Healthcare Data Analysis Case Study

Project Title

Healthcare Data Analysis Using Patient-Level Dataset

Project Overview

This case study presents a **Healthcare Data Analysis project** performed using a structured dataset containing **7,157 patient records**. The analysis focuses on understanding **patient trends, disease patterns, hospital stay behaviour, billing metrics, and follow-up patterns** through exploratory data analysis and visualization.

All details included in this document are **strictly derived from the provided dataset, Python analysis, visualizations, and SQL queries**. No assumptions, external benchmarks, or additional insights beyond the source file have been introduced.

Objective of the Analysis

The objectives of this project were to:

- Analyse patient admission and discharge trends over time
- Study disease-wise and bed occupancy patterns
- Examine billing and health insurance amounts across different categories
- Evaluate patient stay duration and follow-up behaviour
- Summarize healthcare performance metrics using SQL queries

Dataset Summary

- **Total Records:** 7,157 patients
- **Time Period Covered:** December 2022 to March 2024
- **Key Columns Used:**
    - Patient_ID
    - Admit_Date
    - Discharge_Date

- Diagnosis
- Bed Occupancy (General, ICU, Private)
- Test
- Doctor
- Follow-up Date
- Feedback
- Billing Amount
- Health Insurance Amount
- Month, Year

The dataset contained **no missing values** across the core analytical columns.

Tools & Technologies Used

- **Python**
  - **Pandas for data cleaning, grouping, and aggregation**
  - **NumPy for numerical calculations**
  - **Matplotlib for visualizations**
  - **Seaborn for statistical and comparative visualizations**
- **SQL**
  - **Used for patient counts, billing summaries, rankings, percentages, and window function analysis**
- **Microsoft Excel**
  - **Used for tabular summaries and quick validation of aggregated results**
- **Power BI**
  - **Used for creating interactive dashboards and visual representations based on analyzed data**

Data Preparation & Feature Engineering

- Converted admission and discharge dates to datetime format
- Derived patient stay duration using date differences

- Extracted month and year from admission dates
- Removed unused columns before aggregation

Exploratory Data Analysis Performed

1. Patient Admission Trends

- Month-wise patient counts were calculated using admission dates
- Year-wise patient distribution was visualized using aggregated counts

2. Disease & Diagnosis Analysis

- Analysed patient distribution across diagnosis categories
- Compared diagnosis types with associated tests and billing amounts

3. Bed Occupancy Analysis

- Studied patient distribution across General, ICU, and Private beds
- Calculated total billing amounts per bed occupancy type

4. Billing & Insurance Analysis

- Analysed total billing amounts by diagnosis, bed type, and test
- Compared billing amounts with health insurance claim amounts using scatter plots
- Identified top patients based on insurance claim totals

5. Patient Stay Duration Analysis

- Calculated total hospital stay duration per patient
- Visualized patients with longer cumulative stay durations

6. Doctor & Feedback Analysis

- Analysed feedback ratings associated with doctors
- Examined doctor-wise patient handling using SQL queries

7. Follow-Up Analysis

- Identified follow-up date patterns
- Counted follow-ups scheduled per patient and per month

SQL-Based Analysis

The project includes SQL queries to:

- Count total number of patients
- Calculate billing totals by diagnosis and bed type
- Compute average feedback ratings per doctor
- Identify ICU-only patients
- Rank doctors using window functions
- Generate pivot-style diagnosis vs bed occupancy summaries

Strategy, Decision-Making, and Outcomes

Based on the analysis outputs and visual patterns:

- Disease-wise and bed occupancy summaries support **clinical resource planning**
- Billing and insurance comparisons support **financial monitoring and revenue assessment**
- Patient stay duration analysis supports **hospital capacity and discharge planning**
- Doctor and feedback analysis supports **performance review and service quality monitoring**

