Healthcare Analysis Report

Database & Big Data Analytics

Prof. Orazio Tomarchio

Usama ALI **1000053161**
S M Khizar ALAM **1000055853**
Ameer Hamza Iftikhar **1000056053**

Contents

Introduction

The healthcare industry generates vast amounts of data that hold valuable insights for improving patient care, optimizing operations, and making informed decisions. To harness the power of this data, healthcare organizations are increasingly turning to Business Intelligence (BI) dashboards and analytics reports.

In this analytics report, we will explore the application of BI dashboards in the healthcare industry. We will leverage a comprehensive dataset to extract meaningful information, visualize trends, and provide actionable insights. The report aims to empower healthcare professionals, administrators, and decision-makers with the tools and knowledge to drive data-informed strategies and improve patient outcomes.

Our healthcare BI dashboards and analytics report will focus on several key areas:

1. Patient Demographics and Characteristics: We will analyze patient data to gain insights into demographics, including age, gender, and location. Understanding the patient population will help identify target groups, address specific healthcare needs, and tailor services accordingly.
2. Medical Condition Analysis: Which are the most prevalent medical conditions within the dataset? Are there any correlations between specific conditions and patient demographics or treatments?
3. Financial Analysis: Financial data plays a crucial role in healthcare decision-making. We will explore revenue trends, costs, reimbursement patterns, and insurance claims to identify opportunities for cost savings, revenue growth, and efficient resource allocation.

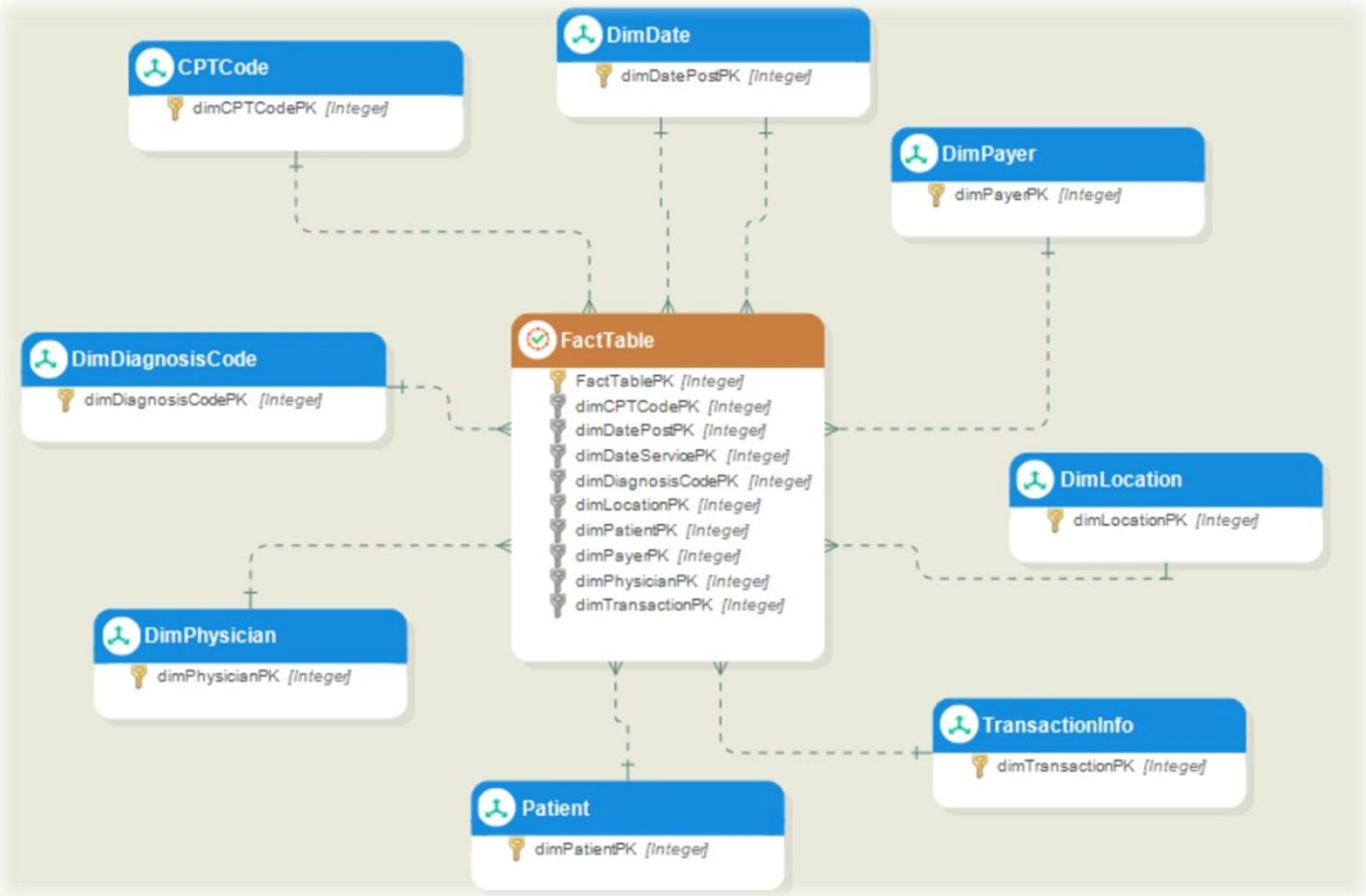
By leveraging BI dashboards and analytics tools, this report aims to transform complex healthcare data into actionable insights. The visualizations and analysis presented will enable healthcare organizations to make data-driven decisions, enhance patient care, optimize operations, and drive overall improvement in the delivery of healthcare services.

Throughout the report, we will present key findings, visual representations, and recommendations based on the data analysis. The goal is to provide stakeholders with a comprehensive understanding of the healthcare landscape, empowering them to take informed actions and drive positive change.

Please note that the specific findings, visualizations, and recommendations will be tailored to the dataset and the objectives of the healthcare organization.

Introducing the Dataset

The healthcare dataset https://www.kaggle.com/datasets/tomaslui/healthcare-dataset available on Kaggle provides comprehensive information on various aspects of healthcare. It includes data on patients, physicians, diagnoses, treatments, payments, and more. The dataset contains multiple files, such as FactTable.csv, which likely serves as the central table for storing key information related to patient visits, services rendered, and financial transactions. Here is the ER-Diagram for the dataset.



1. **dimCPT:**

CPT (Current Procedural Terminology) codes are a standardized system of medical codes used to describe and report healthcare procedures and services performed by healthcare providers. CPT codes are developed and maintained by the American Medical Association (AMA). These codes help ensure accurate and consistent communication of medical procedures for billing, insurance claims, and statistical analysis.

CPT codes are numeric and consist of five digits. Each code represents a specific medical procedure, service, or test.

CptCode	Five digit Current Procedural Terminology used to assign services rendered to the patient
CptDesc	Specific CPT description
CptGrouping	Groups the CPT codes into types of services

2. **dimPatient**:

The file seems to contain patient-related information as shown in the table below. This type of data can be useful for various purposes in healthcare, including patient management, communication, analysis, and reporting.

FirstName	First Name of the patient
LastName	Last Name of the patient
Email	Patients email address
PatientGender	Gender/Sex of the patient
PatientAge	Patients age
City	City that the patient lives in
State	State that the patient lives in

3. **Payer:**

The payer table used in a data is to provide a standardized and centralized reference for various payers involved in healthcare transactions. The dimPayerPK column serves as a unique identifier for each payer, while the PayerName column provides a human-readable representation of the payer's name or description.

dimPayerPK	Unique identifier that joins the payer table to the fact table
dimTransactionPK	Unique identifier that joins the transaction table to the fact table

4. **Physician**:

The physician table contains information about individual physicians, including their unique identifier (dimPhysicianPK), NPI number, name, specialty, and employment status (FTE). This data can be used for various purposes, such as physician management, scheduling, billing, reporting, and analysis in healthcare organizations

ProviderNpi	Unique number assigned to each physician upon completion of medical school
ProviderName	Physicians last name
ProviderSpecialty	Area which the physician has specialized
ProviderFTE	Physician time spent working on a weekly basis

5. Transaction:

The transaction table captures information about various transactions, including the transaction identifier (dimTransactionPK), the type of transaction (TransactionType), specific details of the transaction (Transaction), and any adjustment reasons (AdjustmentReason) associated with the transaction. This data can be used for purposes such as financial analysis, auditing, reconciliation, or reporting in various industries, including healthcare, finance, and retail.

dimTransactionPK	Unique identifier that joins the transaction table to the fact table
TransactionType	Clarifies whether the transaction was a charge, payment, adjustment, etc.
Transaction	Specific transaction descriptions
AdjustmentReason	Groups adjustments into operational reason groups

6. DateDim:

The date dimension or date table provides a comprehensive set of attributes related to dates in the context of transactions or services. These attributes can be used for organizing, analyzing, and reporting data based on various temporal aspects, such as year, month, day, and day of the week. This information can be useful for time-based analysis, trend analysis, and aggregations within the billing system or service data.

dimDatePostPK	Text formatted date that specifies the day the transaction was posted to the billing system
Date	Date format - either date service was rendered or date of a transaction
Year	Year - either date service was rendered or date of a transaction
Month	Month - either date service was rendered or date of a transaction
MonthPeriod	YYYYMM Text Format - either date service was rendered or date of a transaction
MonthYear	Month and Year - either date service was rendered or date of a transaction
Day	Day of Month Number - either date service was rendered or date of a transaction
DayName	Day of week name - either date service was rendered or date of a transaction

7. Diagnosis:

The diagnosis code table contains information about specific diagnosis codes used to represent patients' diagnoses. Each diagnosis code is associated with a unique identifier (dimDiagnosisCodePK), a code itself (DiagnosisCode), a description explaining the code (DiagnosisCodeDescription), and a group or category to which it belongs (DiagnosisCodeGroup). This information is essential in healthcare settings for recording, tracking, and analyzing diagnoses, as well as for reporting, billing, and research purposes.

