



# Hospital Patients Data Analysis

...

# Introduction

- Understand patient demographics and diagnoses
- Evaluate hospital operations (branches, rooms) and departments loads
- Billing impacts
- Explore patterns

# Data Set

- 1,000,000 Patient records
- 26 Columns: Age, Gender, Doctor\_assigned, Billing etc.

|   | Patient_ID                           | Name             | Gender | Age | Date_of_Admission | Date_of_Discharge | Diagnosis     | Treatment    | Doctor_Assigned | Department | ... | Phone              |
|---|--------------------------------------|------------------|--------|-----|-------------------|-------------------|---------------|--------------|-----------------|------------|-----|--------------------|
| 0 | bdd640fb-0667-4ad1-9c80-317fa3b1799d | Daniel Doyle     | Other  | 15  | 2024-10-24        | 2025-05-09        | Diabetes      | Therapy      | Javier Johnson  | Neurology  | ... | 538.990.8386       |
| 1 | a0ee89ae-d453-4d32-8b0d-bb418d5288f1 | Shannon Ray      | Male   | 72  | 2024-11-14        | 2024-07-03        | Hypertension  | ICU          | Dylan Miller    | General    | ... | 516.647.5255       |
| 2 | e27a984d-6548-41d0-bfcd-9eb1a7cad415 | Judy Baker       | Female | 14  | 2023-12-15        | 2025-01-08        | Diabetes      | Chemotherapy | Kevin Hurst     | Cardiology | ... | (353)828-7101x2269 |
| 3 | 9be578c7-81f6-41d4-a392-31a7d777a477 | James Martin     | Other  | 80  | 2025-05-15        | 2025-05-20        | Heart Disease | Therapy      | Linda Burns     | ICU        | ... | 954-330-3911       |
| 4 | 05628059-568c-469b-9064-005c3985c3cf | Crystal Robinson | Female | 21  | 2023-08-08        | 2025-02-02        | Covid-19      | Therapy      | Taylor Heath    | Neurology  | ... | 983-547-3829       |

5 rows × 26 columns

# Data Preprocessing

## Handle missing values

- Check missing values  
`df.isnull.sum()`
- Fill missing values  
`df['col'].fillna('value')`

## Remove duplicates

Use  
`df.drop_duplicates(inplace=True)`

## Convert data types

Use  
`pd.to_datetime(df['col'])`

# Explore The Data (EDA)

# Analysis

## Average age of patient

- Average age: 50 years
- More middle-aged patients

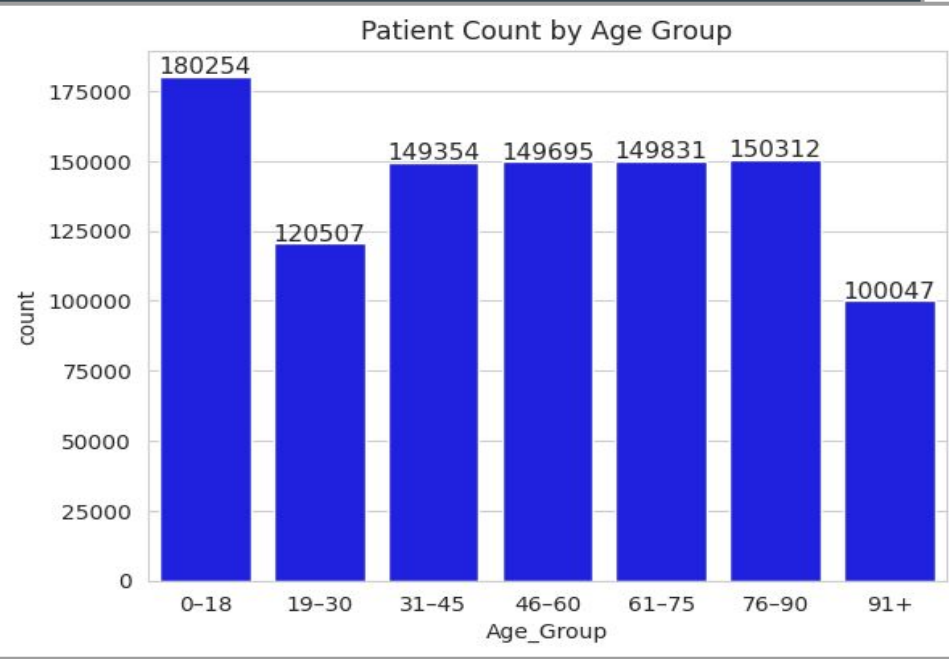
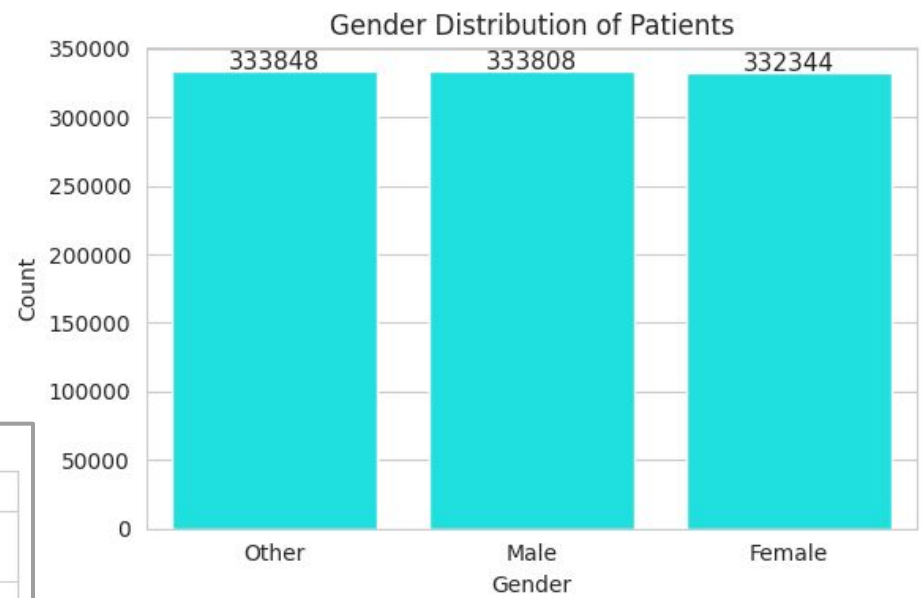
## Top 3 diagnoses

