

HealthCare Insights Dashboard

Project Documentation

Krishna Raj S

INTRODUCTION

Problem Statement:

- In modern healthcare, data is critical for improving patient care and operational efficiency. However, analyzing large volumes of patient and treatment data remains challenging.

Objectives:

- Preprocess a healthcare dataset, integrate it into a database.
- Analyze sql queries and visualize the data.
- A Streamlit-based dashboard application.

Technologies & Tools Used

- Programming: Python
- Framework: Streamlit
- Database: SQL
- Libraries: Pandas, Matplotlib, Seaborn, NumPy, Pymysql

Dataset Overview

- **Dataset Source:** Available in Project Documentation
- **Key Columns :** Patient Id, Admit date, Diagnosis, Bed Occupancy, Billing amount, Insurance amount.
- **Data Pre-processing Steps:** Loading the dataset into VS Code.

Approach & Methodology

- **Understanding & Cleaning of Data:**

1. Getting information about the data.
2. Finding Null Values in the data.
3. Dropping the null values.
4. Assigning the correct data types to the columns.

- **Database Integration:**

1. Using Pymysql library establishing a connection between MySQL workbench server and VS code.

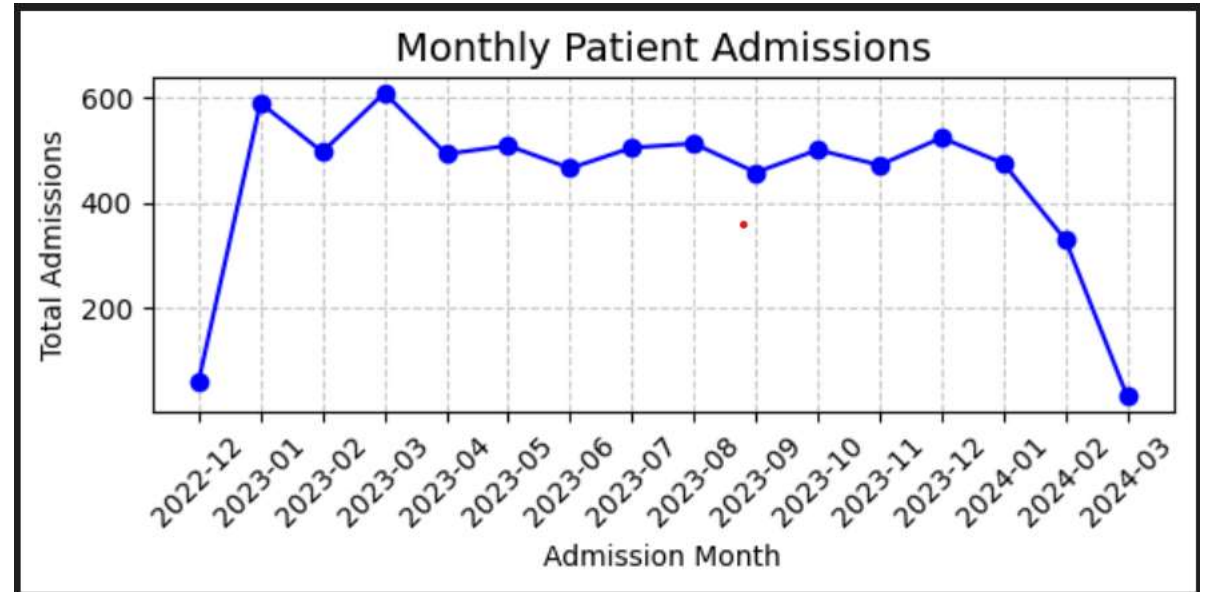
Database Name: healthcare_insights

2. Creating a Table(**Table Name : health**) and Insert the Dataset columns and values into the table.

SQL Analysis:

1) Analyze monthly patient admissions to identify trends over time:

```
SELECT
    DATE_FORMAT(Admit_Date,'%Y-%m') AS admission_month,
    COUNT(Patient_ID) as total_admissions
FROM
    health
GROUP BY
    DATE_FORMAT(Admit_Date,'%Y-%m')
ORDER BY
    admission_month;
```

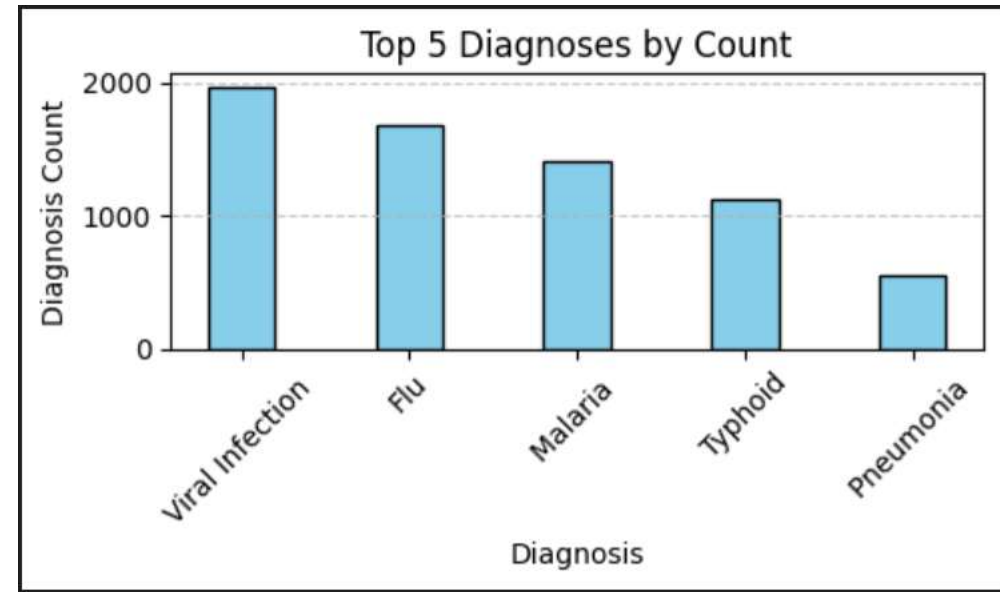


Insights:

- With the above visual we can identify the number of patients admitted over the years on month wise, which will help the admins to allocate the resources.

2) Identify the top 5 most common diagnoses:

```
SELECT
|   Diagnosis,
|   COUNT(*) as diagnosis_count
FROM
|   health
GROUP BY
|   Diagnosis
ORDER BY
|   diagnosis_count DESC
LIMIT 5;
```

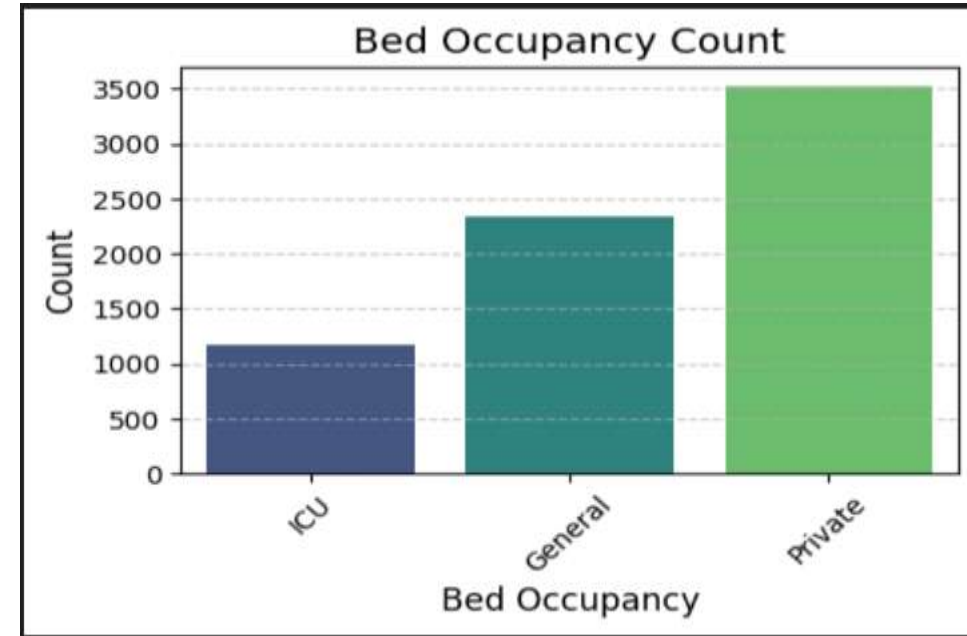


Insights:

- With the help of the above visual, the management can organise a separate camp in the demographic area where the hospital is located and based on the diseases.

