Public health surveillance

Ahmed Abdelaziz Hussein

”STOP DREAMING AND START DOING”

content

1. - Introduction

2. - Python

3. - SQL

4. - Tableau

5. - Summary

PROJECT INTRODUCTION

Public health surveillance is the systematic collection, analysis, and interpretation of health-related data to monitor and prevent disease spread. It helps identify public health threats, track disease patterns, and guide response efforts. By enabling timely interventions, it plays a crucial role in protecting community health.

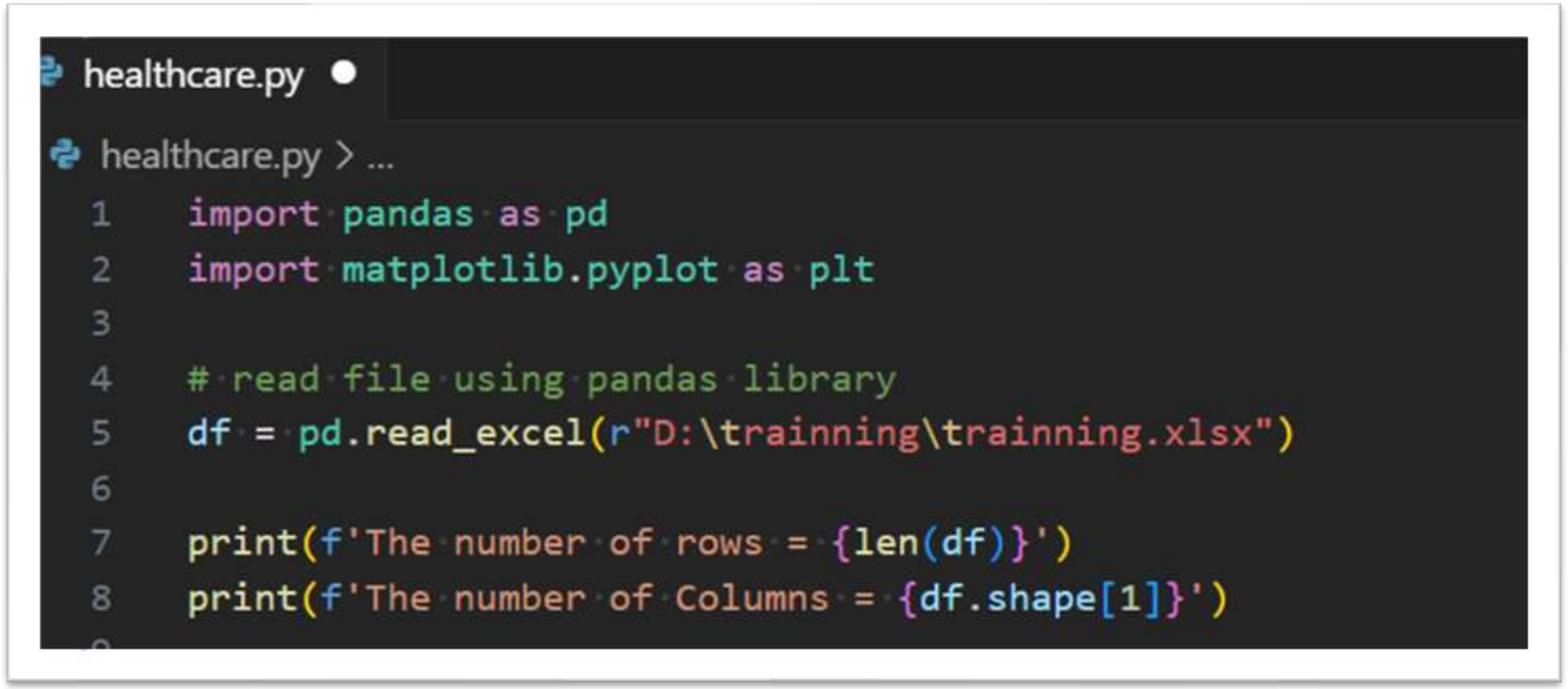
PYTHON ANALYSIS

I used Python to clean the data, take a quick look at it, and to create simple dashboard.
Libraries: Pandas and Matplotlip

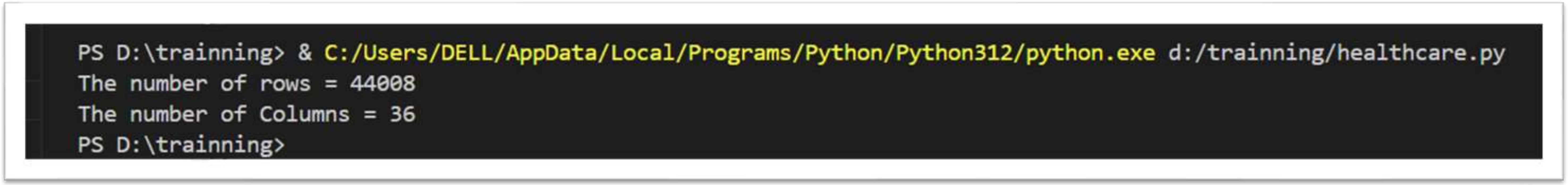
Cleaning data

Count the number of rows and columns in data

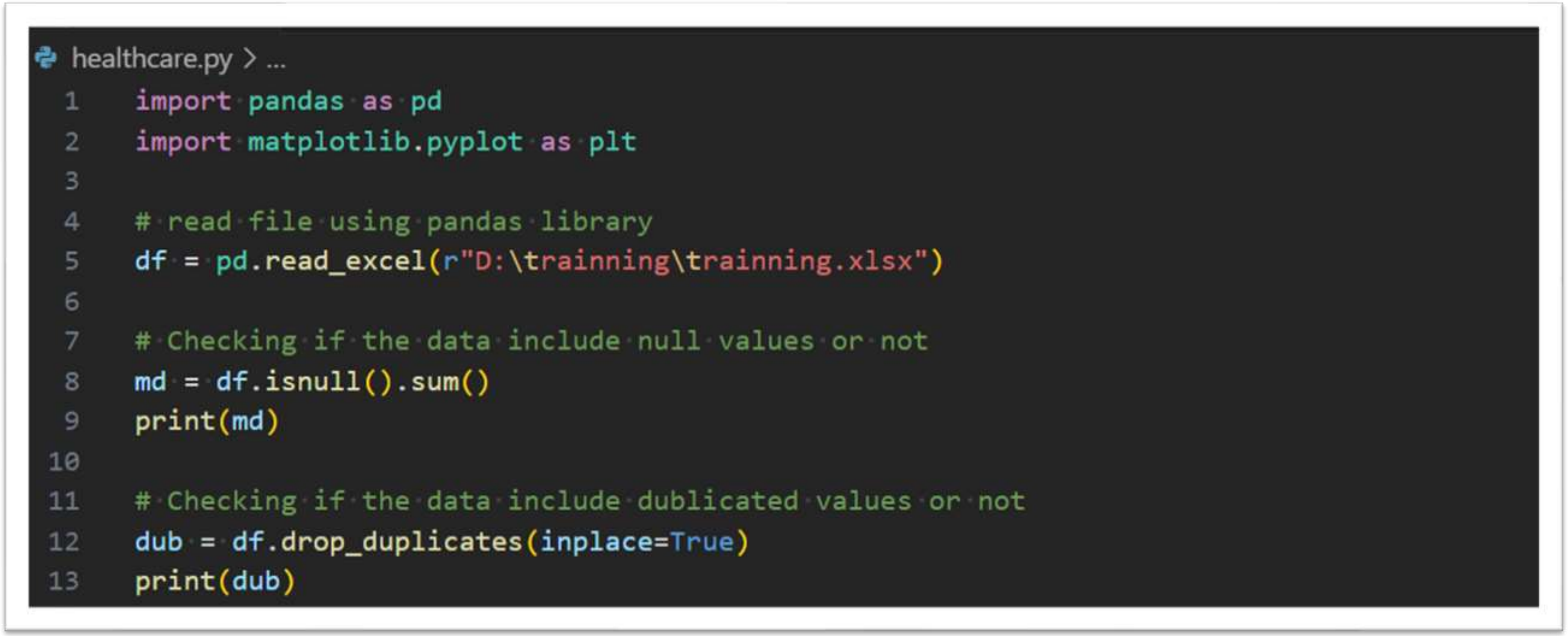
Result: 36 Columns, and 44008 Rows



```python
import pandas as pd
import matplotlib.pyplot as plt

# read file using pandas library
df = pd.read_excel(r"D:\trainning\trainning.xlsx")

print(f'The number of rows = {len(df)}')
print(f'The number of Columns = {df.shape[1]}')
```



```
PS D:\trainning> & C:/Users/DELL/AppData/Local/Programs/Python/Python312/python.exe d:/trainning/healthcare.py
The number of rows = 44008
The number of Columns = 36
PS D:\trainning>
```



```python
import pandas as pd
import matplotlib.pyplot as plt

# read file using pandas library
df = pd.read_excel(r"D:\trainning\trainning.xlsx")

# Checking if the data include null values or not
md = df.isnull().sum()
print(md)

# Checking if the data include dublicated values or not
dub = df.drop_duplicates(inplace=True)
print(dub)
```

I used Pandas library to check if the data contains null values or not.

