Healthcare Dataset Analysis Report

- **Prepared by:** Anuja Mahajan
- **Guide by:** Vishal sir

Introduction:

In this dataset consist of 55,500 rows and & 15 column in this dataset.

Related to the healthcare including patient like age, gender , medical condition, admission & discharge biling & test result .

Executive Summary:

Key finding:

a comprehensive healthcare dataset containing patient information, medical conditions, medications, insurance coverage, and procedural data. Through various analyses and visualizations, the goal is to identify key trends and relationships within the data. This information can be used to better understand patient demographics, optimize healthcare delivery, and enhance insurance plans based on patient needs.

Conclusion*: The dataset provides valuable insights into healthcare trends, showing that middle-aged patients have the highest prevalence of medical conditions and all major insurance providers offer coverage for a wide range of conditions. The distribution of medical procedures (urgent, elective, and emergency) which indicates an equal need for various types of healthcare services. Blood types show minor correlations with specific conditions, but further research is needed to confirm these trends.*

Objective:

The primary objective of this analysis is to:

1. *Identify trends and patterns across medical conditions, age groups, and insurance providers.*
2. *Explore correlations between blood types and medical conditions.*
3. *Understand the distribution of healthcare services (e.g., emergency, urgent, elective) and assess if these services are linked to specific patient demographics.*
4. *Provide actionable insights and recommendations for healthcare providers and insurers.*

Data Overview:

Data Source:

The dataset appears to be sourced from a healthcare provider or consortium that collects patient-level information related to medical conditions, insurance coverage, and procedural types.

Data Description:

The dataset contains information on the following fields:

- ***Age Groups**: Categorized into < 30, 31 to 60, and > 61.*
- ***Blood Types**: Different blood types such as A+, O-, AB+, etc.*
- ***Medical Conditions**: Conditions like Arthritis, Asthma, Cancer, Diabetes, Hypertension, and Obesity.*
- ***Insurance Providers**: Providers such as Cigna, Blue Cross, Medicare, Aetna, and UnitedHealthcare.*
- ***Procedures:** Categorized as Emergency, Elective, and Urgent.*
- ***Medications**: Various medications linked to specific medical conditions.*

Methodology:

1. ***Data Cleaning and Preprocessing:***
    - *Checked for missing values and handled inconsistencies in the dataset.*
    - *Standardized medical condition labels and patient demographic information.*
2. ***Data Exploration and Visualization:***
    - *Created bar plots, pie charts, and heatmaps to analyze the relationship between insurance providers, blood types, age groups, and medical conditions.*
3. ***Bivariate Analysis:***
    - *Performed correlation analysis between blood types and medical conditions .*
    - *Analyzed the distribution of medications prescribed across age groups.*

4. ***Key Analysis:***
    - *Explored whether specific insurance providers cater to particular medical conditions.*
    - *Examined the variation of medical conditions across gender and age demographics.*

Analysis:

- ***Age Distribution****: Middle-aged individuals (31 to 60 years old) represent the largest patient group, followed by individuals aged > 61.*
- ***Insurance Providers vs Medical Conditions:*** *There is no distinct specialization by insurance providers for specific conditions, though slight variations were noted in the focus on conditions such as hypertension and diabetes.*
- ***Blood Type and Medical Conditions****: Some minor correlations between blood type and conditions like cancer and diabetes were observed, but no strong patterns emerged.*
- ***Gender Analysis:*** *Small differences were seen in condition prevalence between males and females, with some conditions more common in one gender.*
- ***Healthcare Services****: Emergency, elective, and urgent services were nearly equally distributed across the dataset.*

Key Findings:

1. ***Age and Medical Conditions****: The majority of patients fall in the 31-60 age range, highlighting a need for focused healthcare interventions for middle-aged individuals.*
2. ***Insurance Coverage****: All major insurance providers cover a wide range of medical conditions without strong specialization in any particular condition.*

3. ***Blood Type Correlation****: While there is some variation in conditions like cancer and diabetes among blood types, the correlation is weak.*
4. ***Healthcare Services:*** *The demand for elective, urgent, and emergency procedures is fairly balanced, indicating an equal need for all types of care.*

