: tzdata>=2022.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas->-r requirements.txt (line 1)) (2023.3)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (1.2.0)

Requirement already satisfied: cyclor>=0.10 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (0.11.0)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (4.51.0)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (1.4.4)

Requirement already satisfied: packaging>=20.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from matplotlib->-r requirements.txt (line 3)) (24.1)

Requirement already satisfied: pillow>=6.2.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (10.4.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (3.1.2)

Requirement already satisfied: scipy<1.14,>=1.4.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (1.11.4)

Requirement already satisfied: pydantic>=2 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (2.8.2)

Requirement already satisfied: PyYAML<6.1,>=5.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (6.0.1)

Requirement already satisfied: jinja2<3.2,>=2.11.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (3.1.4)

Requirement already satisfied: visions<0.7.7,>=0.7.5 in c:\users\pghat\anaconda3\lib\site-packages (from visions[type\_image\_path]<0.7.7,>=0.7.5->ydata-profiling->-r requirements.txt (line 4)) (0.7.6)

Requirement already satisfied: htmlmin==0.1.12 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (0.1.12)

Requirement already satisfied: phik<0.13,>=0.11.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (0.12.4)

Requirement already satisfied: requests<3,>=2.24.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (2.32.3)

Requirement already satisfied: tqdm<5,>=4.48.2 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (4.66.5)

Requirement already satisfied: multimethod<2,>=1.4 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (1.12)

Requirement already satisfied: statsmodels<1,>=0.13.2 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (0.14.2)

Requirement already satisfied: typeguard<5,>=3 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (4.4.1)

Requirement already satisfied: imagehash==4.3.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (4.3.1)

Requirement already satisfied: wordcloud>=1.9.3 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (1.9.3)

Requirement already satisfied: dacite>=1.8 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (1.8.1)

Requirement already satisfied: numba<1,>=0.56.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling->-r requirements.txt (line 4)) (0.60.0)

Requirement already satisfied: PyWavelets in c:\users\pghat\anaconda3\lib\site-packages (from imagehash==4.3.1->ydata-profiling->-r requirements.txt (line 4)) (1.7.0)

Requirement already satisfied: MarkupSafe>=2.0 in c:\users\pghat\anaconda3\lib\site-packages (from jinja2<3.2,>=2.11.1->ydata-profiling->-r requirements.txt (line 4)) (2.1.3)

Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in c:\users\pghat\anaconda3\lib\site-packages (from numba<1,>=0.56.0->ydata-profiling->-r requirements.txt (line 4)) (0.43.0)

Requirement already satisfied: joblib>=0.14.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from phik<0.13,>=0.11.1->ydata-profiling->-r requirements.txt (line 4)) (1.3.2)

Requirement already satisfied: annotated-types>=0.4.0 in c:\users\pghat\anaconda3\lib\site-packages (from pydantic>=2->ydata-profiling->-r requirements.txt (line 4)) (0.6.0)

Requirement already satisfied: pydantic-core==2.20.1 in c:\users\pghat\anaconda3\lib\site-packages (from pydantic>=2->ydata-profiling->-r requirements.txt (line 4)) (2.20.1)

Requirement already satisfied: typing-extensions>=4.6.1 in c:\users\pghat\anaconda3\lib\site-packages (from pydantic>=2->ydata-profiling->-r requirements.txt (line 4)) (4.11.0)

Requirement already satisfied: six>=1.5 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from python-dateutil>=2.8.2->pandas->-r requirements.txt (line 1)) (1.16.0)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling->-r requirements.txt (line 4)) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling->-r requirements.txt (line 4)) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling->-r requirements.txt (line 4)) (2.2.3)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling->-r requirements.txt (line 4)) (2024.8.30)

Requirement already satisfied: patsy>=0.5.6 in c:\users\pghat\anaconda3\lib\site-packages (from statsmodels<1,>=0.13.2->ydata-profiling->-r requirements.txt (line 4)) (0.5.6)

Requirement already satisfied: colorama in c:\users\pghat\appdata\roaming\python\python312\site-packages (from tqdm<5,>=4.48.2->ydata-profiling->-r requirements.txt (line 4)) (0.4.6)

Requirement already satisfied: attrs>=19.3.0 in c:\users\pghat\anaconda3\lib\site-packages (from visions<0.7.7,>=0.7.5->visions[type\_image\_path]<0.7.7,>=0.7.5->ydata-profiling->-r requirements.txt (line 4)) (23.1.0)

Requirement already satisfied: networkx>=2.4 in c:\users\pghat\anaconda3\lib\site-packages (from visions<0.7.7,>=0.7.5->visions[type\_image\_path]<0.7.7,>=0.7.5->ydata-profiling->-r requirements.txt (line 4)) (3.3)

```
In [7]: import warnings
warnings.filterwarnings('ignore')
warnings.simplefilter('ignore')
warnings.filterwarnings('ignore', category=DeprecationWarning)
warnings.filterwarnings('ignore', category=FutureWarning)
```

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from ydata_profiling import ProfileReport

# Configure matplotlib to display plots inline
%matplotlib inline

# Set style for better-looking visualizations
# plt.style.use('seaborn')
sns.set_palette('deep')

```

Data Loading and Initial Inspection -- We'll load the sentencing dataset and perform initial exploratory data analysis (EDA).

