HEALTHCARE DATA ANALYSIS AND VISUALIZATION: Analyzing patient care, outcomes, and hospital costs using Power BI

PROJECT TITLE: HEALTHCARE DATA ANALYSIS AND VISUALIZATION

NAME: POOJA K

BATCH: DA & DS MAY -2025

ONLINE / OFFLINE: OFFLINE

TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| Sno | Content | Page no |
| 1. | Objective | 03 |
| 2. | Data overview | 03 |
| 3. | Data cleaning | 03 |
| 4. | Dashboard overview | 04 |
| 5. | DAX Queries | 06 |
| 6. | Key insight | 09 |
| 7. | Recommendation | 09 |
| 8. | Conclusion | 10 |

1. Objective :

Health data analytics helps in identifying trends, improving decision-making, and addressing healthcare challenges. This project, **Health Analytics Dashboard**, built in Power BI, transforms complex health data into interactive visualizations. The objective is to provide clear insights into health indicators, enabling professionals and policymakers to monitor trends, compare performance, and make informed decisions effectively.

1. Data overview:

The dataset used in the Health Analytics Dashboard provides key health-related data across various periods. It mainly tracks health conditions, reported cases, and mortality rates. By organizing the data by time and disease indicators it helps clearly show how health outcomes evolve over the years.

The data was imported into Power BI and processed through **Power Query** for cleaning and transformation. During this stage, missing values were handled, data types were corrected, and unnecessary fields were removed to ensure reliability. The final structured dataset supports accurate calculations and meaningful comparisons.

Key features of the dataset include:

* **Year** – Timeline of recorded health data.
* **Health Conditions** – Categories of diseases or disorders.
* **Cases** – Reported number of cases for each condition.
* **Mortality Rate** – Percentage of deaths associated with the condition.
* **Other Indicators** – Additional metrics relevant to the analysis.

By building relationships within the data model, these attributes were linked in a way that enables deeper exploration and dynamic reporting. This dataset serves as the foundation for all measures, KPIs, and visualizations within the dashboard.

1. Data Cleaning :

Before building the dashboard, the dataset underwent a thorough cleaning process in **Power Query** to ensure accuracy and consistency. Proper data preparation was essential to avoid errors in calculations and visualizations.

The following steps were applied:

1. **Handling Missing Values**
   * Identified and replaced blank or null entries in critical fields such as cases and mortality rates.
   * For non-critical fields, missing values were left as blank to avoid introducing bias.
2. **Removing Duplicates**
   * Checked for duplicate rows and removed repeated entries to maintain unique records.
3. **Correcting Data Types**
   * Ensured numerical fields (e.g., cases, mortality rate) were set as **Whole Number** or **Decimal**.
   * Verified that year was set as **Date/Integer** for accurate time-based analysis.
4. **Standardizing Column Names**
   * Renamed columns to meaningful labels (e.g., “Mortality (%)” instead of generic names) for clarity.
5. **Filtering Irrelevant Data**
   * Removed unnecessary columns and records not required for the analysis.
6. **Creating Derived Columns**
   * Added calculated columns (e.g., mortality percentage, total cases trend) to support better insights.

Through these steps, the dataset was transformed into a clean and structured format, ready for **data modeling, DAX calculations, and dashboard visualization** in Power BI.

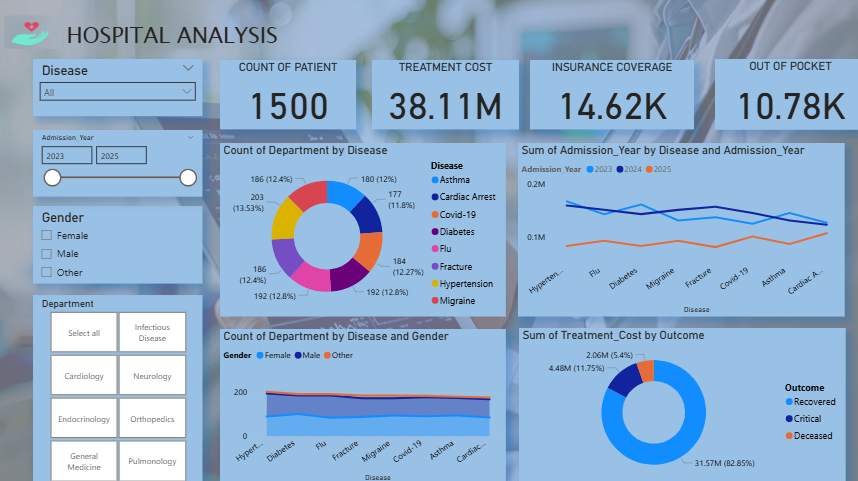
1. Dashboard overview:

The **Hospital Analysis Dashboard** provides a comprehensive view of patient admissions, treatment costs, insurance coverage, and departmental performance. It is designed to help healthcare administrators and analysts monitor hospital operations effectively.