3) Analyze the distribution of bed occupancy types:

```
SELECT
    Bed_Occupancy,
    COUNT(*) as bed_occupancy
FROM
    health
GROUP BY
    Bed_Occupancy
ORDER BY
    bed_occupancy;
```

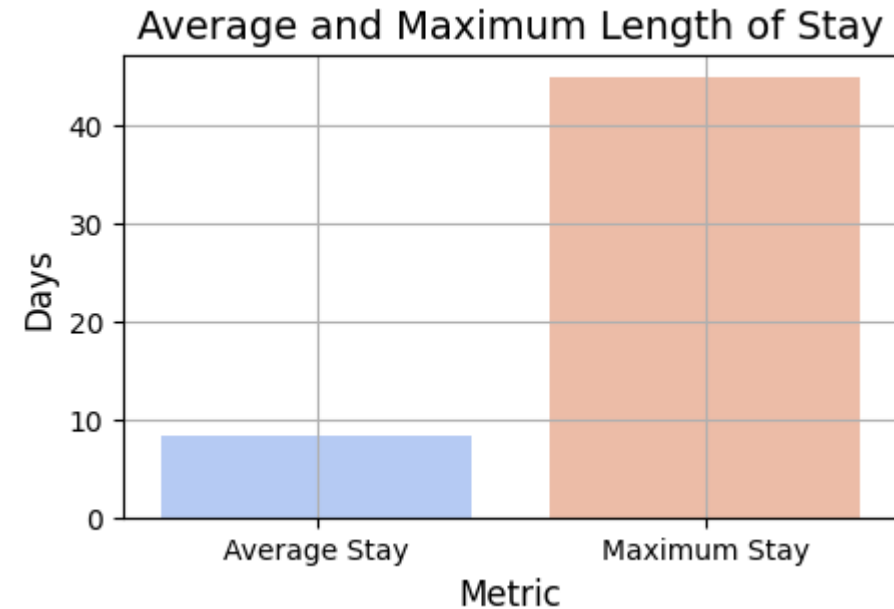


Insights:

- From the above visual we can identify, Private bed occupancy is most common.
- And also with the help of the visual, healthcare facility can spot the underutilized and over utilized bed occupancy and can allocate the resources accordingly.

4) Analyze the average and maximum length of stay for patients:

```
SELECT
    AVG(DATEDIFF(Discharge_Date,Admit_Date)) AS avg_length_stay,
    MAX(DATEDIFF(Discharge_Date,Admit_Date)) AS max_length_stay
FROM
    health;
```

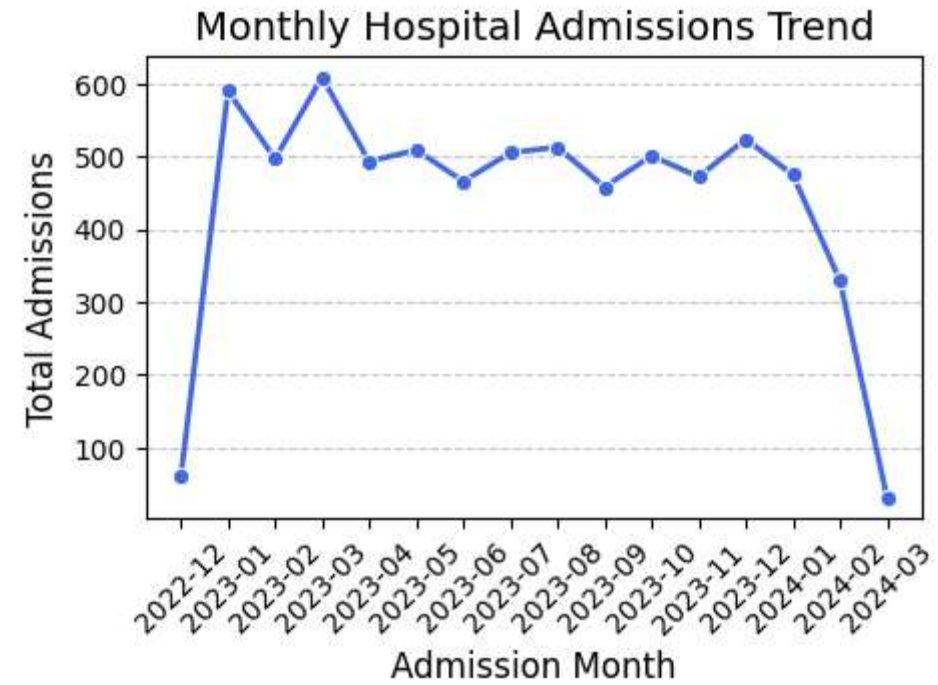


Insights:

- From the visual, the average stay of the patients is 8.5 days for all the diagnoses. So with the help of this, the healthcare facility can identify the trend and allocate the resource accordingly.

5) Identify the seasonality in admissions based on the month:

```
SELECT
    DATE_FORMAT(Admit_Date,'%Y-%m') as admission_month,
    COUNT(*) as total_admissions
FROM
    health
GROUP BY
    admission_month
ORDER BY
    admission_month;
```

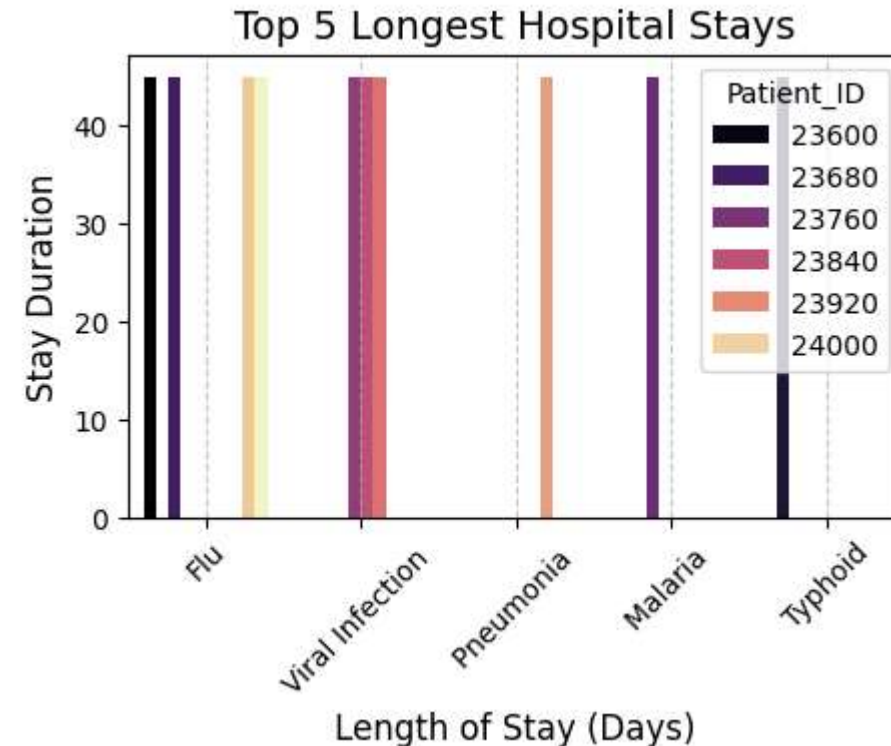


Insights:

- From the above visual, the admissions in the month of Jan, Feb 2023 sees a peak. Then there is consistent admission throughout the seasons. But there is dip in admissions on the month Jan, Feb, Mar of 2024.
- Healthcare management can run a campaign over that months and also they can do staffing (High or low) based upon the admission pattern.