```

In [8]: # Load the dataset
df = pd.read_csv('data\Sentencing_20241024.csv')

# Display basic information about the dataset
print("Dataset Shape:", df.shape)
print("\nFirst few rows:")
display(df.head())

# Display data types and non-null counts
print("\nDataset Info:")
df.info()

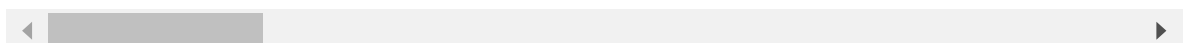
```

Dataset Shape: (303963, 41)

First few rows:

	CASE_ID	CASE_PARTICIPANT_ID	RECEIVED_DATE	OFFENSE_CATEGORY	PRIMARY_
0	149352426337	414373090914	08/15/1984 12:00:00 AM	PROMIS Conversion	
1	149352426337	414373090914	08/15/1984 12:00:00 AM	PROMIS Conversion	
2	149352426337	414373090914	08/15/1984 12:00:00 AM	PROMIS Conversion	
3	149352426337	414373090914	08/15/1984 12:00:00 AM	PROMIS Conversion	
4	149352426337	414373090914	08/15/1984 12:00:00 AM	PROMIS Conversion	

5 rows × 41 columns



Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 303963 entries, 0 to 303962

Data columns (total 41 columns):

#	Column	Non-Null Count	Dtype
0	CASE_ID	303963 non-null	int64
1	CASE_PARTICIPANT_ID	303963 non-null	int64
2	RECEIVED_DATE	303963 non-null	object
3	OFFENSE_CATEGORY	303963 non-null	object
4	PRIMARY_CHARGE_FLAG	303963 non-null	bool
5	CHARGE_ID	303963 non-null	int64
6	CHARGE_VERSION_ID	303963 non-null	int64
7	DISPOSITION_CHARGED_OFFENSE_TITLE	303963 non-null	object
8	CHARGE_COUNT	303963 non-null	int64
9	DISPOSITION_DATE	303963 non-null	object
10	DISPOSITION_CHARGED_CHAPTER	303963 non-null	object
11	DISPOSITION_CHARGED_ACT	298521 non-null	object
12	DISPOSITION_CHARGED_SECTION	298521 non-null	object
13	DISPOSITION_CHARGED_CLASS	303944 non-null	object
14	DISPOSITION_CHARGED_AOIC	303939 non-null	object
15	CHARGE_DISPOSITION	303963 non-null	object
16	CHARGE_DISPOSITION_REASON	1324 non-null	object
17	SENTENCE_JUDGE	303221 non-null	object
18	SENTENCE_COURT_NAME	302566 non-null	object
19	SENTENCE_COURT_FACILITY	301669 non-null	object
20	SENTENCE_PHASE	303963 non-null	object
21	SENTENCE_DATE	303963 non-null	object
22	SENTENCE_TYPE	303963 non-null	object
23	CURRENT_SENTENCE_FLAG	303963 non-null	bool
24	COMMITMENT_TYPE	302131 non-null	object
25	COMMITMENT_TERM	302097 non-null	object
26	COMMITMENT_UNIT	302097 non-null	object
27	LENGTH_OF_CASE_in_Days	284296 non-null	float64
28	AGE_AT_INCIDENT	299461 non-null	float64
29	RACE	302329 non-null	object
30	GENDER	302918 non-null	object
31	INCIDENT_CITY	282097 non-null	object
32	INCIDENT_BEGIN_DATE	300418 non-null	object
33	INCIDENT_END_DATE	25859 non-null	object
34	LAW_ENFORCEMENT_AGENCY	303633 non-null	object
35	LAW_ENFORCEMENT_UNIT	87294 non-null	object
36	ARREST_DATE	297286 non-null	object
37	FELONY_REVIEW_DATE	218346 non-null	object
38	FELONY_REVIEW_RESULT	218346 non-null	object
39	ARRAIGNMENT_DATE	284296 non-null	object
40	UPDATED_OFFENSE_CATEGORY	303963 non-null	object

dtypes: bool(2), float64(2), int64(5), object(32)

memory usage: 91.0+ MB

## Data Inspection and Data Cleaning

```
In [9]: # Checking for missing values
df.isnull().sum()

# which values are in the unit column
df['COMMITMENT_UNIT'].value_counts()
```

```
Out[9]: COMMITMENT_UNIT
Year(s)      219237
Months      68123
Days        11348
Term         2374
Natural Life   841
Dollars       94
Hours         47
Weeks         24
Pounds         5
Grams          2
Ounces         1
Kilos          1
Name: count, dtype: int64
```

```
In [10]: # Create a copy of data with only 'Year(s)' sentences
years_df = df[df['COMMITMENT_UNIT'] == 'Year(s)'].copy()

# Convert commitment terms to numeric values
years_df['COMMITMENT_TERM'] = pd.to_numeric(years_df['COMMITMENT_TERM'], errors=

# Display summary statistics of commitment terms
print("Summary Statistics for Commitment Terms (Years):")
print(years_df['COMMITMENT_TERM'].describe())
```

Summary Statistics for Commitment Terms (Years):

```
count    2.192340e+05
mean      5.544476e+01
std       1.640220e+04
min        0.000000e+00
25%        2.000000e+00
50%        2.000000e+00
75%        4.000000e+00
max        7.122023e+06
```

Name: COMMITMENT\_TERM, dtype: float64

Select Relevant Columns: We'll choose the most important columns for our analysis.