Result: No Null Values.

And to check if the data contains duplicated values.

Result: No Duplicated Values.

ANALYSIS QUESTIONS

- **No Travel**: 34,800
- **Domestic**: 6,667
- **International**: 2,222

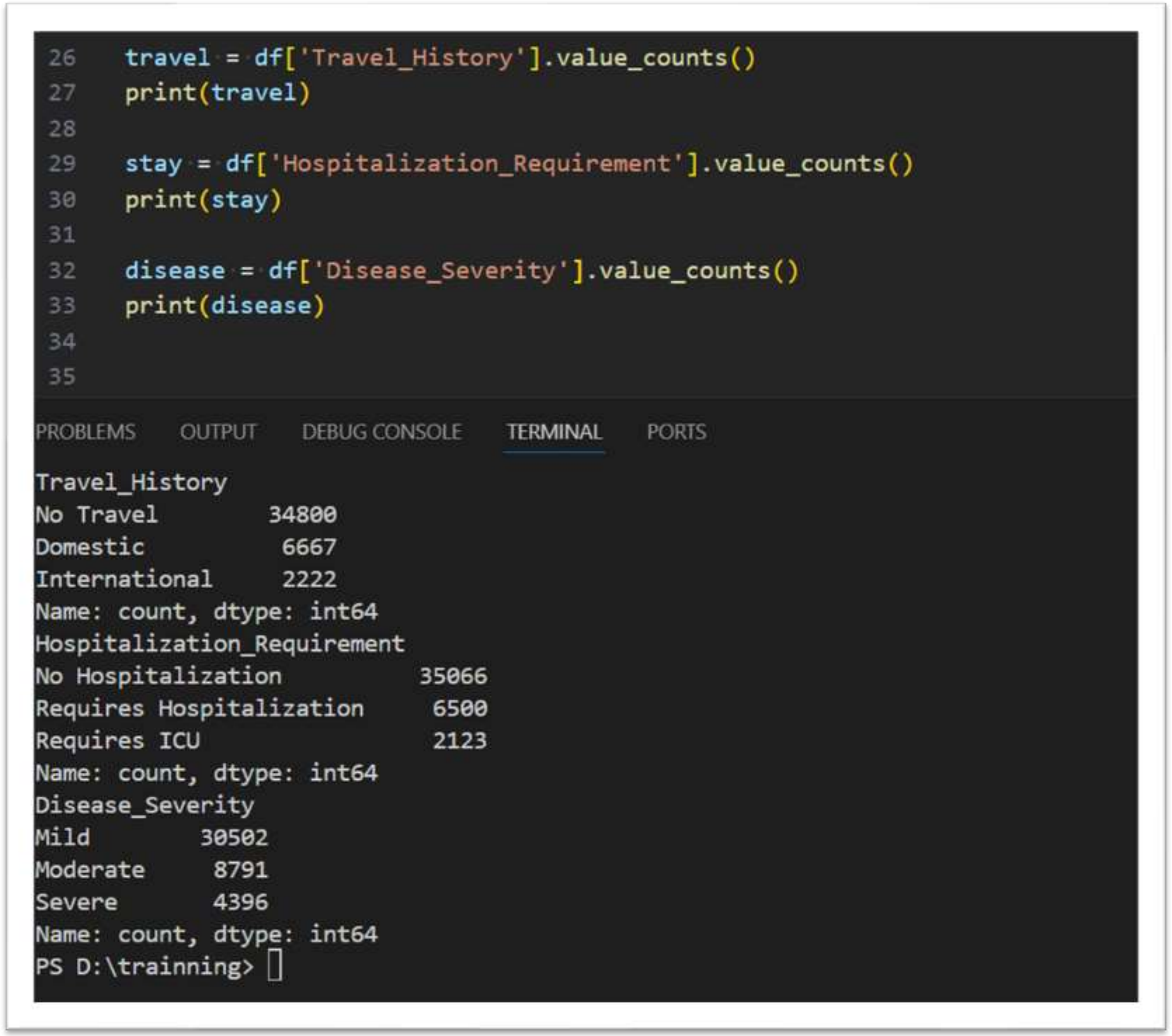
- This indicates that most patients have no travel history, while a smaller number traveled domestically or internationally.

- **No Hospitalization**: 35,066
- **Requires Hospitalization**: 6,500
- **Requires ICU**: 2,123

- The majority of cases don’t require hospitalization, with a smaller number needing regular hospitalization or ICU care.

- **Mild**: 30,502
- **Moderate**: 8,791
- **Severe**: 4,396

- This shows that most cases in the dataset are mild, while a few are severe.



```
travel = df['Travel_History'].value_counts()
print(travel)

stay = df['Hospitalization_Requirement'].value_counts()
print(stay)

disease = df['Disease_Severity'].value_counts()
print(disease)
```

```
Travel_History
No Travel        34800
Domestic          6667
International     2222
Name: count, dtype: int64
Hospitalization_Requirement
No Hospitalization           35066
Requires Hospitalization      6500
Requires ICU                  2123
Name: count, dtype: int64
Disease_Severity
Mild         30502
Moderate      8791
Severe        4396
Name: count, dtype: int64
PS D:\trainning> 
```

- **Urban**: 30,557
- **Rural**: 8,846
- **Suburban**: 4,286

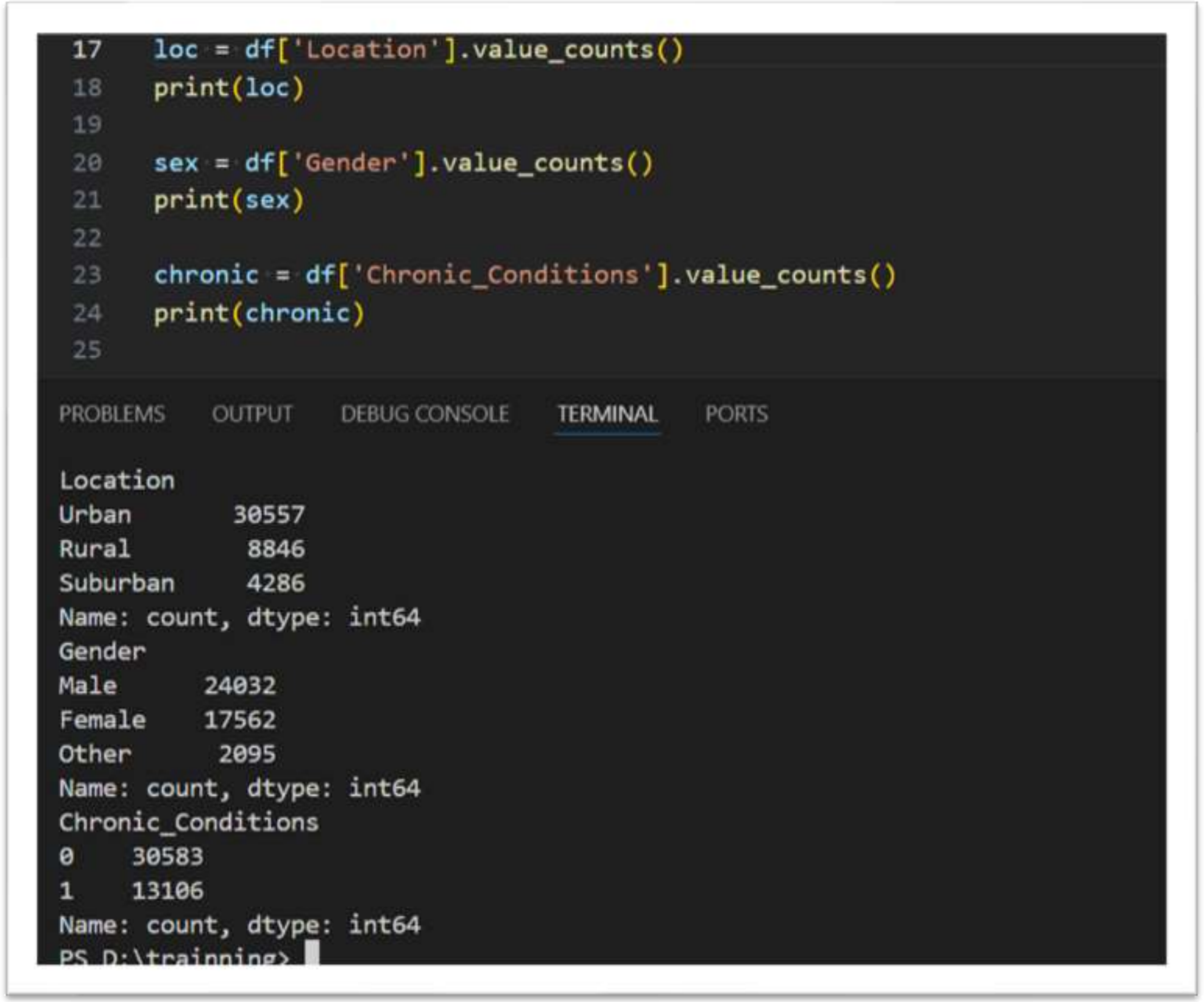
- This indicates that most patients are located in urban areas, with fewer in rural and suburban areas.