dimDiagnosisCodePK	Unique identifier that joins the patient table to the fact table
DiagnosisCode	Code assigned by the physician to specific the patients diagnosis
DiagnosisCodeDescription	Description to clarify the diagnosis code
DiagnosisCodeGroup	Group of Diagnosis Codes

8. Hospital Info:

The hospital table contains essential information about various hospitals within a healthcare system. It provides a snapshot of key attributes that help identify and differentiate hospitals.

dimLocationPK	Unique identifier that joins the location table to the fact table
LocationName	Field that specifics where the service was rendered

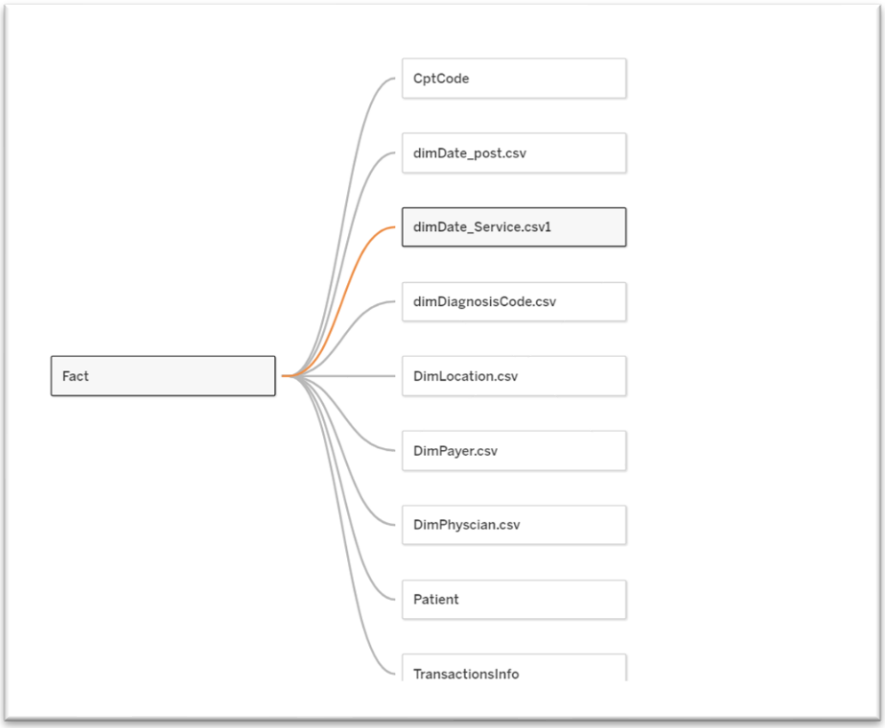
9. FactTable:

The fact table combines data from various dimensions (such as patient, physician, date, CPT code, payer, transaction, location, and diagnosis code) and includes measures (such as CPT units, gross charge, payment, adjustment, and accounts receivable) to provide a comprehensive view of the business process or event being analyzed. It serves as a foundation for reporting, analysis, and decision-making in healthcare organizations.

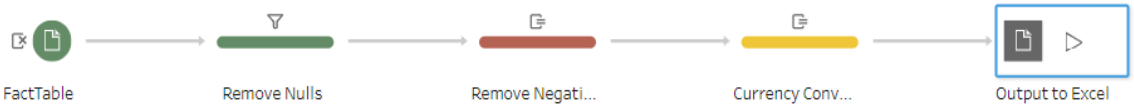
dimTransactionPK	Unique identifier that joins the transaction table to the fact table
dimLocationPK	Unique identifier that joins the location table to the fact table
PatientNumber	Unique Patient identifier. This is often times called a Medical Record Number (MRN)
dimDiagnosisCodePK	Unique identifier that joins the patient table to the fact table
CPTUnits	The number of times the CPT Code has been performed.
GrossCharge	The gross charge for the CPT Code. This includes the contractual adjustment
Payment	Payment received from payer and patient
Adjustment	Any dollars from the gross charge that will not be collected
AR	Outstanding accounts receivable that have yet to be collected or written off.

ETL Operations

We have observed some of the tables in this dataset have some columns that needs transformation and cleaning. For these ETL operations we have used Tableau Prep. The ETL pipelines for these tables are as follows:



Fact table:



In the above ETL pipeline we have performed the following transformations/ cleaning methods:

1. When performing data cleaning during the ETL process, replacing/removing null values is an important step to ensure data quality and accuracy. In this step we added a filter which removes the records from fact table which has two or more null values in the same row.

2. Next, we convert US dollars into Euros for the respected columns. This is achieved by multiplying the field with current conversion rate (0.91).

[GrossCharge]*0.91

[Payment]*0.91

[Adjustment]*0.91

[AR]*0.91

3. Next, we have observed that the values of specific columns are negative which are invalid, hence we convert these negative values into positive values using some expressions as follows:

   i. Cpt units
   IF [CPTUnits]<0 then [CPTUnits]*-1 ELSE CPTUnits END

   If the value in the [CPTUnits] field is less than 0, the condition multiplies it by -1 to convert it into a positive value. If the value is greater than or equal to 0, the condition leaves it unchanged. Similarly, we do this for all the required fields.

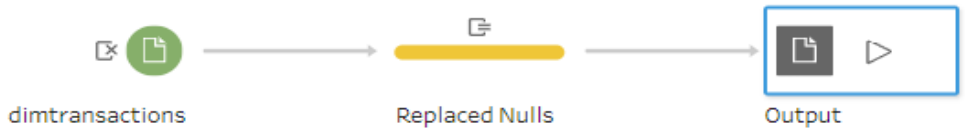
   ii. Gross charge
   IF [GrossCharge]<0 then [GrossCharge]*-1 ELSE [GrossCharge] END

   iii. Payment:
   IF [payment]<0 then [payment]*-1 ELSE [payment] END

   iv. Adjustment
   IF [adjustment]<0 then [adjustment]*-1 ELSE [adjustment] END

   v. AR
   IF [AR]<0 then [AR]*-1 ELSE [AR] END

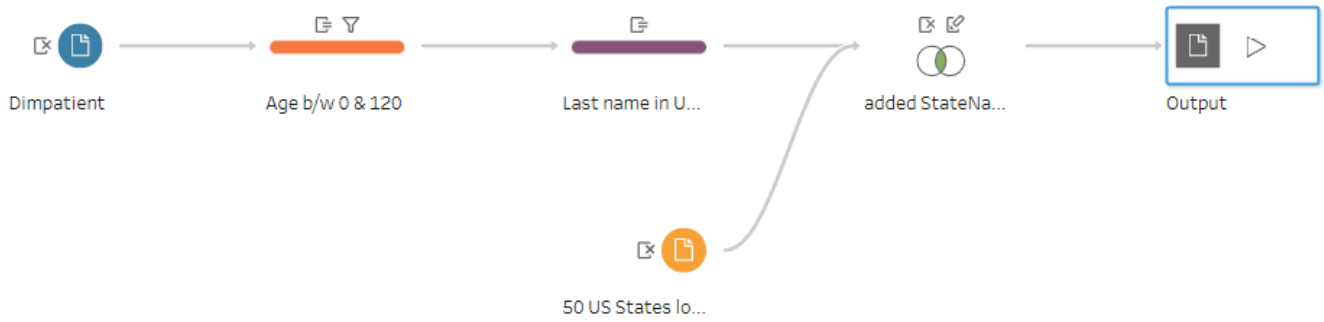
dimTransactions:



In the above ETL pipeline we have added an expression that replaces null values with hardcoded value “undefined”. We used the following Expression for this:

IFNULL([Transaction],"Undefined")

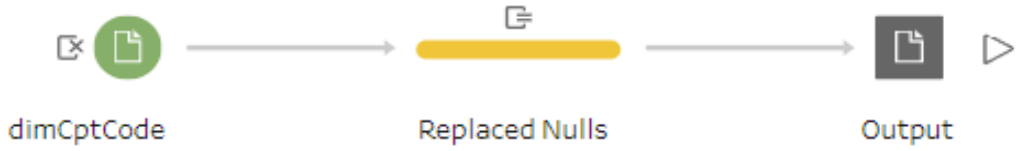
DimPatient:



In the above ETL pipeline we have performed the following transformations/ cleaning steps:

1. In this we have added filter/ rule that passes only those records which has age between 0 and 120.
2. We have added an expression that transform the “last name” of patient to upper case.
3. As we have given the abbrevited code of states of USA but not their full form as we need it to represent the world map in the dashboard. So for this we have join the patients table with the lookup table “50 US States” to get the required full form of the States.

DimCPTCode:



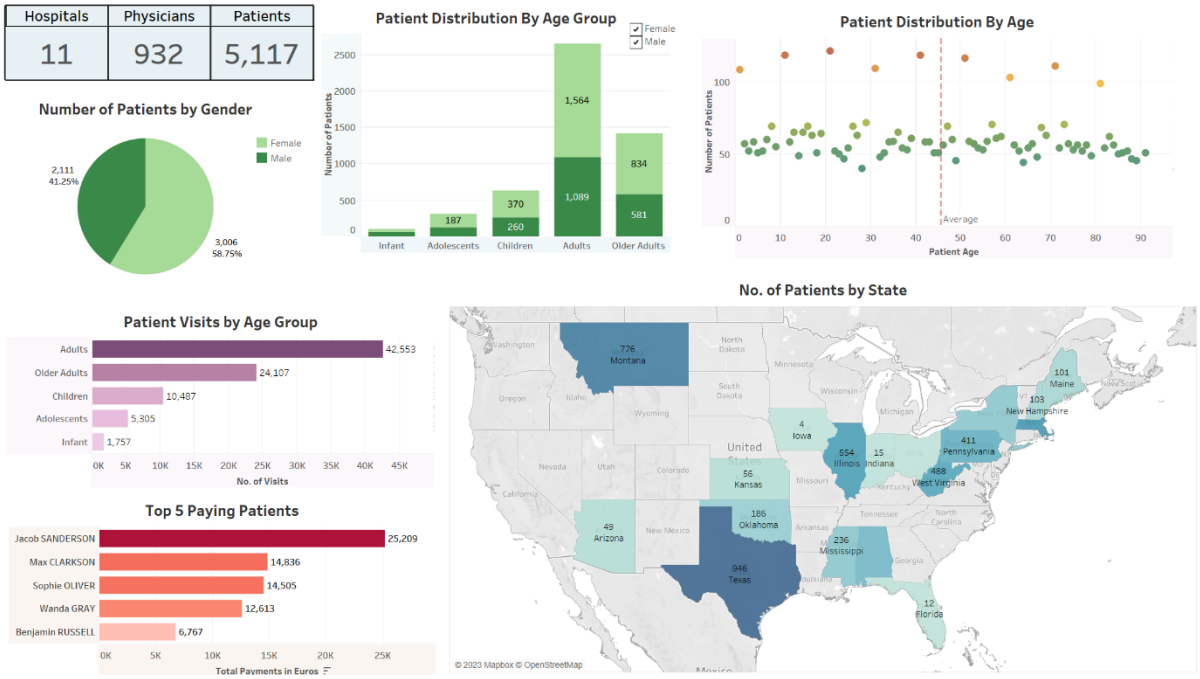
In the above ETL pipeline we added an expression that replaces null values with hardcoded value “undefined”. We used the following Expression for this:

IFNULL([CptDesc],"Undefined")

Dashboard 1 - Patient Demographics

This dashboard can answer the following question:

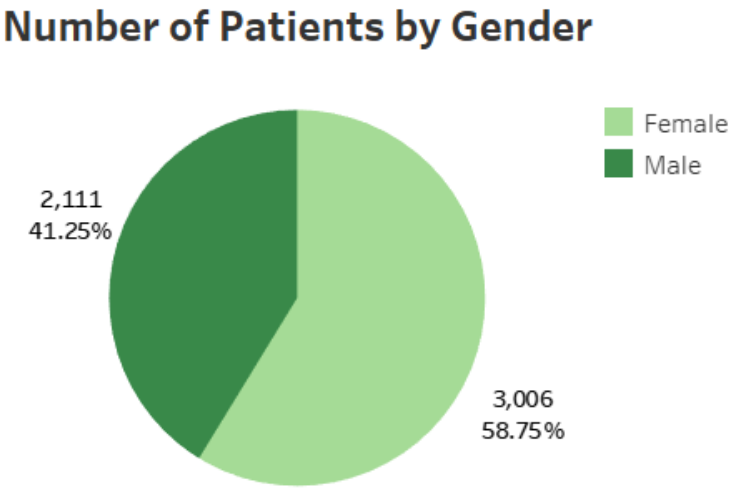
- The total number of physicians, hospital and patients?
- The distribution of male and female by their age group?
- Number of patients by their geographical location?
- Who are the top paying patients?



This dashboard specifically gives information about patients demographics. We will explain in detail:

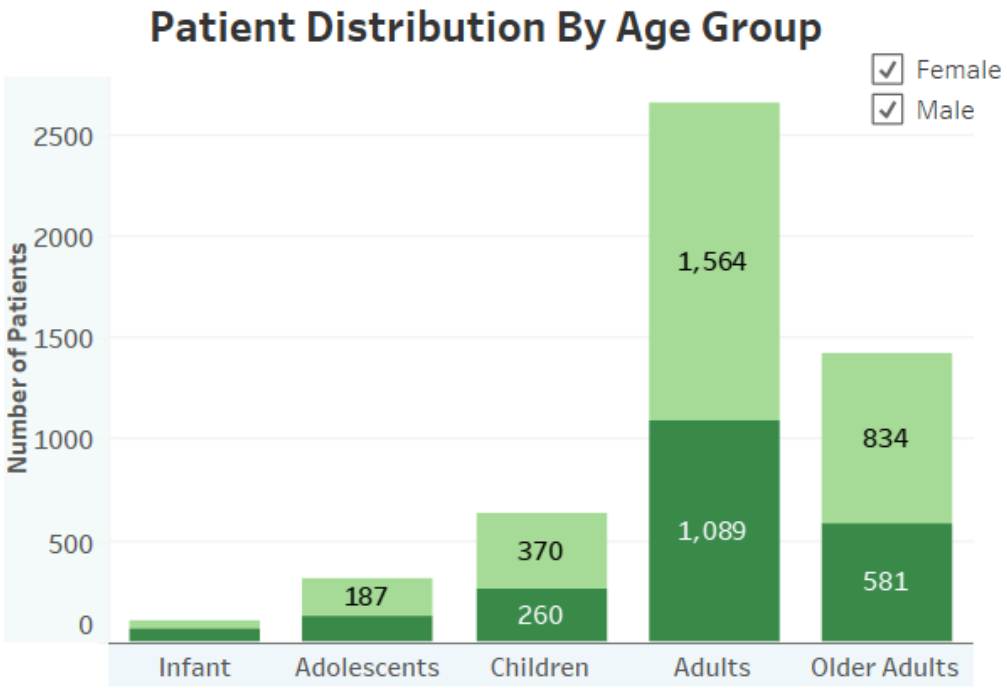
1. **Patients by gender:**

The pie chart can help us understand the relative proportions of male, female, within the patient population. By analysing the sizes of the slices, we can quickly identify which gender category has the largest or smallest representation. This indicates there are nearly 17% more male patients than female.



2. Patients by age group:

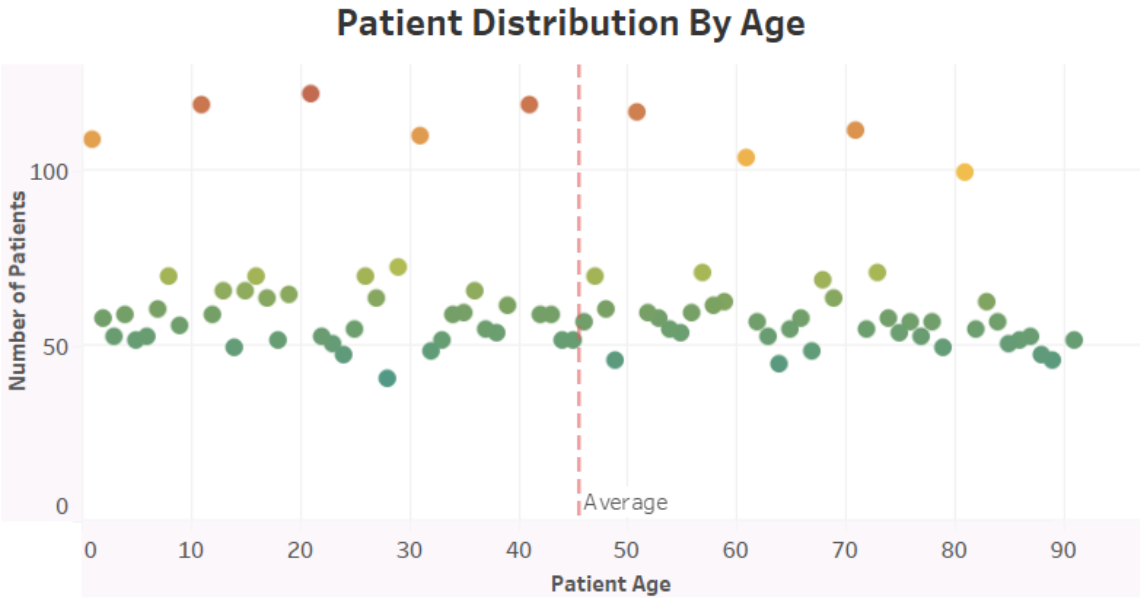
By examining the bar chart, we can easily identify the relative distribution of patients across different age groups. It provides insights into the age composition of the patient population, highlighting which age groups have a higher or lower representation.



3. Patients – Age Group & Gender:

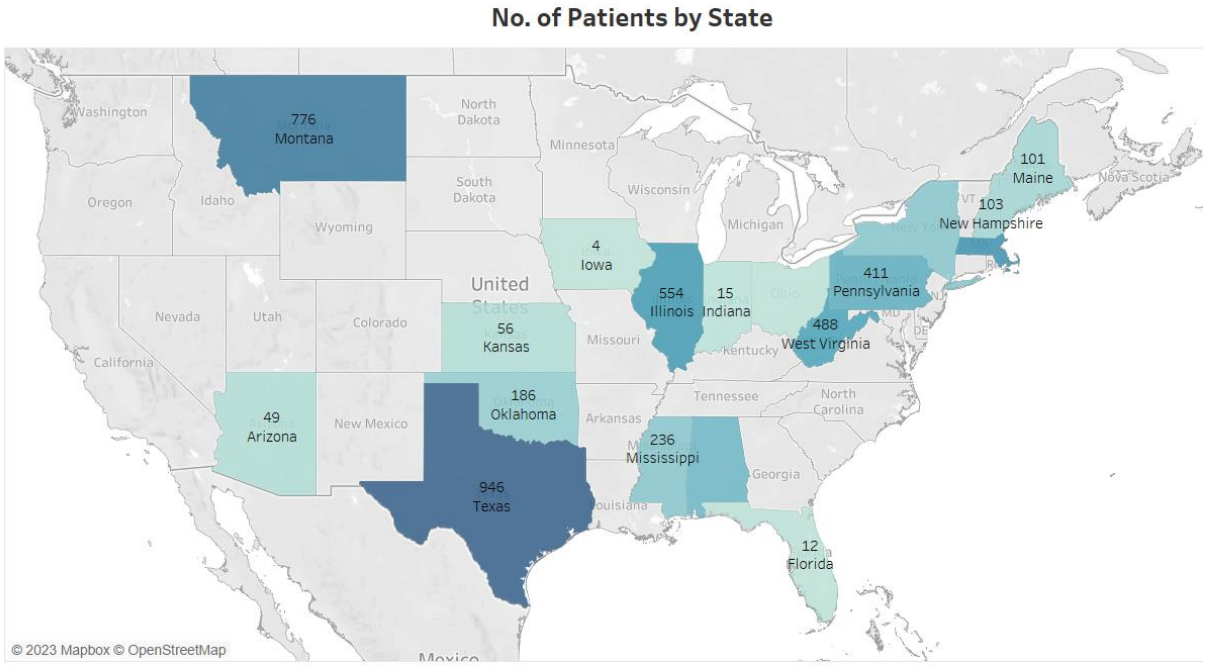
The graph provides a visual depiction of the patient population's age distribution and the gender breakdown within each age group. It allows for easy comparison of the number of male and

female patients across different age groups. The varying heights indicate the relative patient counts in each age group, while the colour distinction helps differentiate between male and female patients.