6) Longest stay of patient in hospital:

```
SELECT
    Patient_ID, Admit_Date, Discharge_Date, Diagnosis,
    DATEDIFF(Discharge_Date, Admit_Date) AS stay_duration
FROM
    health
ORDER BY
    stay_duration DESC
LIMIT 10;
```

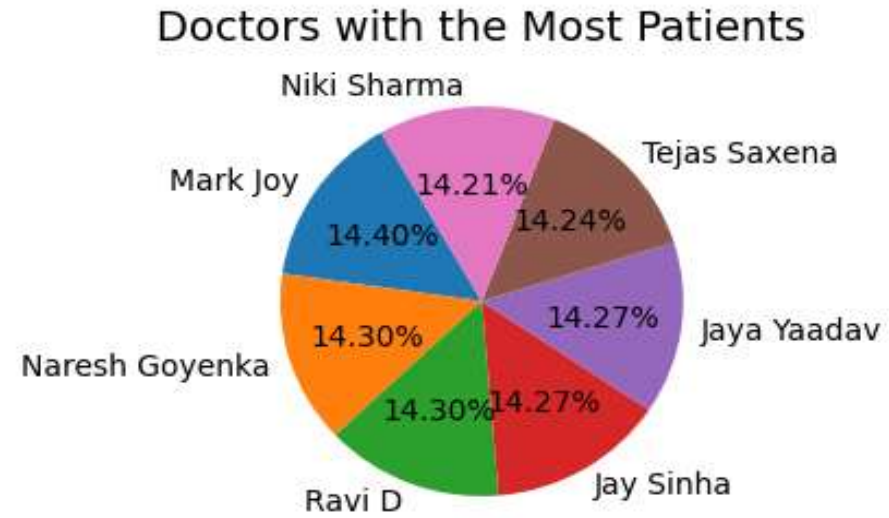


Insights:

- From the above visual, the highest duration for all diagnosis is 45 days. This will help the case reviews to understand prolonged hospitalization reasons.
- Longer stay costs higher for healthcare also, so hospital admins can make discharging plans efficiently.

7)Doctors with most patients:

```
SELECT
  Doctor,
  COUNT(Patient_ID) AS patient_count
FROM
  health
GROUP BY
  Doctor
ORDER BY
  patient_count DESC;
"""
```

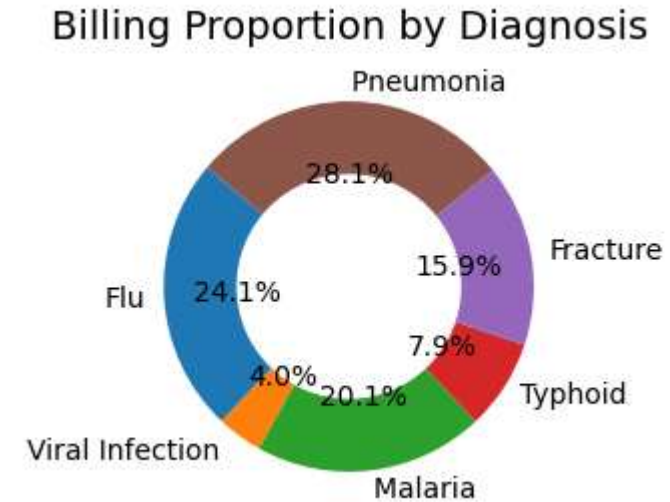


Insights:

- From the visual, the pie chart indicates the percentage of patients count allocated to the doctors. This will help the management to allocate the resource efficiently.
- And in visuals it is clearly showing each doctors are assigned consistently avoiding over assigning.

8) Total Billing_Amount and Health_Insurance_Amount by diagnosis:

```
SELECT
    Diagnosis,
    year(Admit_Date) as Years,
    SUM(Billing_Amount) AS total_billing,
    SUM(Health_Insurance_Amount) AS insurance_billing
FROM
    health
GROUP BY
    Diagnosis, Years
ORDER BY
    total_billing DESC, Years DESC;
```

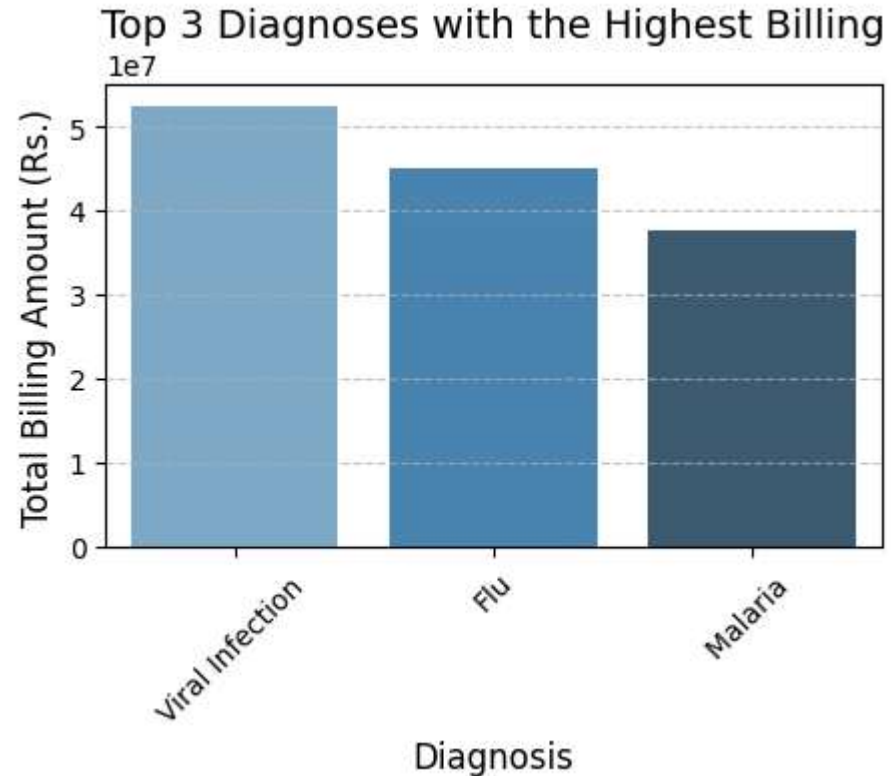


Insights:

- A Donut Chart is used for the visual, that shows which diagnosis contributes the most billing amount.
- Helps in budget planning and resource allocation.
- Identifies major cost drivers in healthcare expenses.

9) Top 3 Diagnosis with highest billing:

```
SELECT
    Diagnosis,
    SUM(Billing_Amount) AS total_billing
FROM
    health
GROUP BY
    Diagnosis
ORDER BY
    total_billing DESC
LIMIT 3;
```



Insights:

- The chart provides a clear view of the top 3 diagnoses with the highest billing amounts.
- The chart highlights high-cost diagnoses, which can inform resource allocation, ensuring that hospitals have the right staff and equipment to manage these cases.

