Healthcare Patient Data Analysis

Prepared By: HITESH ANANTH.UC

Role: Data Analyst

Tools & Technologies: Python, Excel, Pandas

Dataset: Healthcare Patient Dataset

Abstract:

➢ Healthcare organizations generate large volumes of patient data related to admissions, medical conditions, billing, and demographics. Analyzing this data is essential to improve patient care, operational efficiency, and financial planning.

➢ This project focuses on analyzing healthcare patient data to identify key patterns related to patient demographics, admission types, medical conditions, hospital stay duration, and billing behavior. Python was used for data preprocessing, and Excel was used to build an interactive dashboard.

➢ The outcomes of this analysis provide actionable insights that help hospitals understand patient trends, optimize resource allocation, and support data-driven healthcare decision-making.

Introduction:

➢ The healthcare industry relies heavily on data to improve clinical outcomes, manage hospital operations, and control costs. However, real-world healthcare data is often inconsistent, unstructured, and contains missing or invalid values, making data preprocessing a critical step before analysis.

➢ This project aims to analyze healthcare patient data by cleaning and transforming raw data, identifying key trends, and visualizing insights through an interactive dashboard. By combining Python-based data preprocessing with Excel-based visualization, the project delivers a clear and practical approach to healthcare data analysis.

Problem Statement:

Healthcare organizations face several challenges, including:

• Lack of clear visibility into patient demographics

• High emergency admission rates affecting hospital workload

• Variation in billing amounts across insurance providers

• Longer hospital stays for certain medical conditions

• Difficulty in analyzing large patient datasets without proper visualization

Without systematic data analysis, these issues remain unidentified. This project addresses these challenges by transforming raw healthcare data into meaningful insights using analytics and visualization.

Objectives of the Project:

The key objectives of this project are:

• To preprocess and clean healthcare patient data using Python

• To analyze patient demographics and admission patterns

• To study medical conditions and hospital stay duration

• To analyze billing trends across insurance providers

• To build an interactive Excel dashboard with KPIs and slicers

• To provide actionable insights for healthcare decision-making.

Dataset Description:

The dataset represents healthcare patient records collected from a hospital system.

Dataset Characteristics:

• Total records: Approximately **54,860 patients**

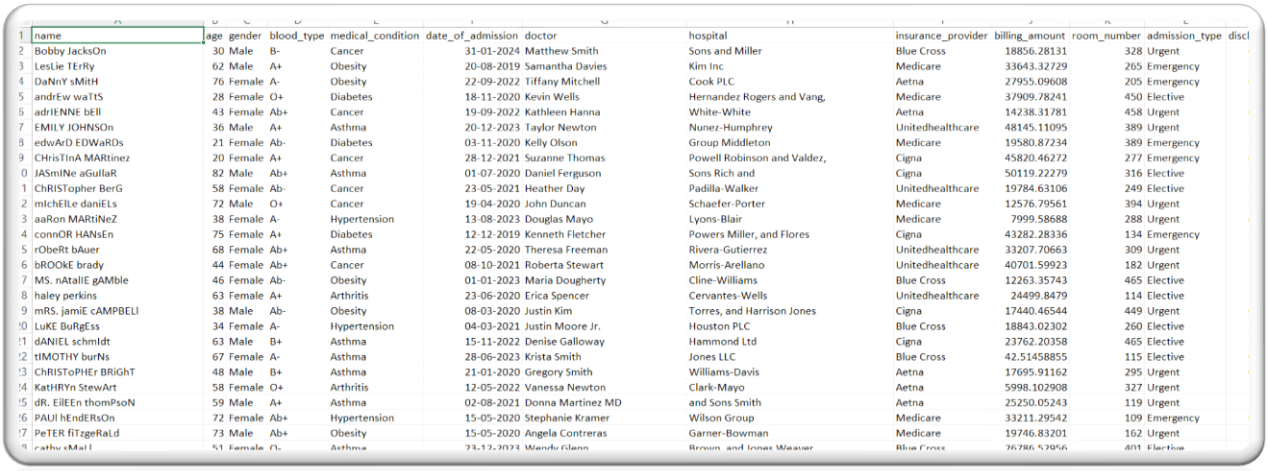
• Total columns: **15+** (including derived features)