```
In [11]: # Define columns to keep
important_columns = [
    # Temporal information
    'RECEIVED_DATE', 'DISPOSITION_DATE', 'SENTENCE_DATE',

    # Case details
    'OFFENSE_CATEGORY', 'DISPOSITION_CHARGED_OFFENSE_TITLE',
    'CHARGE_COUNT', 'CHARGE_DISPOSITION',

    # Sentencing information
    'SENTENCE_JUDGE', 'SENTENCE_COURT_NAME',
    'SENTENCE_TYPE', 'COMMITMENT_TYPE', 'COMMITMENT_TERM',

    # Demographics
    'AGE_AT_INCIDENT', 'RACE', 'GENDER',

    # Location
    'INCIDENT_CITY', 'LAW_ENFORCEMENT_AGENCY'
]
```

```
# Create filtered dataset
df_filtered = df[important_columns].copy()
```

Handle Dates: Convert date strings to datetime objects for better analysis.

```
In [12]: # Convert date columns to datetime
date_columns = ['RECEIVED_DATE', 'DISPOSITION_DATE', 'SENTENCE_DATE']

for col in date_columns:
    df_filtered[col] = pd.to_datetime(df_filtered[col], errors='coerce')

# Add year column for easier analysis
df_filtered['SENTENCE_YEAR'] = df_filtered['SENTENCE_DATE'].dt.year

# Filter to recent years (2010-2023)
df_filtered = df_filtered[
    (df_filtered['SENTENCE_YEAR'] >= 2010) &
    (df_filtered['SENTENCE_YEAR'] <= 2023)
]
```

```
In [13]: # Ensure columns are numeric
df_filtered['COMMITMENT_TERM'] = pd.to_numeric(df_filtered['COMMITMENT_TERM'], e
df_filtered['AGE_AT_INCIDENT'] = pd.to_numeric(df_filtered['AGE_AT_INCIDENT'], e

# Calculate summary statistics by race
summary_stats = df_filtered.groupby('RACE').agg({
    'COMMITMENT_TERM': ['mean', 'median', 'std', 'count'],
    'AGE_AT_INCIDENT': ['mean', 'median', 'std']
}).round(2)

print("Summary Statistics by Race:")
display(summary_stats)
```

Summary Statistics by Race:



	COMMITMENT_TERM				AGE_AT_INCIDENT		
	mean	median	std	count	mean	median	std
RACE							
ASIAN	10.95	8.0	12.25	66	32.61	31.0	10.74
American Indian	14.25	4.0	36.29	134	33.60	30.0	10.38
Asian	14.79	3.0	40.64	1762	34.29	32.0	11.59
Biracial	10.24	3.0	18.37	37	28.51	26.0	10.33
Black	69.59	3.0	17417.22	194426	32.11	29.0	11.96
HISPANIC	13.74	6.0	22.64	6111	32.25	31.0	9.94
Middle Eastern/North African	5.14	2.0	8.34	7	29.43	26.0	8.52
Unknown	11.86	3.0	29.23	453	33.34	30.5	11.66
White	13.00	3.0	32.56	39753	34.77	33.0	11.80
White [Hispanic or Latino]	11.96	3.0	32.96	45553	31.26	29.0	10.47
White/Black [Hispanic or Latino]	12.85	3.0	36.50	1628	31.39	29.0	10.93

Clean Categorical Variables: Standardize race categories and other categorical variables.

```
In [14]: # Standardize race categories
race_mapping = {
    'ASIAN': 'Asian',
    'White [Hispanic or Latino]': 'Hispanic',
    'HISPANIC': 'Hispanic',
    'WHITE': 'White',
    'BLACK': 'Black'
}

df_filtered['RACE'] = df_filtered['RACE'].replace(race_mapping)

# Remove ambiguous categories
df_filtered = df_filtered[df_filtered['RACE'] != 'White/Black [Hispanic or Latino]']

# Display race distribution
print("Race Distribution:")
print(df_filtered['RACE'].value_counts(normalize=True).round(3))
```

Race Distribution:

```
RACE
Black                0.674
Hispanic             0.179
White                0.138
Asian                0.006
Unknown              0.002
American Indian      0.000
Biracial             0.000
Middle Eastern/North African 0.000
Name: proportion, dtype: float64
```

```
In [17]: # Define a regex pattern to match any string containing spaces, letters, commas,
pattern = '[\sA-Za-z,`]+'

# Use the regex pattern to filter rows in the 'SENTENCE_DATE' column that contain spaces
# The 'na=False' parameter ensures that NaN values are not included in the filter
df_filtered.loc[df['SENTENCE_DATE'].str.contains(pattern, na=False), ['SENTENCE_
```

Out[17]:

	SENTENCE_DATE
4	2014-10-16
6	2014-10-16
9	2014-10-16
12	2014-08-06
14	2014-08-06
...	...
301623	2023-12-07
301649	2023-12-15
301843	2023-12-19
303614	2021-08-06
303615	2021-08-06

291588 rows × 1 columns