Recommendations:

Gender-specific Interventions*: Develop healthcare programs that address conditions more common in one gender, such as asthma in females or cancer in males.*

Targeted Healthcare Programs*: Focus on middle-aged patients as they represent the largest group with prevalent conditions.*

Insurance Plan Customization*: Consider tailoring insurance plans to cover chronic conditions like hypertension and diabetes, which are prevalent across all age groups.*

Blood Type Awareness*: While the correlation is weak, blood type-specific awareness campaigns for certain conditions might still be beneficial.*

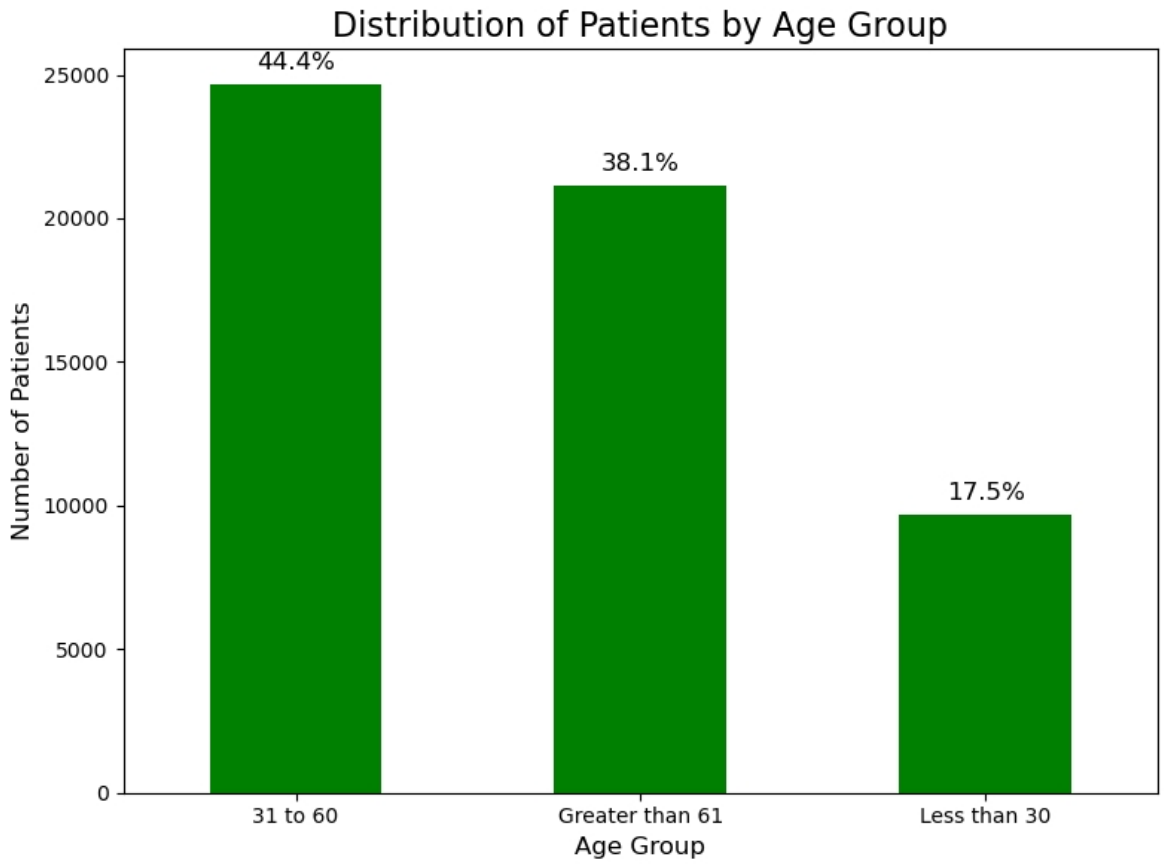
Distribustion of age:

Key finding :