- **Male**: 24,032
- **Female**: 17,562
- **Other**: 2,095

- This shows that the dataset has more male patients than female, with a small number categorized under "Other.“

- **No Chronic Conditions**: 30,583
- **Has Chronic Conditions**: 13,106

- This shows that the majority of patients in the dataset do not have chronic conditions.

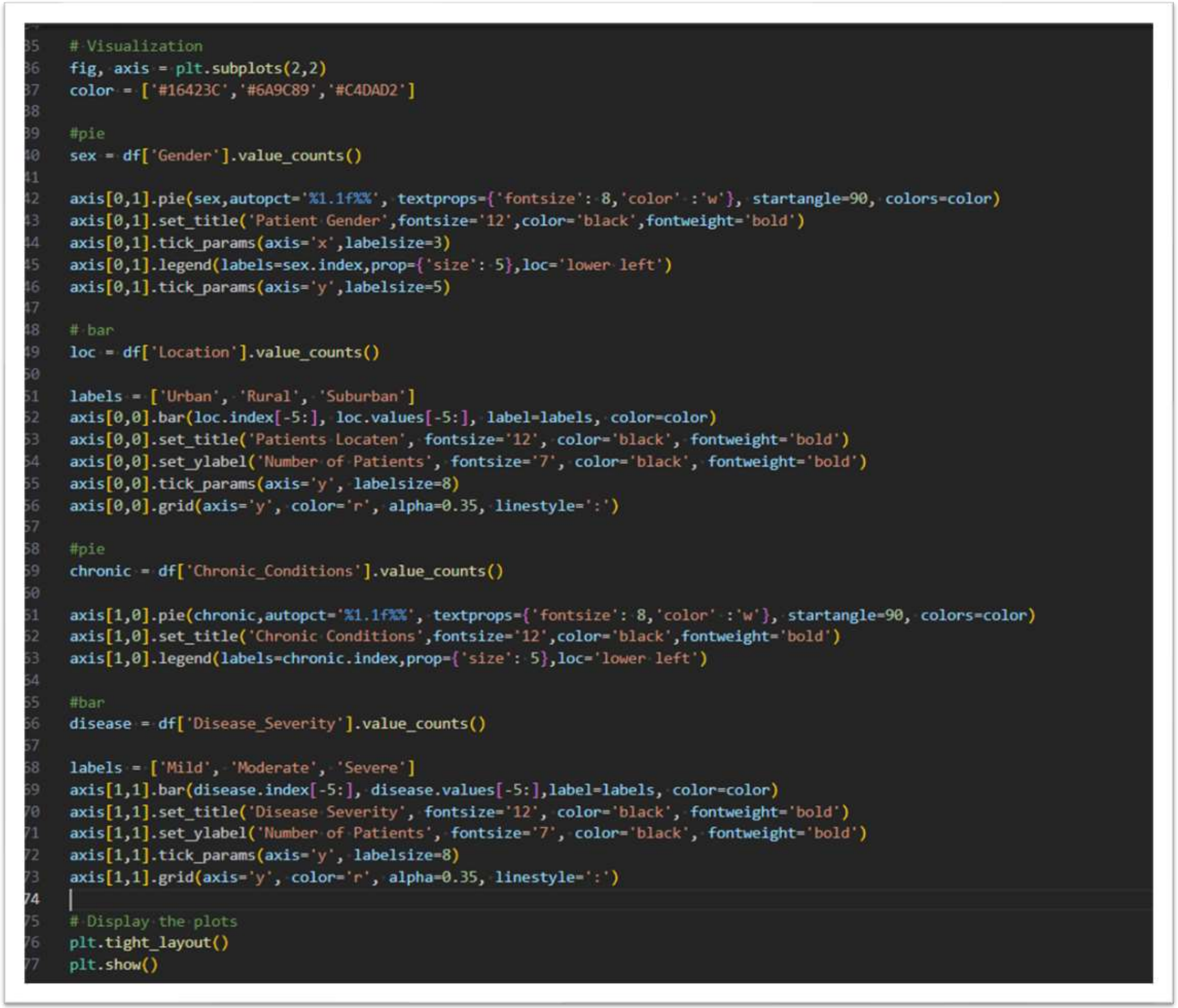


```
loc = df['Location'].value_counts()
print(loc)

sex = df['Gender'].value_counts()
print(sex)

chronic = df['Chronic_Conditions'].value_counts()
print(chronic)

```

```
Location
Urban       30557
Rural        8846
Suburban     4286
Name: count, dtype: int64
Gender
Male      24032
Female    17562
Other      2095
Name: count, dtype: int64
Chronic_Conditions
0    30583
1    13106
Name: count, dtype: int64
PS D:\trainning>
```

Dashboard code

I used Matplotlip library to visualize the date, 2*2 Dashboard, and Bar and Pie charts in Visualization.