```
In [16]: df_filtered['SENTENCE_DATE'].dt.year.value_counts()
```

Out[16]:

SENTENCE_DATE	
2014	31135
2013	30840
2012	30338
2015	27541
2011	27177
2016	25984
2017	23747
2018	20593
2019	19087
2022	16233
2023	15791
2021	14618
2020	8119
2010	385

Name: count, dtype: int64

```
In [18]: df_filtered.loc[(df_filtered['SENTENCE_DATE'].dt.year >= 2010.0) & (df_filtered[
```

Out[18]:

	RECEIVED_DATE	DISPOSITION_DATE	SENTENCE_DATE	OFFENSE_CATEGORY	DI
<b>4</b>	1984-08-15	2014-12-17	2014-10-16	PROMIS Conversion	
<b>6</b>	1984-08-15	2014-12-17	2014-10-16	PROMIS Conversion	
<b>9</b>	1984-08-15	2014-12-17	2014-10-16	PROMIS Conversion	
<b>12</b>	1984-08-23	2014-08-06	2014-08-06	PROMIS Conversion	
<b>14</b>	1984-08-23	2014-08-06	2014-08-06	PROMIS Conversion	
...	...	...	...	...	...
<b>301623</b>	2023-11-16	2023-12-07	2023-12-07	Narcotics	
<b>301649</b>	2023-11-17	2023-12-15	2023-12-15	Violation Order Of Protection	
<b>301843</b>	2023-11-29	2023-12-19	2023-12-19	Aggravated Battery Police Officer	
<b>303614</b>	2024-05-01	2024-08-14	2021-08-06	Forgery	
<b>303615</b>	2024-05-01	2024-08-14	2021-08-06	Forgery	

291588 rows × 18 columns



In [19]: `df_filtered['SENTENCE_year']=df_filtered['SENTENCE_DATE'].dt.year`  
`print(df_filtered['SENTENCE_year'])`

```

4      2014
6      2014
9      2014
12     2014
14     2014
...
301623 2023
301649 2023
301843 2023
303614 2021
303615 2021

```

Name: SENTENCE\_year, Length: 291588, dtype: int32

In [20]: `# df_filtered.groupby('RACE')['SENTENCE_year'].median()`  
`df_filtered['RACE'].value_counts()`

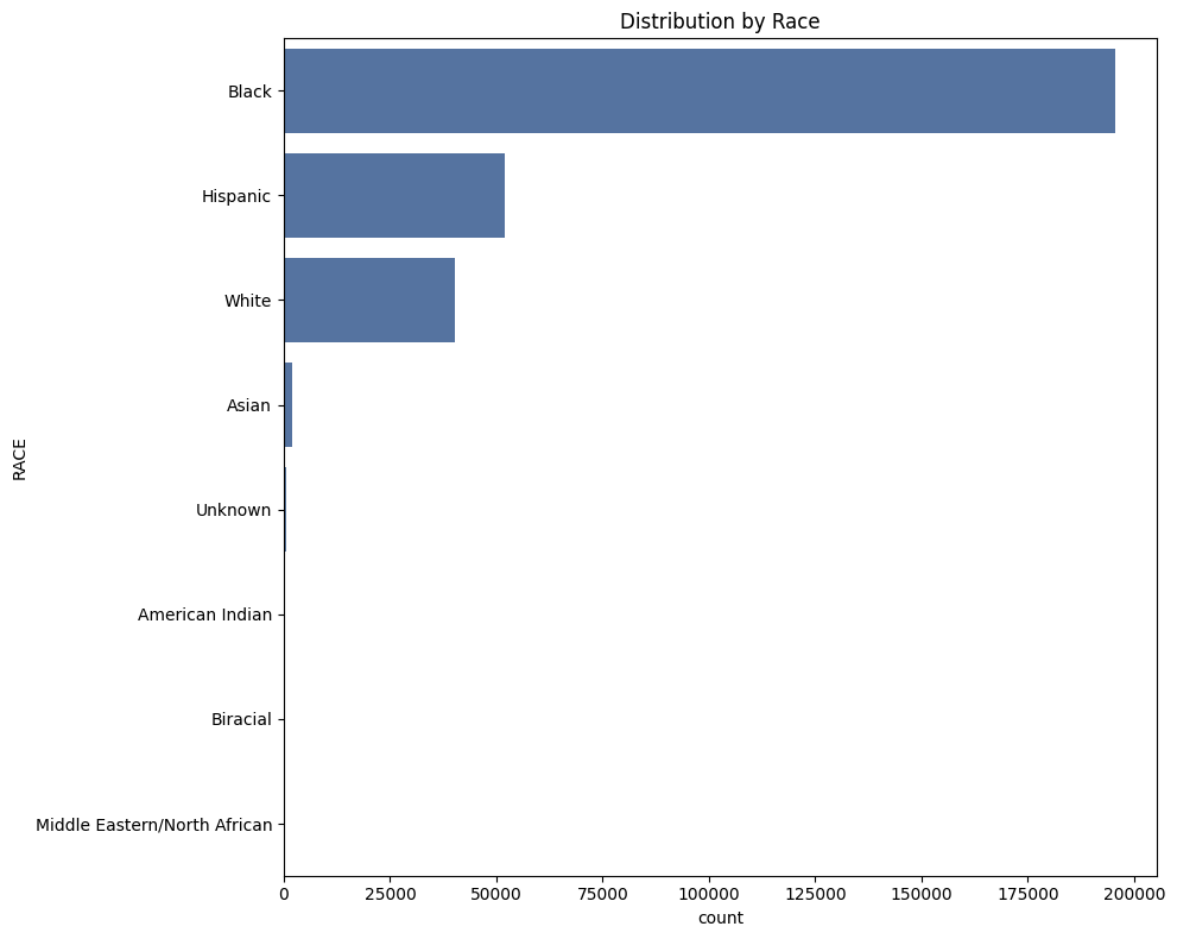
```
Out[20]: RACE
Black      195542
Hispanic   51916
White      40129
Asian      1852
Unknown    453
American Indian  136
Biracial    37
Middle Eastern/North African  7
Name: count, dtype: int64
```

