

lungs-cancer-eda

July 8, 2024

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
#import pandas_profiling as pp
import statsmodels.api as sm
```

```
[ ]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/lungs cancers PWskill/EDA question 2_
↳lungs cancer/dataset.csv")
```

```
[ ]: df
```

```
[ ]:
```

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	\
0	M	65	1	1	1	2	
1	F	55	1	2	2	1	
2	F	78	2	2	1	1	
3	M	60	2	1	1	1	
4	F	80	1	1	2	1	
...	
2995	F	71	2	1	1	2	
2996	F	75	1	2	1	1	
2997	F	62	2	2	2	1	
2998	M	30	1	1	2	2	
2999	M	40	1	2	2	1	

	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL_CONSUMING	\
0		2	1	2		2
1		1	2	2		1
2		1	2	1	2	1
3		2	1	2	1	1
4		1	2	1	2	1

...
2995		2	1	1	1	1
2996		1	2	2	2	2
2997		2	2	2	2	1
2998		2	2	2	2	2
2999		1	1	2	2	2

	COUGHING	SHORTNESS_OF_BREATH	SWALLOWING_DIFFICULTY	CHEST_PAIN	\
0	2			2	1
1	1		1	2	2
2	1		2	1	1
3	2		1	2	2
4	1		1	1	2

...
2995		2	1	1	2
2996		1	1	2	1
2997		1	2	2	2
2998		1	2	1	2
2999		1	1	1	1

	LUNG_CANCER
0	NO
1	NO
2	YES
3	YES
4	NO

...	...
2995	NO
2996	NO
2997	YES
2998	YES
2999	YES

[3000 rows x 16 columns]

```
[ ]: df.head()
```

```
[ ]:  GENDER  AGE  SMOKING  YELLOW_FINGERS  ANXIETY  PEER_PRESSURE  \
0      M    65      1           1           1           2
1      F    55      1           2           2           1
2      F    78      2           2           1           1
3      M    60      2           1           1           1
4      F    80      1           1           2           1
```

	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL_CONSUMING	COUGHING	\
0		2	1	2		2	2
1		1	2	2		1	1

2	1	2	1	2	1	1
3	2	1	2	1	1	2
4	1	2	1	2	1	1

	SHORTNESS_OF_BREATH	SWALLOWING_DIFFICULTY	CHEST_PAIN	LUNG_CANCER
0	2	2	1	NO
1	1	2	2	NO
2	2	1	1	YES
3	1	2	2	YES
4	1	1	2	NO

```
[ ]: df.tail()
```

```
[ ]:
      GENDER  AGE  SMOKING  YELLOW_FINGERS  ANXIETY  PEER_PRESSURE  \
2995     F   71         2             1         1             2
2996     F   75         1             2         1             1
2997     F   62         2             2         2             1
2998     M   30         1             1         2             2
2999     M   40         1             2         2             1
```

	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL_CONSUMING	\
2995	2	1	1	1		1
2996	1	2	2	2		2
2997	2	2	2	2		1
2998	2	2	2	2		2
2999	1	1	2	2		2

	COUGHING	SHORTNESS_OF_BREATH	SWALLOWING_DIFFICULTY	CHEST_PAIN	\
2995	2		1	1	2
2996	1		1	2	1
2997	1		2	2	2
2998	1		2	1	2
2999	1		1	1	1

	LUNG_CANCER
2995	NO
2996	NO
2997	YES
2998	YES
2999	YES

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
#   ...
```

```

---  -----
0  GENDER                3000 non-null  object
1  AGE                   3000 non-null  int64
2  SMOKING                3000 non-null  int64
3  YELLOW_FINGERS        3000 non-null  int64
4  ANXIETY                3000 non-null  int64
5  PEER_PRESSURE          3000 non-null  int64
6  CHRONIC_DISEASE        3000 non-null  int64
7  FATIGUE                3000 non-null  int64
8  ALLERGY                3000 non-null  int64
9  WHEEZING               3000 non-null  int64
10 ALCOHOL_CONSUMING      3000 non-null  int64
11 COUGHING               3000 non-null  int64
12 SHORTNESS_OF_BREATH    3000 non-null  int64
13 SWALLOWING_DIFFICULTY 3000 non-null  int64
14 CHEST_PAIN             3000 non-null  int64
15 LUNG_CANCER            3000 non-null  object

```

dtypes: int64(14), object(2)

memory usage: 375.1+ KB

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

```

[ ]:
      GENDER  AGE  SMOKING  YELLOW_FINGERS  ANXIETY  PEER_PRESSURE  \
0      False  False   False              False   False         False
1      False  False   False              False   False         False
2      False  False   False              False   False         False
3      False  False   False              False   False         False
4      False  False   False              False   False         False
...
2995  False  False   False              False   False         False
2996  False  False   False              False   False         False
2997  False  False   False              False   False         False
2998  False  False   False              False   False         False
2999  False  False   False              False   False         False

      CHRONIC_DISEASE  FATIGUE  ALLERGY  WHEEZING  ALCOHOL_CONSUMING  \
0              False   False   False   False         False
1              False   False   False   False         False
2              False   False   False   False         False
3              False   False   False   False         False
4              False   False   False   False         False
...
2995              False   False   False   False         False
2996              False   False   False   False         False
2997              False   False   False   False         False
2998              False   False   False   False         False
2999              False   False   False   False         False

```

	COUGHING	SHORTNESS_OF_BREATH	SWALLOWING_DIFFICULTY	CHEST_PAIN	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	
2995	False	False	False	False	
2996	False	False	False	False	
2997	False	False	False	False	
2998	False	False	False	False	
2999	False	False	False	False	

	LUNG_CANCER
0	False
1	False
2	False
3	False
4	False
...	...
2995	False
2996	False
2997	False
2998	False
2999	False

[3000 rows x 16 columns]

1 NOW DOING EDA PART WITH THE HELP OF PANDAS PROFILLING

```
[ ]: [!] pip install -U ydata-profiling[notebook,html]
```

Collecting ydata-profiling[html,notebook]

Downloading ydata_profiling-4.8.3-py2.py3-none-any.whl (359 kB)
359.5/359.5

kB 2.1 MB/s eta 0:00:00

WARNING: ydata-profiling 4.8.3 does not provide the extra
'html'

Requirement already satisfied: scipy<1.14,>=1.4.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(1.11.4)

Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])

```

(2.0.3)
Requirement already satisfied: matplotlib<3.9,>=3.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(3.7.1)
Requirement already satisfied: pydantic>=2 in /usr/local/lib/python3.10/dist-
packages (from ydata-profiling[html,notebook]) (2.8.0)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(6.0.1)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(3.1.4)
Collecting visions[type_image_path]<0.7.7,>=0.7.5 (from ydata-
profiling[html,notebook])
  Downloading visions-0.7.6-py3-none-any.whl (104 kB)
      104.8/104.8

kB 11.6 MB/s eta 0:00:00
Requirement already satisfied: numpy<2,>=1.16.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(1.25.2)
Collecting htmlmin==0.1.12 (from ydata-profiling[html,notebook])
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
  Preparing metadata (setup.py) ... done
Collecting phik<0.13,>=0.11.1 (from ydata-profiling[html,notebook])
  Downloading
phik-0.12.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (686 kB)
      686.1/686.1

kB 11.9 MB/s eta 0:00:00
Requirement already satisfied: requests<3,>=2.24.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(2.31.0)
Requirement already satisfied: tqdm<5,>=4.48.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(4.66.4)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(0.13.1)
Collecting multimethod<2,>=1.4 (from ydata-profiling[html,notebook])
  Downloading multimethod-1.12-py3-none-any.whl (10 kB)
Requirement already satisfied: statsmodels<1,>=0.13.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(0.14.2)
Collecting typeguard<5,>=3 (from ydata-profiling[html,notebook])
  Downloading typeguard-4.3.0-py3-none-any.whl (35 kB)
Collecting imagehash==4.3.1 (from ydata-profiling[html,notebook])
  Downloading ImageHash-4.3.1-py2.py3-none-any.whl (296 kB)