10) Analyze the month wise billing amount:

```
SELECT
    YEAR(Admit_Date) as Year,
    MONTH(Admit_Date) as Month,
    SUM(Billing_Amount) AS Total_Billing_Amount
FROM
    health
GROUP BY
    Year,Month
ORDER BY
    Year DESC,Month DESC
```

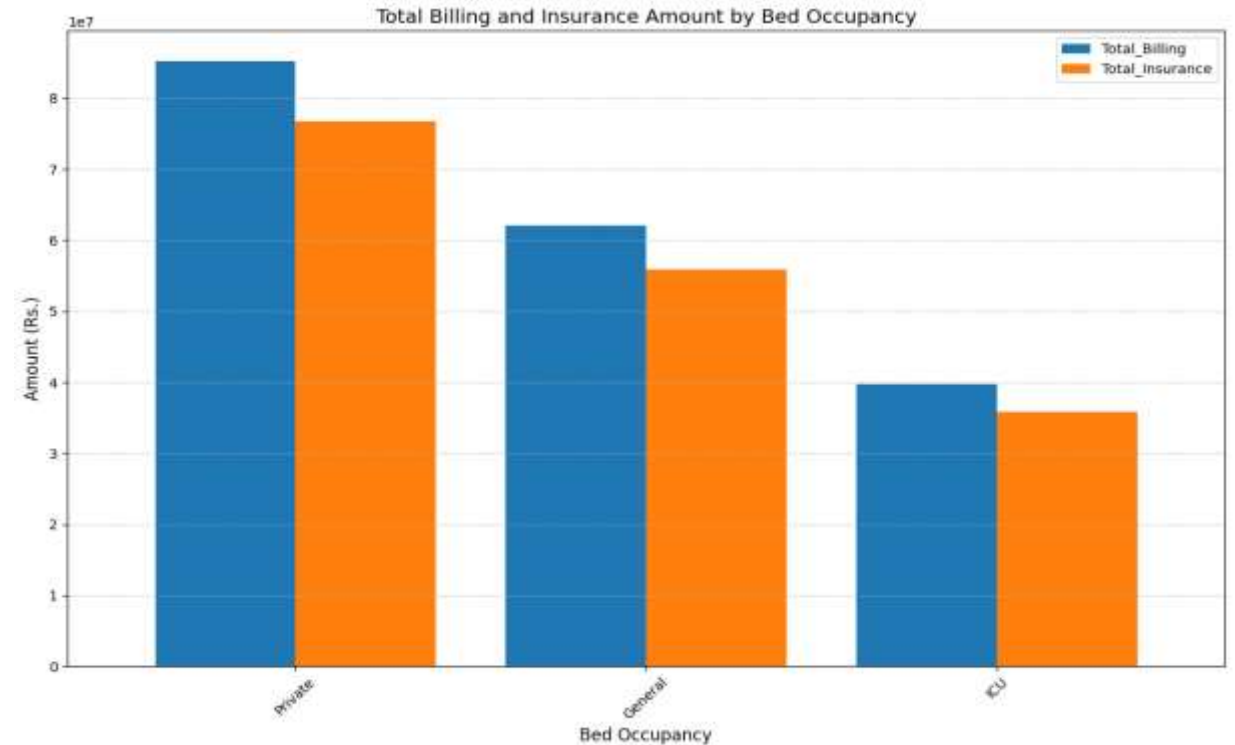


Insights:

- This line chart shows the yearly month wise billing amount. In the year 2023, there is 2 peak months and remaining months are consistent in the payments.
- In the year 2024, it is exponentially decreasing. So healthcare management should efficiently allocate the budget in the coming months.

11) Total Billing and Insurance amount by bed occupancy:

```
SELECT
    Bed_Occupancy,
    SUM(Billing_Amount) AS Total_Billing,
    SUM(Health_Insurance_amount) AS Total_Insurance
FROM
    health
GROUP BY
    Bed_Occupancy
ORDER BY
    Total_Billing DESC;
```

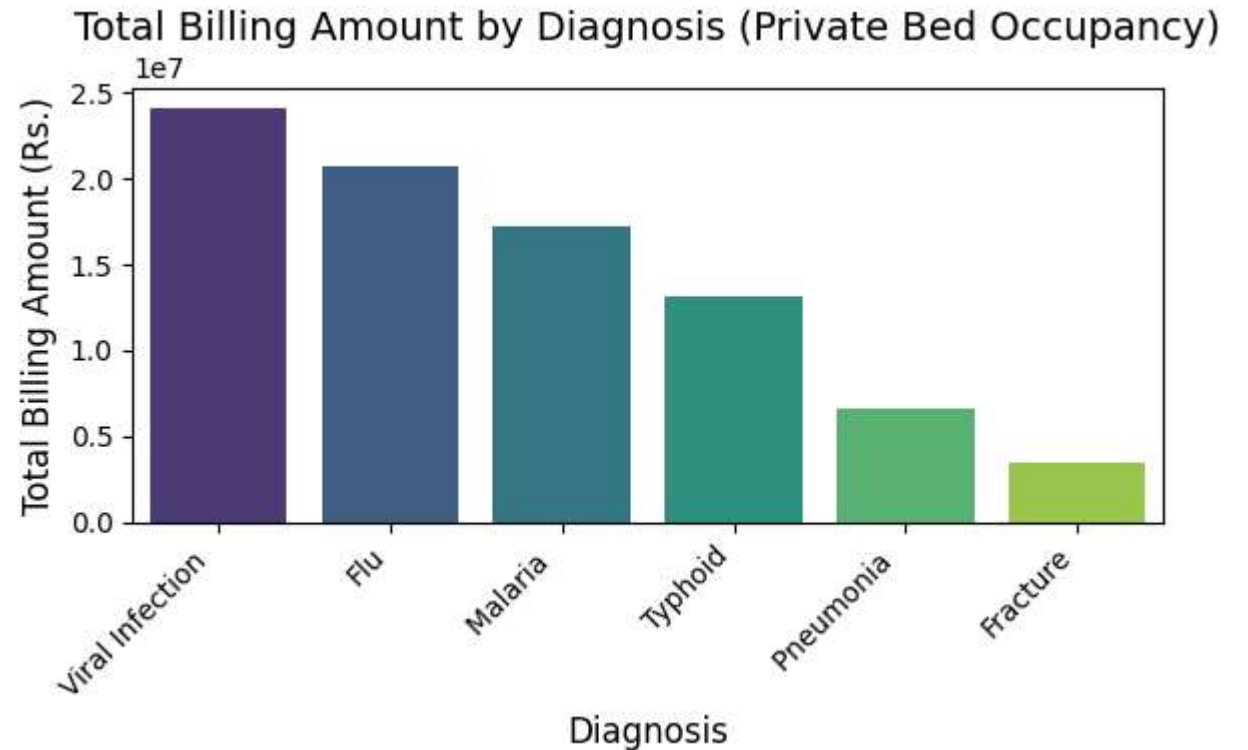


Insights:

- This side by side bar chart gives the comparison of Billing and Insurance amount based on bed occupancies.
- It clearly shows Private bed occupancy generates higher billing amount and ICU generates lower.
- This is help to allocate the resource and improve healthcare facilities.

12) Diagnosis with total billing amount where bed occupancy = 'Private':

```
SELECT
    Diagnosis,
    Bed_Occupancy,
    SUM(Billing_Amount) AS Total_Billing
FROM
    health
Where
    Bed_Occupancy = 'Private'
GROUP BY
    Diagnosis, Bed_Occupancy
ORDER BY
    Total_Billing DESC;
```