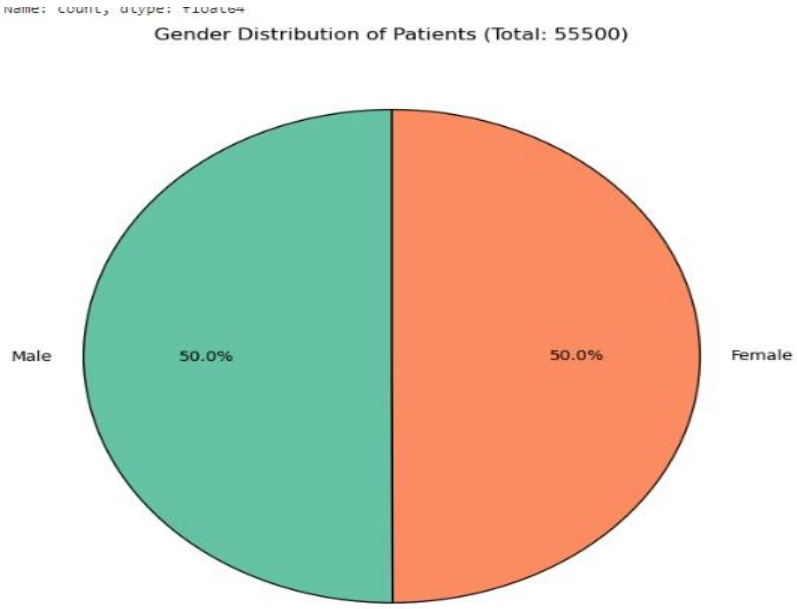
 The dataset reveals that the 31 to 60 age group represents the largest proportion of patients, followed by individuals aged > 61. In this middle-aged and

older adults are more frequently diagnosed with chronic conditions such as hypertension, diabetes, and arthritis.

 The < 30 age group has a smaller representation, indicating that younger individuals may either have fewer chronic medical conditions or may not seek medical care as frequently.



❑ Gender Distribution of patients



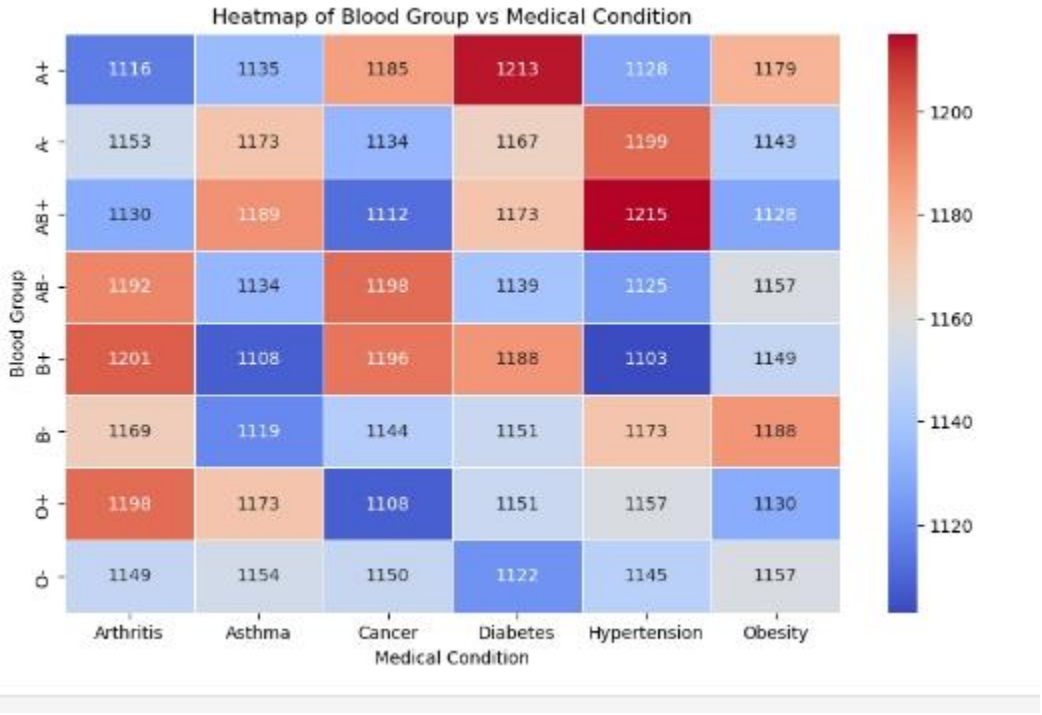
Key finding :

1.They are equally distributed The gender distribution of the dataset shows an **equal proportion** of male and female patients, with each gender comprising exactly **50%** of the total population (55,500 patients).

