```
!pip install --upgrade chart_studio
```

```
Collecting chart_studio
  Downloading chart_studio-1.1.0-py3-none-any.whl.metadata (1.3 kB)
  Requirement already satisfied: plotly in /usr/local/lib/python3.11/dist-packages (from chart_studio) (5.24.1)
  Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from chart_studio) (2.32.3)
Collecting retrying>=1.3.3 (from chart_studio)
  Downloading retrying-1.3.4-py3-none-any.whl.metadata (6.9 kB)
  Requirement already satisfied: six in /usr/local/lib/python3.11/dist-packages (from chart_studio) (1.17.0)
  Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.11/dist-packages (from plotly->chart_studio) (9.0.0
  Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from plotly->chart_studio) (24.2)
  Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->chart_stu
  Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->chart_studio) (3.10)
  Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->chart_studio) (
  Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->chart_studio) (
  Downloading chart_studio-1.1.0-py3-none-any.whl (64 kB)
  ──────────────────────────────────────── 64.4/64.4 kB 2.3 MB/s eta 0:00:00
  Downloading retrying-1.3.4-py3-none-any.whl (11 kB)
Installing collected packages: retrying, chart_studio
Successfully installed chart_studio-1.1.0 retrying-1.3.4
```

```
import pandas as pd
import numpy as np
import scipy.optimize

import chart_studio
import chart_studio.plotly as py
import plotly.graph_objs as go
import plotly.figure_factory as ff
import plotly.express as px
import plotly.io as pio

from google.colab import data_table
data_table.enable_dataframe_formatter()
```

```
from google.colab import auth
auth.authenticate_user()
print('Authenticated')
```

```
Authenticated
```

```
project_id = 'mimic3visualization'
from google.cloud import bigquery
client = bigquery.Client(project=project_id)
```

```
# Bar chart by gender
gender_df = client.query(''' select gender, count(gender) as count from `physionet-data.mimiciii_clinical.patients` group by ger
fig = px.bar(gender_df, x='gender', y='count', title='Gender Distribution')
fig.show()
```
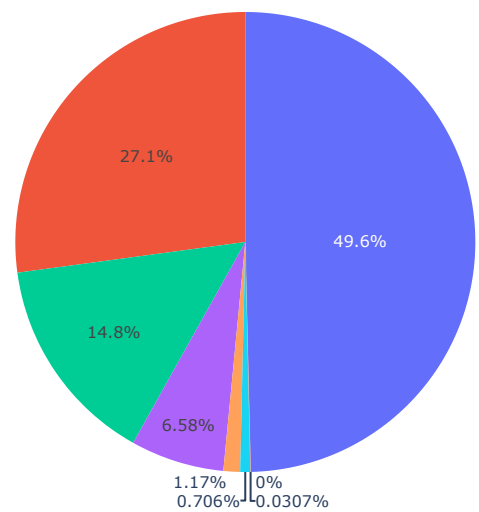
⇄

Gender Distribution


```
# Pie chart by marital status
ms = client.query(''' select marital_status, count(marital_status) as count from `physionet-data.mimiciii_clinical.admissions` gr