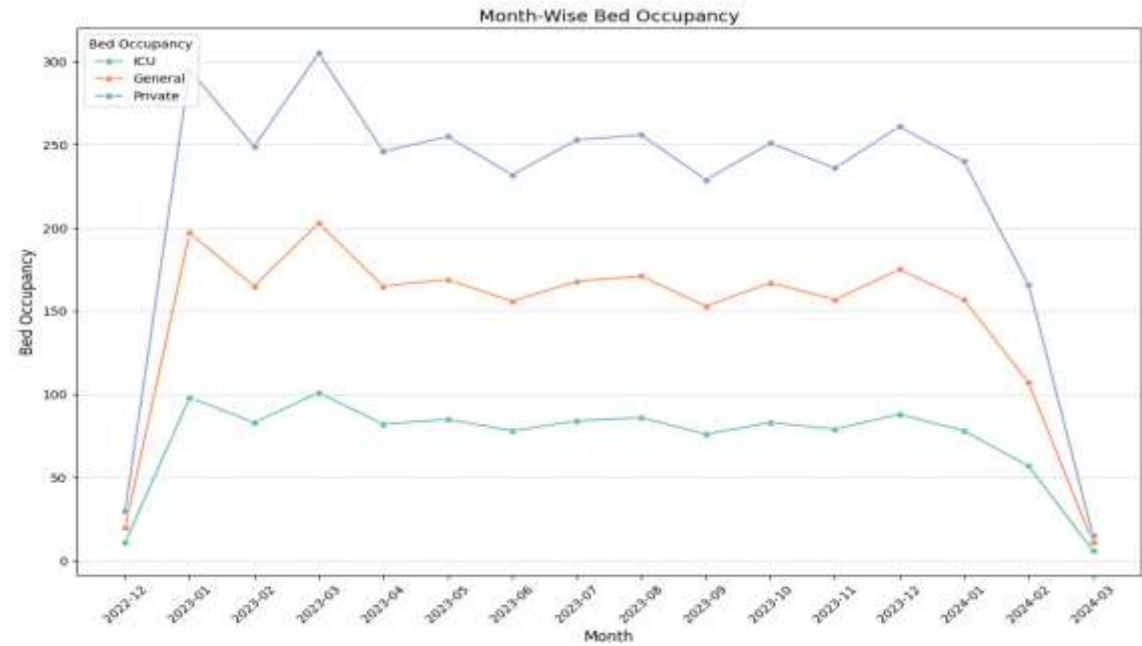


Insights:

- This vertical bar chart will highlight which diagnoses lead to the highest billing amounts for patients in Private bed occupancy. By observing the chart, you can quickly identify the most expensive diagnoses.
- The chart allows you to compare the relative costs of different diagnoses, making it easier to detect trends in healthcare spending, especially for patients in Private rooms. This could guide future decisions on resource management.

13)Month wise bed occupancy list:

```
SELECT
    DATE_FORMAT(Admit_Date,'%Y-%m') AS admission_month,
    Bed_Occupancy,
    COUNT(*) as bed_occupancy
FROM
    health
GROUP BY
    admission_month,Bed_Occupancy
ORDER BY
    admission_month,bed_occupancy
```

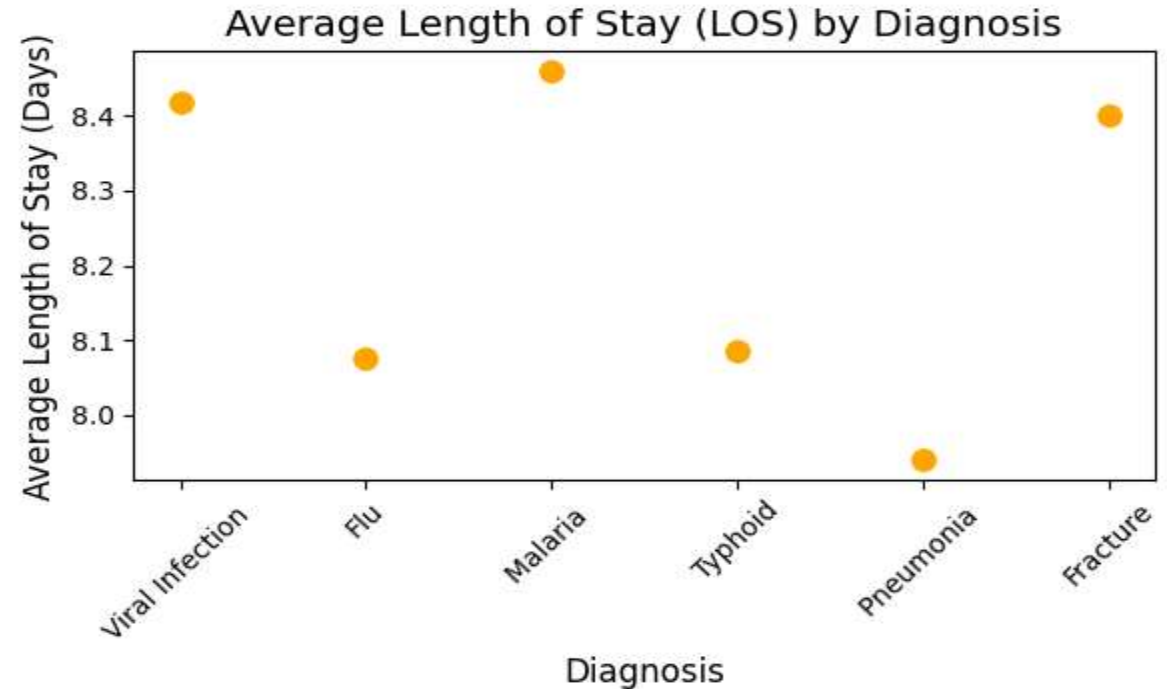


Insights:

- The line chart will allow you to see the trends in bed occupancy over time, identifying which months have higher demand for each bed type (ICU, Private, General).
- The demand for the Private Beds fluctuates across months, the hospital can adjust the pricing or availability of beds based on the chart trends.

14) Patient count with Average length of stay by diagnosis:

```
SELECT
    Diagnosis,
    COUNT(Patient_ID) AS Patient_Count,
    AVG(DATEDIFF(Discharge_Date, Admit_Date)) AS AVG_LOS
FROM
    health
GROUP BY
    Diagnosis
ORDER BY
    Patient_Count DESC;
```



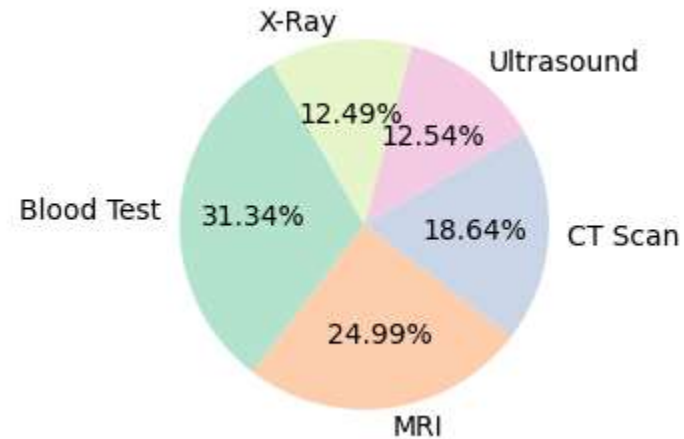
Insights:

- The scatter plot will provide insight into the average length of stay for patients with different diagnoses.
- In the plot all diagnosis have same average length of stays, the management can plan the resources accordingly.

15) Number of Patients who underwent each type of Tests:

```
SELECT
    Test,
    COUNT(*) AS Test_Count
FROM
    health
GROUP BY
    Test
ORDER BY
    Test_Count DESC;
```

Test Type Distribution by Number of Patients



Insights:

- The pie chart provides a proportional breakdown of the test types, giving a clear understanding of the distribution of test frequency. It can help identify the dominant tests and whether any types of tests are being over or under-utilized.
- By identifying the most commonly performed tests, healthcare facilities can better plan for staffing, equipment availability, and supply chain needs.

Dashboard Development Using Streamlit

A Dashboard is developed using the Framework called Streamlit which showcases the data visualization in a webpage .

BUSINESS USE CASES:

- With the help of the business use cases we showcasing the data visualizations for the following.