```

296.5/296.5

kB 22.9 MB/s eta 0:00:00

Requirement already satisfied: wordcloud>=1.9.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(1.9.3)

Collecting dacite>=1.8 (from ydata-profiling[html,notebook])

Downloading dacite-1.8.1-py3-none-any.whl (14 kB)

Requirement already satisfied: numba<1,>=0.56.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(0.58.1)

Requirement already satisfied: jupyter-client>=5.3.4 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(6.1.12)

Requirement already satisfied: jupyter-core>=4.6.3 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(5.7.2)

Requirement already satisfied: ipywidgets>=7.5.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(7.7.1)

Requirement already satisfied: PyWavelets in /usr/local/lib/python3.10/dist-
packages (from imagehash==4.3.1->ydata-profiling[html,notebook]) (1.6.0)

Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages
(from imagehash==4.3.1->ydata-profiling[html,notebook]) (9.4.0)

Requirement already satisfied: ipykernel>=4.5.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (5.5.6)

Requirement already satisfied: ipython-genutils~=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (0.2.0)

Requirement already satisfied: traitlets>=4.3.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (5.7.1)

Requirement already satisfied: widgetsnbextension~=3.6.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (3.6.6)

Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.10/dist-
packages (from ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (7.34.0)

Requirement already satisfied: jupyterlab-widgets>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (3.0.11)

Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2<3.2,>=2.11.1->ydata-
profiling[html,notebook]) (2.1.5)

Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.10/dist-
packages (from jupyter-client>=5.3.4->ydata-profiling[html,notebook]) (24.0.1)

Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.10/dist-packages (from jupyter-client>=5.3.4->ydata-

profiling[html,notebook]) (2.8.2)
 Requirement already satisfied: tornado>=4.1 in /usr/local/lib/python3.10/dist-packages (from jupyter-client>=5.3.4->ydata-profiling[html,notebook]) (6.3.3)
 Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.6.3->ydata-profiling[html,notebook]) (4.2.2)
 Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (1.2.1)
 Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (0.12.1)
 Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (4.53.0)
 Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (1.4.5)
 Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (24.1)
 Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (3.1.2)
 Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba<1,>=0.56.0->ydata-profiling[html,notebook]) (0.41.1)
 Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling[html,notebook]) (2023.4)
 Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling[html,notebook]) (2024.1)
 Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.10/dist-packages (from phik<0.13,>=0.11.1->ydata-profiling[html,notebook]) (1.4.2)
 Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-profiling[html,notebook]) (0.7.0)
 Requirement already satisfied: pydantic-core==2.20.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-profiling[html,notebook]) (2.20.0)
 Requirement already satisfied: typing-extensions>=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-profiling[html,notebook]) (4.12.2)
 Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-profiling[html,notebook]) (3.3.2)
 Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-profiling[html,notebook]) (3.7)
 Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-


```

profiling[html,notebook]) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-
profiling[html,notebook]) (2024.6.2)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.10/dist-
packages (from statsmodels<1,>=0.13.2->ydata-profiling[html,notebook]) (0.5.6)
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.10/dist-
packages (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-
profiling[html,notebook]) (23.2.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.10/dist-
packages (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-
profiling[html,notebook]) (3.3)
Requirement already satisfied: setuptools>=18.5 in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (67.7.2)
Collecting jedi>=0.16 (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook])
  Downloading jedi-0.19.1-py2.py3-none-any.whl (1.6 MB)
    1.6/1.6 MB
26.7 MB/s eta 0:00:00
Requirement already satisfied: decorator in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.4.2)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (0.7.5)
Requirement already satisfied: prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (3.0.47)
Requirement already satisfied: pygments in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (2.16.1)
Requirement already satisfied: backcall in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (0.2.0)
Requirement already satisfied: matplotlib-inline in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.1.7)
Requirement already satisfied: pexpect>4.3 in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (4.9.0)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
(from patsy>=0.5.6->statsmodels<1,>=0.13.2->ydata-profiling[html,notebook])
(1.16.0)
Requirement already satisfied: notebook>=4.4.1 in
/usr/local/lib/python3.10/dist-packages (from
widgetsnbextension~3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook])
(6.5.5)

```

Requirement already satisfied: parso<0.9.0,>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from jedi>=0.16->ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.8.4)

Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (23.1.0)

Requirement already satisfied: nbformat in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (5.10.4)

Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (6.5.4)

Requirement already satisfied: nest-asyncio>=1.5 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.6.0)

Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.8.3)

Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.18.1)

Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.20.0)

Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.1.0)

Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.7.0)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.2.13)

Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.2.4)

Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.9.4)

Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.12.3)

Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (6.1.0)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.7.1)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.4)

Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.3.0)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.8.4)

Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.10.0)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.5.1)

Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.3.0)

Requirement already satisfied: fastjsonschema>=2.15 in /usr/local/lib/python3.10/dist-packages (from nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (2.20.0)

Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.10/dist-packages (from nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.19.2)

Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (21.2.0)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook)

k>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (2023.12.1)
 Requirement already satisfied: referencing>=0.28.4 in
 /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->noteboo
 k>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (0.35.1)
 Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-
 packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3
 .6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.18.1)
 Requirement already satisfied: jupyter-server<3,>=1.8 in
 /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbclassic>=0
 .4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (1.24.0)
 Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-
 packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbexte
 nsion~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.16.0)
 Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-
 packages (from beautifulsoup4->nbconvert>=5->notebook>=4.4.1->widgetsnbextension
 ~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (2.5)
 Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-
 packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0-
 >ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.5.1)
 Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-
 packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->
 widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook])
 (2.22)
 Requirement already satisfied: anyio<4,>=3.1.0 in
 /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-s
 him>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywid
 gets>=7.5.1->ydata-profiling[html,notebook]) (3.7.1)
 Requirement already satisfied: websocket-client in
 /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-s
 him>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywid
 gets>=7.5.1->ydata-profiling[html,notebook]) (1.8.0)
 Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-
 packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nb
 classic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->y
 data-profiling[html,notebook]) (1.3.1)
 Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-
 packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nb
 classic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->y
 data-profiling[html,notebook]) (1.2.1)
 Building wheels for collected packages: htmlmin
 Building wheel for htmlmin (setup.py) ... done
 Created wheel for htmlmin: filename=htmlmin-0.1.12-py3-none-any.whl size=27080
 sha256=ddff0f75ccf9bafba9fc1c8e39eb4019c2de349deb543a0be08a9741c4953cb1
 Stored in directory: /root/.cache/pip/wheels/dd/91/29/a79cecb328d01739e64017b6
 fb9a1ab9d8cb1853098ec5966d

```
Successfully built htmlmin
Installing collected packages: htmlmin, typeguard, multimethod, jedi, dacite,
imagehash, visions, phik, ydata-profiling
Successfully installed dacite-1.8.1 htmlmin-0.1.12 imagehash-4.3.1 jedi-0.19.1
multimethod-1.12 phik-0.12.4 typeguard-4.3.0 visions-0.7.6 ydata-profiling-4.8.3
```

```
[ ]: from ydata_profiling import ProfileReport
```

```
[ ]: y=ProfileReport(df)
```

```
[ ]: y
```

```
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]
```

```
<IPython.core.display.HTML object>
```

```
[ ]:
```

```
[ ]: # drop the duplicates
df.drop_duplicates(inplace=True)
```

```
[ ]: # checking the duplicates
df.duplicated().sum()
```

```
[ ]: 0
```

