**Milestone 2**

**(a)**

We have planned to use a healthcare dataset from kaggle which has data similar to real-world medical industry data. Source is https://www.kaggle.com/datasets/prasad22/healthcare-dataset.In our dataset we have 10,000 entries with each row representing a patient’s healthcare record.

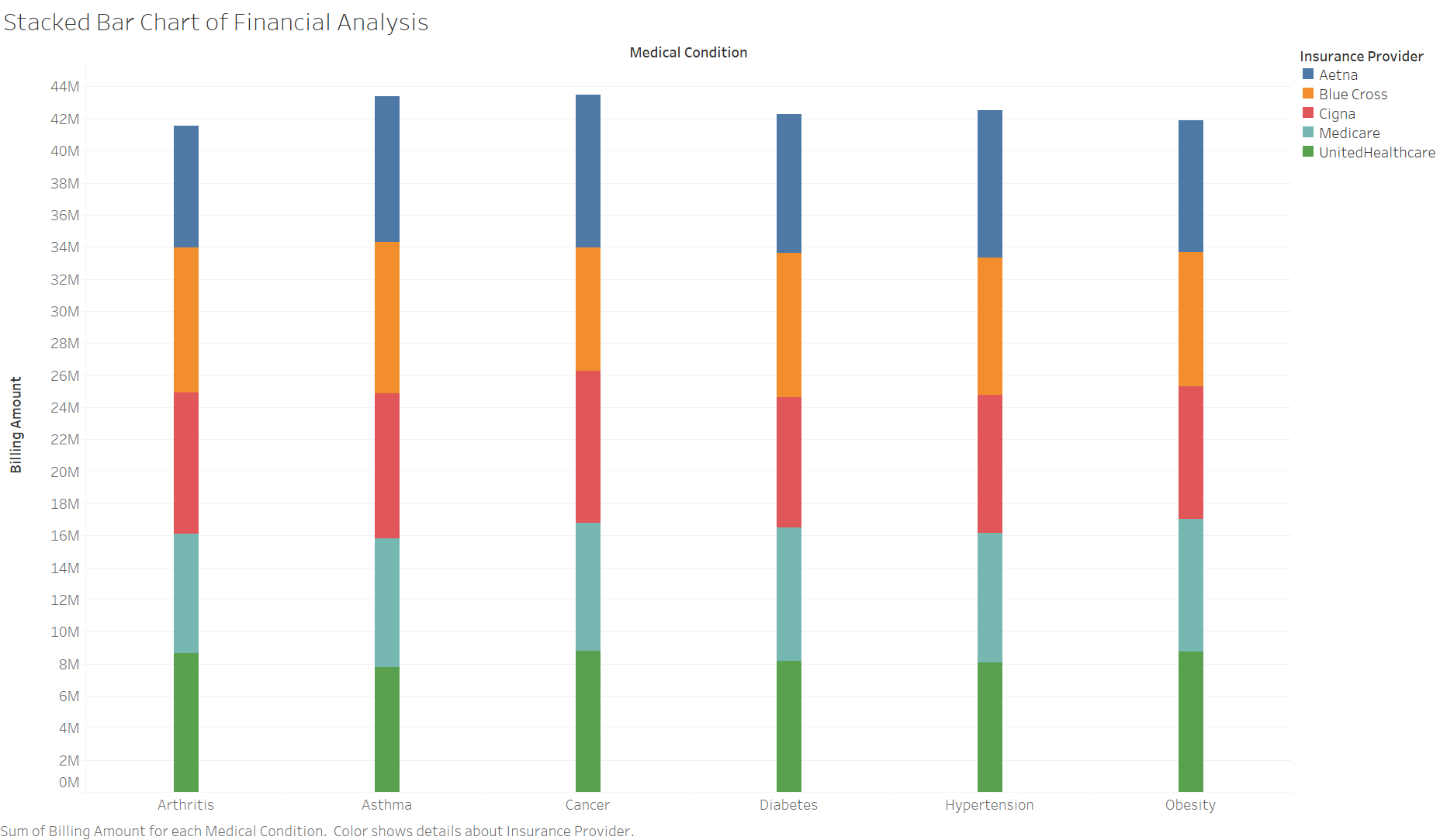
2 Numerical Variables- Age, Billing Amount.