1) Treatment and Service Trends

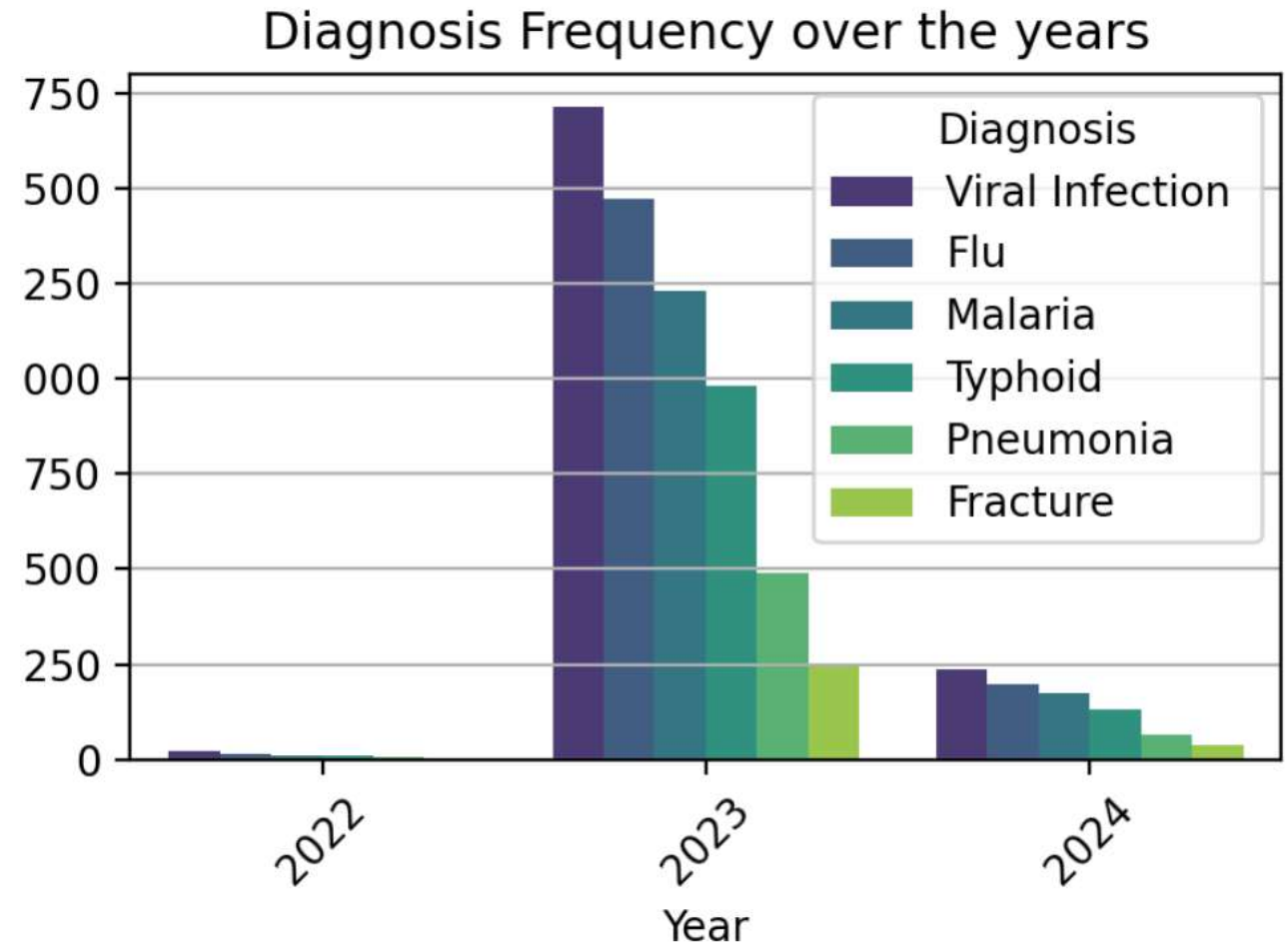
(i) Understand the frequency and popularity of treatments or procedures.

Insights:

- **Viral Infection** is the most frequent diagnosis for all three years (2022, 2023, and 2024). This shows that viral infections have been a significant health concern over the years.

- **Malaria** and **Typhoid** have consistent appearances.

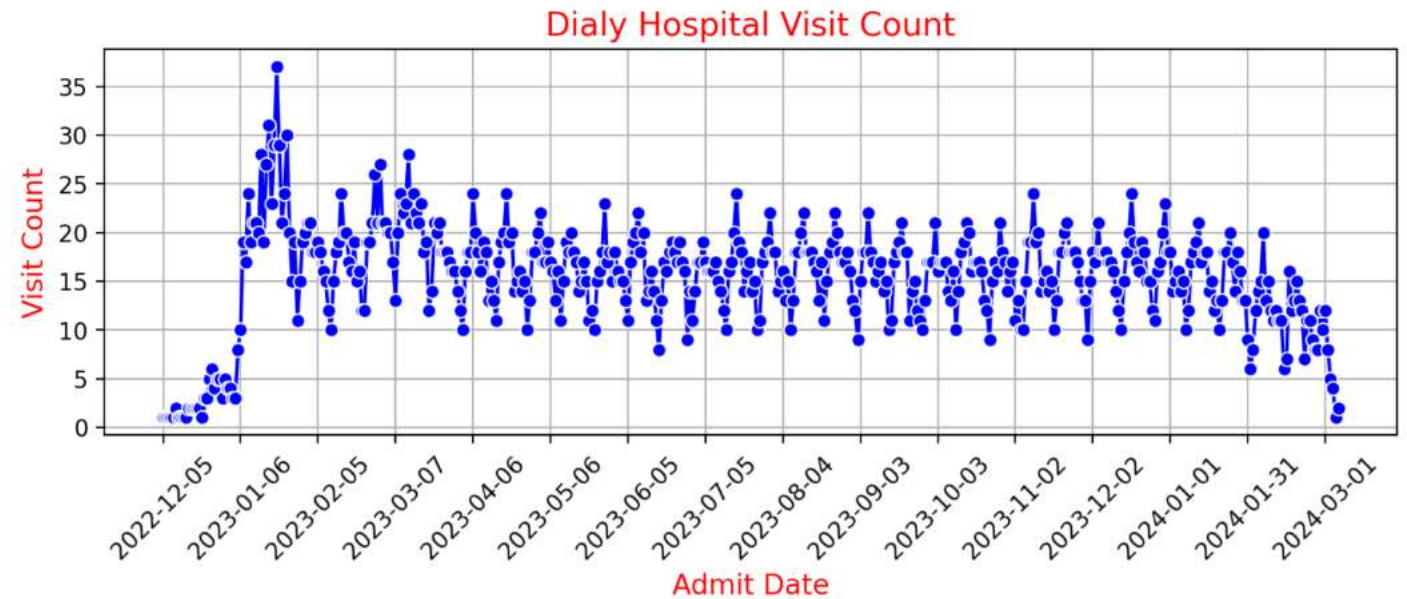
- Since there is rise in **Viral Infection** and **Flu**, resources can be dedicated to the treatment and preventive care.



(ii) Identify the peak times for patient visits to optimize resource allocation.

Insights:

- It is evident from the line chart is certain dates have high visits count(37,31,30 etc), indicating that these specific dates were likely associated with increased healthcare demand
- Optimization of staffing can be better planned for these peak periods to avoid overcrowding and ensure that resources like doctors, nurses, and hospital beds are adequately allocated.