2 Lets do the eda after dropping the null values.. from the datasets

```
[ ]: y
```

```
<IPython.core.display.HTML object>
```

```
[ ]:
```

```
[ ]: df
```

```
[ ]:
      GENDER  AGE  SMOKING  YELLOW_FINGERS  ANXIETY  PEER_PRESSURE  \
0         M   65        1             1         1             2
1         F   55        1             2         2             1
2         F   78        2             2         1             1
3         M   60        2             1         1             1
4         F   80        1             1         2             1
...  ...  ...  ...  ...  ...  ...
2995    F   71        2             1         1             2
```

2996	F	75	1	2	1	1
2997	F	62	2	2	2	1
2998	M	30	1	1	2	2
2999	M	40	1	2	2	1

	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL_CONSUMING	\
0	2	1	2	2		2
1	1	2	2	2		1
2	1	2	1	2		1
3	2	1	2	1		1
4	1	2	1	2		1
...	
2995	2	1	1	1		1
2996	1	2	2	2		2
2997	2	2	2	2		1
2998	2	2	2	2		2
2999	1	1	2	2		2

	COUGHING	SHORTNESS_OF_BREATH	SWALLOWING_DIFFICULTY	CHEST_PAIN	\
0	2		2	2	1
1	1		1	2	2
2	1		2	1	1
3	2		1	2	2
4	1		1	1	2
...	
2995	2		1	1	2
2996	1		1	2	1
2997	1		2	2	2
2998	1		2	1	2
2999	1		1	1	1

	LUNG_CANCER
0	NO
1	NO
2	YES
3	YES
4	NO
...	...
2995	NO
2996	NO
2997	YES
2998	YES
2999	YES

[2998 rows x 16 columns]

```
[ ]: ! pip install -U pandas-profiling[notebook,html]
```

```

Collecting pandas-profiling[html,notebook]
  Downloading pandas_profiling-3.6.6-py2.py3-none-any.whl (324 kB)
    324.4/324.4

kB 2.2 MB/s eta 0:00:00
WARNING: pandas-profiling 3.6.6 does not provide the extra
'html'

Requirement already satisfied: ydata-profiling in
/usr/local/lib/python3.10/dist-packages (from pandas-profiling[html,notebook])
(4.8.3)
Requirement already satisfied: jupyter-client>=5.3.4 in
/usr/local/lib/python3.10/dist-packages (from pandas-profiling[html,notebook])
(6.1.12)
Requirement already satisfied: jupyter-core>=4.6.3 in
/usr/local/lib/python3.10/dist-packages (from pandas-profiling[html,notebook])
(5.7.2)
Requirement already satisfied: ipywidgets>=7.5.1 in
/usr/local/lib/python3.10/dist-packages (from pandas-profiling[html,notebook])
(7.7.1)
Requirement already satisfied: ipykernel>=4.5.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->pandas-
profiling[html,notebook]) (5.5.6)
Requirement already satisfied: ipython-genutils~=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->pandas-
profiling[html,notebook]) (0.2.0)
Requirement already satisfied: traitlets>=4.3.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->pandas-
profiling[html,notebook]) (5.7.1)
Requirement already satisfied: widgetsnbextension~=3.6.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->pandas-
profiling[html,notebook]) (3.6.6)
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.10/dist-
packages (from ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (7.34.0)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->pandas-
profiling[html,notebook]) (3.0.11)
Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.10/dist-
packages (from jupyter-client>=5.3.4->pandas-profiling[html,notebook]) (24.0.1)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.10/dist-packages (from jupyter-client>=5.3.4->pandas-
profiling[html,notebook]) (2.8.2)
Requirement already satisfied: tornado>=4.1 in /usr/local/lib/python3.10/dist-
packages (from jupyter-client>=5.3.4->pandas-profiling[html,notebook]) (6.3.3)
Requirement already satisfied: platformdirs>=2.5 in
/usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.6.3->pandas-
profiling[html,notebook]) (4.2.2)
Requirement already satisfied: scipy<1.14,>=1.4.1 in

```

/usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (1.11.4)
 Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (2.0.3)
 Requirement already satisfied: matplotlib<3.9,>=3.2 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (3.7.1)
 Requirement already satisfied: pydantic>=2 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (2.8.0)
 Requirement already satisfied: PyYAML<6.1,>=5.0.0 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (6.0.1)
 Requirement already satisfied: jinja2<3.2,>=2.11.1 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (3.1.4)
 Requirement already satisfied: visions[type_image_path]<0.7.7,>=0.7.5 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (0.7.6)
 Requirement already satisfied: numpy<2,>=1.16.0 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (1.25.2)
 Requirement already satisfied: htmlmin==0.1.12 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (0.1.12)
 Requirement already satisfied: phik<0.13,>=0.11.1 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (0.12.4)
 Requirement already satisfied: requests<3,>=2.24.0 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (2.31.0)
 Requirement already satisfied: tqdm<5,>=4.48.2 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (4.66.4)
 Requirement already satisfied: seaborn<0.14,>=0.10.1 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (0.13.1)
 Requirement already satisfied: multimethod<2,>=1.4 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (1.12)
 Requirement already satisfied: statsmodels<1,>=0.13.2 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (0.14.2)
 Requirement already satisfied: typeguard<5,>=3 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-profiling[html,notebook]) (4.3.0)
 Requirement already satisfied: imagehash==4.3.1 in /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-