- Diabetes
- Fracture
- Hypertension

## Patient with or without medical history

|                 | count  |
|-----------------|--------|
| Medical_History |        |
| Yes             | 500317 |
| No              | 499683 |

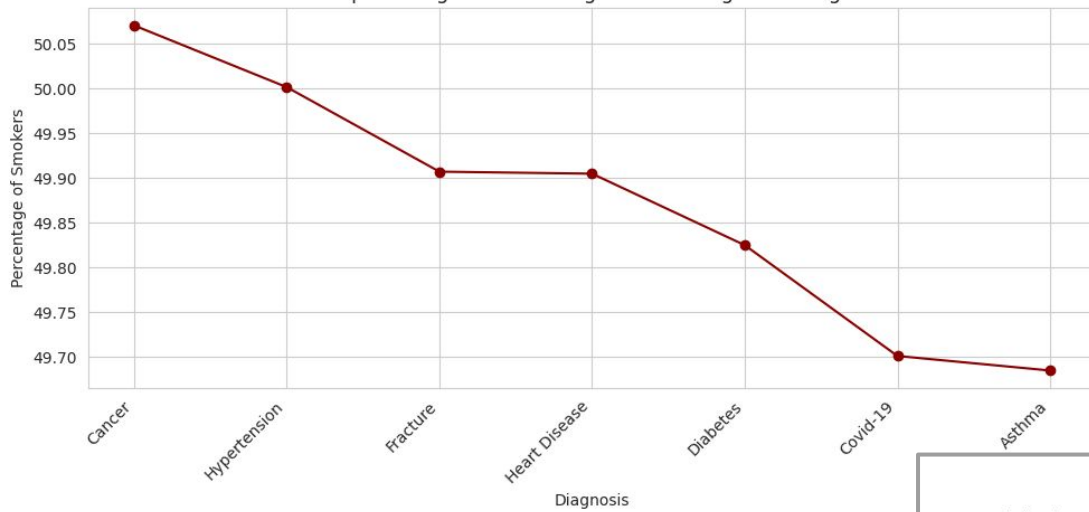
# Gender distribution of patient



Highest number of hospital admissions by age group

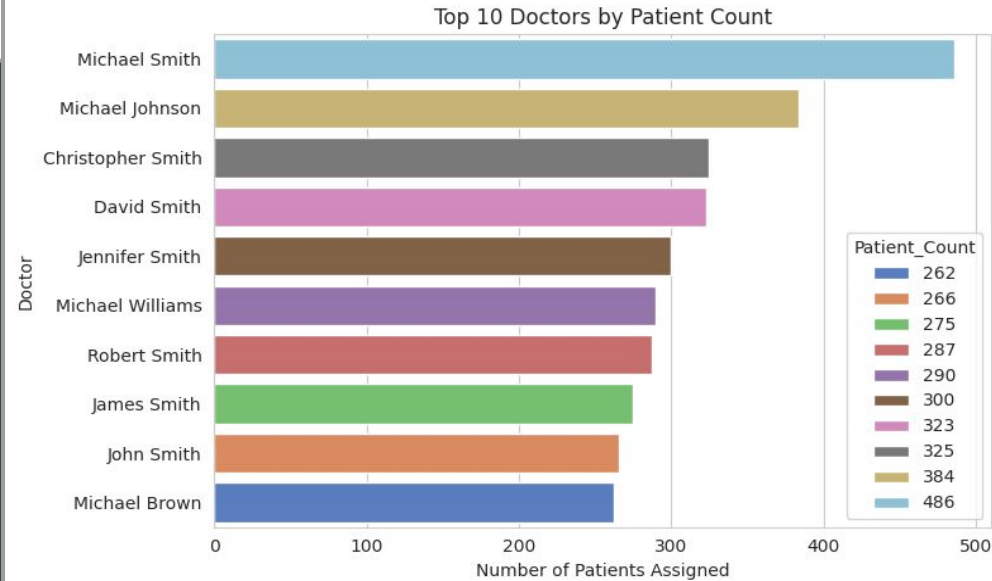
---

Top 10 Diagnoses with Highest Smoking Percentage



## Smoker vs Diagnosis

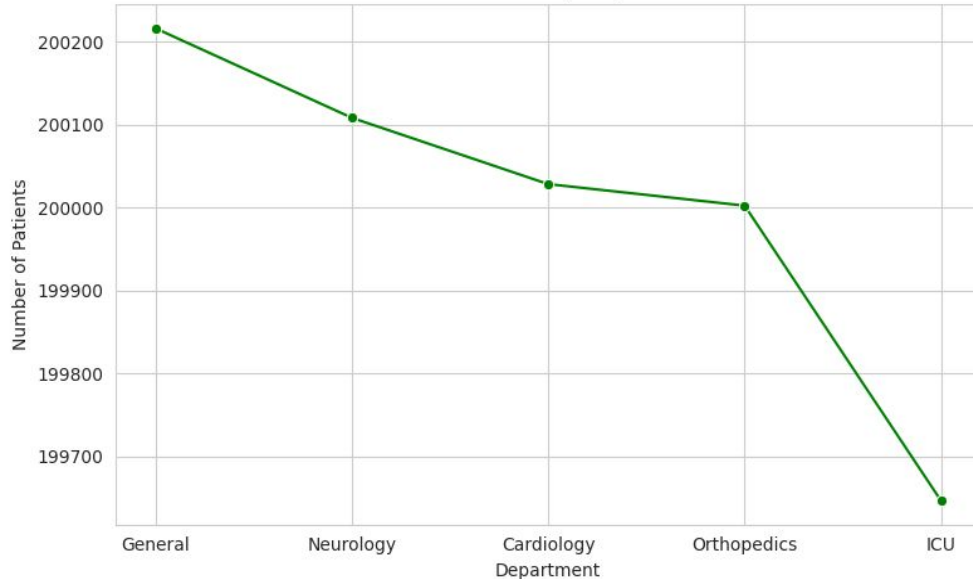
## Top performing doctor by patient volume