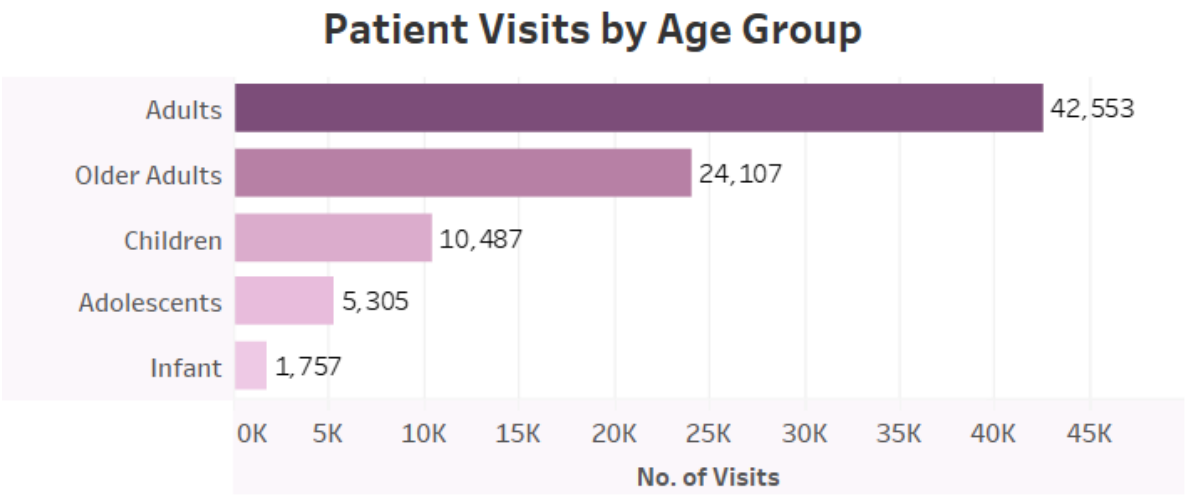
4. Patients by State

The map visualization allows for easy identification of the states with the highest concentration of patients, highlighting geographic patterns and variations in the patient population.



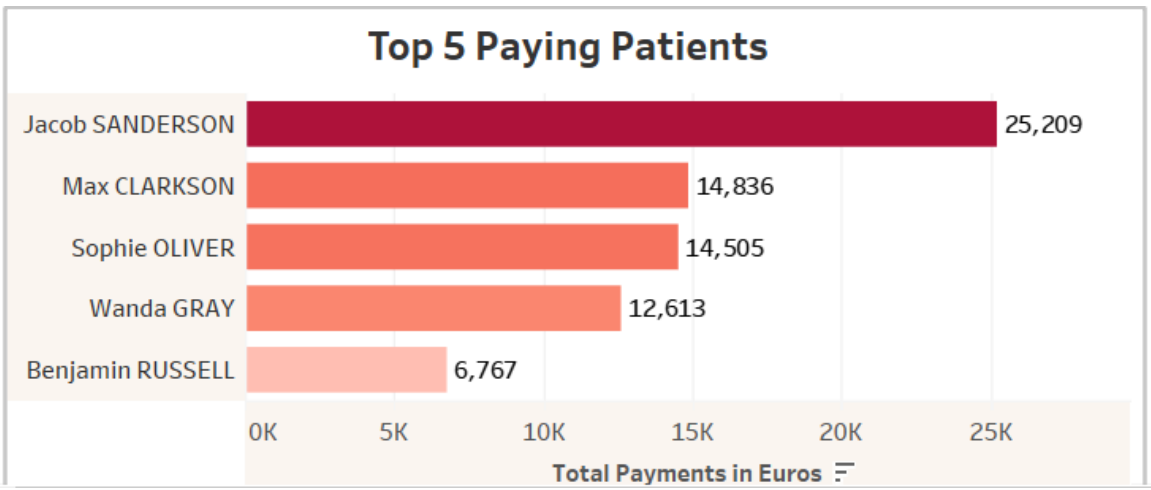
5. Patients visits by Age Group:

This bar chart displays the distribution of patient visits across different age groups. It uses bar chart to illustrate the number or proportion of visits in each age category. The chart offers interactivity for exploring specific age groups or comparing different time periods, aiding in the analysis of healthcare utilization patterns. This widget facilitates data-driven decision-making and optimization of healthcare services for specific age demographics.



6. Top 5 Paying patients:

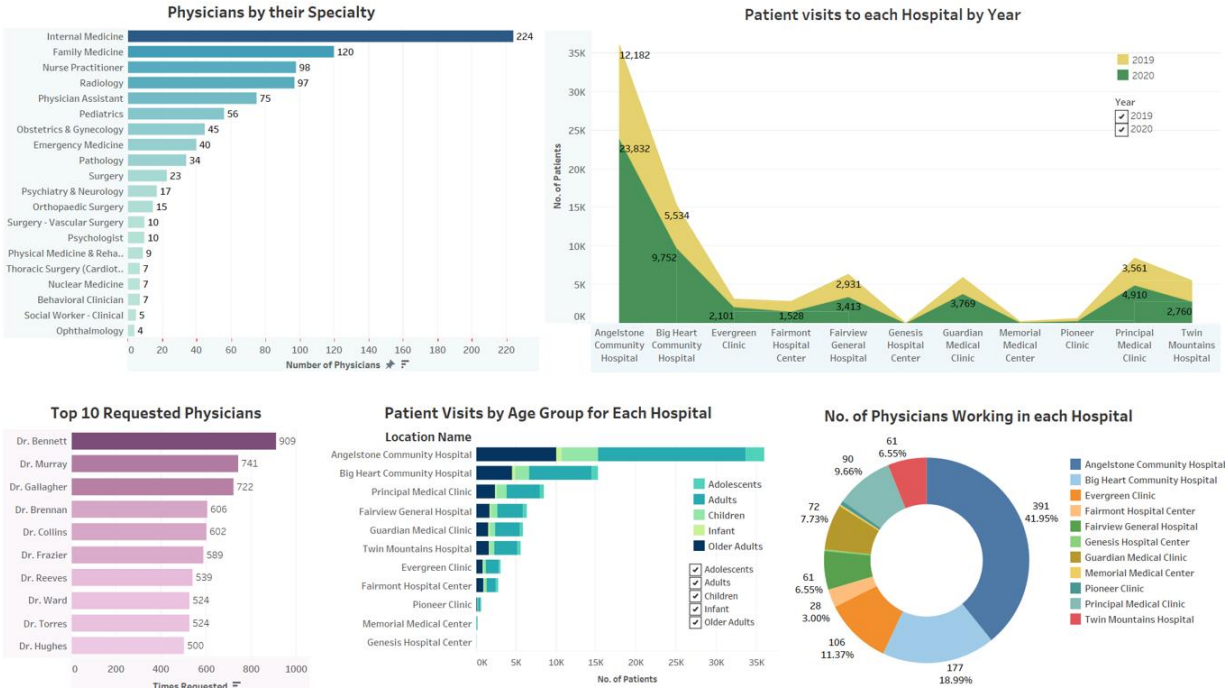
By examining the graph, it becomes evident which patients have made the largest financial contributions to the healthcare facility. This information can be valuable for revenue analysis, identifying high-value patients, and assessing the financial impact of individual patients on the healthcare organization.



Dashboard 2 – Physicians & Hospitals

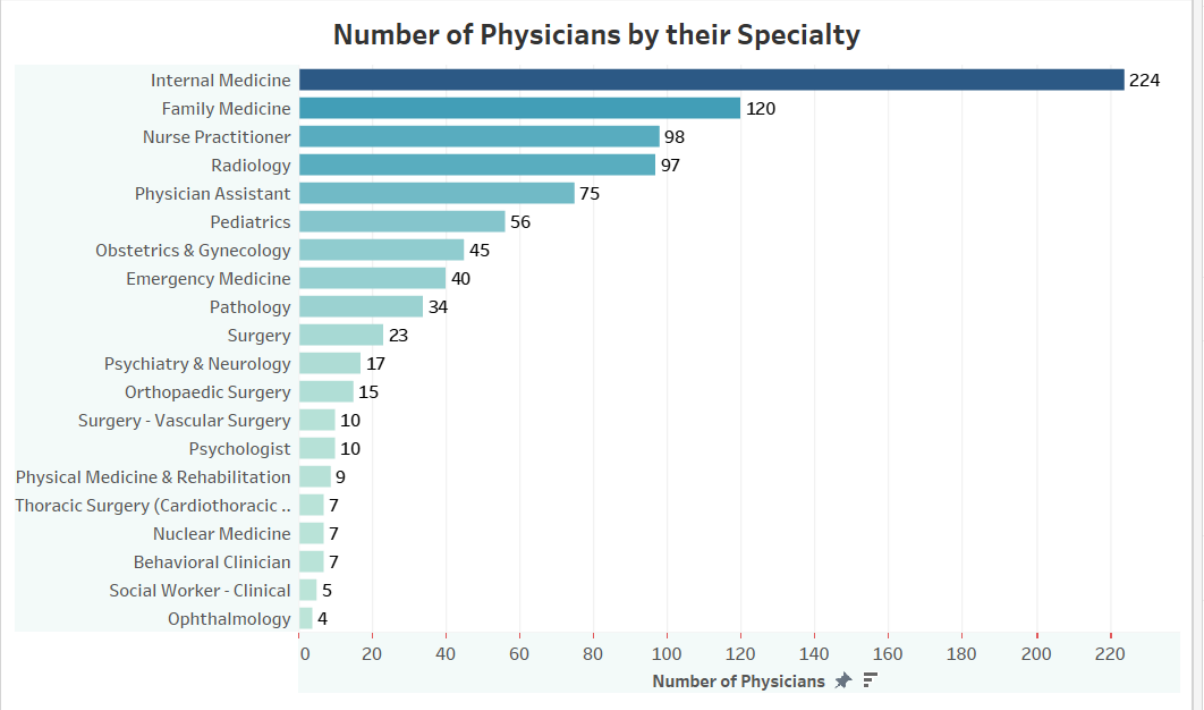
This dashboard can answer the following question:

- The distributions of physicians as per their specialty?
- Number of patients visits across different hospitals yearly.
- Most demanded physicians?
- Number of Physicians in each hospital?