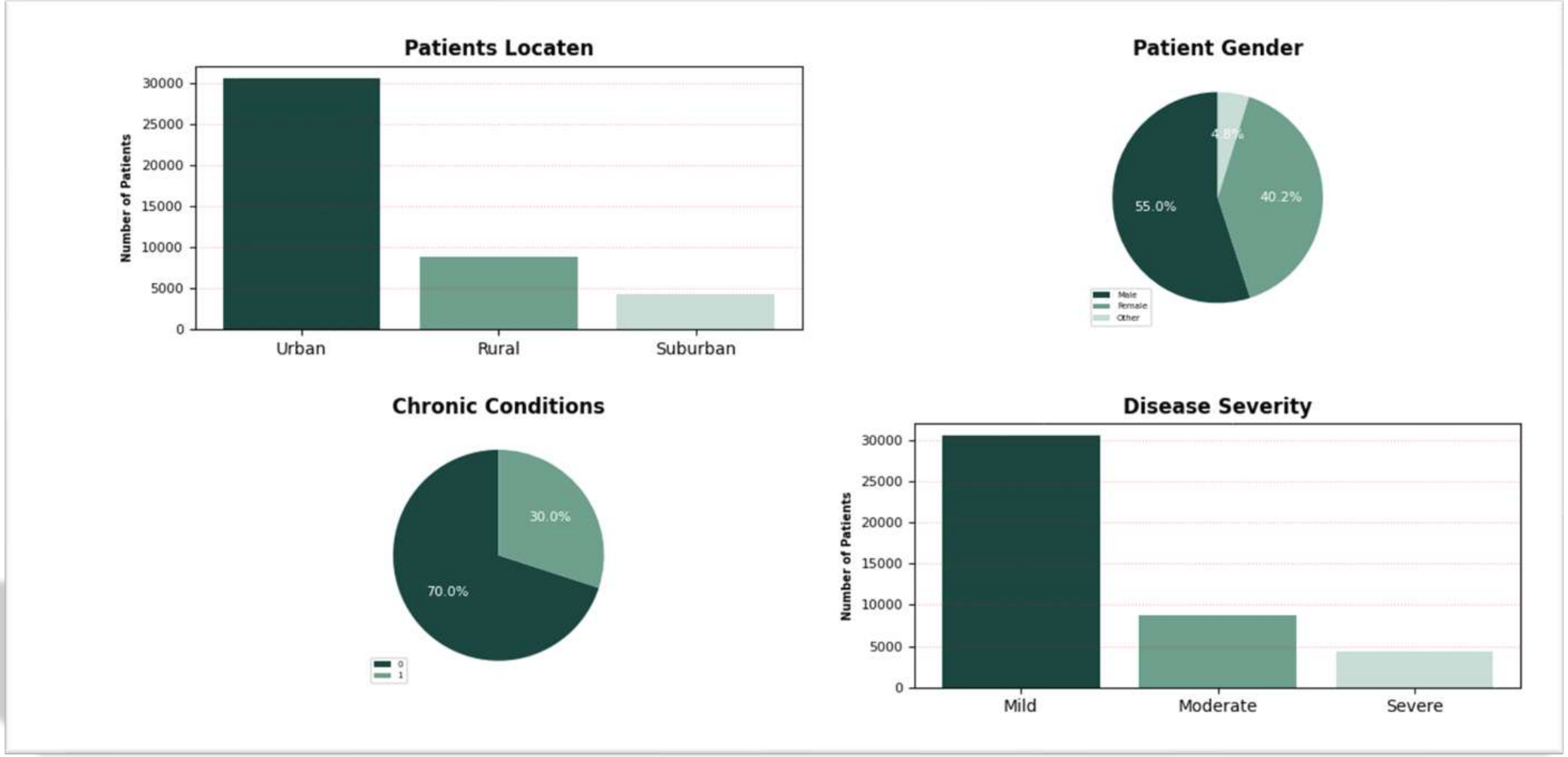
```python
# Visualization
fig, axis = plt.subplots(2,2)
color = ['#16423C','#6A9C89','#C4DAD2']

#pie
sex = df['Gender'].value_counts()

axis[0,1].pie(sex,autopct='%1.1f%%', textprops={'fontsize': 8,'color' :'w'}, startangle=90, colors=color)
axis[0,1].set_title('Patient Gender',fontsize='12',color='black',fontweight='bold')
axis[0,1].tick_params(axis='x',labelsize=3)
axis[0,1].legend(labels=sex.index,prop={'size': 5},loc='lower left')
axis[0,1].tick_params(axis='y',labelsize=5)

# bar
loc = df['Location'].value_counts()

labels = ['Urban', 'Rural', 'Suburban']
axis[0,0].bar(loc.index[-5:], loc.values[-5:], label=labels, color=color)
axis[0,0].set_title('Patients Locaten', fontsize='12', color='black', fontweight='bold')
axis[0,0].set_ylabel('Number of Patients', fontsize='7', color='black', fontweight='bold')
axis[0,0].tick_params(axis='y', labelsize=8)
axis[0,0].grid(axis='y', color='r', alpha=0.35, linestyle=':')

#pie
chronic = df['Chronic_Conditions'].value_counts()

axis[1,0].pie(chronic,autopct='%1.1f%%', textprops={'fontsize': 8,'color' :'w'}, startangle=90, colors=color)
axis[1,0].set_title('Chronic Conditions',fontsize='12',color='black',fontweight='bold')
axis[1,0].legend(labels=chronic.index,prop={'size': 5},loc='lower left')

#bar
disease = df['Disease_Severity'].value_counts()

labels = ['Mild', 'Moderate', 'Severe']
axis[1,1].bar(disease.index[-5:], disease.values[-5:],label=labels, color=color)
axis[1,1].set_title('Disease Severity', fontsize='12', color='black', fontweight='bold')
axis[1,1].set_ylabel('Number of Patients', fontsize='7', color='black', fontweight='bold')
axis[1,1].tick_params(axis='y', labelsize=8)
axis[1,1].grid(axis='y', color='r', alpha=0.35, linestyle=':')

# Display the plots
plt.tight_layout()
plt.show()
```

Python Dashboard

The final dashboard with python

Bar Chart: to describe the number of patients in different locations(Urban, Rural, and Suburban).
And the Severity of the disease on patients.

Pie Chart: to describe the percentage of patient gender and the percentage of patient the have chronic disease or not.



FROM THE PREVIOUS DASHBOARD

- **Patients Location** (Top Left):
  - This bar chart displays the number of patients in each location type: Urban, Rural, and Suburban.
  - The chart shows that the majority of patients are located in urban areas, followed by rural and then suburban.
- **Patient Gender** (Top Right):
  - This pie chart illustrates the gender distribution of the patients: Male, Female, and Other.
  - The largest portion is male patients (55.0%), followed by female (40.2%), with a smaller percentage in the "Other" category (4.8%).
- **Chronic Conditions** (Bottom Left):
  - This pie chart shows the proportion of patients with chronic conditions (1) versus those without (0).
  - About 30% of patients have chronic conditions, while 70% do not.
- **Disease Severity** (Bottom Right):
  - This bar chart shows the severity levels of the disease among the patients: Mild, Moderate, and Severe.
  - Most cases are mild, with fewer cases classified as moderate and severe.

SQL ANALYSIS

In the Analysis Questions phase using SQL, data is queried to extract relevant insights through filtering, aggregation, and joining tables. This phase focuses on answering specific business or research questions for informed decision-making.

1.Average Age by Gender:

•The first query calculates the average age of patients based on gender.

•The results display the average age for each gender category (Male, Female, Other).

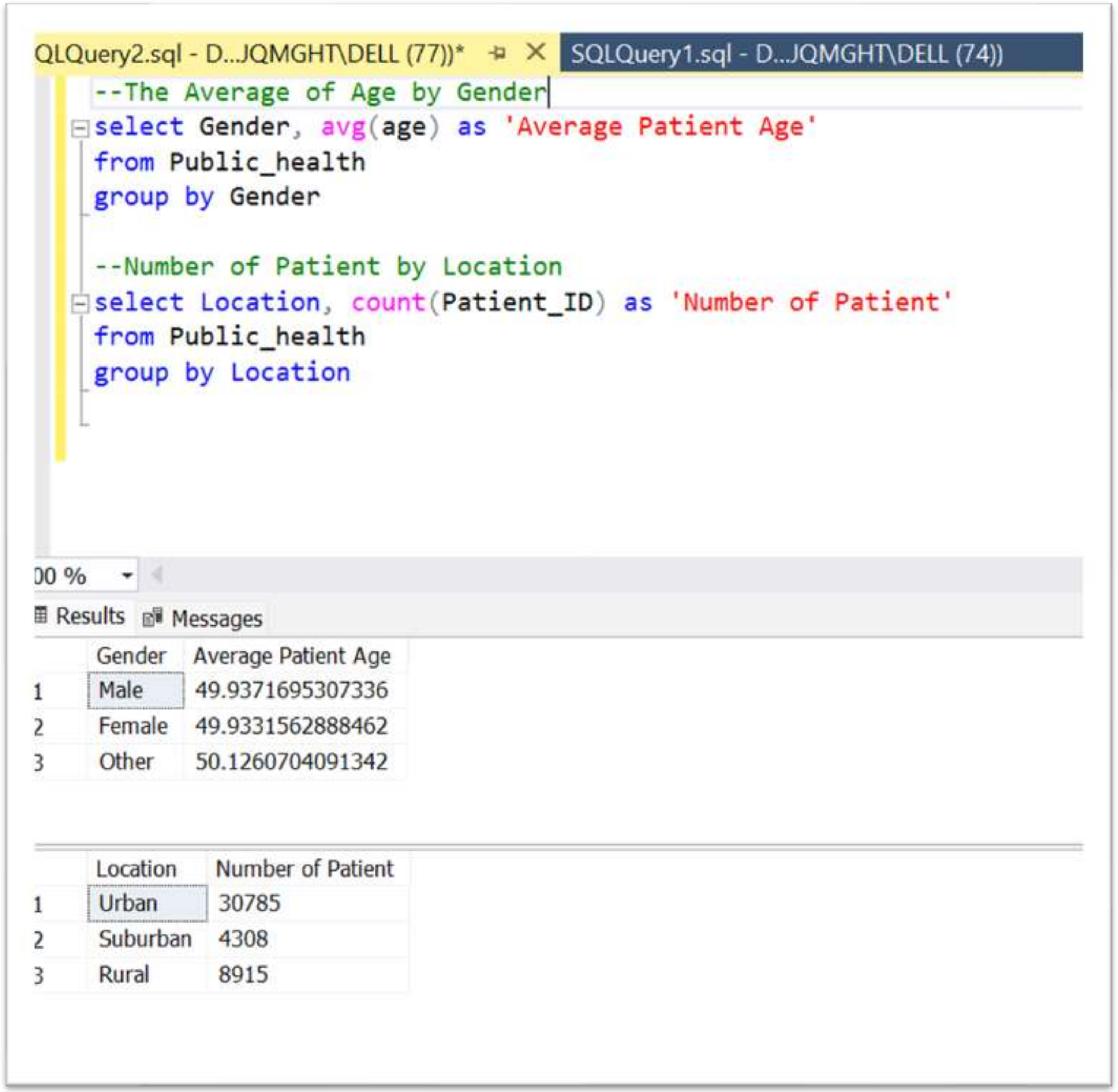
Male (49.94), Female (49.93), Other (50.13).