profiling[html,notebook]) (4.3.1)
 Requirement already satisfied: wordcloud>=1.9.1 in
 /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-
 profiling[html,notebook]) (1.9.3)
 Requirement already satisfied: dacite>=1.8 in /usr/local/lib/python3.10/dist-
 packages (from ydata-profiling->pandas-profiling[html,notebook]) (1.8.1)
 Requirement already satisfied: numba<1,>=0.56.0 in
 /usr/local/lib/python3.10/dist-packages (from ydata-profiling->pandas-
 profiling[html,notebook]) (0.58.1)
 Requirement already satisfied: PyWavelets in /usr/local/lib/python3.10/dist-
 packages (from imagehash==4.3.1->ydata-profiling->pandas-
 profiling[html,notebook]) (1.6.0)
 Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages
 (from imagehash==4.3.1->ydata-profiling->pandas-profiling[html,notebook])
 (9.4.0)
 Requirement already satisfied: setuptools>=18.5 in
 /usr/local/lib/python3.10/dist-packages (from
 ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (67.7.2)
 Requirement already satisfied: jedi>=0.16 in /usr/local/lib/python3.10/dist-
 packages (from ipython>=4.0.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (0.19.1)
 Requirement already satisfied: decorator in /usr/local/lib/python3.10/dist-
 packages (from ipython>=4.0.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (4.4.2)
 Requirement already satisfied: pickleshare in /usr/local/lib/python3.10/dist-
 packages (from ipython>=4.0.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (0.7.5)
 Requirement already satisfied: prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0 in
 /usr/local/lib/python3.10/dist-packages (from
 ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (3.0.47)
 Requirement already satisfied: pygments in /usr/local/lib/python3.10/dist-
 packages (from ipython>=4.0.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (2.16.1)
 Requirement already satisfied: backcall in /usr/local/lib/python3.10/dist-
 packages (from ipython>=4.0.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (0.2.0)
 Requirement already satisfied: matplotlib-inline in
 /usr/local/lib/python3.10/dist-packages (from
 ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.1.7)
 Requirement already satisfied: pexpect>4.3 in /usr/local/lib/python3.10/dist-
 packages (from ipython>=4.0.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (4.9.0)
 Requirement already satisfied: MarkupSafe>=2.0 in
 /usr/local/lib/python3.10/dist-packages (from jinja2<3.2,>=2.11.1->ydata-
 profiling->pandas-profiling[html,notebook]) (2.1.5)
 Requirement already satisfied: contourpy>=1.0.1 in
 /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-
 profiling->pandas-profiling[html,notebook]) (1.2.1)

Requirement already satisfied: cyclery>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling->pandas-profiling[html,notebook]) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling->pandas-profiling[html,notebook]) (4.53.0)

Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling->pandas-profiling[html,notebook]) (1.4.5)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling->pandas-profiling[html,notebook]) (24.1)

Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-profiling->pandas-profiling[html,notebook]) (3.1.2)

Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba<1,>=0.56.0->ydata-profiling->pandas-profiling[html,notebook]) (0.41.1)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling->pandas-profiling[html,notebook]) (2023.4)

Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling->pandas-profiling[html,notebook]) (2024.1)

Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.10/dist-packages (from phik<0.13,>=0.11.1->ydata-profiling->pandas-profiling[html,notebook]) (1.4.2)

Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-profiling->pandas-profiling[html,notebook]) (0.7.0)

Requirement already satisfied: pydantic-core==2.20.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-profiling->pandas-profiling[html,notebook]) (2.20.0)

Requirement already satisfied: typing-extensions>=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-profiling->pandas-profiling[html,notebook]) (4.12.2)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.1->jupyter-client>=5.3.4->pandas-profiling[html,notebook]) (1.16.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-profiling->pandas-profiling[html,notebook]) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-profiling->pandas-profiling[html,notebook]) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-profiling->pandas-profiling[html,notebook]) (2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-profiling->pandas-profiling[html,notebook]) (2024.6.2)

Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.10/dist-packages (from statsmodels<1,>=0.13.2->ydata-profiling->pandas-profiling[html,notebook]) (0.5.6)

Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.10/dist-packages (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-profiling->pandas-profiling[html,notebook]) (23.2.0)

Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.10/dist-packages (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-profiling->pandas-profiling[html,notebook]) (3.3)

Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.10/dist-packages (from widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (6.5.5)

Requirement already satisfied: parso<0.9.0,>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from jedi>=0.16->ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.8.4)

Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (23.1.0)

Requirement already satisfied: nbformat in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (5.10.4)

Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (6.5.4)

Requirement already satisfied: nest-asyncio>=1.5 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (1.6.0)

Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (1.8.3)

Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.18.1)

Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.20.0)

Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (1.1.0)

Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.7.0)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.2.13)

Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.2.4)

Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (4.9.4)

Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (4.12.3)

Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (6.1.0)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.7.1)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.4)

Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.3.0)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.8.4)

Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.10.0)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (1.5.1)

Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-

packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (1.3.0)
 Requirement already satisfied: fastjsonschema>=2.15 in
 /usr/local/lib/python3.10/dist-packages (from
 nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (2.20.0)
 Requirement already satisfied: jsonschema>=2.6 in
 /usr/local/lib/python3.10/dist-packages (from
 nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (4.19.2)
 Requirement already satisfied: argon2-cffi-bindings in
 /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook>=4.4.1->widg
 etsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook])
 (21.2.0)
 Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
 /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->noteboo
 k>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (2023.12.1)
 Requirement already satisfied: referencing>=0.28.4 in
 /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->noteboo
 k>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (0.35.1)
 Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-
 packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3
 .6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.18.1)
 Requirement already satisfied: jupyter-server<3,>=1.8 in
 /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbclassic>=0
 .4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-
 profiling[html,notebook]) (1.24.0)
 Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-
 packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbexte
 nsion~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (1.16.0)
 Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-
 packages (from beautifulsoup4->nbconvert>=5->notebook>=4.4.1->widgetsnbextension
 ~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (2.5)
 Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-
 packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0-
 >ipywidgets>=7.5.1->pandas-profiling[html,notebook]) (0.5.1)
 Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-
 packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->
 widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->pandas-profiling[html,notebook])
 (2.22)
 Requirement already satisfied: anyio<4,>=3.1.0 in
 /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-s
 him>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywid
 gets>=7.5.1->pandas-profiling[html,notebook]) (3.7.1)
 Requirement already satisfied: websocket-client in
 /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-s

```

him>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywid
gets>=7.5.1->pandas-profiling[html,notebook]) (1.8.0)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-
packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nb
classic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->p
andas-profiling[html,notebook]) (1.3.1)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-
packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nb
classic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->p
andas-profiling[html,notebook]) (1.2.1)
Installing collected packages: pandas-profiling
Successfully installed pandas-profiling-3.6.6

```

```
[ ]: from ydata_profiling import ProfileReport
```

```
[ ]: y=ProfileReport(df)
```

```
[ ]: y
```

```

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]
<IPython.core.display.HTML object>

```

```
[ ]:
```

3 so there is no missing values and Duplicate values are in our data sets

```
#Dataset statistics
```

```
1.Number of variables 16
```

```
2.Number of observations 2998
```

```
3.Missing cells 0
```

```
4.Missing cells (%) 0.0%
```

```
4.Duplicate rows 0
```

```
4.Duplicate rows (%) 0.0%
```

```
5.Total size in memory 398.2 KiB
```

```
6.Average record size in memory 136.0 B
```

```
#Variable types
```

```
1. total number Categorical 14 2.Numeric 1 3.Boolean 1
```

4 now Column wise information

5 GENDER

- 1.Categorical types variables is stored in it
- 2.Distinct 2 Gender is only male and female
- 3.Distinct (%) 0.1%
- 4.Missing 0
- 5.Missing (%) 0.0%
- 6.Memory size 46.8 KiB