### ****Key Highlights:****

1. **KPIs (Top Cards):**
   * **Count of Patients:** Total admitted patients (1500).
   * **Treatment Cost:** Total expenditure on treatments (38.11M).
   * **Insurance Coverage:** Total insurance-covered amount (14.62K).
   * **Out of Pocket:** Total amount paid directly by patients (10.78K).
2. **Filters & Slicers:**
   * Disease type, Admission Year (2023–2025), Gender, and Department.
   * These allow users to interact with the dashboard and analyze specific categories.
3. **Visualizations:**
   * **Donut Chart (Count of Department by Disease):** Shows the distribution of patients across diseases such as Asthma, Cardiac Arrest, Covid-19, Diabetes, Flu, Fracture, Hypertension, and Migraine.
   * **Line Chart (Admission Year by Disease):** Tracks the number of admissions over time for different diseases.
   * **Stacked Bar (Count of Department by Gender):** Displays patient distribution across diseases, segmented by gender (Female, Male, Other).
   * **Donut Chart (Treatment Cost by Outcome):** Shows treatment costs categorized by patient outcomes (Recovered, Critical, Deceased).
4. **Insights Enabled:**
   * Monitor patient counts by disease and department.
   * Understand cost distribution across treatment outcomes.
   * Identify trends in admissions over the years.
   * Compare disease prevalence across genders.

This interactive dashboard offers a **holistic view of hospital operations**, making it easier for decision-makers to track performance, allocate resources, and improve patient care outcomes.

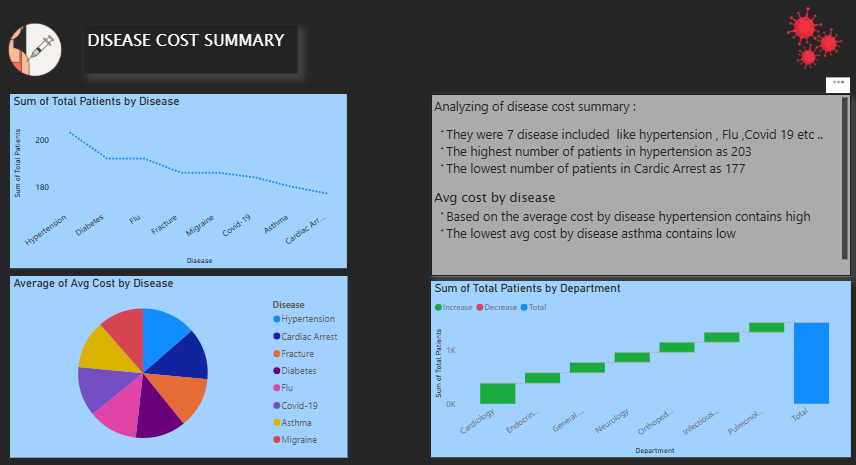


## **Disease Cost Summary – Dashboard Explanation**

The **Disease Cost Summary** page focuses on analyzing the number of patients and average treatment costs across different diseases and departments. It provides both patient distribution and cost-related insights.

### ****Visualizations & Insights****

1. **Line Chart – Sum of Total Patients by Disease**
   * Shows the patient count across different diseases.
   * **Hypertension** has the highest patient count (203).
   * **Cardiac Arrest** has the lowest patient count (177).
2. **Pie Chart – Average Cost by Disease**
   * Displays the average treatment cost per disease.
   * **Hypertension** shows the highest average cost.
   * **Asthma** has the lowest average cost.
3. **Waterfall Chart – Sum of Total Patients by Department**
   * Breaks down patient numbers across departments.
   * Highlights increases/decreases in patient counts per department.
   * Ends with the **Total patient count** across all departments.
4. **Text Box – Summary Insights**
   * Lists key findings for quick reference:
     + Total 7 major diseases considered (Hypertension, Flu, Covid-19, Diabetes, Fracture, Asthma, Migraine, Cardiac Arrest).
     + Highest patient count: **Hypertension (203)**.
     + Lowest patient count: **Cardiac Arrest (177)**.
     + Cost analysis shows **Hypertension** has the highest average cost, while **Asthma** has the lowest.



1. DAX Queries :

There are a few DAX queries performed in healthcare analysis to improve the analysis of the data based on measures, creating new columns and tables based on the dataset

1. **Total Patients**

Total\_Patients = COUNTROWS('Patient')

👉 This measure counts the total number of records in the **Patient** table, which represents the total admitted patients in the hospital.

1. **Total Cost**

Total\_Cost = SUM('Patient'[Treatment Cost])

👉 Sums up the values in the **Treatment Cost** column, giving the overall cost incurred for all patients.