```
In [21]: #collapse race categories
df_filtered['RACE'].replace("ASIAN", "Asian", inplace = True)
df_filtered['RACE'].replace("White [Hispanic or Latino]", "Hispanic", inplace =
df_filtered['RACE'].replace("HISPANIC", "Hispanic", inplace = True)

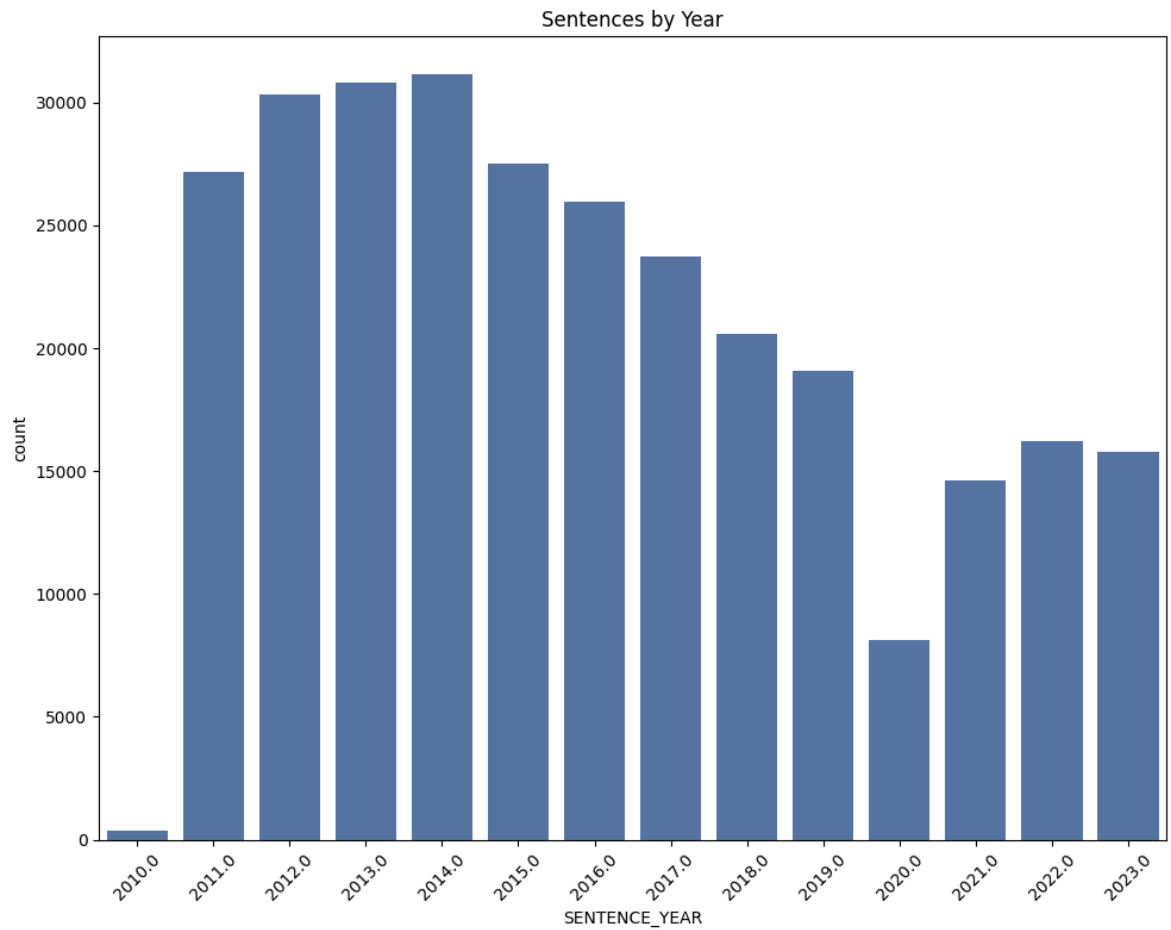
keep_df = df_filtered[df_filtered['RACE'] != "White/Black [Hispanic or Latino]"]
keep_df['RACE'].value_counts()
```