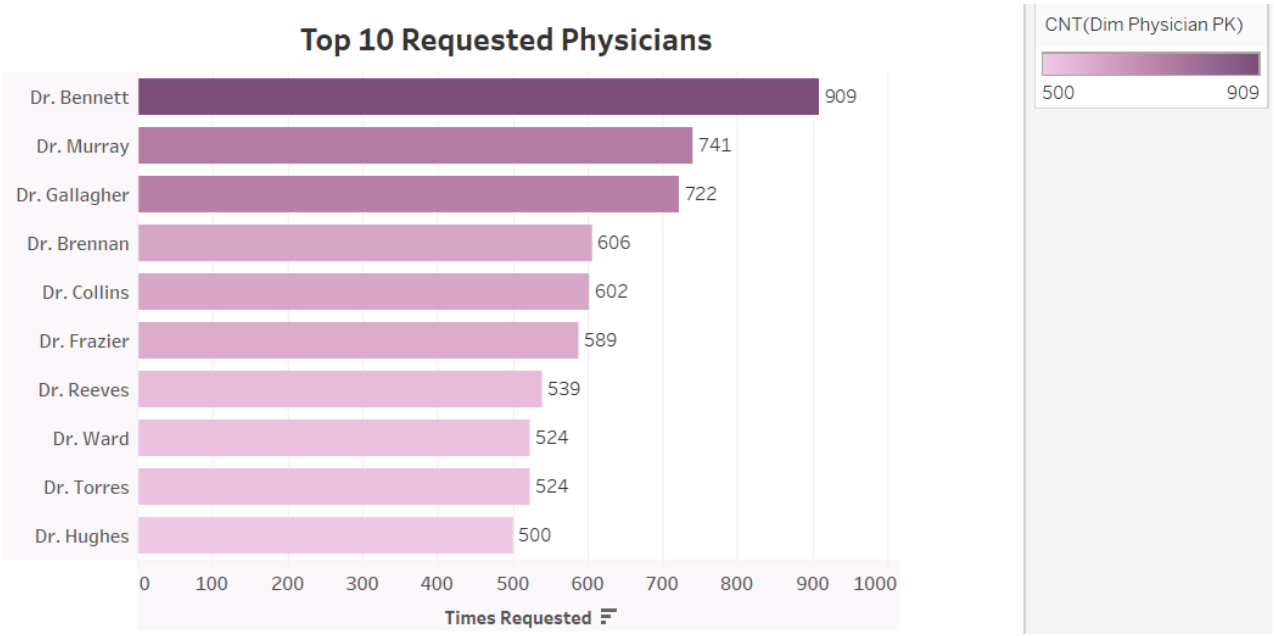
1. Physicians by Specialty

This visualization can assist healthcare administrators, policymakers, and researchers in resource planning, workforce allocation, and identifying potential gaps or surpluses in specific specialties. It provides valuable information for strategic decision-making, such as identifying areas that may require additional recruitment efforts or targeting training programs to address specific specialty needs.



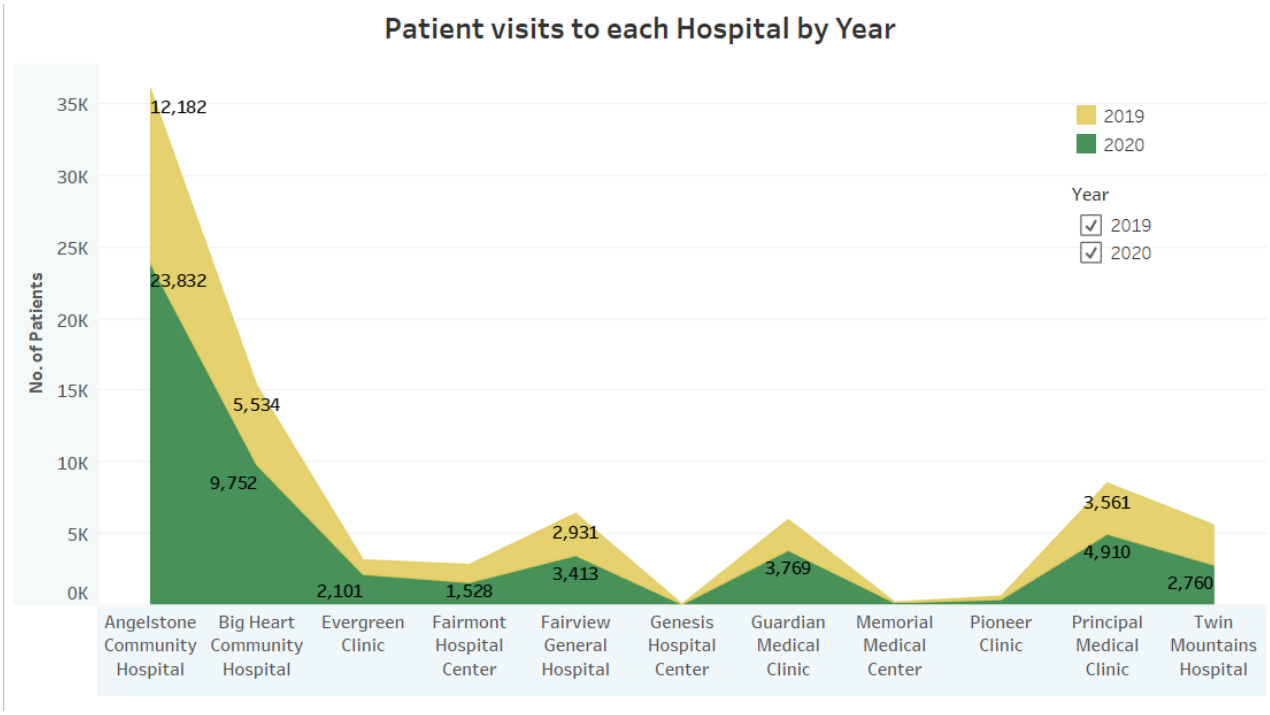
2. Most Demanded Physicians

The graph provides a visual representation of the popularity and demand for different physicians. It allows for easy comparison of the demand among the top 10 physicians and provides insights into the physicians that are in high demand among patients.



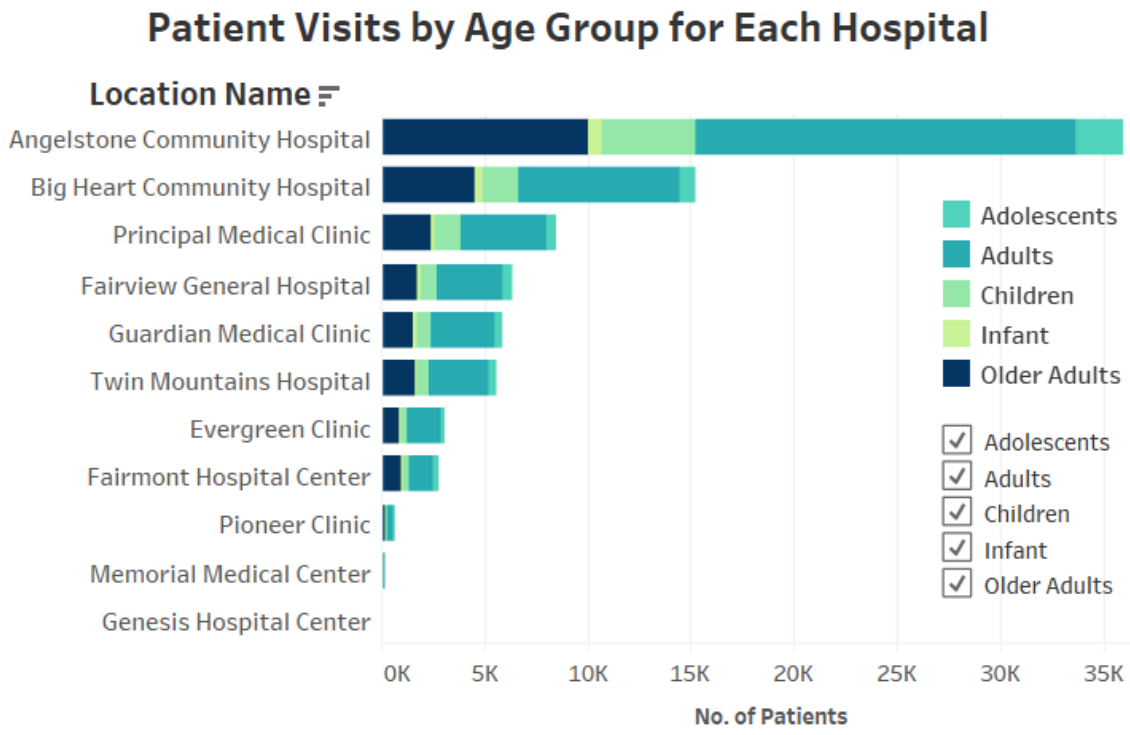
3. Patients visits per year.

The trend line graph provides a visual representation of the number of patient visits to each hospital over time, offering insights into the trends and patterns of patient visitation. It serves as a valuable tool for healthcare administrators, enabling them to monitor and analyse patient visitation data to optimize healthcare services and make data-driven decisions.