1. **Average Stay**

Avg\_Stay = AVERAGE('Patient'[Length of Stay])

👉 Calculates the average length of hospital stay for patients, useful for analyzing patient recovery times.

1. **Recovery Rate %**

Recovery\_Rate = DIVIDE(

COUNTROWS(FILTER('Patient','Patient'[Outcome] = "Recovered")),

[Total\_Patients]

)

👉 Divides the number of patients whose outcome is **Recovered** by the total patient count. This shows the percentage of patients who recovered.

1. **Count of Age Group**

AgeGroup\_Count = DISTINCTCOUNT('Patient'[Age Group])

👉 Returns the number of distinct age groups in the dataset, which helps analyze patient distribution across different age categories.

1. **Mortality Rate %**

Mortality\_Rate = DIVIDE(

COUNTROWS(FILTER('Patient','Patient'[Outcome] = "Deceased")),

[Total\_Patients]

)

👉 Calculates the percentage of patients who died during treatment. It divides the count of **Deceased** outcomes by the total number of patients.

1. **Critical %**

Critical\_Rate = DIVIDE(

COUNTROWS(FILTER('Patient','Patient'[Outcome] = "Critical")),

[Total\_Patients]

)

👉 Measures the proportion of patients in **Critical condition** compared to the total patient count.

1. **Average Cost per Patient**

Avg\_Cost\_Per\_Patient = DIVIDE([Total\_Cost], [Total\_Patients])

👉 Divides the total treatment cost by the number of patients, showing how much cost on average is spent per patient.

1. **Total Insurance**

Total\_Insurance = SUM('Patient'[Insurance Coverage])

👉 Sums the **Insurance Coverage** column to calculate the total amount of treatment costs covered by insurance providers.

1. **Insurance Coverage %**

Insurance\_Coverage\_Percent = DIVIDE([Total\_Insurance], [Total\_Cost])

👉 Divides the **Total Insurance amount** by the **Total Cost** to calculate the percentage of hospital costs that were covered by insurance.

## **Disease Cost Summary – DAX Explanation (Table)**

**DAX Query** groups data by **Disease** and calculates:

* **Total Patients** → Number of patient records per disease.
* **Total Cost** → Sum of treatment costs.
* **Avg Cost** → Average treatment cost per patient.

**Insights:**

* **Hypertension** → Highest patients (203) & cost (~27.6K avg).
* **Cardiac Arrest** → Lowest patients (177).
* **Asthma** → Lower average cost (~24.4K).

This helps compare **disease-wise patient load and expenses** for better hospital cost management.

## **Patients by Department – DAX Explanation**

**DAX Query** groups data by **Department** and calculates:

* **Total Patients** → Number of patients in each department.
* **Avg Stay** → Average length of stay (days).

**Insights:**

* **Cardiology** → Highest patient count (380).
* **Endocrinology** → Shortest average stay (~6.45 days).
* **Pulmonology** → Longest stay (~7.88 days).

This analysis shows **patient load and hospitalization time** per department for resource planning.

1. Key insights :

 **Patient Overview**: 1,500 patients analyzed with an **average stay of ~7.3 days**.

 **Costs**: Total treatment cost ≈ **38.11M**, with an **average cost per patient ~25.4K**.

 **Recovery & Mortality**: Recovery rate is **84%**, mortality rate **5%**, critical cases **11%**.

 **Insurance**: Total insurance ≈ **21.94M** with **58% coverage**, showing partial financial support.

 **By Disease**:

* **Hypertension & Migraine** have high treatment costs.
* **Fracture** and **Cardiac Arrest** show higher average costs per patient.

 **By Department**:

* **Cardiology** has the **highest patient load (380)**.
* **Pulmonology** shows the **longest average stay (7.88 days)**.
* **Endocrinology** has the **shortest stay (6.45 days)**.

1. Recommendation:

 Improve recovery rate with better care & follow-up.

 Reduce average stay via faster diagnosis & treatment.

 Control high treatment costs (fracture, cardiac).

 Expand insurance coverage (>58%).

 Strengthen Cardiology (high demand).

 Optimize Pulmonology (long stays).

 Promote preventive care (diabetes, hypertension).

 Use digital health (EHR, AI, telemedicine).

 Train staff & allocate resources effectively.

1. Conclusion:

 Hospital serves **1500+ patients** across multiple departments.

 Avg. stay ~7 days; **recovery rate 84%**, mortality 5%.

 High costs in **Cardiology & Orthopedics** need control.

 **Insurance covers only 58%** of total expenses.

 **Pulmonology has longest stays** → needs process improvement.

 **Cardiology has highest patient load** → requires more resources.

 Preventive care & efficiency improvements are essential.

 **Digital health adoption** can boost accuracy & reduce delays.