```
Out[21]: RACE
Black      195542
Hispanic   51916
White      40129
Asian      1852
Unknown    453
American Indian  136
Biracial    37
Middle Eastern/North African  7
Name: count, dtype: int64
```

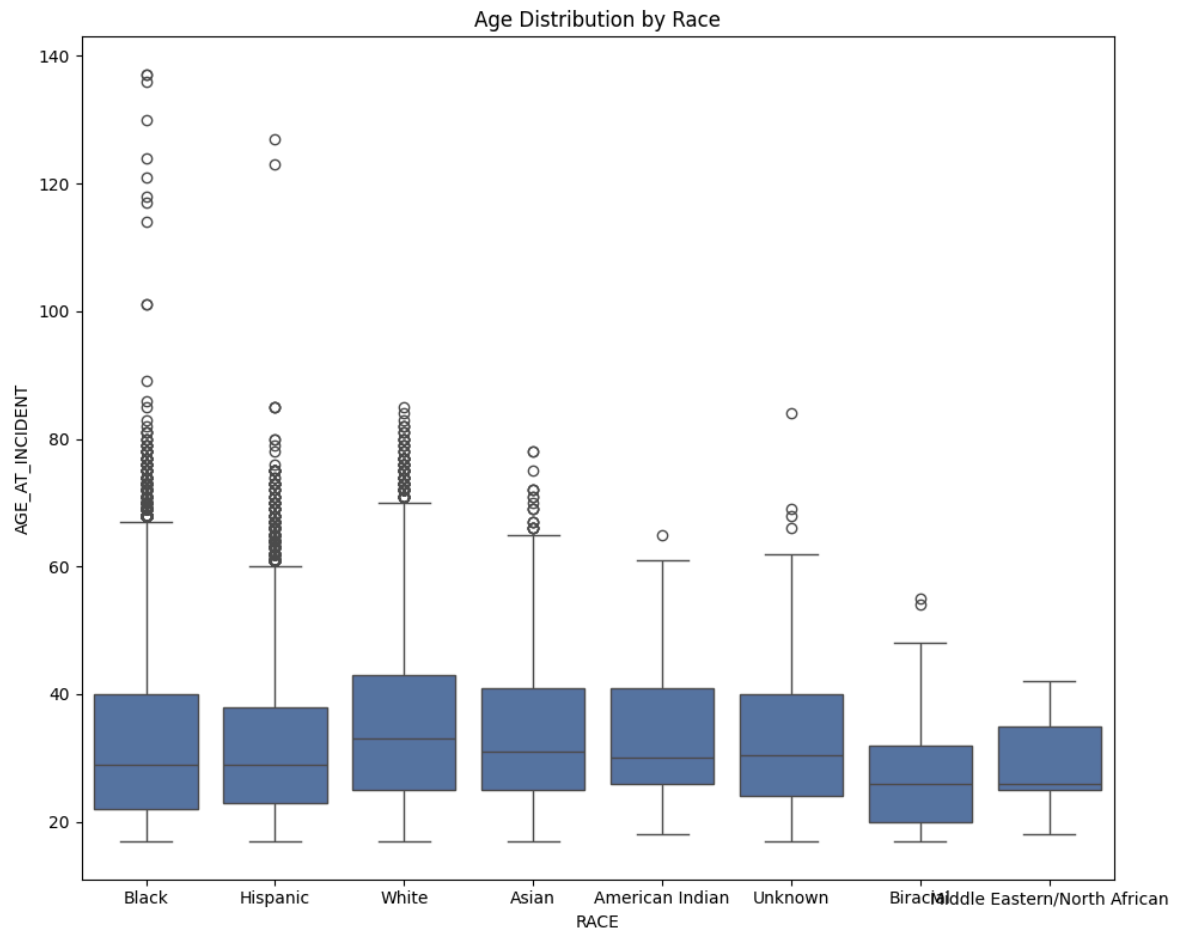
```
In [22]: # Race Distribution
plt.figure(figsize=(10, 8))
sns.countplot(data=df_filtered, y='RACE',
              order=df_filtered['RACE'].value_counts().index)
plt.title('Distribution by Race')
plt.tight_layout()
plt.show()
```



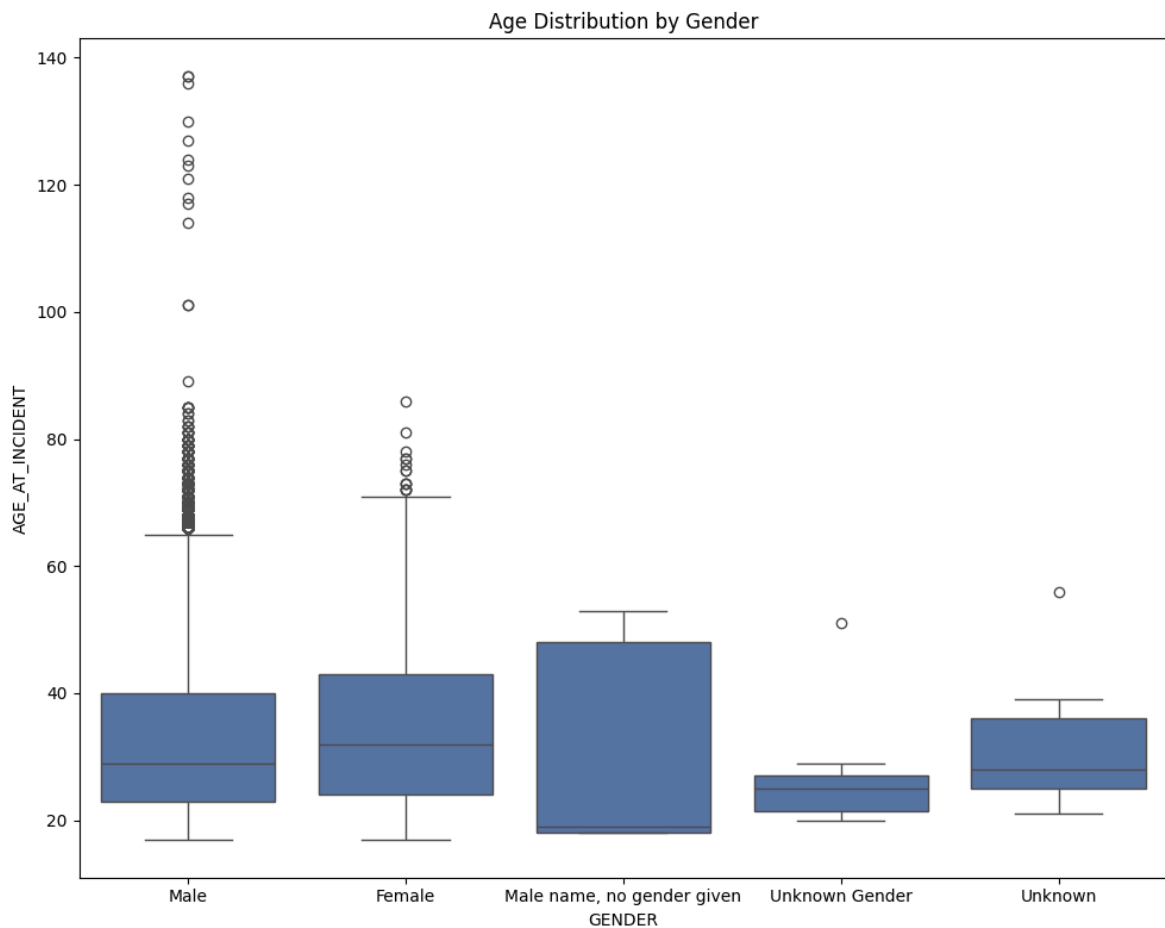
```
In [23]: # Sentences by Year
plt.figure(figsize=(10, 8))
sns.countplot(data=df_filtered, x='SENTENCE_YEAR')
plt.xticks(rotation=45)
plt.title('Sentences by Year')
plt.tight_layout()
plt.show()
```