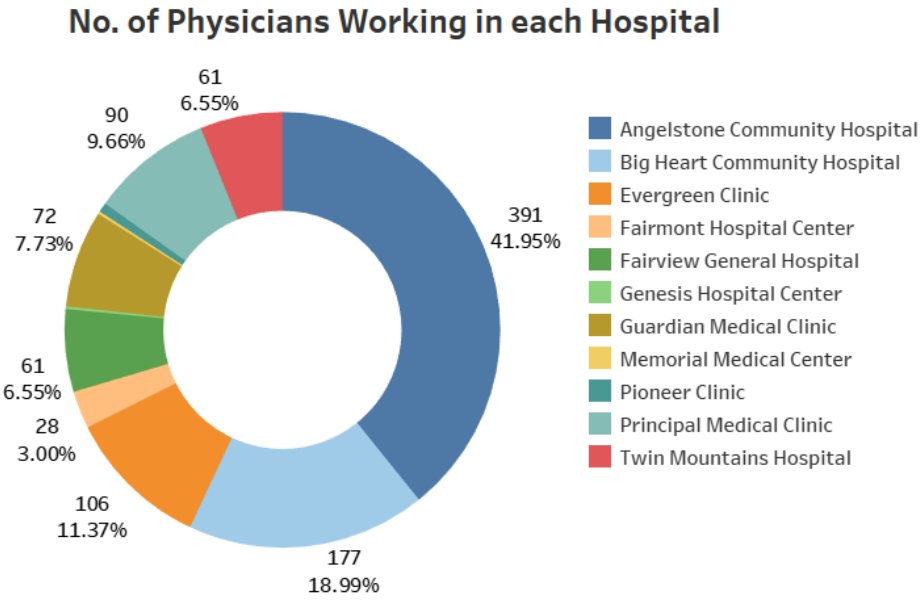
4. Patients visits by age group for each hospital:

The graph allows for a comparison of patient visits across different age groups within each hospital, providing insights into the age demographics of patients seeking medical care at each facility. Additionally, the graph facilitates a comparison between hospitals, allowing for an assessment of any differences in patient age distributions. This information can be valuable for healthcare planning, targeted marketing, and understanding the unique needs of patient populations at each hospital.



5. Physician in each hospital

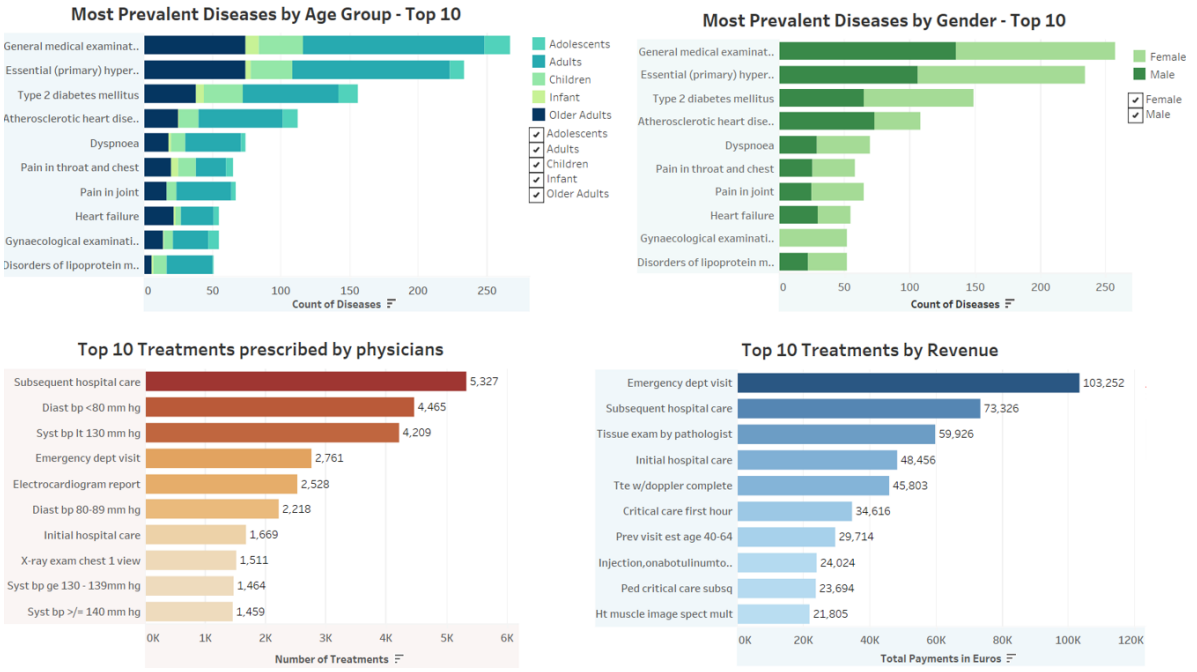
The donut chart represents the distribution of physicians in each hospital. Each slice of the donut corresponds to a specific hospital, and the size of the slice represents the proportion of physicians working at that hospital. By analysing the donut chart, it becomes possible to understand the distribution of physicians among different hospitals. It helps in identifying hospitals with a larger or smaller number of physicians, as well as the relative proportions of physicians working at each hospital.



Dashboard 3 - Treatment and disease

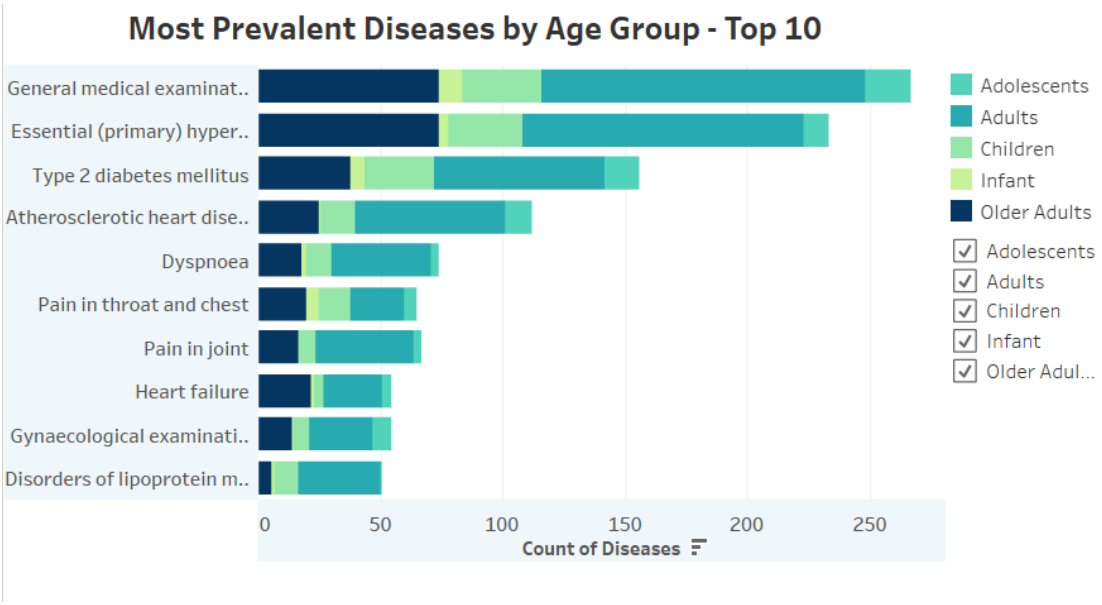
This dashboard can answer the following question:

- Top disease with respect to age group and gender?
- Most prescribed treatment?



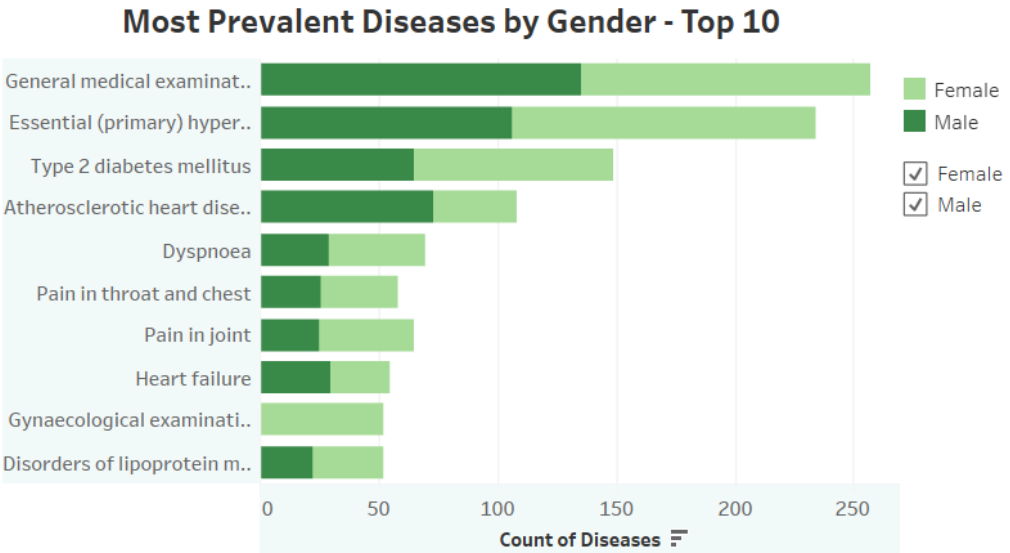
1. Most prevalent disease by age group:

This bar chart provides valuable insights into the distribution of diseases across different age groups in a healthcare dataset. It uses data analysis and visualization techniques to rank diseases by prevalence within each age group. The widget presents the information in an interactive bar chart, allowing users to explore specific age groups or diseases of interest. This widget helps users understand the most common diseases in different demographics, making it easier to identify patterns and trends in healthcare data.