2.Number of Patients by Location:

•The second query counts the number of patients in each location category

•The results show the total count of patients for each location (Urban, Suburban, Rural).

Urban (30,785), Suburban (4,308), Rural (8,915).


QLQuery2.sql - D...JQMGHT\DELL (77))*
SQLQuery1.sql - D...JQMGHT\DELL (74))
00 %
Results
Messages

```
--The Average of Age by Gender
select Gender, avg(age) as 'Average Patient Age'
from Public_health
group by Gender

--Number of Patient by Location
select Location, count(Patient_ID) as 'Number of Patient'
from Public_health
group by Location
```

	Gender	Average Patient Age
1	Male	49.9371695307336
2	Female	49.9331562888462
3	Other	50.1260704091342

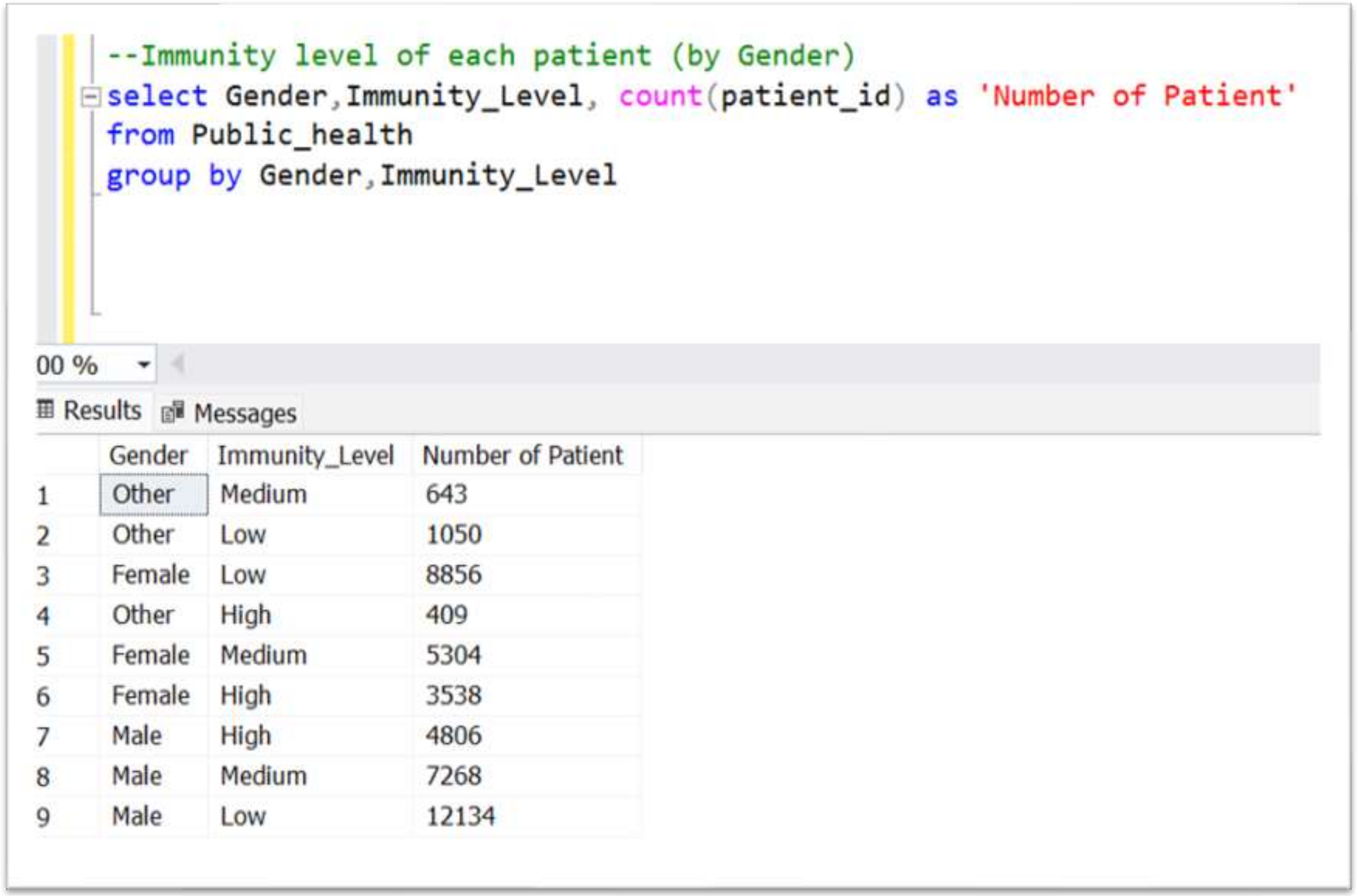
	Location	Number of Patient
1	Urban	30785
2	Suburban	4308
3	Rural	8915

Immunity Level of Each Patient by Gender:

- This query counts the number of patients by their immunity level and gender.
- It groups patients by gender and immunity level (Low, Medium, High) and counts them.

The output table reflect:

- Number of Patients by Immunity Level and Gender:
- For example, Male patients with Low immunity: 12,134.
- Female patients with High immunity: 3,538.



```sql
--Immunity level of each patient (by Gender)
select Gender,Immunity_Level, count(patient_id) as 'Number of Patient'
from Public_health
group by Gender,Immunity_Level
```

	Gender	Immunity_Level	Number of Patient
1	Other	Medium	643
2	Other	Low	1050
3	Female	Low	8856
4	Other	High	409
5	Female	Medium	5304
6	Female	High	3538
7	Male	High	4806
8	Male	Medium	7268
9	Male	Low	12134

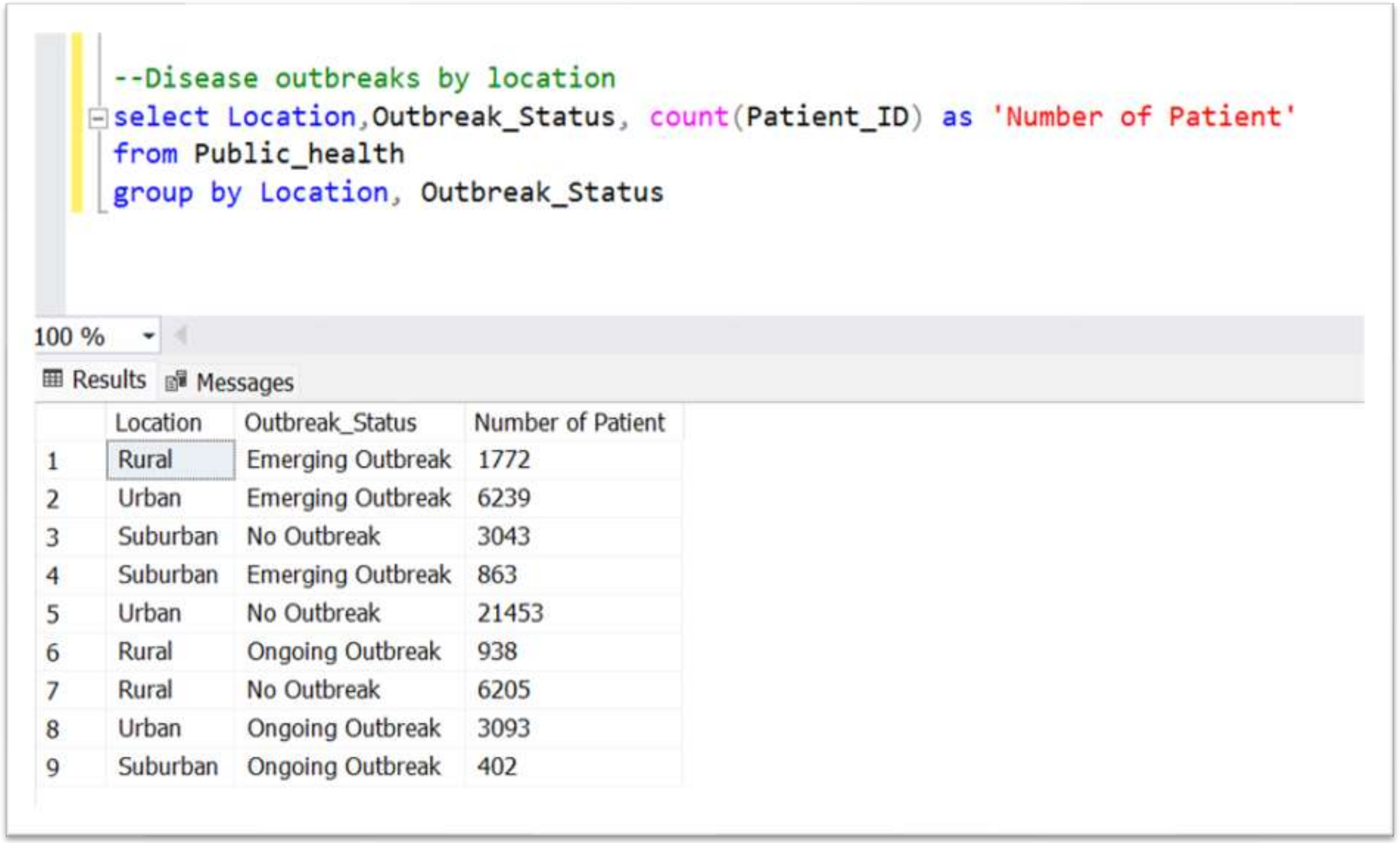
Query Explanation:

- The query counts the number of patients in each location based on the outbreak status.
- This groups the data by Location (e.g., Urban, Suburban, Rural) and Outbreak_Status (e.g., Emerging Outbreak, Ongoing Outbreak, No Outbreak), counting the number of patients in each category.