```
In [24]: # Age Distribution by Race
plt.figure(figsize=(10, 8))
sns.boxplot(data=df_filtered, x='RACE', y='AGE_AT_INCIDENT')
plt.title('Age Distribution by Race')
plt.tight_layout()
plt.show()
```



```
In [25]: # Age Distribution by Gender
plt.figure(figsize=(10, 8))
sns.boxplot(data=df_filtered, x='GENDER', y='AGE_AT_INCIDENT')
plt.title('Age Distribution by Gender')
plt.tight_layout()
```



```
In [ ]: # It will take 7 minutes to run the this code to generate the profiling report
# Generate the report
# profile = ProfileReport(df, title="Sentencing Data Profile", explorative=True)
# profile.to_file("ydata_profiling_report.html")
```

```
In [38]: !pip install ydata-profiling
```



Requirement already satisfied: ydata-profiling in c:\users\pghat\anaconda3\lib\site-packages (4.12.0)

Requirement already satisfied: scipy<1.14,>=1.4.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ydata-profiling) (1.11.4)

Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ydata-profiling) (2.1.4)

Requirement already satisfied: matplotlib<3.10,>=3.5 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from ydata-profiling) (3.7.5)

Requirement already satisfied: pydantic>=2 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (2.8.2)

Requirement already satisfied: PyYAML<6.1,>=5.0.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (6.0.1)

Requirement already satisfied: jinja2<3.2,>=2.11.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (3.1.4)

Requirement already satisfied: visions<0.7.7,>=0.7.5 in c:\users\pghat\anaconda3\lib\site-packages (from visions[type\_image\_path]<0.7.7,>=0.7.5->ydata-profiling) (0.7.6)

Requirement already satisfied: numpy<2.2,>=1.16.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (1.26.4)

Requirement already satisfied: htmlmin==0.1.12 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (0.1.12)

Requirement already satisfied: phik<0.13,>=0.11.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (0.12.4)

Requirement already satisfied: requests<3,>=2.24.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (2.32.3)

Requirement already satisfied: tqdm<5,>=4.48.2 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (4.66.5)

Requirement already satisfied: seaborn<0.14,>=0.10.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (0.13.2)

Requirement already satisfied: multimethod<2,>=1.4 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (1.12)

Requirement already satisfied: statsmodels<1,>=0.13.2 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (0.14.2)

Requirement already satisfied: typeguard<5,>=3 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (4.4.1)

Requirement already satisfied: imagehash==4.3.1 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (4.3.1)

Requirement already satisfied: wordcloud>=1.9.3 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (1.9.3)

Requirement already satisfied: dacite>=1.8 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (1.8.1)

Requirement already satisfied: numba<1,>=0.56.0 in c:\users\pghat\anaconda3\lib\site-packages (from ydata-profiling) (0.60.0)

Requirement already satisfied: PyWavelets in c:\users\pghat\anaconda3\lib\site-packages (from imagehash==4.3.1->ydata-profiling) (1.7.0)

Requirement already satisfied: pillow in c:\users\pghat\anaconda3\lib\site-packages (from imagehash==4.3.1->ydata-profiling) (10.4.0)

Requirement already satisfied: MarkupSafe>=2.0 in c:\users\pghat\anaconda3\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\pghat\anaconda3\lib\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (1.2.0)

Requirement already satisfied: cycler>=0.10 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (0.11.0)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (4.51.0)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (1.4.4)

Requirement already satisfied: packaging>=20.0 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (24.1)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\pghat\anaconda3\lib\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (3.1.2)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from matplotlib<3.10,>=3.5->ydata-profiling) (2.9.0.post0)

Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in c:\users\pghat\anaconda3\lib\site-packages (from numba<1,>=0.56.0->ydata-profiling) (0.43.0)

Requirement already satisfied: pytz>=2020.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling) (2024.1)

Requirement already satisfied: tzdata>=2022.1 in c:\users\pghat\anaconda3\lib\site-packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling) (2023.3)

Requirement already satisfied: joblib>=0.14.1 in c:\users\pghat\appdata\roaming\python\python312\site-packages (from phik<0.13,>=0.11.1->ydata-profiling) (1.3.2)

Requirement already satisfied: annotated-types>=0.4.0 in c:\users\pghat\anaconda3\lib\site-packages (from pydantic>=2->ydata-profiling) (0.6.0)

Requirement already satisfied: pydantic-core==2.20.1 in c:\users\pghat\anaconda3\lib\site-packages (from pydantic>=2->ydata-profiling) (2.20.1)

Requirement already satisfied: typing-extensions>=4.6.1 in c:\users\pghat\anaconda3\lib\site-packages (from pydantic>=2->ydata-profiling) (4.11.0)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling) (2.2.3)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\pghat\anaconda3\lib\site-packages (from requests<3,>=2.24.0->ydata-profiling) (2024.8.30)

Requirement already satisfied: patsy>=0.5.6 in c:\users\pghat\anaconda3\lib\site-packages (from statsmodels<1,>=0.13.2->ydata-profiling) (0.5.6)

Requirement already satisfied: colorama in c:\users\pghat\appdata\roaming\python\python312\site-packages (from tqdm<5,>=4.48.2->ydata-profiling) (0.4.6)

Requirement already satisfied: attrs>=19.3.0 in c:\users\pghat\anaconda3\lib\site-packages (from visions<0.7.7,>=0.7.5->visions[type\_image\_path]<0.7.7,>=0.7.5->ydata-profiling) (23.1.0)

Requirement already satisfied: networkx>=2.4 in c:\users\pghat\anaconda3\lib\site-packages (from visions<0.7.7,>=0.7.5->visions[type\_image\_path]<0.7.7,>=0.7.5->ydata-profiling) (3.3)

Requirement already satisfied: six in c:\users\pghat\appdata\roaming\python\python312\site-packages (from patsy>=0.5.6->statsmodels<1,>=0.13.2->ydata-profiling) (1.16.0)

## Import pandas.

## Import the ProfileReport class from the ydata\_profiling library.

## Create a DataFrame using your data.

## Use the ProfileReport() class and pass the DataFrame.

```
In [39]: import ydata_profiling as yp
```

```
In [40]: import pandas as pd
from ydata_profiling import ProfileReport

df = pd.read_csv('data\Sentencing_20241024.csv')
profile = ProfileReport(df, title="Profiling Report")
```

Here's the straightforward code following the steps outlined above. First, we import the necessary libraries and then read the CSV file using the `read_csv()` function. In this case, we're using the CSV file of Setencing Reviews. Next, we use the `ProfileReport` class and pass our `DataFrame` inside it.

Additionally, we're setting a new title, "ProfileReport". By default, the title is something else, but if you want to customize it, use the title variable within the class. Finally, to generate and display the report, you can either use `profile` or `profile.to_notebook_iframe()`.

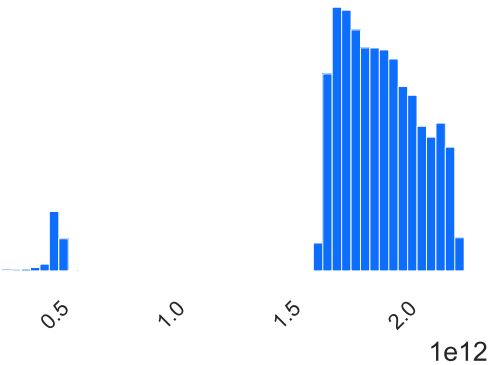
The report will be generated in the following sequence: First, the entire dataset will be summarized. Then, the report structure will be generated. Finally, it will display the report, which you can save as an HTML file and use for further analysis.

```
In [42]: from ydata_profiling import ProfileReport
import pandas as pd
df = pd.read_csv('data\Sentencing_20241024.csv')
profile = ProfileReport(df, title="Sentencing_20241024")
profile.to_notebook_iframe()
```

```
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]
```

Sentencing\_20241024

Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	2.3 MiB



More details

RECEIVED\_DATE

Date

Distinct	7067
Distinct (%)	2.3%
Missing	0
Missing (%)	0.0%