2. Most Prevalent disease by gender:

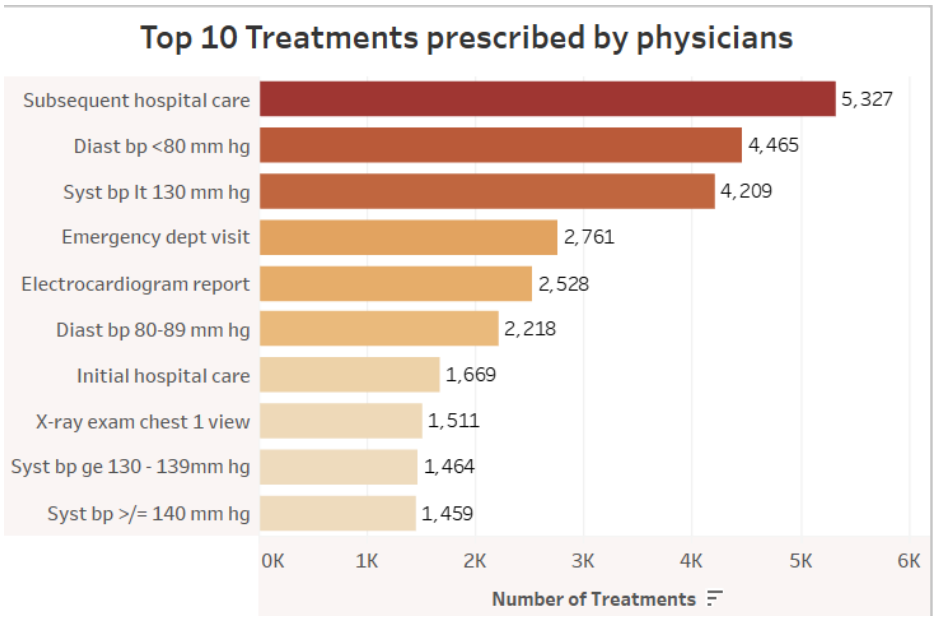
The horizontal bar graph showcasing the prevalent diseases by gender provides a clear visual representation of the gender-specific disease patterns within a population or healthcare setting. It helps in understanding the gender disparities in disease prevalence, guiding targeted interventions, and improving healthcare outcomes for both males and females.



3. Top Treatments prescribed by physicians:

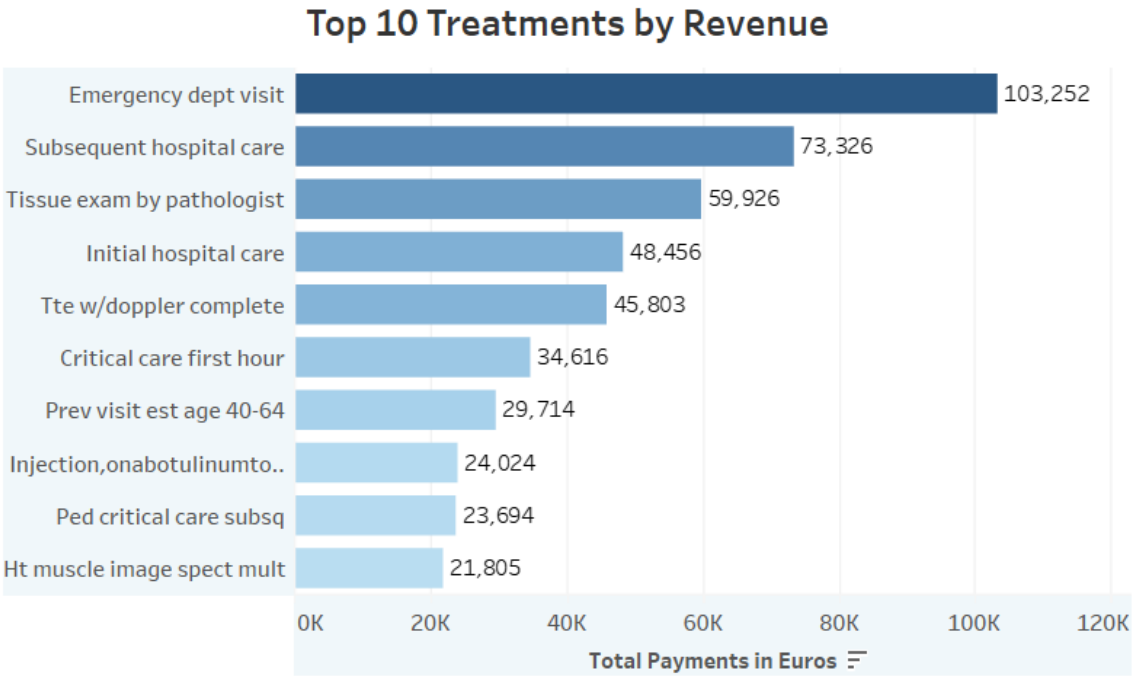
The horizontal bar graph illustrates the treatments prescribed by physicians, showcasing the frequency or number of times each treatment is prescribed. By analyzing the graph, we can gain insights into the prevalence and popularity of different treatments prescribed by physicians. It provides a clear visual comparison, allowing us to identify the most

commonly prescribed treatments. This information is valuable for understanding the treatment landscape within a particular healthcare context.



4. Top Treatments by Revenue

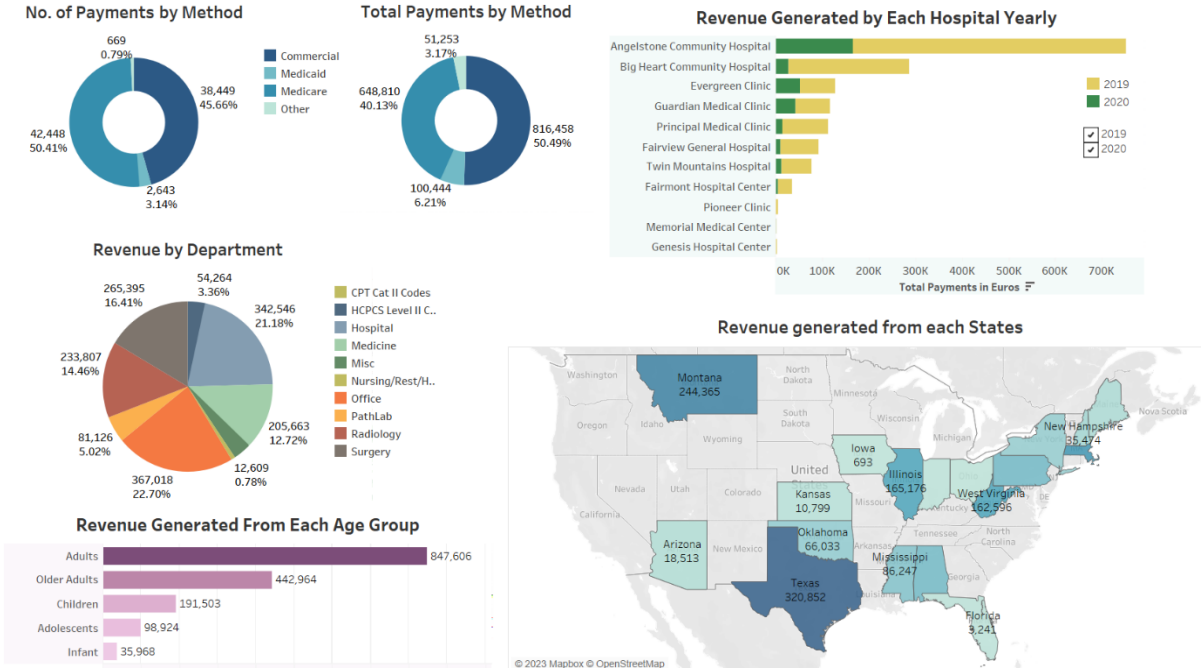
The horizontal bar graph can also highlight any variations in revenue generation among different treatments. This information can be valuable for identifying areas of growth, potential profitability, and the effectiveness of pricing strategies for different treatments.



Dashboard 4 - Payment Analysis:

We can have the following financial insights from this dashboard:

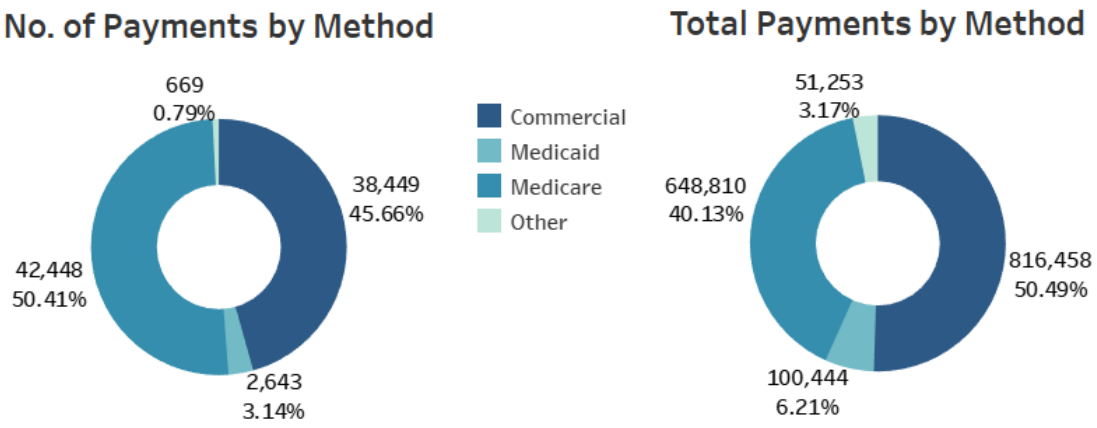
- Amount generated from each payment method.
- Revenue generated from the hospital each year.
- Revenue generated by department and age group.
- Revenue generated by geographical location.