Result Explanation:

The output table shows the number of patients across different outbreak statuses by location:

- For instance:
- Urban location with “Emerging Outbreak”: 6,239 patients.•
Suburban location with “No Outbreak”: 3,043 patients.
- Rural location with “Ongoing Outbreak”: 938 patients.



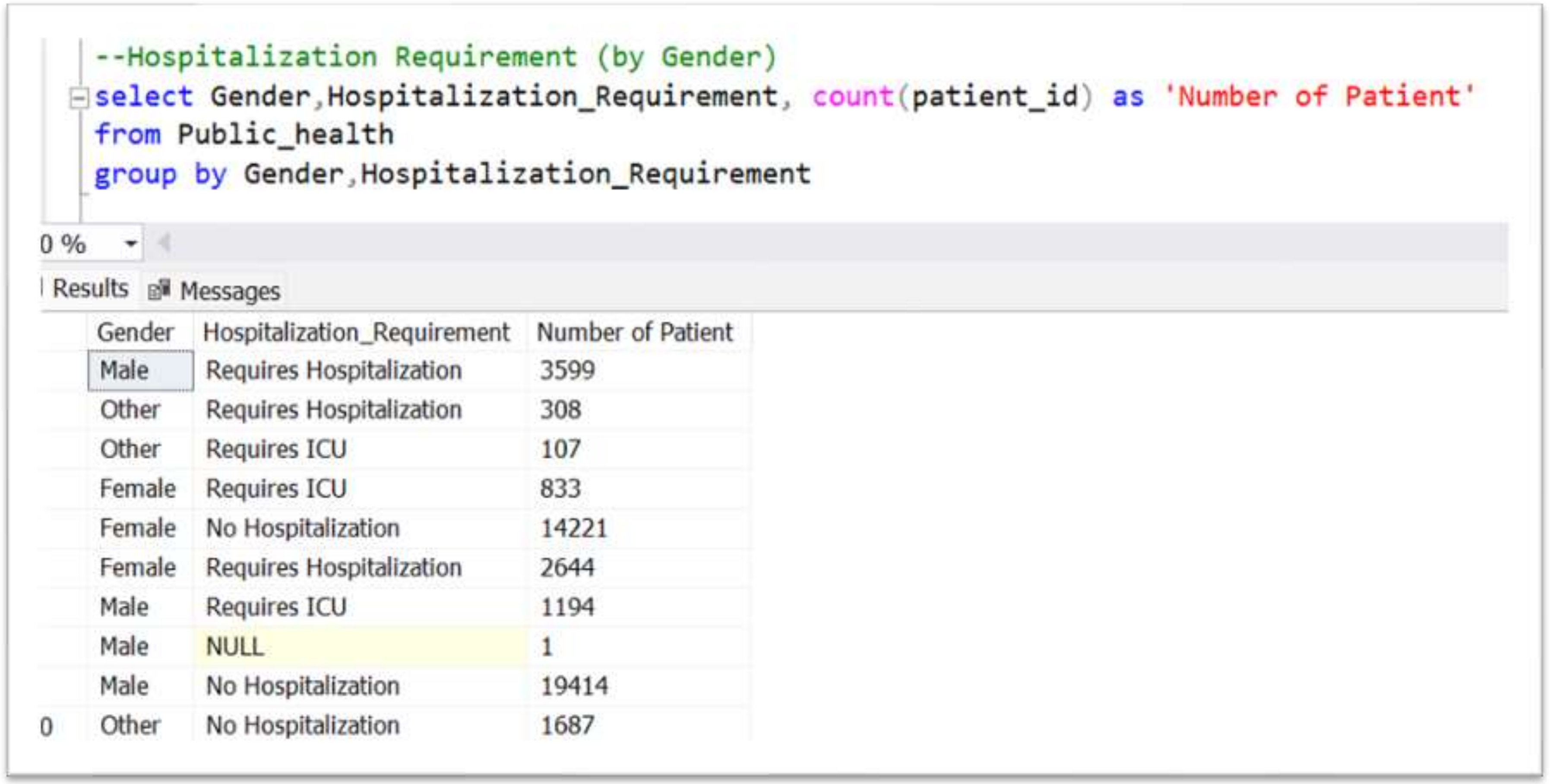
```
--Disease outbreaks by location
select Location,Outbreak_Status, count(Patient_ID) as 'Number of Patient'
from Public_health
group by Location, Outbreak_Status
```

	Location	Outbreak_Status	Number of Patient
1	Rural	Emerging Outbreak	1772
2	Urban	Emerging Outbreak	6239
3	Suburban	No Outbreak	3043
4	Suburban	Emerging Outbreak	863
5	Urban	No Outbreak	21453
6	Rural	Ongoing Outbreak	938
7	Rural	No Outbreak	6205
8	Urban	Ongoing Outbreak	3093
9	Suburban	Ongoing Outbreak	402

- **Hospitalization Requirements include**:
  - **Requires Hospitalization**: Patients who needed general hospitalization.
  - **Requires ICU**: Patients who needed intensive care.
  - **No Hospitalization**: Patients who did not require any form of hospitalization.

Result:

1. **Male Patients**:
   1. Majority are in the "No Hospitalization" category (19,414).
   2. 3,599 required hospitalization, and 1,194 required ICU.
   3. There is one record with a missing (NULL) hospitalization requirement.
2. **Female Patients**:
   1. The majority also did not require hospitalization (14,221).
   2. 2,644 required hospitalization, and 833 required ICU.
3. **Other Gender**:
   1. Smaller counts across all categories.
   2. 308 required hospitalization, 107 required ICU, and 1,687 had no hospitalization requirement.



```
--Hospitalization Requirement (by Gender)
select Gender,Hospitalization_Requirement, count(patient_id) as 'Number of Patient'
from Public_health
group by Gender,Hospitalization_Requirement
```