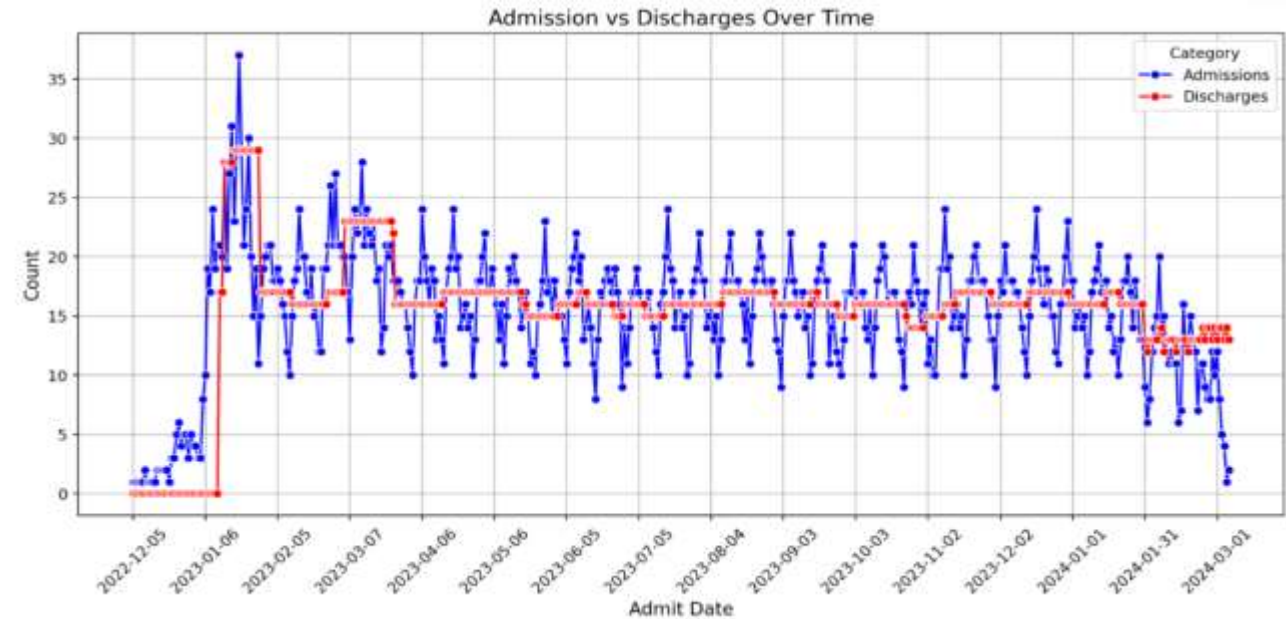
- Understanding high-traffic days can help in budgeting for extra operational costs, as hospitals may need to increase staff shifts, utilize temporary resources, or prepare for emergency situations that may arise on such days.

2) Healthcare Facility Utilization:

(i) Measure admission and discharge trends over time.

Insights:

- This line chart provides the trends of Admissions vs Discharges, where it can help to optimize the allocation of resources.
- The admission spikes indicate that management can plan on staffing, allocating beds and medical supplies.
- A low discharge rate over several consecutive days suggests that patients remain admitted for longer durations.



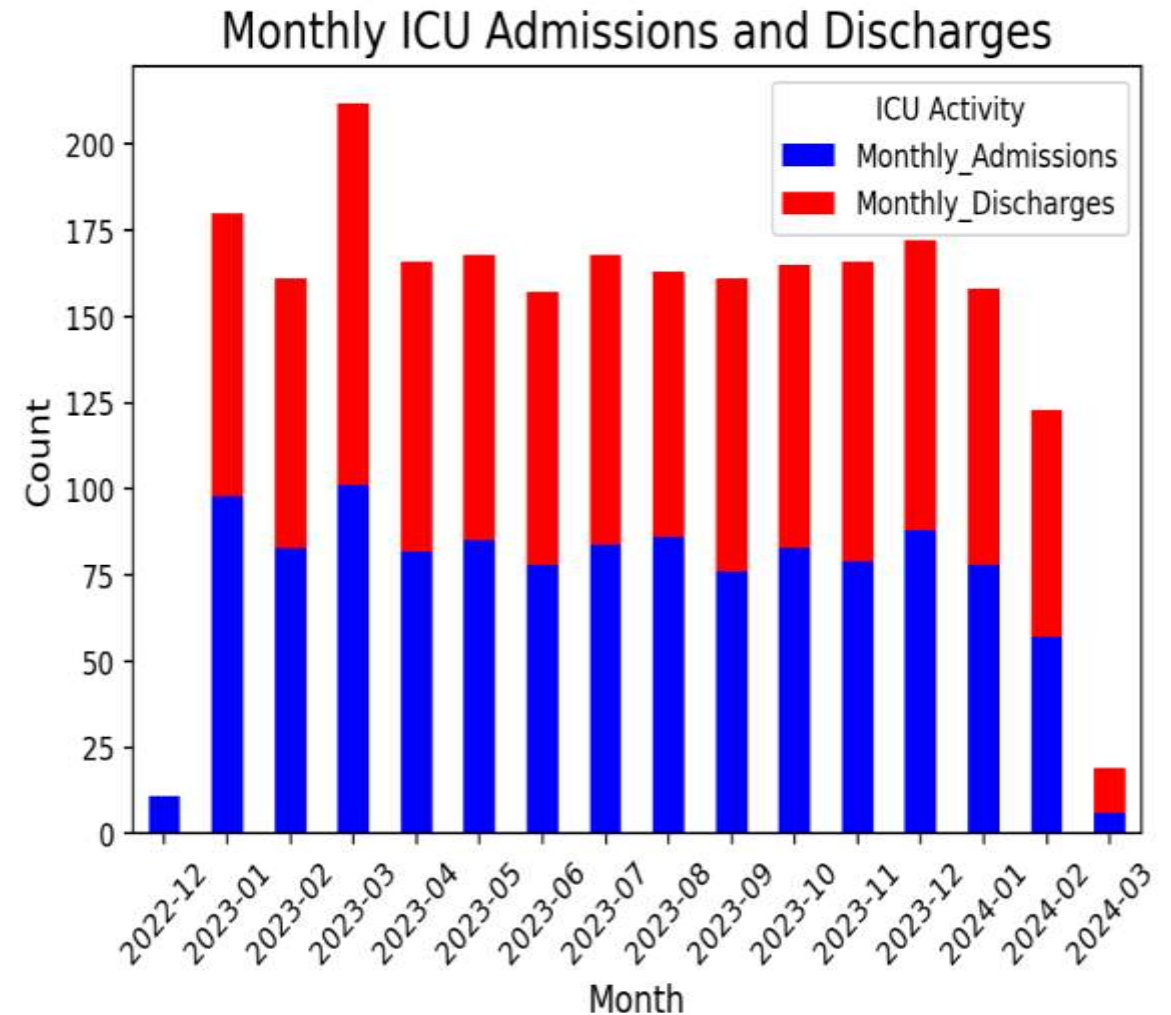
- Proper planning can help prevent overcrowding, staff burnout, and delays in patient care.

(ii) Monitor facility usage to prevent overcrowding and ensure efficient operations.

1) For Bed Occupancy = ICU

Insights:

- The visual is the monthly hospital admission and discharge trends of Bed occupancy: 'ICU' reveal patterns in patient influx and hospital capacity utilization.
- Admissions fluctuated throughout 2023, peaking in March (101 admissions) and remaining relatively stable between April and December.
- Discharges follow a similar pattern, with some months experiencing more discharges than admissions, suggesting a reduction in patient backlog.