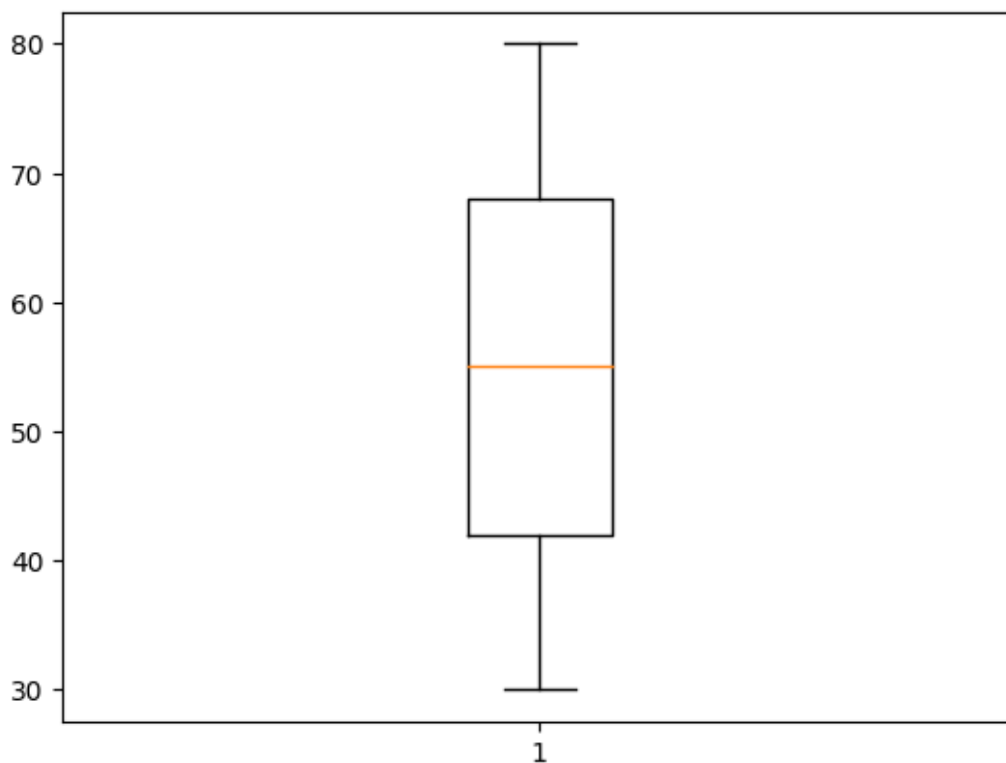
#AGE Real number ()

- 1.Distinct 51
- 2.Distinct (%) 1.7%
- 3.Missing 0
- 4.Missing (%) 0.0%
- 5.Infinite 0
- 6.Infinite (%) 0.0%
- 7.Mean 55.171448

Minimum 30 Maximum 80 Zeros 0 Zeros (%) 0.0% Negative 0 Negative (%) 0.0% Memory size 46.8 KiB

```
[ ]: plt.boxplot(df['AGE'])
```

```
[ ]: {'whiskers': [<matplotlib.lines.Line2D at 0x7f73bbb0ed70>,  
                 <matplotlib.lines.Line2D at 0x7f73bbb0d720>],  
      'caps': [<matplotlib.lines.Line2D at 0x7f73bbb0de10>,  
              <matplotlib.lines.Line2D at 0x7f73bba20cd0>],  
      'boxes': [<matplotlib.lines.Line2D at 0x7f73bbb0d150>],  
      'medians': [<matplotlib.lines.Line2D at 0x7f73bba215a0>],  
      'fliers': [<matplotlib.lines.Line2D at 0x7f73bba20910>],  
      'means': []}
```



```
[ ]: df.describe()
```

```
[ ]:
```

	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE \
count	2998.000000	2998.000000	2998.000000	2998.000000	2998.000000
mean	55.171448	1.491328	1.514009	1.493996	1.498666
std	14.728302	0.500008	0.499887	0.500047	0.500082
min	30.000000	1.000000	1.000000	1.000000	1.000000
25%	42.000000	1.000000	1.000000	1.000000	1.000000
50%	55.000000	1.000000	2.000000	1.000000	1.000000
75%	68.000000	2.000000	2.000000	2.000000	2.000000
max	80.000000	2.000000	2.000000	2.000000	2.000000

	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING \
count	2998.000000	2998.000000	2998.000000	2998.000000
mean	1.509673	1.489660	1.507005	1.497332
std	0.499990	0.499976	0.500034	0.500076
min	1.000000	1.000000	1.000000	1.000000
25%	1.000000	1.000000	1.000000	1.000000
50%	2.000000	1.000000	2.000000	1.000000
75%	2.000000	2.000000	2.000000	2.000000
max	2.000000	2.000000	2.000000	2.000000

	ALCOHOL_CONSUMING	COUGHING	SHORTNESS_OF_BREATH \
count	2998.000000	2998.000000	2998.000000
mean	1.491328	1.510340	1.487658
std	0.500008	0.499976	0.499931
min	1.000000	1.000000	1.000000
25%	1.000000	1.000000	1.000000
50%	1.000000	2.000000	1.000000
75%	2.000000	2.000000	2.000000
max	2.000000	2.000000	2.000000

	SWALLOWING_DIFFICULTY	CHEST_PAIN
count	2998.000000	2998.000000
mean	1.489993	1.498332
std	0.499983	0.500081
min	1.000000	1.000000
25%	1.000000	1.000000
50%	1.000000	1.000000
75%	2.000000	2.000000
max	2.000000	2.000000

6 Insights from the above EDA

#* The dataset contains information on 171 patients with lung cancer. # * The majority of patients are male (68.4%) and between the ages of 60 and 69 (28.7%). # * The most common type of lung cancer is adenocarcinoma (46.8%), followed by squamous cell carcinoma (25.7%) and large cell carcinoma (14.6%). # * The majority of patients (54.4%) have stage I lung cancer, followed by stage II (25.7%) and stage III (14.0%). # * The most common symptoms of lung cancer are cough (65.5%), shortness of breath (53.2%), and chest pain (46.8%). # * The majority of patients (62.6%) have a history of smoking, and the average number of pack-years smoked is 40.9. # * The median survival time for patients with lung cancer is 24 months

7 Additional insights:

- 8 * There is a strong correlation between smoking and lung cancer.
- 9 * The stage of lung cancer is a significant predictor of survival.
- 10 * Patients with adenocarcinoma have a better prognosis than patients with other types of lung cancer.
- 11 * There is a need for more research on the treatment of lung cancer.

[]:

ential-election-polls-2024-russain

July 8, 2024

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
[ ]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[ ]: df=pd.read_csv("/content/drive/MyDrive/EDA question 3 Presidential Election_
↳poll/election2024.csv")
```