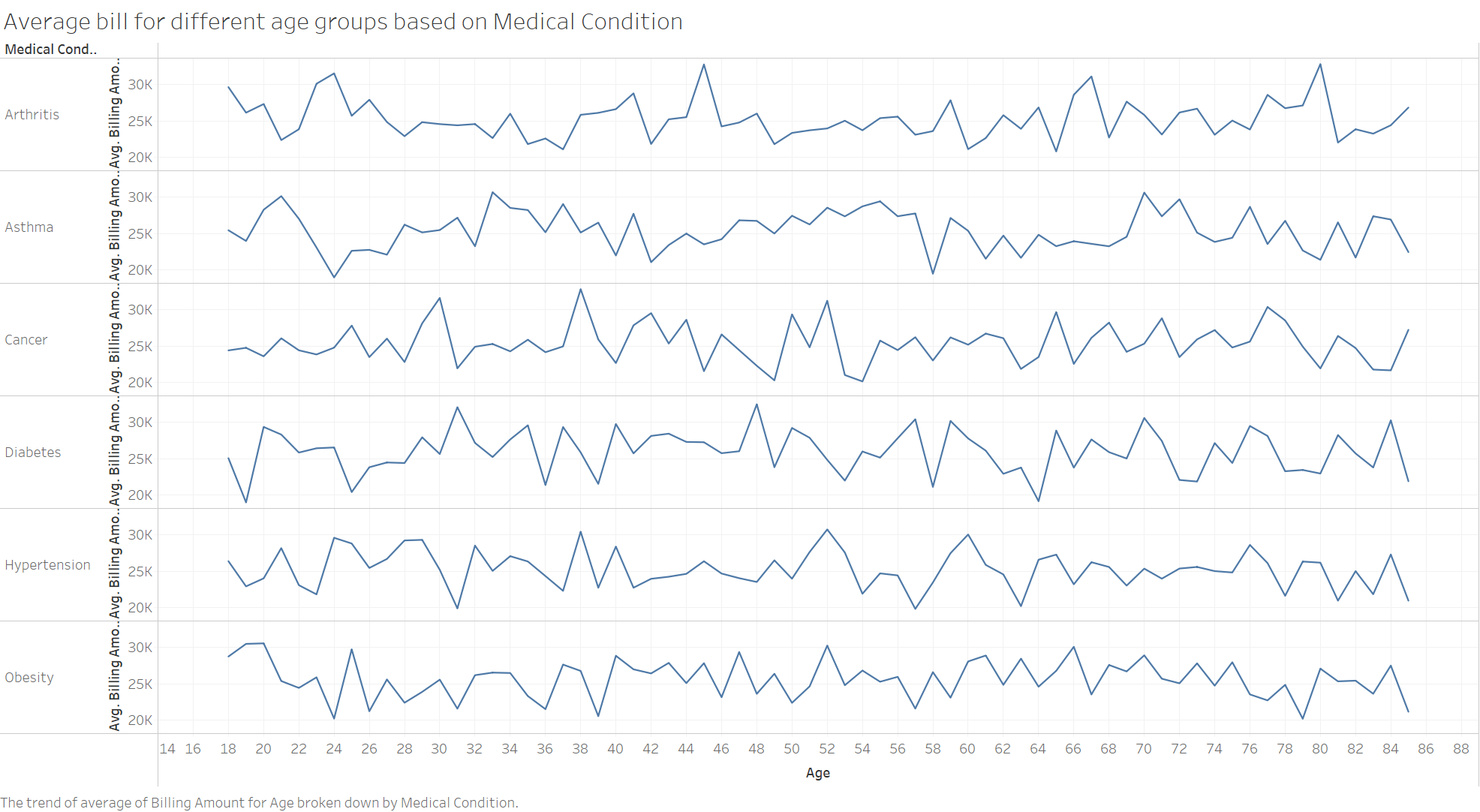
11 Categorical Variables- Name, Gender, Blood Type, Medical Condition, Doctor, Hospital, Insurance Provider, Admission Type, Medication, Test Results, Room Number.