This dashboard is more useful for the finance department of hospital. The working of each widget is given as follows:

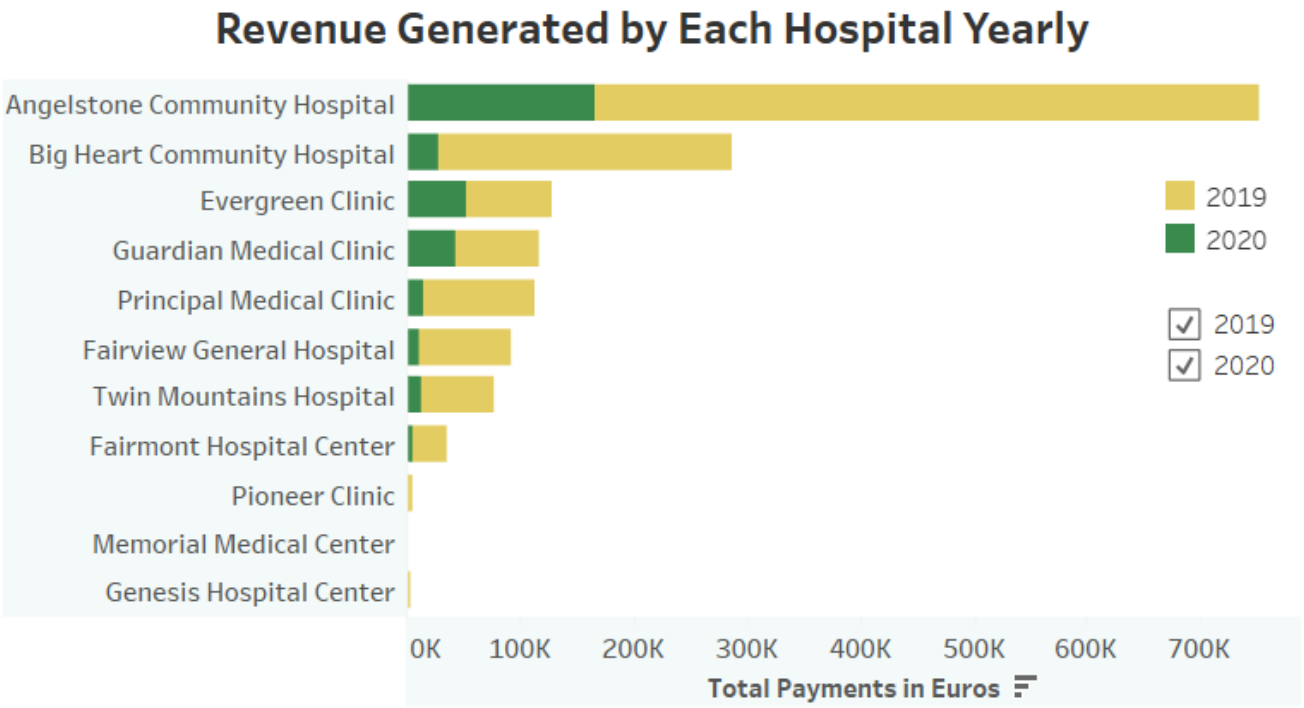
1. Payment Methods:

The two doughnut charts showcase both; the number of payments made through each payment method and the amount of payment for each method used in hospitals. The chart provides a visual representation of the proportion of each payment method in relation to the total payments received. By examining the doughnut chart, we can quickly identify the predominant payment methods utilized by patients in hospitals. It allows us to understand the relative popularity and usage of each payment method, providing insights into the financial transactions within the healthcare system.



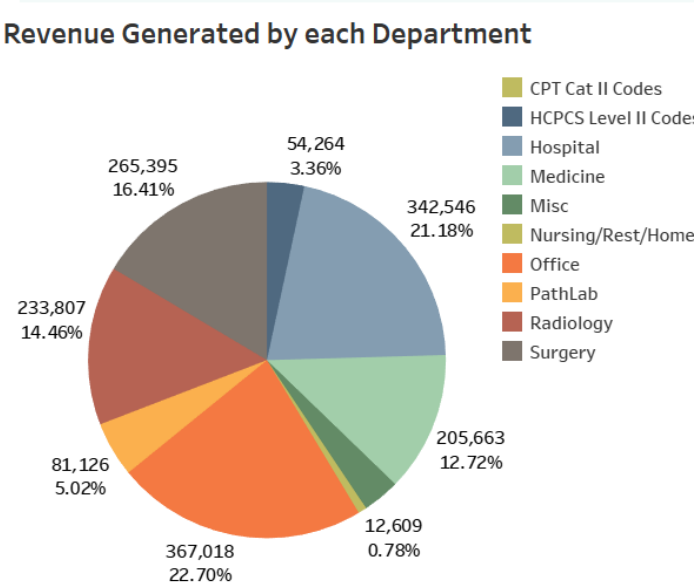
2. Revenue Generated by Each Hospital:

The horizontal bar graph represents the revenue generated by each hospital. Each bar in the graph corresponds to a specific hospital, and the length of the bar represents the revenue generated by that hospital. By analysing the horizontal bar graph, we can understand the financial performance of different hospitals within the healthcare system. It helps in identifying the hospitals that contribute significantly to the overall revenue and financial success of the organization.



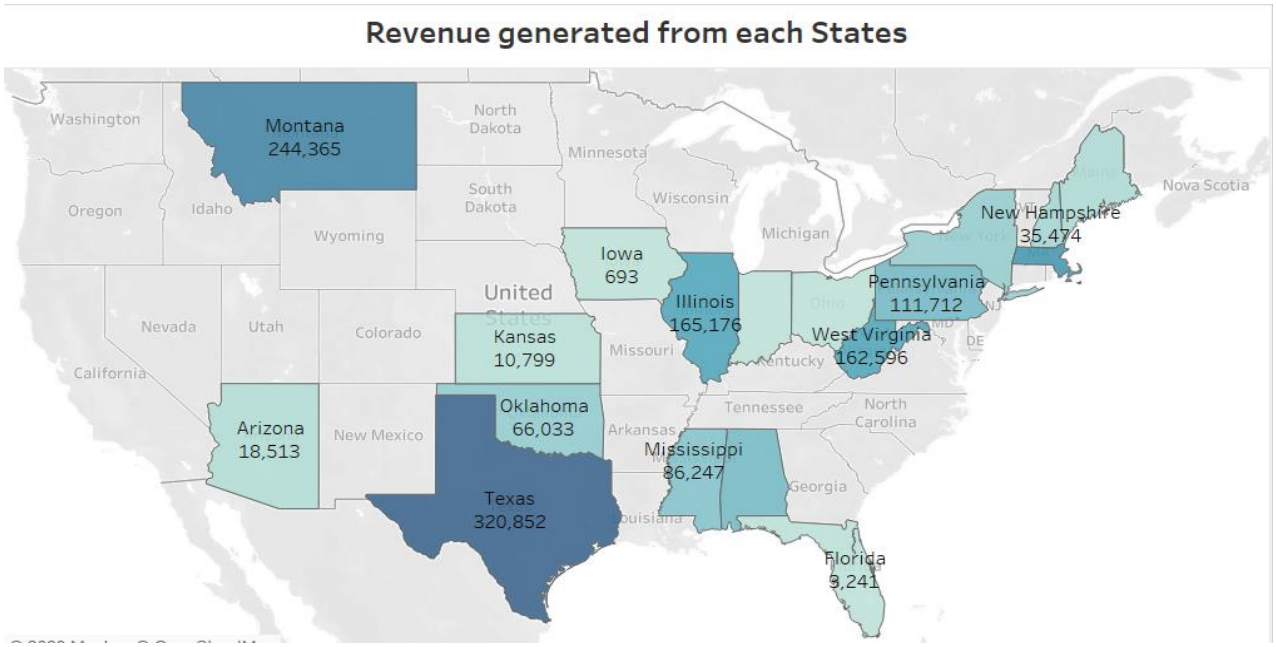
3. Revenue Generated by Each Department of Hospital:

The pie chart effectively communicates the distribution of sales from each department within the hospital, providing a concise summary of their contribution to the hospital's revenue. It helps in understanding the financial dynamics of the hospital, identifying key revenue drivers, and guiding strategic decision-making.



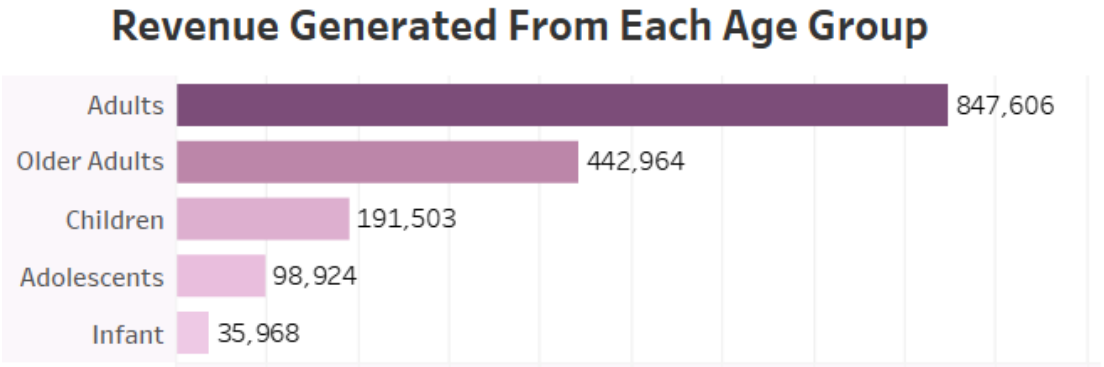
4. Revenue Generated from Each State:

The world map visualization represents the revenue generated from each state within a country. It provides a geographic representation of the revenue distribution, allowing for a comprehensive understanding of the financial performance across different states.



5. **Revenue Generated from Age Group:**

The graph provides a visual overview of the revenue distribution across different age groups, allowing for a quick comparison of the financial contributions from each group. By analyzing the width of the bars, we can easily identify which age groups contribute the most revenue and which age groups generate relatively less revenue.



Conclusion

In conclusion, the use of BI dashboards and analytics in healthcare provides valuable insights and opportunities for improvement. It enables data-informed decision-making, enhanced patient care, optimized operations, and improved healthcare outcomes.

The report focused on key areas such as patient demographics, clinical performance, financial analysis, operational efficiency, and disease patterns. Visualizations like pie charts, bar graphs, and scatter plots effectively presented the data, making it easy to interpret and gain insights.

Patient demographics help understand the patient population and tailor healthcare services accordingly. Clinical performance metrics identify areas for improvement, enhancing patient experiences and outcomes. Financial analysis drives decision-making by identifying trends and optimizing resource allocation.

By adopting BI dashboards and analytics, healthcare organizations can make data-driven decisions, improve performance, and deliver high-quality care. Continuous monitoring and refinement ensure ongoing effectiveness. Data-driven approaches drive continuous improvement in healthcare outcomes and operational efficiency.

Overall, leveraging BI dashboards and analytics empowers the healthcare industry to deliver better care, make informed decisions, and enhance overall healthcare services.