2. This balance in gender representation indicates that the medical conditions covered in the dataset are affecting both genders almost equally, providing

a well-rounded basis for further analysis without a gender bias.

➢ Blood Group vs Medical Condition

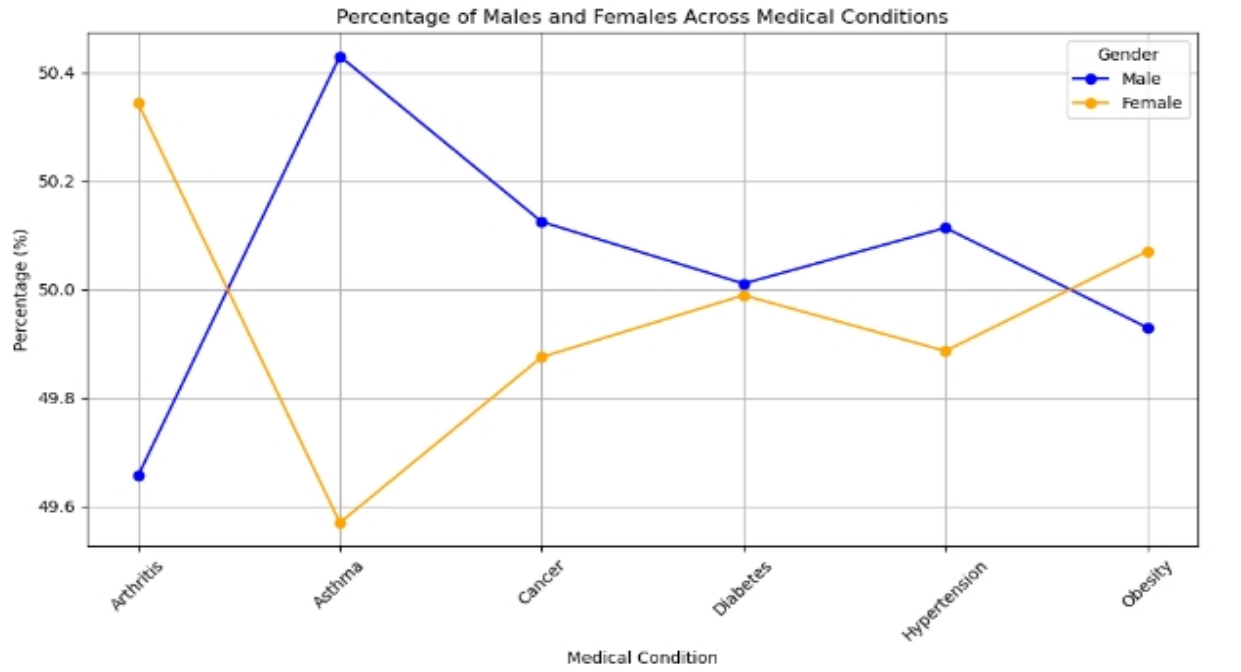


1. **Blood Type** :

- Blood groups show varying frequencies across different medical conditions.
- **AB+** has the highest number of patients for **Diabetes** (1215 cases).
- **A-** has a relatively low number of patients for **Cancer** (1134 cases), while **A+** has a slightly higher number (1185 cases).

2. **Condition Prevalence**:
   - **Hypertension** seems evenly distributed among different blood groups, with no extreme highs or lows.
   - **Obesity** is more frequent in certain blood types like **B-** and **O+**, compared to others like **A+**.
3. **Balanced Occurrence**:
   - Conditions like **Arthritis** and **Asthma** appear fairly balanced across blood groups, without significant disparities.