2 Date Variables - Date of Admission, Discharge Date.

**(b)**

**Direction 1 – Based on Financial (Billing Related)**





**Direction 2 -Based on Medical Records (medical related)**

Age Distribution Bar Graph:

A graph of a bar chart

Description automatically generated with medium confidence

Explanation:

The above bar graph shows how many people are in different age groups. It covers ages from 15 to 85 years old. The highest number of patients are around 60, indicating a significant presence in this particular age group. And there are fewer patients around the age of 15. Overall, the graph helps us see which age groups have more people and which ones have fewer in our dataset.

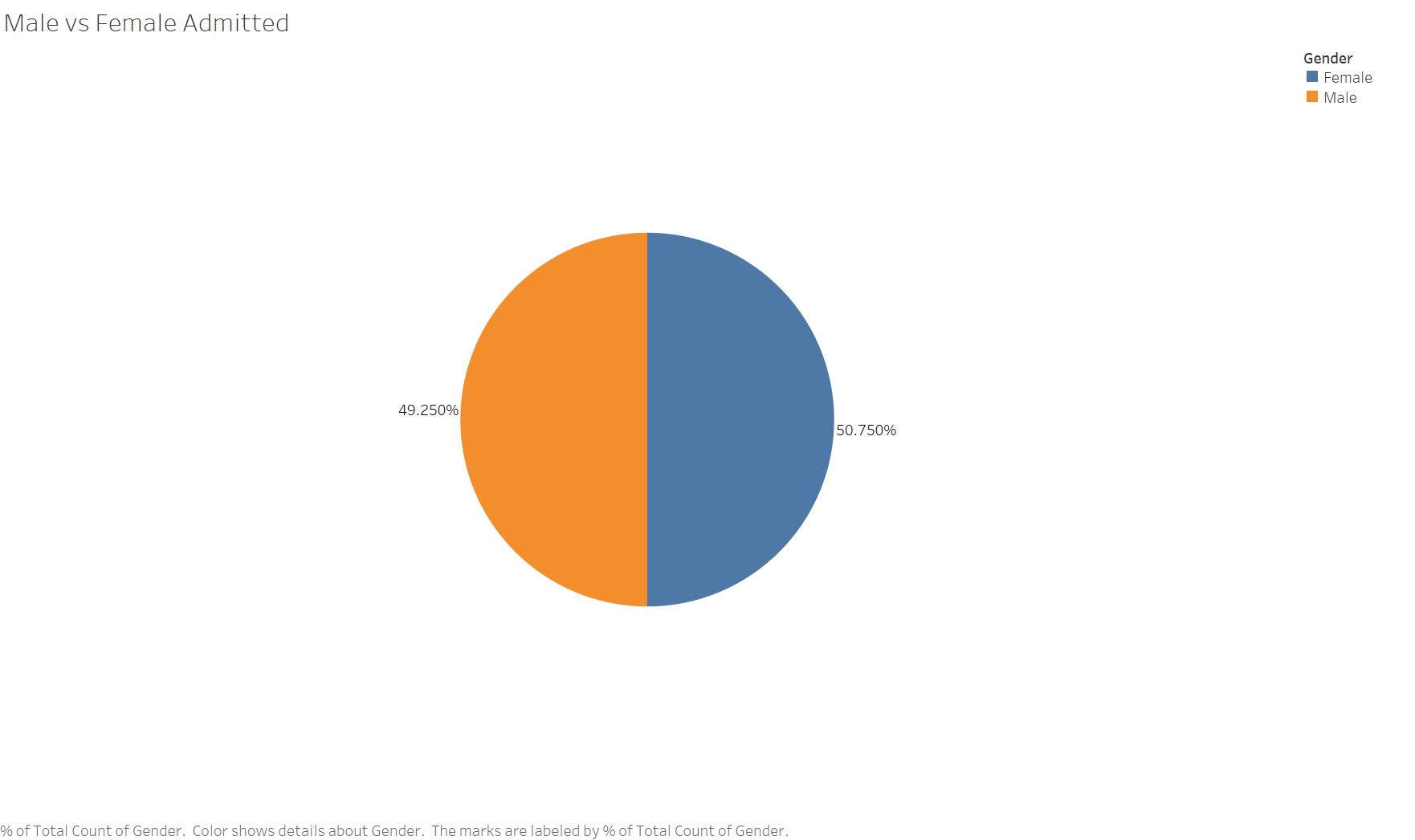
Distribution of Medical Condition:

A graph of different colored bars

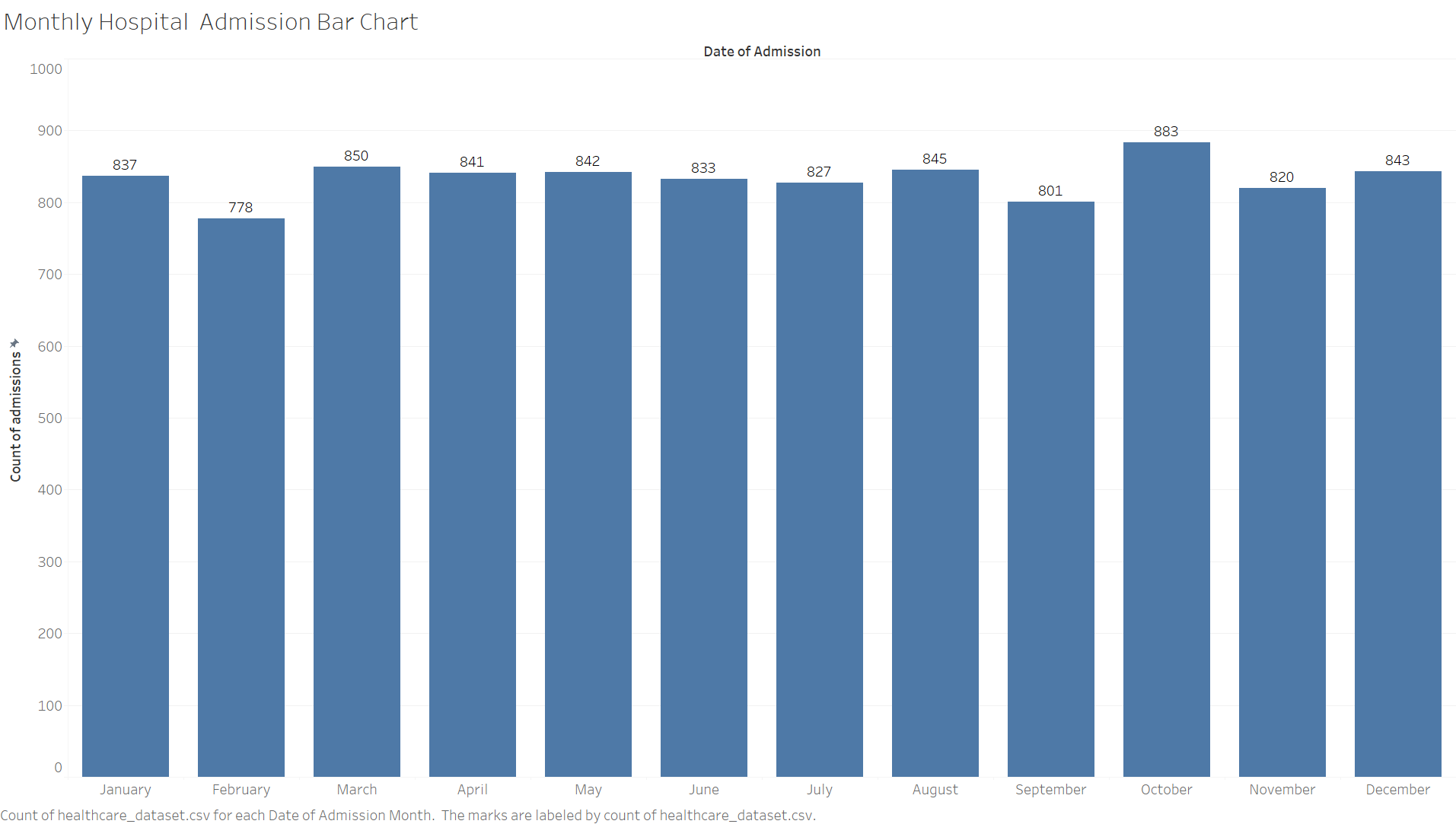
Description automatically generated with medium confidence

**Description:** The above bar graph provides a visual representation of the distribution of each medical condition within the dataset. From the analysis, it is noticeable that asthma, cancer, and hypertension emerge as the most frequent medical conditions among the patients, suggesting a significant proportion of individuals are affected by these health issues.

