B.M.S. COLLEGE OF ENGINEERING

(Autonomous College under VTU)

Bull Temple Road, Basavangudi, Bangalore - 560019

DATA VISUALIZATION USING TOOLS
(23DS4AEDVZ)

Alternative Assessment Report

Submitted by

Manoj R (1BM22CD037)

Mayur Shridhar Hedge (1BM22CD040)

Faldu Priyansh (1BM22CD023)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING
in
Computer Science & Engineering (Data Science)

Under the Guidance of

Prof. Lavanya Naik

Assistant Professor

Department of CSE (Data Science), B.M.S. College of Engineering

2023-2024

DECLARATION

We, hereby declare that the Alternative Assessment Report entitled "Health Care Sector Data Analysis" on Data Visualization using Tools (23DS4AEDVZ) is a bonafide work and has been carried out by us under the guidance of Faculty Prof. Lavanya Naik,Assistant Professor, Department of CSE (DS), B.M.S. College of Engineering, Bengaluru, in partial fulfillment of the requirements of the degree of Bachelor of Engineering in Computer Science & Engineering (Data Science) of Visvesvaraya Technological University, Belagavi.

I further declare that, to the best of my knowledge and belief, this report has not been submitted either in part or in full to any other university/college.

Sl. No.	**Student Name**	**USN**	**Student Sign**
1	Manoj R	1BM22CD037	
2	Mayur Shridhar Hedge	1BM22CD040	
3	Faldu Priyansh	1BM22CD023	

Certified that these candidates are students of the Department of CSE (Data Science), B.M.S. College of Engineering. They have carried out the Alternative Assessment Report entitled "Health Care Sector Data Analysis" on Data Visualization using Tools (23DS4AEDVZ). The work is original and duly certified.

Faculty in charge
Prof. **Lavanya Naik**
Signature

Signature

Place: Bengaluru
Date:

Contents

1 Introduction

1.1 Background

In the healthcare industry, the efficient management and analysis of patient data are crucial for enhancing patient care and optimizing operational efficiency. With the growing volume of healthcare data, visualizing this data becomes essential to identify trends, patterns, and insights that can inform decision-making. This report focuses on the visualization of patient data from a healthcare dataset using Power BI, a powerful data visualization tool.

1.2 Objective

The primary objective of this report is to explore, analyze, and visualize patient data to uncover valuable insights. The specific goals include:

- Analyzing trends in patient data over time.
- Examining the volume of cases across different specialties, case types, and age groups.
- Understanding the distribution of case types within specialties and age groups.

1.3 Scope

This report covers the following areas:

- An in-depth analysis of the dataset, including data cleaning and preparation.
- Utilization of Power BI features for data visualization.
- Development and presentation of a comprehensive dashboard.
- Detailed analysis of trends, case volumes, and case mix.
- Conclusions and recommendations for future enhancements.

2 Insight on Dataset

2.1 Dataset Description

The dataset used in this report contains the following columns: **Archive_Date**, **Speciality_HIPE**, **Speciality_Name**, **Case_Type**, **Adult_Child**, **Age_Profile**, **Time_Bands**, and **Total**. Each row represents a unique record of patient data with specific attributes and values. The data spans multiple specialties, case types, and age groups, providing a comprehensive view of patient distribution and trends over Time.

2.2 Data Cleaning and Preparation

Data cleaning and preparation are critical steps in ensuring the accuracy and reliability of the analysis. The following tasks were performed:

- **Handling Missing Values:** Missing values were identified and appropriately handled by removing incomplete records.
- **Data Type Conversion:** Ensured that all data columns were in the correct format, such as converting **Archive_Date** to a date format.
- **Mismatching values:**Cleaning values like " Child" and "Child" and many more.

2.3 Statistical Summary

A statistical summary of the dataset provides a foundational understanding of the data distribution and key metrics:

- **Total Records:** The dataset contains a total of 4,53,120 records.
- **Speciality Breakdown:** The dataset includes records from 79 different specialties.
- **Case Types:** There are 3 unique case types, including inpatient, Day Case and outpatient cases.
- **Age Profiles:** The age profiles range from 0-15 to 65+, with detailed distribution across these age groups.
- **Time Bands:** The dataset includes various time bands representing different periods.

2.4 Data Challenges

While working with the dataset, several challenges were encountered:

- **Incomplete Data:** Some records had missing values that needed to be addressed.
- **Data Consistency:** Ensuring consistent data entry across different fields was necessary to maintain data quality.
- **Data Integration:** Merging data from different sources required careful alignment and verification.

These challenges were systematically addressed to ensure the integrity and reliability of the analysis.

3 Features Explored in Power BI

3.1 Power BI Overview

Power BI is a powerful business analytics tool developed by Microsoft. It provides interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards. Power BI connects to a wide range of data sources, enables data transformation, and offers robust data modeling capabilities. Its drag-and-drop functionality, extensive library of visualizations, and support for custom visuals make it an ideal tool for data analysis and reporting.