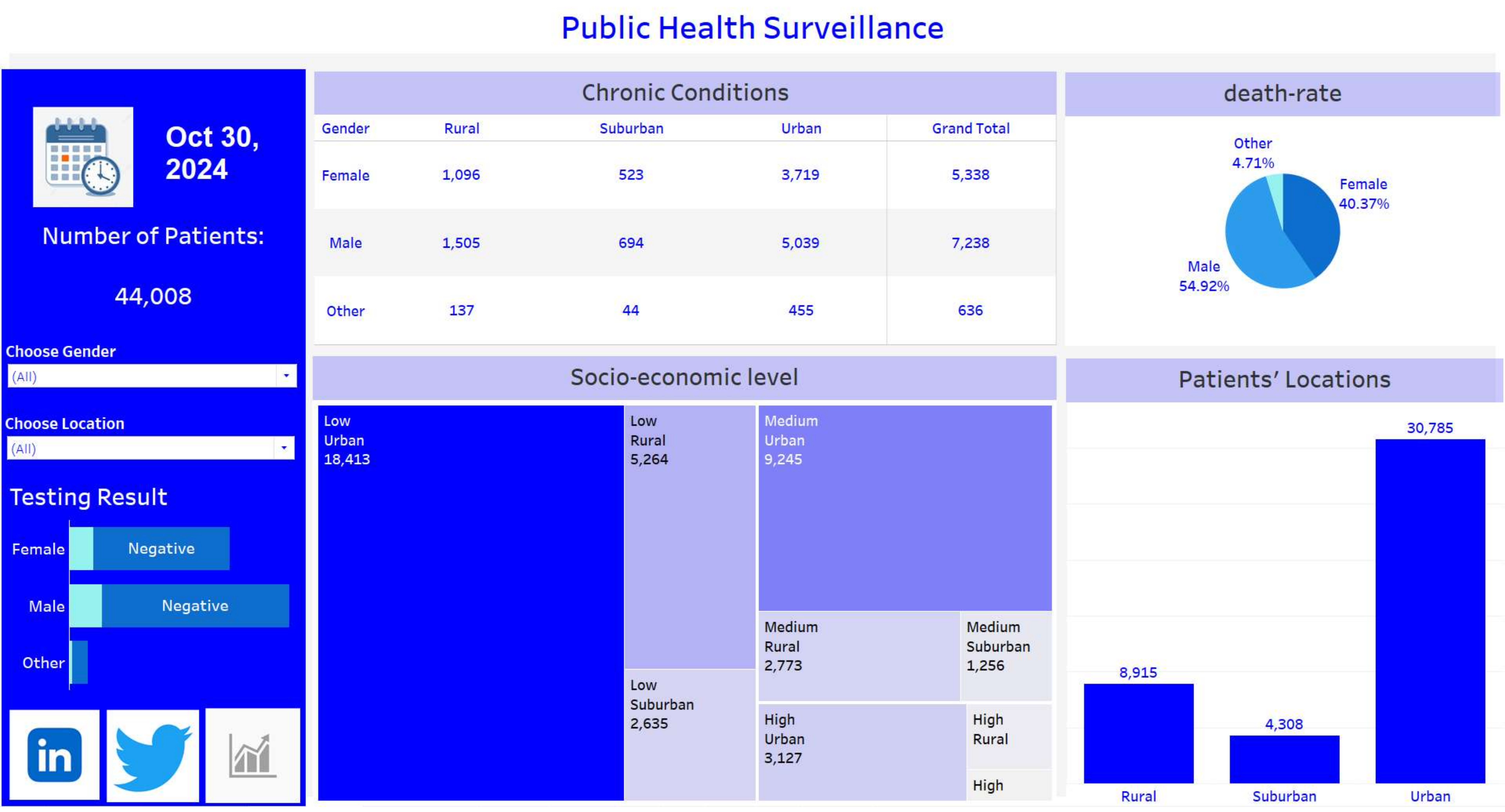
	Gender	Hospitalization_Requirement	Number of Patient
	Male	Requires Hospitalization	3599
	Other	Requires Hospitalization	308
	Other	Requires ICU	107
	Female	Requires ICU	833
	Female	No Hospitalization	14221
	Female	Requires Hospitalization	2644
	Male	Requires ICU	1194
	Male	NULL	1
	Male	No Hospitalization	19414
0	Other	No Hospitalization	1687

Tableau

is a powerful data visualization tool used to create interactive and shareable dashboards. It enables users to analyze large datasets with ease, providing insights through various charts, graphs, and maps. Tableau’s intuitive interface makes it accessible for both technical and non-technical users to visualize and understand data effectively.

VISUALIZATION DASHBOARD

The Visualization Dashboard phase involves creating interactive and informative displays to represent key insights from data. Charts, graphs, and tables are used to highlight trends, patterns, and comparisons. These visualizations support data-driven decision-making and make complex information easier to understand.



Public Health Surveillance

Oct 30, 2024

Number of Patients:

44,008

Choose Gender: (All)

Choose Location: (All)

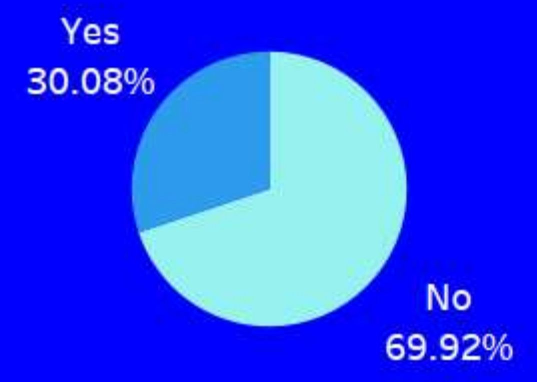
Chronic Conditions

Gender	Rural	Suburban	Urban	Grand Total
Female	1,096	523	3,719	5,338
Male	1,505	694	5,039	7,238
Other	137	44	455	636