Key Attributes:

• Age

• Gender

- Medical Condition
- Admission Type
- Date of Admission
- Discharge Date
- Length of Stay
- Billing Amount
- Insurance Provider



name	age	gender	blood_type	medical_condition	date_of_admission	doctor	hospital	insurance_provider	billing_amount	room_number	admission_type	discl
Bobby JacksOn	30	Male	B-	Cancer	31-01-2024	Matthew Smith	Sons and Miller	Blue Cross	18856.28131	328	Urgent	
LesLie TErRy	62	Male	A+	Obesity	20-08-2019	Samantha Davies	Kim Inc	Medicare	33643.32729	265	Emergency	
DaNnY sMitH	76	Female	A-	Obesity	22-09-2022	Tiffany Mitchell	Cook PLC	Aetna	27955.09608	205	Emergency	
andrEw waTtS	28	Female	O+	Diabetes	18-11-2020	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.78241	450	Elective	
adrIENNE bEll	43	Female	Ab+	Cancer	19-09-2022	Kathleen Hanna	White-White	Aetna	14238.31781	458	Urgent	
EMILY JOHNSOn	36	Male	A+	Asthma	20-12-2023	Taylor Newton	Nunez-Humphrey	Unitedhealthcare	48145.11095	389	Urgent	
edwArD EDWaRDs	21	Female	Ab-	Diabetes	03-11-2020	Kelly Olson	Group Middleton	Medicare	19580.87234	389	Emergency	
CHrisTInA MARtinez	20	Female	A+	Cancer	28-12-2021	Suzanne Thomas	Powell Robinson and Valdez,	Cigna	45820.46272	277	Emergency	
JASmINe aGuIlaR	82	Male	Ab+	Asthma	01-07-2020	Daniel Ferguson	Sons Rich and	Cigna	50119.22279	316	Elective	
ChRISTopher BerG	58	Female	Ab-	Cancer	23-05-2021	Heather Day	Padilla-Walker	Unitedhealthcare	19784.63106	249	Elective	
mIchElLe daniELs	72	Male	O+	Cancer	19-04-2020	John Duncan	Schaefer-Porter	Medicare	12576.79561	394	Urgent	
aaRon MARtiNeZ	38	Female	A-	Hypertension	13-08-2023	Douglas Mayo	Lyons-Blair	Medicare	7999.58688	288	Urgent	
connOR HANsEn	75	Female	A+	Diabetes	12-12-2019	Kenneth Fletcher	Powers Miller, and Flores	Cigna	43282.28336	134	Emergency	
rObeRt bAuer	68	Female	Ab+	Asthma	22-05-2020	Theresa Freeman	Rivera-Gutierrez	Unitedhealthcare	33207.70663	309	Urgent	
bROOkE brady	44	Female	Ab+	Cancer	08-10-2021	Roberta Stewart	Morris-Arellano	Unitedhealthcare	40701.59923	182	Urgent	
MS. nAtalIE gAMble	46	Female	Ab-	Obesity	01-01-2023	Maria Dougherty	Cline-Williams	Blue Cross	12263.35743	465	Elective	
haley perkins	63	Female	A+	Arthritis	23-06-2020	Erica Spencer	Cervantes-Wells	Unitedhealthcare	24499.8479	114	Elective	
mRS. jamiE cAMPBELl	38	Male	Ab-	Obesity	08-03-2020	Justin Kim	Torres, and Harrison Jones	Cigna	17440.46544	449	Urgent	
LuKE BuRgEss	34	Female	A-	Hypertension	04-03-2021	Justin Moore Jr.	Houston PLC	Blue Cross	18843.02302	260	Elective	
dANIEL schmIdt	63	Male	B+	Asthma	15-11-2022	Denise Galloway	Hammond Ltd	Cigna	23762.20358	465	Elective	
tIMOTHY burNs	67	Female	A-	Asthma	28-06-2023	Krista Smith	Jones LLC	Blue Cross	42.51458855	115	Elective	
ChRISToPHEr BRiGhT	48	Male	B+	Asthma	21-01-2020	Gregory Smith	Williams-Davis	Aetna	17695.91162	295	Urgent	
KatHRYn StewArt	58	Female	O+	Arthritis	12-05-2022	Vanessa Newton	Clark-Mayo	Aetna	5998.102908	327	Urgent	
dR. EilEEn thomPsoN	59	Male	A+	Asthma	02-08-2021	Donna Martinez MD	and Sons Smith	Aetna	25250.05243	119	Urgent	
PAUl hEndERsOn	72	Female	Ab+	Hypertension	15-05-2020	Stephanie Kramer	Wilson Group	Medicare	33211.29542	109	Emergency	
PeTER fiTzgeRaLd	73	Male	Ab+	Obesity	15-05-2020	Angela Contreras	Garner-Bowman	Medicare	19746.83201	162	Urgent	
cathy sMaLl	51	Female	O-	Asthma	23-12-2023	Wendy Glenn	Brown, and Jones Weaver	Blue Cross	26786.52956	401	Elective	