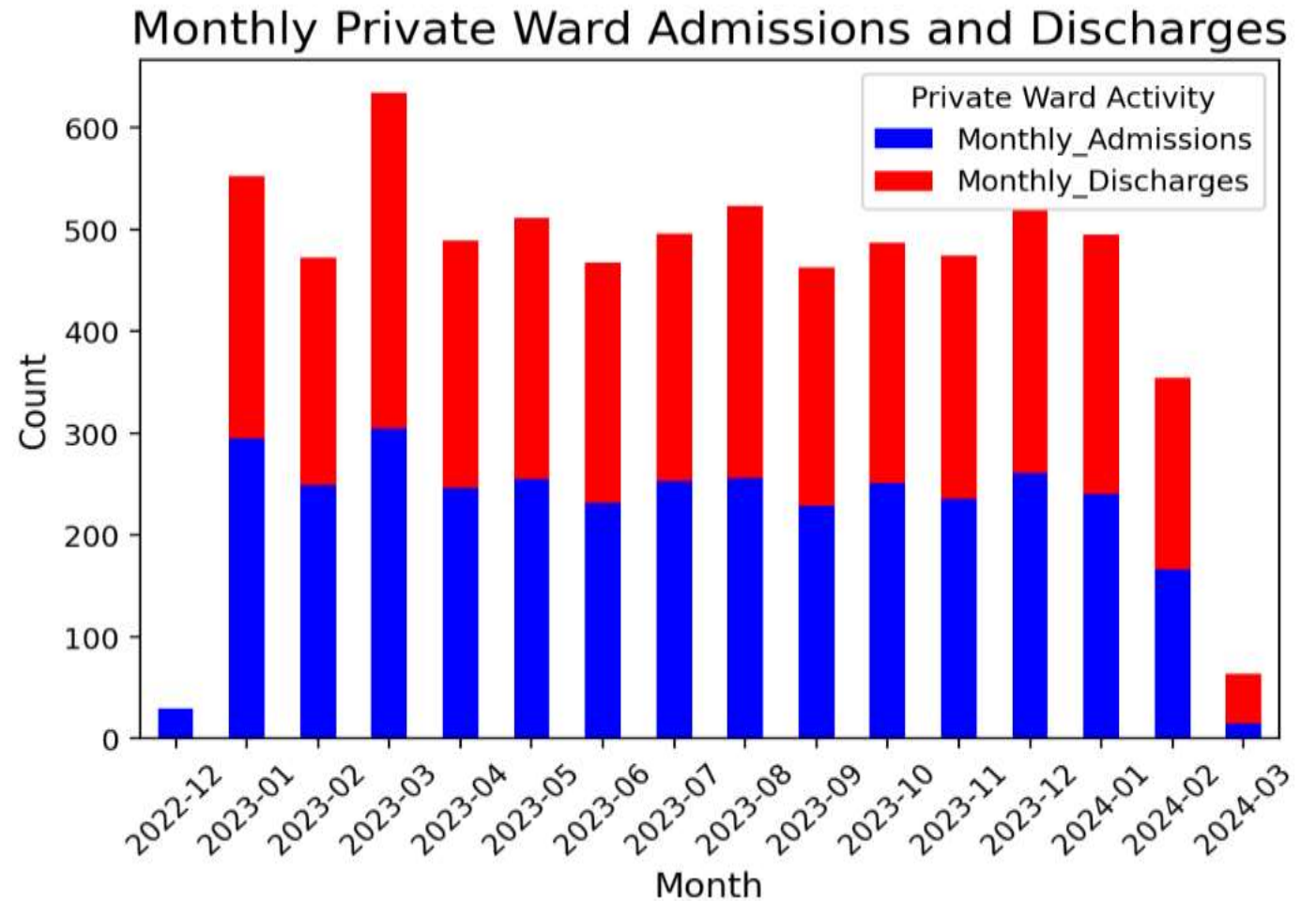
(ii) For Bed_Occupancy = 'Private'

Insights:

- The visual is the monthly hospital admission and discharge trends of Bed occupancy: **Private** reveal patterns in patient influx and hospital capacity utilization.

- The highest number of admissions occurred in March 2023 (305 patients), suggesting a peak in hospital demand.

- From January 2023 to December 2023, admissions remained relatively stable between 229 and 305 patients per month.

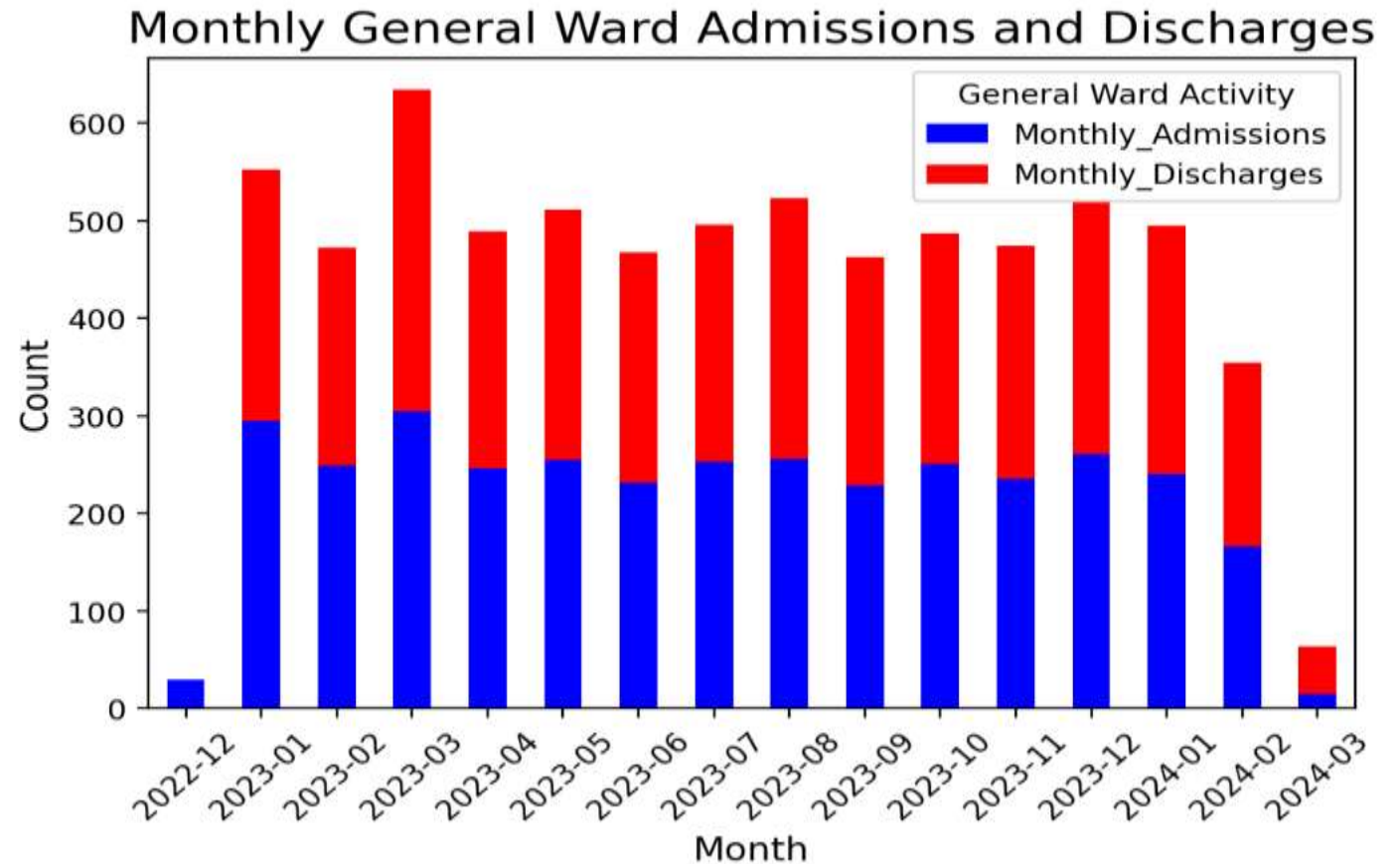


- March 2023 had the highest discharges (330), exceeding admissions (305), indicating efficient patient turnovers and clearing previous month backlogs.

(iii) For Bed_Occupancy = 'General'

Insights:

- The visual is the monthly hospital admission and discharge trends of Bed occupancy: '**General**' reveal patterns in patient influx and hospital capacity utilization.
- Admissions were highest in March 2023 (203) and January 2023 (197), indicating a high seasonal demand.
- Discharges followed a similar trend, with March 2024 having only 33 discharges, reflecting a reduced patient load.



- March 2023 had the highest discharges (205), slightly exceeding admissions (203).

Results & Impacts

- With the help of all the above trends and patterns, the healthcare management can analyze, allocate the proper resources, staffing, equipments, efficient bed_occupancies, doctor allocations are all done.
- With the help of billing and insurance amount, the management can allocate the budget for operations, salaries, equipments etc.
- Also based upon the Admissions vs Discharges count, the management can run campaigns, full health checkup campaigns etc.

Future Scope & Enhancements

- Machine Learning for Admission Prediction
- AI-based Recommendations for Resource Planning