fig = px.pie(ms, values='count', names='marital_status', title='Marital Status Distribution')
fig.show()
```
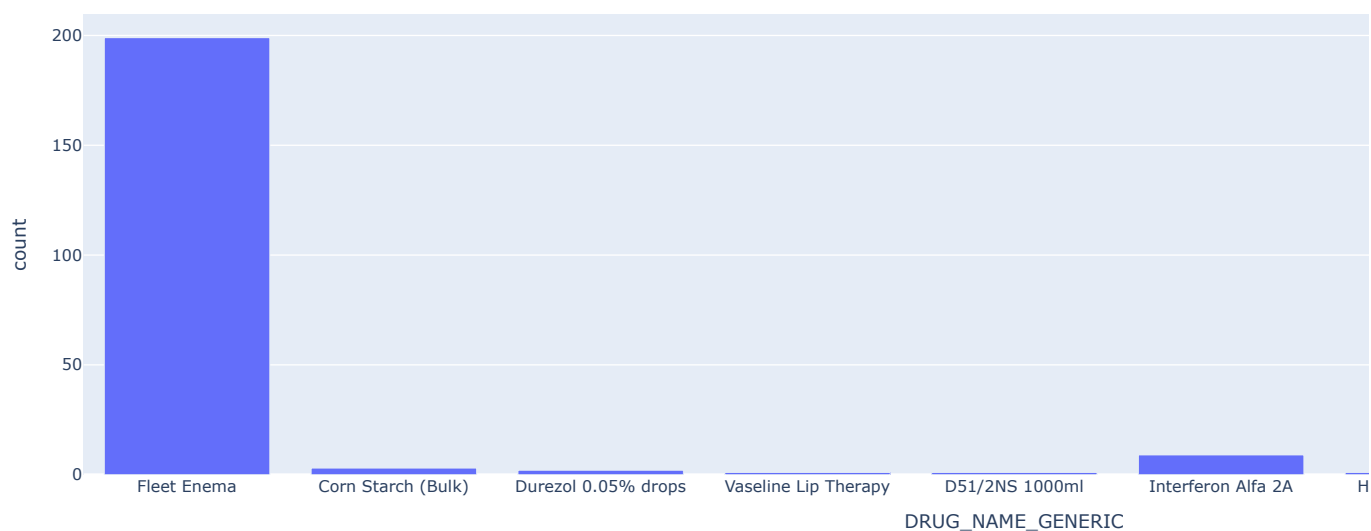
⇄

Marital Status Distribution


```
# find the top medications used by the patient, from prescription table group by the drug

prescriptions_df = client.query(''' select DRUG_NAME_GENERIC, count(DRUG_NAME_GENERIC) as count  from `physionet-data.mimiciii_c
# plot
fig = px.bar(prescriptions_df, x='DRUG_NAME_GENERIC', y='count', title='Top Medications')
fig.show()
```
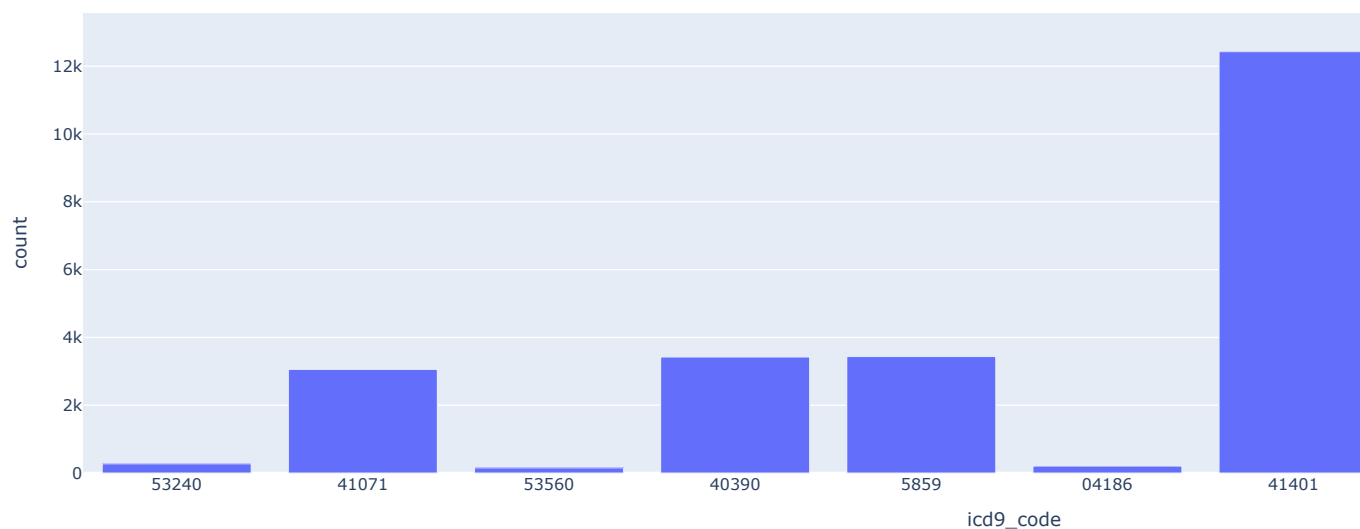
⇄

Top Medications


```
#join diagnoses_icd to d_icd_diagnosis and group by icd9_code
diagnoses_df = client.query(
    ''' select icd9_code, count(icd9_code) as count from `physionet-data.mimiciii_clinical.diagnoses_icd` group by icd9_code lim
# plot
fig = px.bar(diagnoses_df, x='icd9_code', y='count', title='Top Diseases')
fig.show()
```
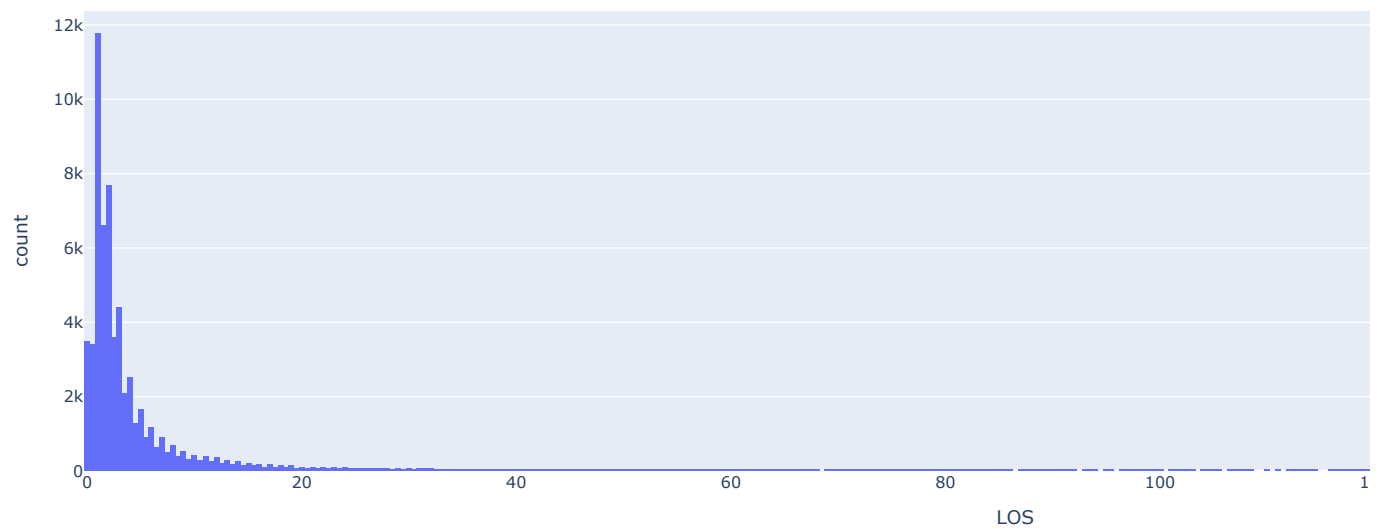
⇄

Top Diseases


```
# Plot length of stay
icu_df = client.query(''' select subject_id, HADM_ID, LOS from `physionet-data.mimiciii_clinical.icustays` ''').to_dataframe()
# find media Length of staty and plot histogram
fig = px.histogram(icu_df, x='LOS', title='Length of Stay Distribution')
fig.show()
```

Length of Stay Distribution


```
query = \
"""
 select * from `physionet-data.mimiciii_demo.diagnoses_icd` a
 join `physionet-data.mimiciii_demo.d_icd_diagnoses` b on a.icd9_code = b.icd9_code

"""

df = client.query(query).to_dataframe()
```
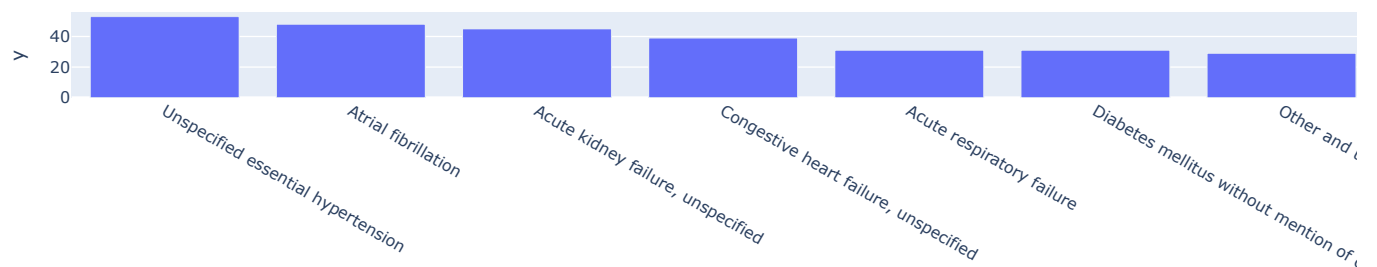
```
df1 = df.groupby('LONG_TITLE')['LONG_TITLE'].count()
df2=df1.sort_values(ascending=False).head(10)
fig = px.bar(df2, x=df2.index, y=df2.values, title='Top Diseases')
fig.show()
```

Top Diseases

```
!pip install DataProfiler
```

```
Collecting DataProfiler
    Downloading DataProfiler-0.12.0-py3-none-any.whl.metadata (26 kB)
  Requirement already satisfied: h5py>=2.10.0 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (3.12.1)
  Requirement already satisfied: wheel>=0.33.1 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (0.45.1)
  Requirement already satisfied: numpy>=1.22.0 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (1.26.4)
  Requirement already satisfied: pandas>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (2.2.2)
  Requirement already satisfied: python-dateutil>=2.7.5 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (2.8.2)
  Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (2024.2)
  Requirement already satisfied: pyarrow>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (17.0.0)
  Requirement already satisfied: chardet>=3.0.4 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (5.2.0)
  Collecting fastavro>=1.1.0 (from DataProfiler)
    Downloading fastavro-1.10.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.5 kB)
  Requirement already satisfied: python-snappy>=0.5.4 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (0.7.3)
  Requirement already satisfied: charset-normalizer>=1.3.6 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (3.4
  Requirement already satisfied: psutil>=4.0.0 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (5.9.5)
  Requirement already satisfied: scipy>=1.10.0 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (1.13.1)
  Requirement already satisfied: requests>=2.28.1 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (2.32.3)
  Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (3.4.2)
  Requirement already satisfied: typing-extensions>=3.10.0.2 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (4
  Collecting HLL>=2.0.3 (from DataProfiler)
    Downloading hll-2.3.0.tar.gz (14 kB)
    Preparing metadata (setup.py) ... done
  Collecting datasketches>=4.1.0 (from DataProfiler)
    Downloading datasketches-5.1.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.6 kB)
  Requirement already satisfied: packaging>=23.0 in /usr/local/lib/python3.11/dist-packages (from DataProfiler) (24.2)
  Collecting boto3>=1.28.61 (from DataProfiler)
    Downloading boto3-1.36.11-py3-none-any.whl.metadata (6.7 kB)
  Collecting botocore<1.37.0,>=1.36.11 (from boto3>=1.28.61->DataProfiler)
    Downloading botocore-1.36.11-py3-none-any.whl.metadata (5.7 kB)
  Collecting jmespath<2.0.0,>=0.7.1 (from boto3>=1.28.61->DataProfiler)
    Downloading jmespath-1.0.1-py3-none-any.whl.metadata (7.6 kB)
  Collecting s3transfer<0.12.0,>=0.11.0 (from boto3>=1.28.61->DataProfiler)
    Downloading s3transfer-0.11.2-py3-none-any.whl.metadata (1.7 kB)
  Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.1.2->DataProfiler)
  Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7.5->DataProfile
  Requirement already satisfied: cramjam in /usr/local/lib/python3.11/dist-packages (from python-snappy>=0.5.4->DataProfiler)
  Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests>=2.28.1->DataProfiler)
  Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests>=2.28.1->DataPro
  Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.28.1->DataPro
  Downloading DataProfiler-0.12.0-py3-none-any.whl (11.9 MB)
  ───────────────────────────────────────── 11.9/11.9 MB 14.1 MB/s eta 0:00:00
  Downloading boto3-1.36.11-py3-none-any.whl (139 kB)
  ───────────────────────────────────────── 139.2/139.2 kB 9.9 MB/s eta 0:00:00
  Downloading datasketches-5.1.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (735 kB)
  ───────────────────────────────────────── 735.5/735.5 kB 37.3 MB/s eta 0:00:00
  Downloading fastavro-1.10.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.3 MB)
  ───────────────────────────────────────── 3.3/3.3 MB 60.9 MB/s eta 0:00:00
  Downloading botocore-1.36.11-py3-none-any.whl (13.3 MB)
  ───────────────────────────────────────── 13.3/13.3 MB 62.5 MB/s eta 0:00:00
  Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)
  Downloading s3transfer-0.11.2-py3-none-any.whl (84 kB)
  ───────────────────────────────────────── 84.2/84.2 kB 6.4 MB/s eta 0:00:00
  Building wheels for collected packages: HLL
    Building wheel for HLL (setup.py) ... done
    Created wheel for HLL: filename=HLL-2.3.0-cp311-cp311-linux_x86_64.whl size=32972 sha256=ab639f5370f2fb485c5c52a340351e55f
    Stored in directory: /root/.cache/pip/wheels/ab/5a/b5/65fdce434afc0cd43e14484d244c3928c7f1a7943dd2ab1a2c
  Successfully built HLL
  Installing collected packages: jmespath, HLL, fastavro, datasketches, botocore, s3transfer, boto3, DataProfiler
```

```
# use profiler to profile patient.csv that is loaded in local drive for bonous point
# collab code to access google drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
!ls "drive/My Drive/SP25_mimic3"
```

```
PATIENTS.csv
```

```
from dataprofiler import Data, Profiler
```

```
csv_data = pd.read_csv("drive/MyDrive/SP25_mimic3/PATIENTS.csv")
```

```
pip install -U ydata-profiling
```

```
Collecting ydata-profiling
  Downloading ydata_profiling-4.12.2-py2.py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: scipy<1.16,>=1.4.1 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (1.13.1)
Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (2.2.