Tools and Technologies Used:

- **Python** – Primary tool for data preprocessing
- **Pandas** – Data manipulation and cleaning
- **Excel** – Dashboard creation and visualization
- **CSV** – Data storage and transfer format

Data Preprocessing Methodology:

Data preprocessing was performed using Python to ensure data accuracy and reliability.

Steps Performed:

1. **Data Loading and Inspection**
    - Imported the dataset using Pandas
    - Reviewed dataset structure using head() and info()
2. **Column Standardization**
    - Converted column names to lowercase
    - Removed extra spaces and replaced spaces with underscores
3. **Date Handling**
    - Converted admission and discharge dates into datetime format
4. **Text Data Cleaning**
    - Standardized categorical columns such as gender, admission type, and medical condition
5. **Data Validation**
    - Removed invalid age and billing values

- Eliminated duplicate patient records

6. **Feature Engineering**
    - Created **Length of Stay** feature
    - Grouped patient ages into meaningful **Age Groups**

After preprocessing, a clean and structured dataset was prepared for analysis and visualization.

Exploratory Data Analysis (EDA):

Exploratory analysis was conducted to identify key patterns in healthcare operations.

Key Areas of Analysis:

• Patient distribution by age group and gender

• Admission type analysis (Emergency, Urgent, Elective)

• Medical condition prevalence

• Average length of hospital stay by condition

• Billing trends across insurance providers

• Admission trends over time

EDA helped uncover patient behavior patterns and operational insights within the healthcare dataset.

Dashboard and Visualization (Excel):

An interactive Excel dashboard was developed to visualize key healthcare metrics.

Dashboard Components:

KPI Cards:

• Total Patients

• Average Billing Amount

• Average Length of Stay

• Average Patient Age

Visual Analysis:

• Patients by Age Group

• Patients by Gender

• Admission Type Distribution

• Patients by Medical Condition

• Average Length of Stay by Condition

• Average Billing by Insurance Provider

• Monthly Admission Trend

Interactive Slicers:

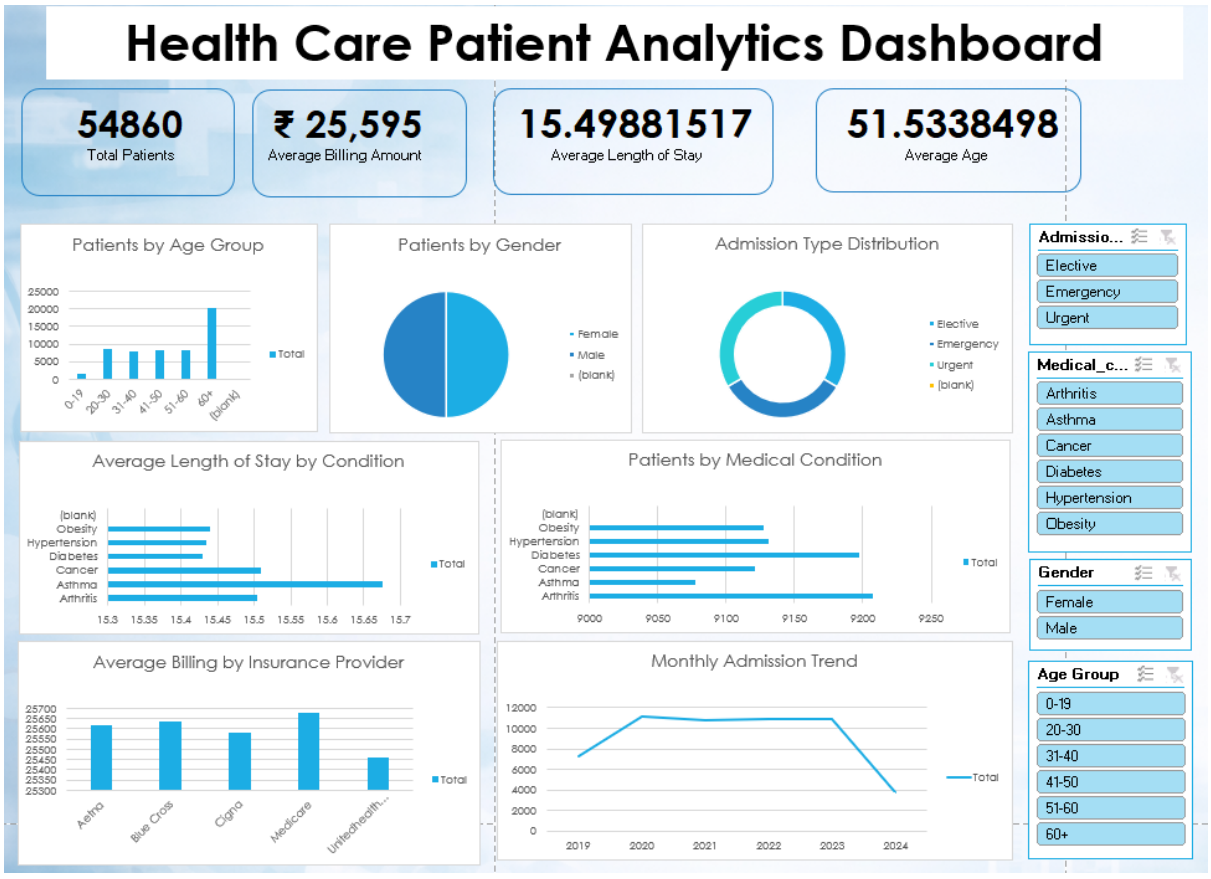
• Age Group

• Gender

• Admission Type

• Medical Condition

The dashboard allows dynamic filtering and real-time insight generation.



Key Findings:

- Senior and adult age groups account for the majority of hospital visits
- Emergency admissions form a significant portion of patient inflow
- Chronic conditions such as diabetes and arthritis show high patient counts
- Certain medical conditions result in longer hospital stays
- Billing amounts vary noticeably across insurance providers
- Patient admissions show time-based trends

Business Impact:

The insights generated from this project help healthcare organizations:

- Improve patient care planning
- Optimize hospital resource utilization
- Understand high-cost medical conditions
- Analyze insurance-based billing patterns
- Support data-driven operational decisions

Conclusion:

➢ This project demonstrates how healthcare data can be effectively cleaned, analyzed, and visualized using Python and Excel. By identifying key patient patterns and translating them into an interactive dashboard, the project provides clear insights into patient demographics, admissions, medical conditions, and billing behavior. The analysis supports better healthcare planning, operational efficiency, and informed decision-making.

Future Scope:

- Integration of real-time patient data
- Predictive analysis for hospital admissions
- Cost forecasting models
- Advanced dashboards using Power BI or SQL
- Automation of data pipelines

References:

- https://github.com/uchitesh8-web/HealthCare-Patient-Analysis