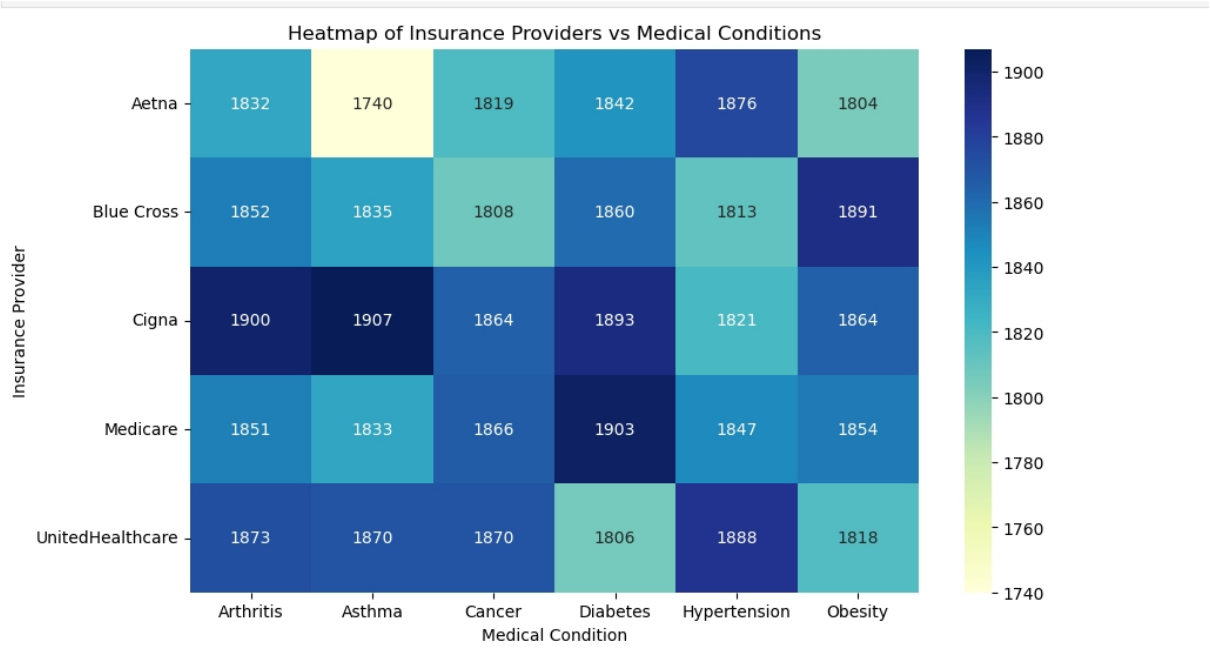
➢ Percentage of males and female in medical condition:



Gender Differences Across Conditions:

- **Asthma** has the highest difference between males and females, with significantly more males (over 50.4%) compared to females (below 49.6%).
- **Obesity** shows a relatively equal distribution between males and females, with a slight increase in the female percentage.

➢Insures provides vs medical condition



 Cigna: The highest numbers are associated with Cigna, especially for *Arthritis* (1900) and *Asthma* (1907). This suggests that Cigna might cover more claims or have a higher prevalence of patients with these conditions compared to others.

 Aetna: Shows a relatively lower number for *Asthma* (1740), which contrasts with higher numbers in *Hypertension* (1876) and *Obesity* (1804).

UnitedHealthcare: Has consistent values across conditions, but slightly higher for *Hypertension*

(1888), indicating a focus or a larger patient base in that category.

Medicare: Stands out with its highest number in *Diabetes* (1903), suggesting a larger impact or coverage for diabetic patients.

Variation by condition: Conditions like *Cancer* and *Asthma* have more variation in the numbers, showing different focuses or patient bases among insurance providers.

- **Cigna** and **UnitedHealthcare** seem to cater to slightly higher numbers of patients across most medical conditions, as indicated by the darker shades.