Gender Pie chart



**Direction 3 – Time Series (Date of Admission, Date of Discharge)**



A blue line graph with numbers

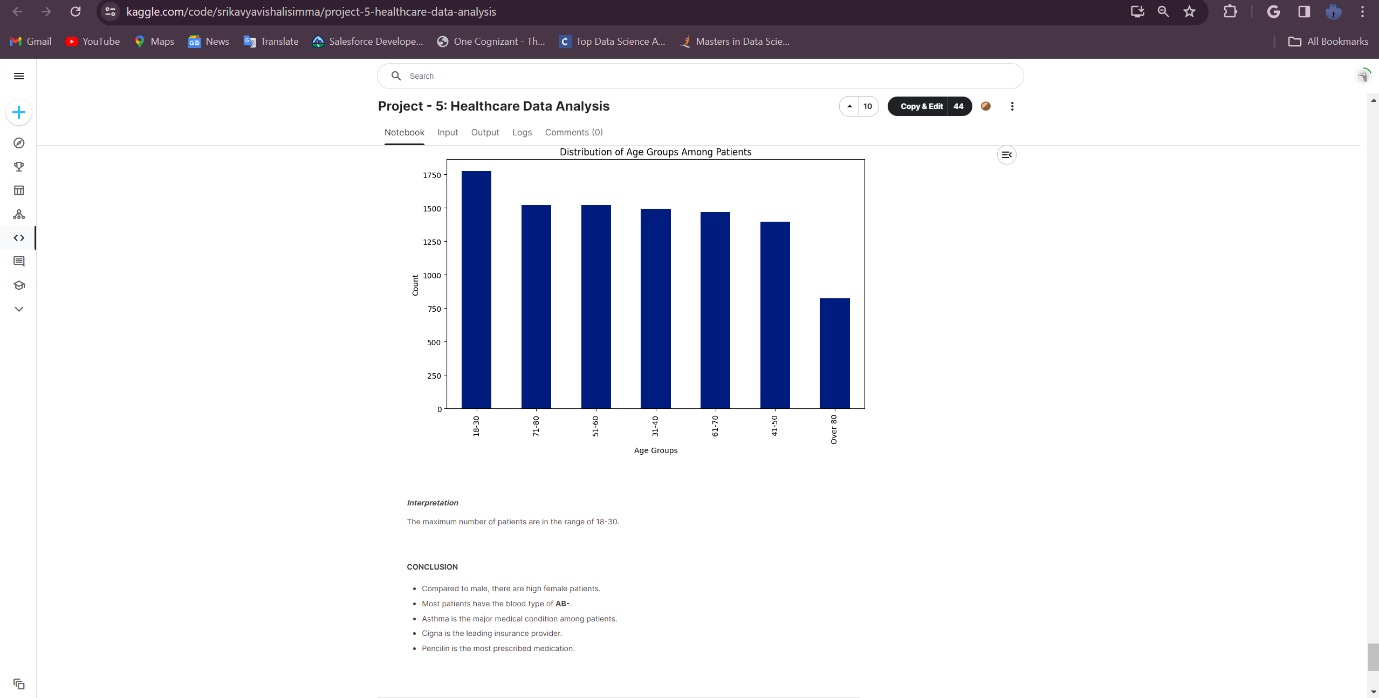
Description automatically generated

Explanation:

The above graph visualizes the number of patients being admitted at different times of the years.

**(c)**

We obtained our dataset from Kaggle. Most of the projects conducted using our dataset have visualized it using counts and have identified the maximum and minimum values of the variables through simple bar charts and pie charts to show the distribution. We have not taken any inspirations so far and we plan to use various visualization techniques to reveal more insightful information, leading us to better stories and outcomes for the data.

****

This is an example of how the data is visualized in a project based on age groups admitted and the conclusion is just the Maximum and Minimum.