Healthcare Exploratory Data Analysis (EDA)

Introduction

Healthcare data provides a crucial foundation for understanding patient demographics, medical conditions, and the effectiveness of healthcare services. This exploratory data analysis project focuses on a synthetic healthcare dataset that emulates real-world scenarios, containing various patient information, admission details, medical conditions, and healthcare services. The goal of this project is to explore, clean, and visualize the data to uncover insights into healthcare trends and medical distributions.

Objectives

1. **Data Cleaning and Preparation:** Identify and resolve inconsistencies or errors in the data.
2. **Exploratory Data Analysis:** Examine the dataset for trends, distributions, and relationships.
3. **Visualization:** Present data-driven insights using clear and intuitive charts.
4. **Actionable Insights:** Provide conclusions to guide healthcare decisions.

Dataset Overview

The dataset contains **10,000 rows** and **15 columns**, covering diverse aspects of healthcare:

- **Patient Information:** Age, gender, blood type.
- **Medical Details:** Medical condition, medication, test results.
- **Admission Details:** Admission type, days hospitalized, billing amount.
- **Healthcare Provider Details:** Doctor, hospital, insurance provider.

Key Dataset Attributes:

- Age ranges from **18 to 85**.
- Billing amounts vary between **$1,000 and $49,995**.
- No missing or duplicate records.
- Key categorical columns include **gender**, **blood type**, **medical condition**, and **insurance provider**.

Data Cleaning

- Verified absence of null values and duplicates.
- Converted Date of Admission and Discharge Date to datetime format.
- Added a new column Days Hospitalized based on admission and discharge dates.
- Removed irrelevant columns like Name and Room Number for focused analysis.

Exploratory Data Analysis (EDA)

Data Distribution

1. **Numerical Columns:**
    - Age, Billing Amount, and Days Hospitalized were analyzed using histograms.
    - Age showed a balanced distribution, with most patients between 18-30 years old.
    - Billing Amount showed no significant skewness, with a mean of ~$25,517.
2. **Categorical Columns:**

- **Gender:** Slightly more female patients (51%) than male.
- **Blood Type:** AB- was the most common blood type.
- **Medical Condition:** Asthma was the leading condition, followed by cancer and hypertension.
- **Insurance Providers:** Cigna topped the list, followed by Blue Cross.

Key Insights and Visualizations

Gender vs. Medical Conditions

- Female patients had higher counts in most medical conditions.
- **Top conditions for females:** Cancer, Asthma.
- **Top conditions for males:** Hypertension.

Correlation Analysis

- A slight positive correlation exists between age and days hospitalized.
- Billing amount negatively correlates with days hospitalized and test results.

Hospital and Doctor Billing

- **Top Hospital:** Smith and Sons generated the highest total billing (~$477,639).
- **Top Doctor:** Dr. Michael Johnson recorded the highest billing (~$181,576).

Billing and Gender Analysis

- For most medical conditions, males incurred higher average billing amounts than females.
- Diabetes generated the highest billing amount among females.

Age Group Analysis

- Most patients were aged **18-30**, followed by **71-80** and **51-60** age groups.

Conclusion

1. Female patients slightly outnumber males across most medical conditions.
2. Asthma is the most prevalent condition, especially among younger age groups.
3. Cigna dominates as the leading insurance provider.
4. Dr. Michael Johnson and Smith and Sons Hospital are the top performers in terms of billing.
5. Young adults (18-30) represent the largest patient group, suggesting the importance of preventive healthcare initiatives for this demographic.

Recommendations

1. **For Healthcare Providers:**
    - Focus on managing asthma and diabetes, as they represent significant health burdens.
    - Tailor preventive care programs for young adults and seniors.
2. **For Insurance Companies:**
    - Address the needs of conditions like asthma and hypertension, which have widespread impact.
3. **For Further Analysis:**
    - Explore the impact of medication on test results and recovery times.
    - Analyze regional trends in healthcare usage for targeted strategies.