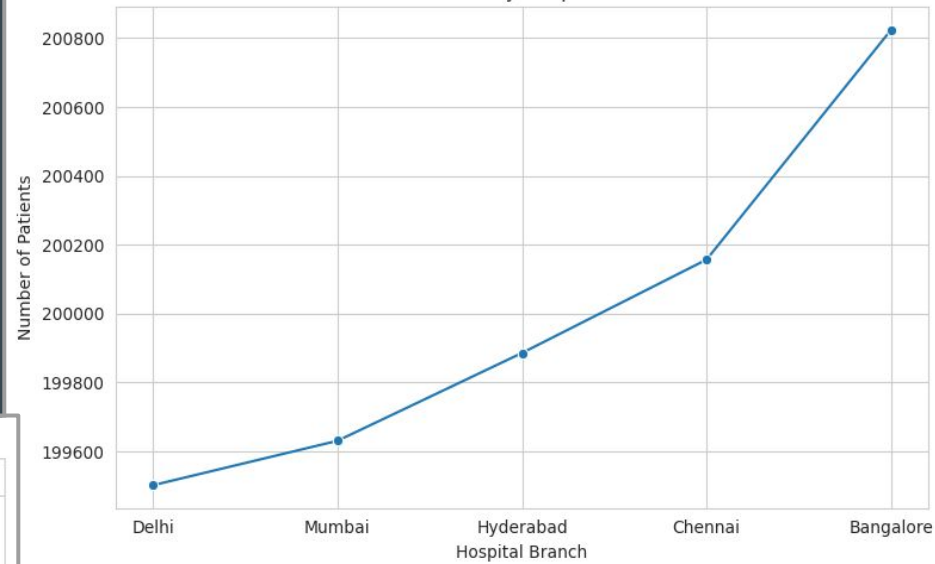


Branch with highest  
patient load

Patient Volume by Department

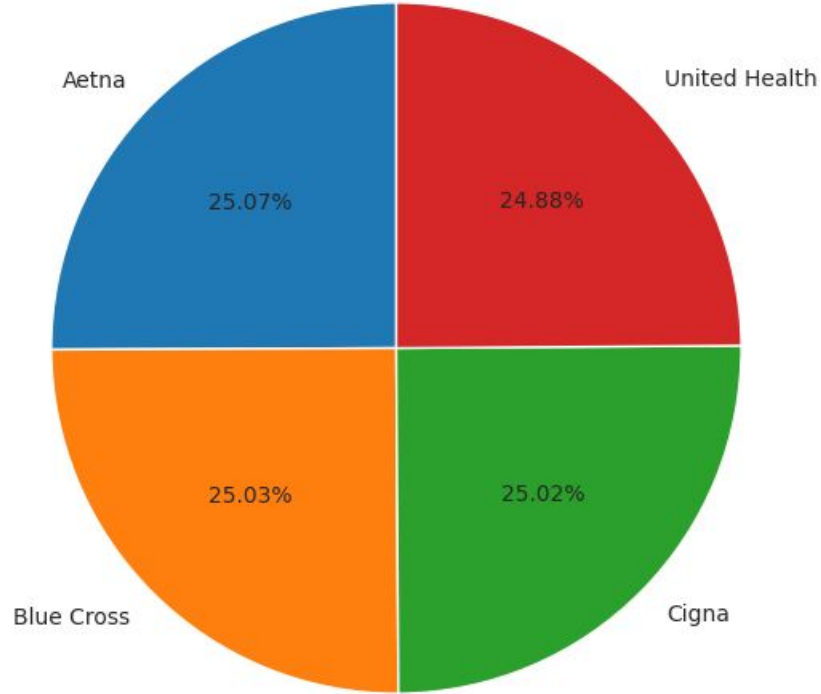


Patient Load by Hospital Branch



Patients handled by each  
department

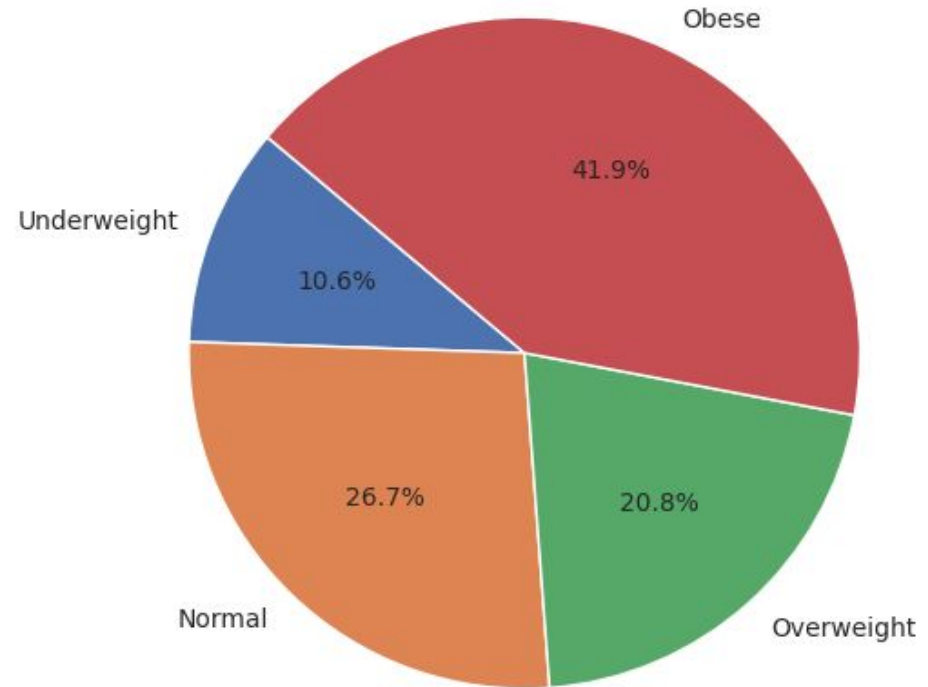
Insurance Providers (In percentage %)