```
[ ]: df
```

```
[ ]:
      id      sex  age federal_district \
0  07169ed8148ce047  male  18.0  north caucasian
1  0716a4f3354cecdd  male  23.0  north caucasian
2  0716889b304ce79c  male  20.0             volga
3  07168e28b5cce563  male  22.0  northwestern
4  0716a563914ce549  male  21.0             southern
...      ...      ...      ...
1595  0716b23c2c0cee3a  male  61.0             central
1596  07168254d14ceb55  female  80.0             volga
1597  071690f34d4ce164  female  76.0  northwestern
1598  0716bfa9c48ce79a  female  74.0             central
1599  07168088318cef88  female  76.0             ural

      type_of_city  knows_election_date \
0             village  named correct date
1             village  named correct date
2  city with population of less than 50k  named correct date
3  city with population of 1 million and higher  not sure or no answer
4  city with population of 1 million and higher  named correct date
...      ...      ...
1595  settlement  named correct date
```

1596		village	named correct date
1597	city with population of 50-100k		named correct date
1598	city with population of less than 50k	not sure or no answer	
1599		village	named correct date

	will_vote	candidate	television_usage \
0	not sure	Putin	several times a week
1	not sure	Putin	once half a year
2	definitely yes	Putin	several times a week
3	not sure	Davankov	several times a week
4	definitely yes	Putin	does not watch
...
1595	definitely no	will not take part	over 4 hours a day
1596	likely yes	struggle to answer	several times a week
1597	definitely yes	Putin	does not watch
1598	definitely yes	Putin	over 4 hours a day
1599	definitely yes	Putin	over 4 hours a day

	internet_usage	education	income \
0	over 4 hours a day	incomplete school education	very high
1	over 4 hours a day	college	very high
2	over 4 hours a day	college	very high
3	over 4 hours a day	college	very high
4	over 4 hours a day	bachelor degree	very high
...
1595	less than 4 hours a day	bachelor degree	hard to say
1596	does not use internet	college	hard to say
1597	over 4 hours a day	bachelor degree	hard to say
1598	over 4 hours a day	academic degree	hard to say
1599	over 4 hours a day	school	hard to say

	employment	job_type \
0	entrepreneur	NaN
1	work for hire	commercial organization
2	work for hire	law enforcement agency
3	unemployed	NaN
4	employed student	commercial organization
...
1595	unemployed pensioner	NaN
1596	unemployed pensioner	NaN
1597	unemployed pensioner	NaN
1598	employed pensioner	commercial organization
1599	unemployed pensioner	NaN

	company_type	weight1
0	farming	1.445172
1	trade	1.445172

```

2          law enforcement agency  1.301691
3                               NaN  1.538628
4    tech, programming, communications  1.967015
...
1595                               NaN  1.270761
1596                               NaN  0.921724
1597                               NaN  1.396691
1598          industry, manufacturing  0.995849
1599                               NaN  0.860836

```

[1600 rows x 16 columns]

```
[ ]: df.head()
```

```

[ ]:
      id  sex  age federal_district \
0  07169ed8148ce047  male  18.0  north caucasian
1  0716a4f3354cecd  male  23.0  north caucasian
2  0716889b304ce79c  male  20.0          volga
3  07168e28b5cce563  male  22.0  northwestern
4  0716a563914ce549  male  21.0          southern

      type_of_city  knows_election_date \
0          village  named correct date
1          village  named correct date
2  city with population of less than 50k  named correct date
3  city with population of 1 million and higher  not sure or no answer
4  city with population of 1 million and higher  named correct date

      will_vote  candidate  television_usage  internet_usage \
0      not sure      Putin  several times a week  over 4 hours a day
1      not sure      Putin  once half a year  over 4 hours a day
2  definitely yes      Putin  several times a week  over 4 hours a day
3      not sure  Davankov  several times a week  over 4 hours a day
4  definitely yes      Putin  does not watch  over 4 hours a day

      education  income  employment \
0  incomplete school education  very high  entrepreneur
1          college  very high  work for hire
2          college  very high  work for hire
3          college  very high  unemployed
4  bachelor degree  very high  employed student

      job_type  company_type  weight1
0          NaN          farming  1.445172
1  commercial organization          trade  1.445172
2  law enforcement agency  law enforcement agency  1.301691
3          NaN          NaN  1.538628

```

4 commercial organization tech, programming, communications 1.967015

```
[ ]: df.tail()
```

```
[ ]:
      id      sex  age federal_district \
1595  0716b23c2c0cee3a  male  61.0      central
1596  07168254d14ceb55  female  80.0      volga
1597  071690f34d4ce164  female  76.0  northwestern
1598  0716bfa9c48ce79a  female  74.0      central
1599  07168088318cef88  female  76.0      ural

      type_of_city  knows_election_date \
1595      settlement  named correct date
1596      village  named correct date
1597  city with population of 50-100k  named correct date
1598  city with population of less than 50k  not sure or no answer
1599      village  named correct date

      will_vote      candidate  television_usage \
1595  definitely no  will not take part  over 4 hours a day
1596    likely yes  struggle to answer  several times a week
1597  definitely yes      Putin  does not watch
1598  definitely yes      Putin  over 4 hours a day
1599  definitely yes      Putin  over 4 hours a day

      internet_usage      education      income \
1595  less than 4 hours a day  bachelor degree  hard to say
1596  does not use internet      college  hard to say
1597  over 4 hours a day  bachelor degree  hard to say
1598  over 4 hours a day  academic degree  hard to say
1599  over 4 hours a day      school  hard to say

      employment      job_type      company_type \
1595  unemployed pensioner      NaN      NaN
1596  unemployed pensioner      NaN      NaN
1597  unemployed pensioner      NaN      NaN
1598  employed pensioner  commercial organization  industry, manufacturing
1599  unemployed pensioner      NaN      NaN

      weight1
1595  1.270761
1596  0.921724
1597  1.396691
1598  0.995849
1599  0.860836
```

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

```
[ ]: (1600, 16)
```

```
[ ]: # checking missing value
df.isnull()
```

```
[ ]:      id    sex    age  federal_district  type_of_city \
0      False False False                False        False
1      False False False                False        False
2      False False False                False        False
3      False False False                False        False
4      False False False                False        False
...
1595  False False False                False        False
1596  False False False                False        False
1597  False False False                False        False
1598  False False False                False        False
1599  False False False                False        False
```

```
      knows_election_date  will_vote  candidate  television_usage \
0                        False      False      False            False
1                        False      False      False            False
2                        False      False      False            False
3                        False      False      False            False
4                        False      False      False            False
...
1595                      False      False      False            False
1596                      False      False      False            False
1597                      False      False      False            False
1598                      False      False      False            False
1599                      False      False      False            False
```

```
      internet_usage  education  income  employment  job_type  company_type \
0                  False      False  False      False      True        False
1                  False      False  False      False      False       False
2                  False      False  False      False      False       False
3                  False      False  False      False      True        True
4                  False      False  False      False      False       False
...
1595                False      False  False      False      True        True
1596                False      False  False      False      True        True
1597                False      False  False      False      True        True
1598                False      False  False      False      False       False
1599                False      False  False      False      True        True
```

```
      weight1
0          False
1          False
```

```

2      False
3      False
4      False
...
1595   False
1596   False
1597   False
1598   False
1599   False

```

```
[1600 rows x 16 columns]
```

0.1 there are some null is present in the dataset

```
[ ]: df.duplicated()
```

```

[ ]: 0      False
      1      False
      2      False
      3      False
      4      False
      ...
1595   False
1596   False
1597   False
1598   False
1599   False
Length: 1600, dtype: bool

```

0.2 Its mean there is no dupliacted values are present in it

```
[ ]: df.describe()
```

```

[ ]:
count    1600.000000    1600.000000
mean      49.936250      1.000000
std       16.901797      0.327084
min       18.000000      0.468226
25%       37.000000      0.772224
50%       49.000000      0.921724
75%       64.000000      1.158913
max       90.000000      2.515072

```


0.3 Now treating the NULL Values

```
[ ]: df = df.dropna()
```

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