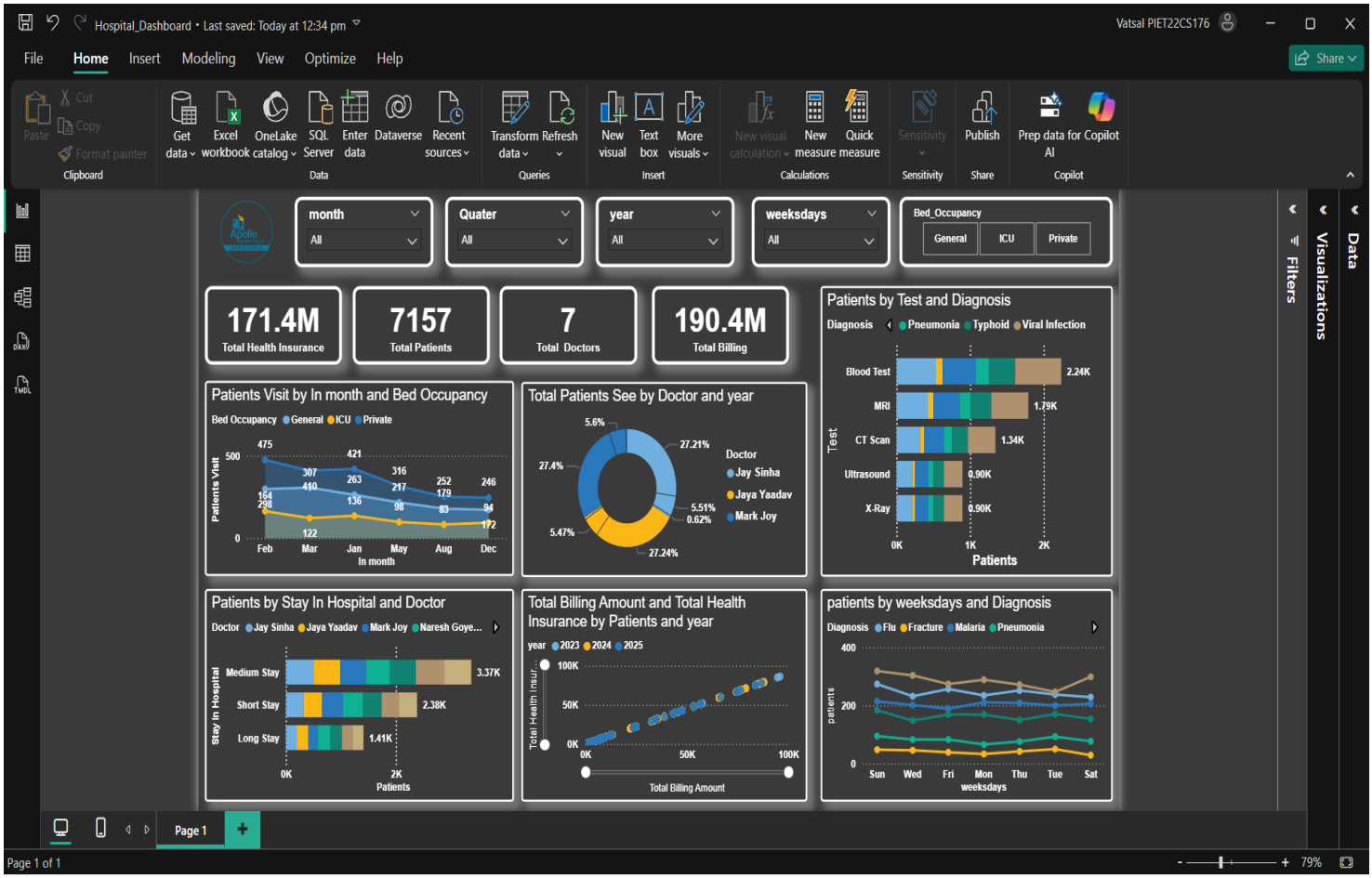
These outcomes are directly aligned with the computed metrics, aggregations, and visualizations in the project.

Resume-Ready Project Summary

Healthcare Data Analysis Project | Python, Pandas, NumPy, Matplotlib, SQL

- Analysed 7,157 patient records to study disease patterns, billing metrics, and hospital stay behaviour
- Performed data cleaning, feature engineering, and exploratory data analysis using Python
- Built visualizations for bed occupancy, billing trends, and patient admissions
- Used SQL queries and window functions to extract healthcare performance metrics

---



Conclusion

This project demonstrates hands-on experience in healthcare data analysis, combining Python-based EDA and SQL analytics to derive structured insights from real-world patient data.