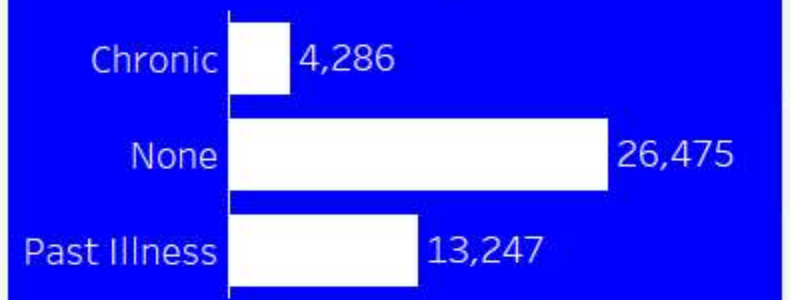
Public Health Surveillance

Oct 30, 2024

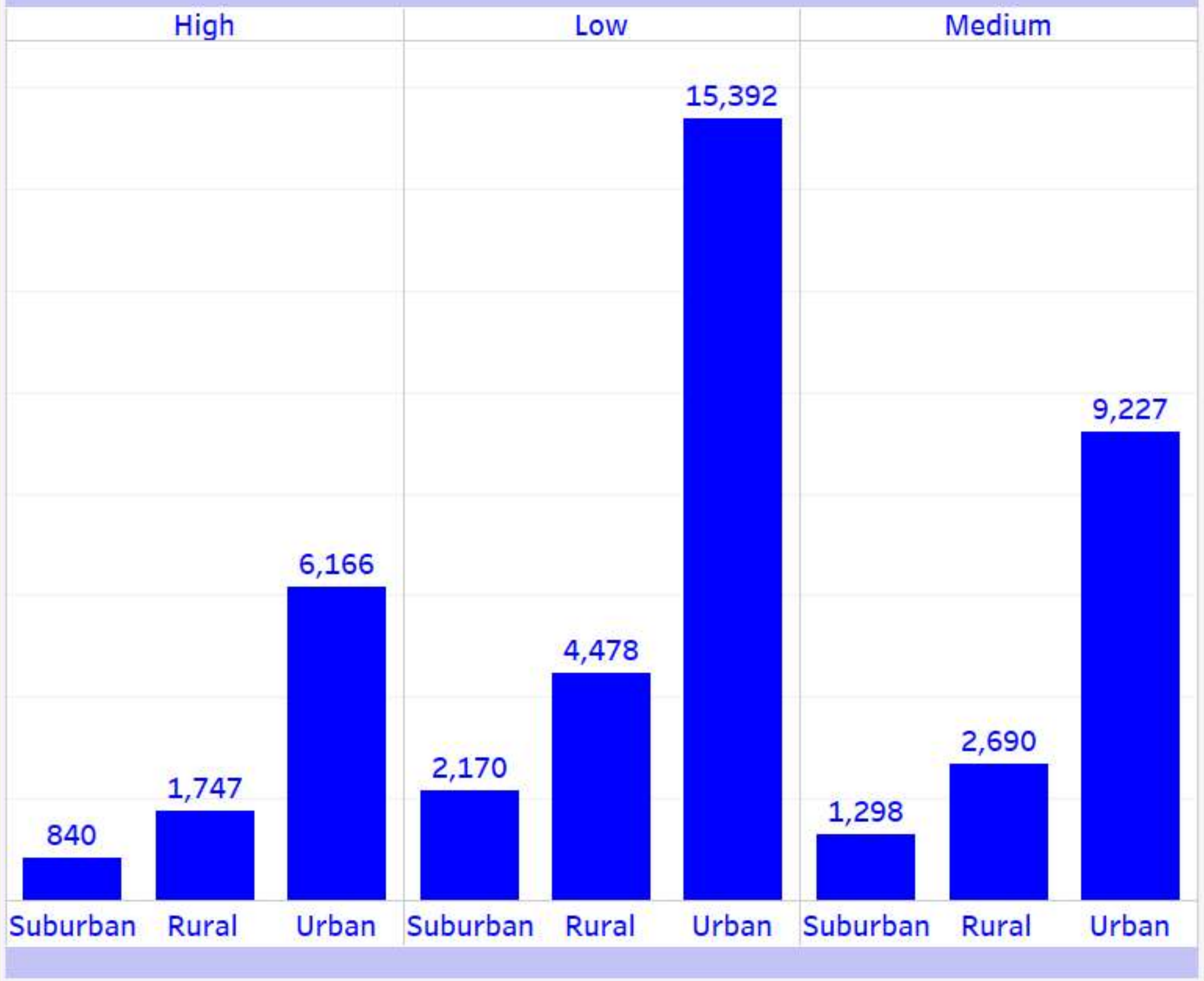
Vaccenation Hesitancy



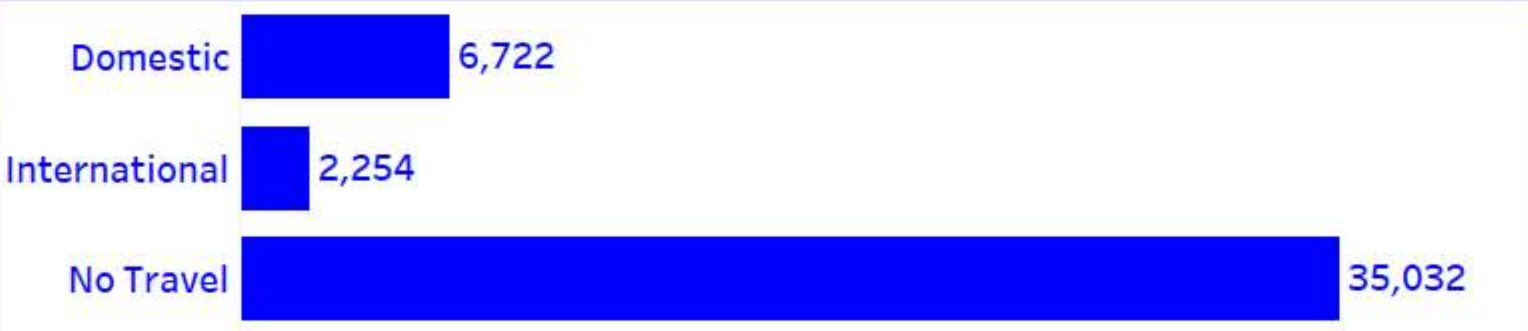
Medical History



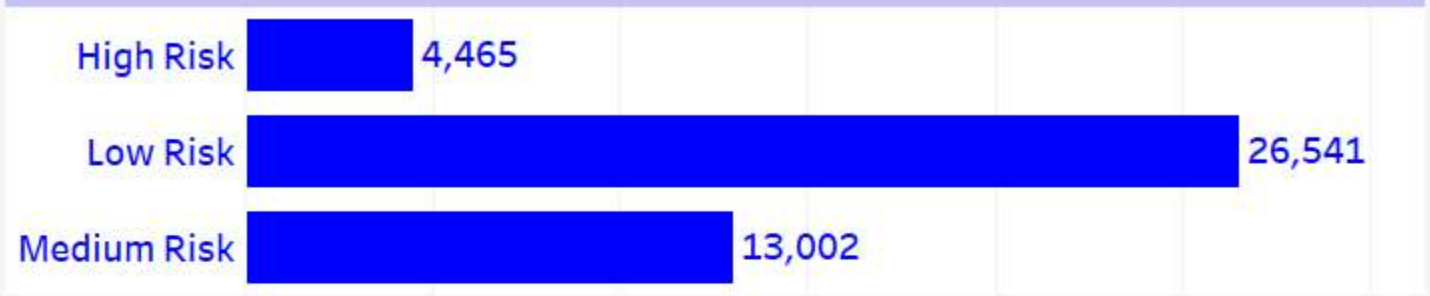
Immunity Level



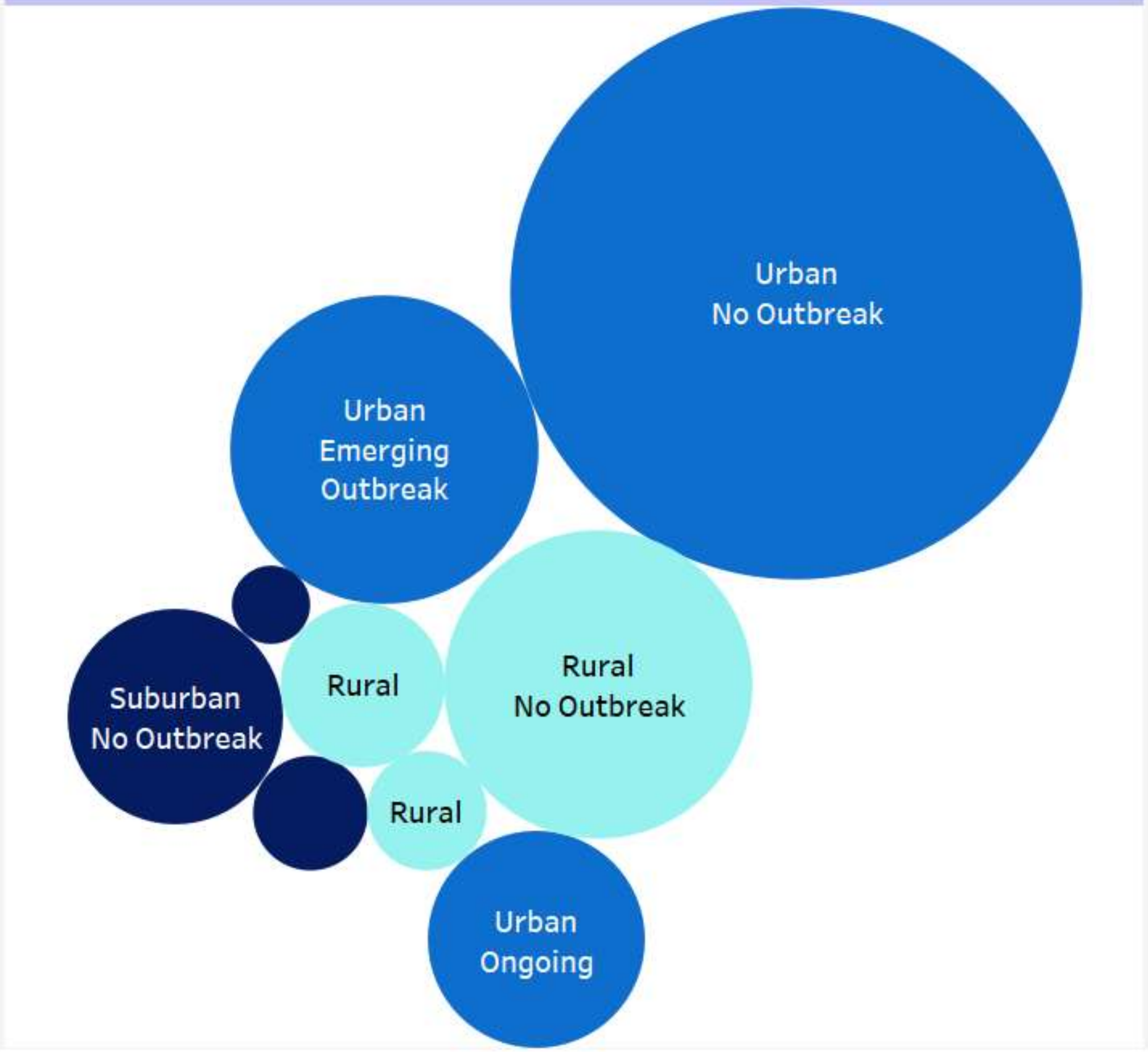
Medical History



Infection Risk Level



Outbreak Status



summary

At a time when public health is more critical than ever, data insights can be a game-changer in keeping our communities safe. Our latest Public Health Surveillance dashboard provides a comprehensive view of key health metrics across urban, suburban, and rural populations. Here are some eye-opening highlights:

- **Chronic Conditions**: A significant number of cases are concentrated in urban areas, especially among male patients.
- **Death Rate**: The majority are male, with urban areas showing a higher patient density.
- **Socio-economic Disparities**: A large portion of patients from low-income backgrounds live in urban areas, highlighting the need for targeted interventions.
- **Testing Results**: The data indicates a higher rate of negative test results among males.

- **Vaccination Hesitancy**: Nearly 1 in 3 individuals are hesitant, indicating a need for more community engagement and education.
- **Immunity Levels**: Urban areas show a higher count of low immunity, emphasizing the need for focused immunization efforts.
- **Infection Risk**: While most of the population is at low risk, a notable portion falls into medium and high-risk categories, especially those with chronic conditions.
- **Outbreak Status**: Urban areas show no major outbreaks currently, but smaller emerging ones highlight the importance of proactive monitoring.
- **Medical History**: Travelers, especially international, represent a group with unique exposure risks.

Key Takeaway: Data empowers us to take targeted actions. From vaccination campaigns to rapid outbreak responses, public health officials can use these insights to make impactful decisions.

THANK YOU