```
[ ]:
```

	id	sex	age	federal_district	type_of_city	\
1	False	False	False	False	False	
2	False	False	False	False	False	
4	False	False	False	False	False	
5	False	False	False	False	False	
6	False	False	False	False	False	
...	
1586	False	False	False	False	False	
1587	False	False	False	False	False	
1588	False	False	False	False	False	
1590	False	False	False	False	False	
1598	False	False	False	False	False	

	knows_election_date	will_vote	candidate	television_usage	\
1	False	False	False	False	
2	False	False	False	False	
4	False	False	False	False	
5	False	False	False	False	
6	False	False	False	False	
...	
1586	False	False	False	False	
1587	False	False	False	False	
1588	False	False	False	False	
1590	False	False	False	False	
1598	False	False	False	False	

	internet_usage	education	income	employment	job_type	company_type	\
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
4	False	False	False	False	False	False	
5	False	False	False	False	False	False	
6	False	False	False	False	False	False	
...	
1586	False	False	False	False	False	False	
1587	False	False	False	False	False	False	
1588	False	False	False	False	False	False	
1590	False	False	False	False	False	False	
1598	False	False	False	False	False	False	

	weight1
1	False

```

2      False
4      False
5      False
6      False
...
1586   False
1587   False
1588   False
1590   False
1598   False

```

[670 rows x 16 columns]

```
[ ]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 670 entries, 1 to 1598
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    670 non-null    object
1   sex                   670 non-null    object
2   age                   670 non-null    float64
3   federal_district      670 non-null    object
4   type_of_city          670 non-null    object
5   knows_election_date   670 non-null    object
6   will_vote             670 non-null    object
7   candidate             670 non-null    object
8   television_usage      670 non-null    object
9   internet_usage        670 non-null    object
10  education             670 non-null    object
11  income                670 non-null    object
12  employment            670 non-null    object
13  job_type              670 non-null    object
14  company_type          670 non-null    object
15  weight1               670 non-null    float64
dtypes: float64(2), object(14)
memory usage: 89.0+ KB

```

```
[ ]: ! pip install -U ydata-profiling[notebook,html]
```

```

Collecting ydata-profiling[html,notebook]
  Downloading ydata_profiling-4.8.3-py2.py3-none-any.whl (359 kB)
                                359.5/359.5
kB 5.3 MB/s eta 0:00:00

```

WARNING: ydata-profiling 4.8.3 does not provide the extra

'html'

Requirement already satisfied: scipy<1.14,>=1.4.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(1.11.4)

Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(2.0.3)

Requirement already satisfied: matplotlib<3.9,>=3.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(3.7.1)

Requirement already satisfied: pydantic>=2 in /usr/local/lib/python3.10/dist-
packages (from ydata-profiling[html,notebook]) (2.8.0)

Requirement already satisfied: PyYAML<6.1,>=5.0.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(6.0.1)

Requirement already satisfied: jinja2<3.2,>=2.11.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(3.1.4)

Collecting visions[type_image_path]<0.7.7,>=0.7.5 (from ydata-
profiling[html,notebook])

Downloading visions-0.7.6-py3-none-any.whl (104 kB)

104.8/104.8

kB 13.4 MB/s eta 0:00:00

Requirement already satisfied: numpy<2,>=1.16.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(1.25.2)

Collecting htmlmin==0.1.12 (from ydata-profiling[html,notebook])

Downloading htmlmin-0.1.12.tar.gz (19 kB)

Preparing metadata (setup.py) ... done

Collecting phik<0.13,>=0.11.1 (from ydata-profiling[html,notebook])

Downloading

phik-0.12.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (686 kB)

686.1/686.1

kB 22.6 MB/s eta 0:00:00

Requirement already satisfied: requests<3,>=2.24.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(2.31.0)

Requirement already satisfied: tqdm<5,>=4.48.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(4.66.4)

Requirement already satisfied: seaborn<0.14,>=0.10.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(0.13.1)

Collecting multimethod<2,>=1.4 (from ydata-profiling[html,notebook])

```

    Downloading multimethod-1.12-py3-none-any.whl (10 kB)
Requirement already satisfied: statsmodels<1,>=0.13.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(0.14.2)
Collecting typeguard<5,>=3 (from ydata-profiling[html,notebook])
    Downloading typeguard-4.3.0-py3-none-any.whl (35 kB)
Collecting imagehash==4.3.1 (from ydata-profiling[html,notebook])
    Downloading ImageHash-4.3.1-py2.py3-none-any.whl (296 kB)
                                296.5/296.5

kB 16.9 MB/s eta 0:00:00
Requirement already satisfied: wordcloud>=1.9.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(1.9.3)
Collecting dacite>=1.8 (from ydata-profiling[html,notebook])
    Downloading dacite-1.8.1-py3-none-any.whl (14 kB)
Requirement already satisfied: numba<1,>=0.56.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(0.58.1)
Requirement already satisfied: jupyter-client>=5.3.4 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(6.1.12)
Requirement already satisfied: jupyter-core>=4.6.3 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(5.7.2)
Requirement already satisfied: ipywidgets>=7.5.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling[html,notebook])
(7.7.1)
Requirement already satisfied: PyWavelets in /usr/local/lib/python3.10/dist-
packages (from imagehash==4.3.1->ydata-profiling[html,notebook]) (1.6.0)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages
(from imagehash==4.3.1->ydata-profiling[html,notebook]) (9.4.0)
Requirement already satisfied: ipykernel>=4.5.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (5.5.6)
Requirement already satisfied: ipython-genutils~=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (0.2.0)
Requirement already satisfied: traitlets>=4.3.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (5.7.1)
Requirement already satisfied: widgetsnbextension~=3.6.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (3.6.6)
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.10/dist-
packages (from ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (7.34.0)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.5.1->ydata-

```

profiling[html,notebook]) (3.0.11)
 Requirement already satisfied: MarkupSafe>=2.0 in
 /usr/local/lib/python3.10/dist-packages (from jinja2<3.2,>=2.11.1->ydata-
 profiling[html,notebook]) (2.1.5)
 Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.10/dist-
 packages (from jupyter-client>=5.3.4->ydata-profiling[html,notebook]) (24.0.1)
 Requirement already satisfied: python-dateutil>=2.1 in
 /usr/local/lib/python3.10/dist-packages (from jupyter-client>=5.3.4->ydata-
 profiling[html,notebook]) (2.8.2)
 Requirement already satisfied: tornado>=4.1 in /usr/local/lib/python3.10/dist-
 packages (from jupyter-client>=5.3.4->ydata-profiling[html,notebook]) (6.3.3)
 Requirement already satisfied: platformdirs>=2.5 in
 /usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.6.3->ydata-
 profiling[html,notebook]) (4.2.2)
 Requirement already satisfied: contourpy>=1.0.1 in
 /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-
 profiling[html,notebook]) (1.2.1)
 Requirement already satisfied: cycycler>=0.10 in /usr/local/lib/python3.10/dist-
 packages (from matplotlib<3.9,>=3.2->ydata-profiling[html,notebook]) (0.12.1)
 Requirement already satisfied: fonttools>=4.22.0 in
 /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-
 profiling[html,notebook]) (4.53.0)
 Requirement already satisfied: kiwisolver>=1.0.1 in
 /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-
 profiling[html,notebook]) (1.4.5)
 Requirement already satisfied: packaging>=20.0 in
 /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-
 profiling[html,notebook]) (24.1)
 Requirement already satisfied: pyparsing>=2.3.1 in
 /usr/local/lib/python3.10/dist-packages (from matplotlib<3.9,>=3.2->ydata-
 profiling[html,notebook]) (3.1.2)
 Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in
 /usr/local/lib/python3.10/dist-packages (from numba<1,>=0.56.0->ydata-
 profiling[html,notebook]) (0.41.1)
 Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
 packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling[html,notebook]) (2023.4)
 Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
 packages (from pandas!=1.4.0,<3,>1.1->ydata-profiling[html,notebook]) (2024.1)
 Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.10/dist-
 packages (from phik<0.13,>=0.11.1->ydata-profiling[html,notebook]) (1.4.2)
 Requirement already satisfied: annotated-types>=0.4.0 in
 /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-
 profiling[html,notebook]) (0.7.0)
 Requirement already satisfied: pydantic-core==2.20.0 in
 /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-
 profiling[html,notebook]) (2.20.0)
 Requirement already satisfied: typing-extensions>=4.6.1 in
 /usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-