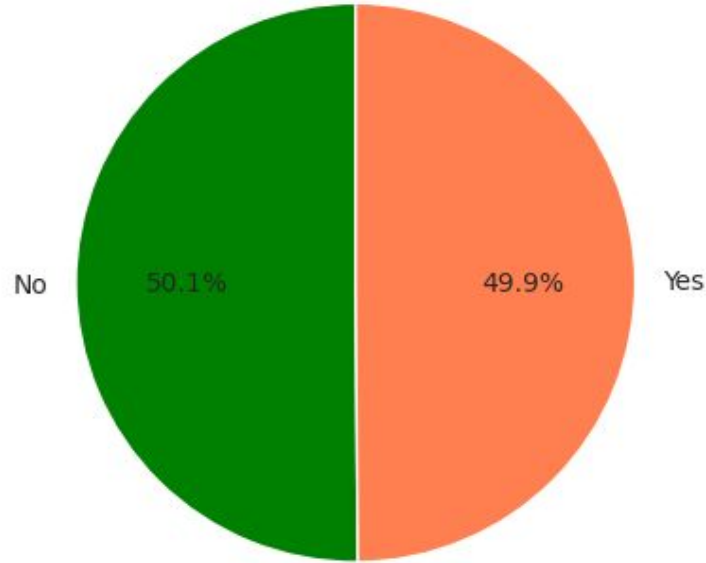
Top insurance provider

## BMI trend among patients

BMI Distribution Categories

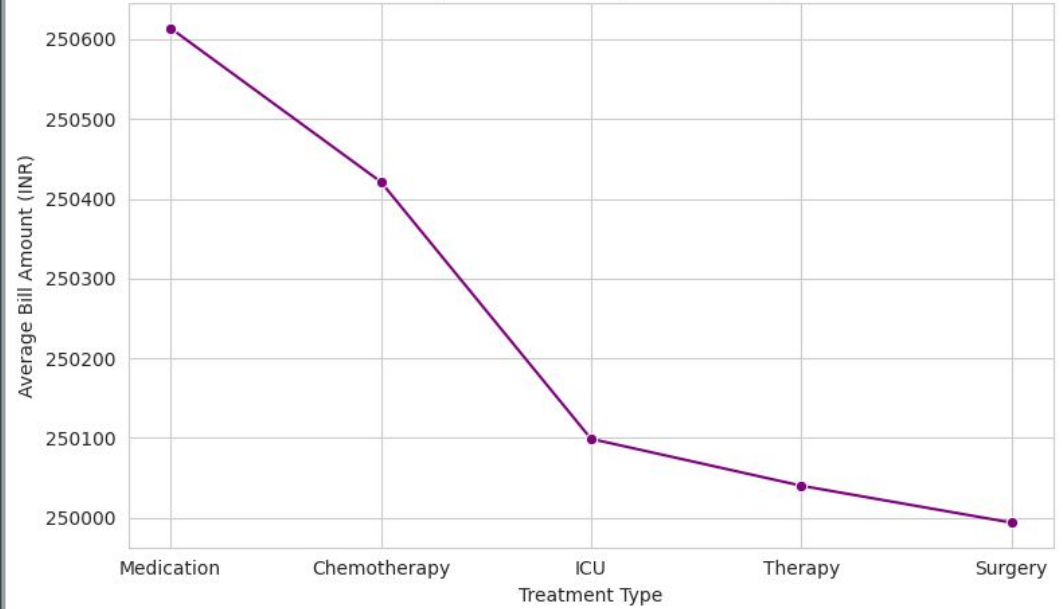


Percentage of Patients Who Smoke



Patients who smoke

Average Bill Amount by Treatment Type



Average bill amount by  
treatment type

### Correlation Between BMI and Diagnosis

- Diabetes, Covid-19
- Average BMI range from 27.9 to 28.04

### Average Length of Stay

- 182 days

### Correlation Between BMI and Bill Amount

- No relationship

### Room Occupancy by Branch

- Bangalore: 223 patients / room
- Delhi has lowest occupancy

THANK YOU