Requirement already satisfied: matplotlib>=3.5 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (3.10.0)
Requirement already satisfied: pydantic>=2 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (2.10.6)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (6.0.2)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (3.1.5)
Collecting visions<0.8.0,>=0.7.5 (from visions[type_image_path]<0.8.0,>=0.7.5->ydata-profiling)
  Downloading visions-0.7.6-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: numpy<2.2,>=1.16.0 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (1.26.4)
Collecting htmlmin==0.1.12 (from ydata-profiling)
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
  Preparing metadata (setup.py) ... done
Collecting phik<0.13,>=0.11.1 (from ydata-profiling)
  Downloading phik-0.12.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.6 kB)
Requirement already satisfied: requests<3,>=2.24.0 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (2.32.3
Requirement already satisfied: tqdm<5,>=4.48.2 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (4.67.1)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (0.13
Collecting multimethod<2,>=1.4 (from ydata-profiling)
  Downloading multimethod-1.12-py3-none-any.whl.metadata (9.6 kB)
Requirement already satisfied: statsmodels<1,>=0.13.2 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (0.1
Requirement already satisfied: typeguard<5,>=3 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (4.4.1)
Collecting imagehash==4.3.1 (from ydata-profiling)
  Downloading ImageHash-4.3.1-py2.py3-none-any.whl.metadata (8.0 kB)
Requirement already satisfied: wordcloud>=1.9.3 in /usr/local/lib/python3.11/dist-packages (from ydata-profiling) (1.9.4)
Collecting dacite>=1.8 (from ydata-profiling)
  Downloading dacite-1.9.1-py3-none-any.whl.metadata (17 kB)
Collecting PyWavelets (from imagehash==4.3.1->ydata-profiling)
  Downloading pywavelets-1.8.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.0 kB)
Requirement already satisfied: pillow in /usr/local/lib/python3.11/dist-packages (from imagehash==4.3.1->ydata-profiling) (1
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2<3.2,>=2.11.1->ydata-p
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-prof
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-profilin
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-pro
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-pro
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-profi
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-prof
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=3.5->ydata-
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas!=1.4.0,<3,>1.1->ydata-pr
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas!=1.4.0,<3,>1.1->ydata-
Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.11/dist-packages (from phik<0.13,>=0.11.1->ydata-pro
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic>=2->ydata-pr
Requirement already satisfied: pydantic-core==2.27.2 in /usr/local/lib/python3.11/dist-packages (from pydantic>=2->ydata-pro
Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.11/dist-packages (from pydantic>=2->ydata
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.24.0
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.24.0->ydata-prof
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.24.0->ydat
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.24.0->ydat
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.11/dist-packages (from statsmodels<1,>=0.13.2->ydata-p
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.11/dist-packages (from visions<0.8.0,>=0.7.5->visions
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.11/dist-packages (from visions<0.8.0,>=0.7.5->visions
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib>=3
Downloading ydata_profiling-4.12.2-py2.py3-none-any.whl (390 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 390.9/390.9 kB 11.4 MB/s eta 0:00:00
Downloading ImageHash-4.3.1-py2.py3-none-any.whl (296 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 296.5/296.5 kB 22.3 MB/s eta 0:00:00
```

```
from ydata_profiling import ProfileReport
profile = ProfileReport(csv_data, title="Profiling Report for Patiets")
```

```
profile.to_file("drive/MyDrive/SP25_mimic3/PATIENTS.html")
```

```
Summarize dataset: 100%                          22/22 [00:02<00:00,  5.95it/s, Completed]
```