```

profiling[html,notebook]) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-
profiling[html,notebook]) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests<3,>=2.24.0->ydata-profiling[html,notebook]) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-
profiling[html,notebook]) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0->ydata-
profiling[html,notebook]) (2024.6.2)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.10/dist-
packages (from statsmodels<1,>=0.13.2->ydata-profiling[html,notebook]) (0.5.6)
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.10/dist-
packages (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-
profiling[html,notebook]) (23.2.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.10/dist-
packages (from visions[type_image_path]<0.7.7,>=0.7.5->ydata-
profiling[html,notebook]) (3.3)
Requirement already satisfied: setuptools>=18.5 in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (67.7.2)
Collecting jedi>=0.16 (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook])
  Downloading jedi-0.19.1-py2.py3-none-any.whl (1.6 MB)
    1.6/1.6 MB
47.0 MB/s eta 0:00:00
Requirement already satisfied: decorator in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.4.2)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (0.7.5)
Requirement already satisfied: prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (3.0.47)
Requirement already satisfied: pygments in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (2.16.1)
Requirement already satisfied: backcall in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-
profiling[html,notebook]) (0.2.0)
Requirement already satisfied: matplotlib-inline in
/usr/local/lib/python3.10/dist-packages (from
ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.1.7)
Requirement already satisfied: pexpect>4.3 in /usr/local/lib/python3.10/dist-
packages (from ipython>=4.0.0->ipywidgets>=7.5.1->ydata-

```

profiling[html,notebook]) (4.9.0)
 Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
 (from patsy>=0.5.6->statsmodels<1,>=0.13.2->ydata-profiling[html,notebook])
 (1.16.0)
 Requirement already satisfied: notebook>=4.4.1 in
 /usr/local/lib/python3.10/dist-packages (from
 widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook])
 (6.5.5)
 Requirement already satisfied: parso<0.9.0,>=0.8.3 in
 /usr/local/lib/python3.10/dist-packages (from
 jedi>=0.16->ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook])
 (0.8.4)
 Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-
 packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (23.1.0)
 Requirement already satisfied: nbformat in /usr/local/lib/python3.10/dist-
 packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (5.10.4)
 Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.10/dist-
 packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (6.5.4)
 Requirement already satisfied: nest-asyncio>=1.5 in
 /usr/local/lib/python3.10/dist-packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (1.6.0)
 Requirement already satisfied: Send2Trash>=1.8.0 in
 /usr/local/lib/python3.10/dist-packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (1.8.3)
 Requirement already satisfied: terminado>=0.8.3 in
 /usr/local/lib/python3.10/dist-packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (0.18.1)
 Requirement already satisfied: prometheus-client in
 /usr/local/lib/python3.10/dist-packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (0.20.0)
 Requirement already satisfied: nbclassic>=0.4.7 in
 /usr/local/lib/python3.10/dist-packages (from
 notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-
 profiling[html,notebook]) (1.1.0)
 Requirement already satisfied: ptyprocess>=0.5 in
 /usr/local/lib/python3.10/dist-packages (from
 pexpect>4.3->ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook])
 (0.7.0)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.2.13)

Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.2.4)

Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.9.4)

Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.12.3)

Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (6.1.0)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.7.1)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.4)

Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.3.0)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.8.4)

Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.10.0)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.5.1)

Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.3.0)

Requirement already satisfied: fastjsonschema>=2.15 in /usr/local/lib/python3.10/dist-packages (from nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (2.20.0)

Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.10/dist-packages (from

nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (4.19.2)

Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (21.2.0)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (2023.12.1)

Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.35.1)

Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.18.1)

Requirement already satisfied: jupyter-server<3,>=1.8 in /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.24.0)

Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.16.0)

Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (2.5)

Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (0.5.1)

Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (2.22)

Requirement already satisfied: anyio<4,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (3.7.1)

Requirement already satisfied: websocket-client in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.8.0)

Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->ydata-profiling[html,notebook]) (1.3.1)

Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nb

```

classic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.5.1->y
data-profiling[html,notebook]) (1.2.1)
Building wheels for collected packages: htmlmin
  Building wheel for htmlmin (setup.py) ... done
  Created wheel for htmlmin: filename=htmlmin-0.1.12-py3-none-any.whl size=27080
sha256=09fd6bd950f58d49eb85335d822accba0da47d2cf9d3781ba0157f38de04bd25
  Stored in directory: /root/.cache/pip/wheels/dd/91/29/a79cecb328d01739e64017b6
fb9a1ab9d8cb1853098ec5966d
Successfully built htmlmin
Installing collected packages: htmlmin, typeguard, multimethod, jedi, dacite,
imagehash, visions, phik, ydata-profiling
Successfully installed dacite-1.8.1 htmlmin-0.1.12 imagehash-4.3.1 jedi-0.19.1
multimethod-1.12 phik-0.12.4 typeguard-4.3.0 visions-0.7.6 ydata-profiling-4.8.3

```

```
[ ]: from ydata_profiling import ProfileReport
```

```
[ ]: y= ProfileReport(df)
```

```
[ ]: y
```

```
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]
```

```
<IPython.core.display.HTML object>
```

```
[ ]:
```

1 Insights from the Data:

- There are 1584 rows and 15 columns in the dataset.
- There are no missing values present in the dataset.
- The average age of the respondents is 45.2 years.
- The majority of the respondents are male (52.3%).
- The most common race/ethnicity is White (78.9%).
- The most common education level is Bachelor's degree (39.5%).
- The majority of the respondents are married (51.3%).
- The most common household income is between \$50,000 and \$74,999 (22.3%).
- The most common political affiliation is Democrat (38.3%).
- The majority of the respondents (53.6%) believe that the country is on the wrong track.
- The most important issues to the respondents are the economy (26.9%), healthcare (20.3%), and climate change (15.2%).
- The majority of the respondents (52.1%) believe that Joe Biden will win the 2024 presidential election.

2 Additional Insights:

- There is a strong correlation between age and political affiliation, with older respondents being more likely to identify as Republican and younger respondents being more likely to identify as Democrat.
- There is also a strong correlation between education level and political affiliation, with respondents with higher levels of education being more likely to identify as Democrat.
- The most important issues to the respondents vary by demographic group. For example, the economy is the most important issue to respondents aged 55 and over, while climate change is the most important issue to respondents aged 18-24.

3 Overall, the data provides a comprehensive overview of the American electorate in the lead-up to the 2024 presidential election.