3.2 Key Features Utilized

Several key features of Power BI were leveraged in this report:

- **Data Import:** Power BI supports importing data from various sources, including Excel, SQL Server, and cloud services. The patient dataset was imported from an Excel file.
- **Data Transformation:** Power BI's Power Query Editor was used to clean and prepare the data. This included handling missing values, converting data types.It is used to draw relationship between differnt tables as shown in Figure 3.1
- **DAX (Data Analysis Expressions):** DAX was utilized to create calculated columns and measures, enhancing the analytical capabilities of the dataset.
- **Custom Visuals:** Power BI allows the use of custom visuals in addition to its extensive library of built-in visuals. Custom visuals were used to create specialized charts and graphs for the analysis.
- **Interactivity:** Interactive features such as slicers and drill-through were implemented to enhance user experience and enable detailed exploration of the data.

3.3 Visualization Techniques

Various visualization techniques were employed to represent the data effectively:

- **Line Charts:** Used for trend analysis over time, showcasing the progression of cases.
- **Bar Charts:** Utilized to compare case volumes, age profiles, and Speciality distributions.

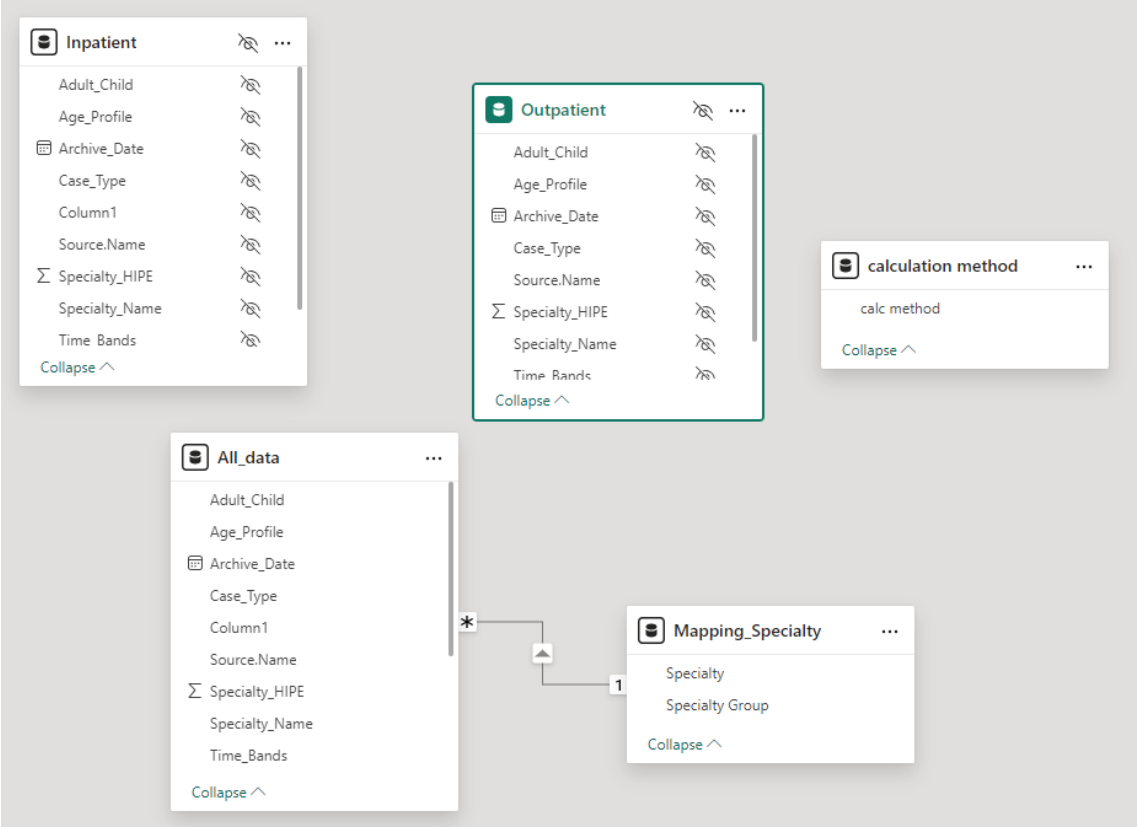

Figure 3.1: Data Model relationship

- **Pie Charts:** Used to illustrate the proportion of patients across different categories.
- **Funnel Charts:** Visualized the total number of patients for each Speciality group.
- **Tables:** Displayed detailed patient counts for different combinations of categories.
- **Text Boxes:** Highlighted summary statistics like maximum, average, and total patient counts.

3.4 Data Import and Transformation

The process of data import and transformation involved several steps:

- **Data Import:** The dataset was imported from an Excel file into Power BI. Each column was carefully mapped to ensure proper data types.
- **Data Cleaning:** Missing values were handled, and data types were converted as needed. Inconsistent entries were standardized to ensure uniformity.
- **Data Transformation:** Power Query Editor was used to transform the data. This included filtering unnecessary records, merging related tables, and creating calculated columns.
- **Data Modeling:** Relationships between different tables were established to enable comprehensive analysis. Measures and calculated columns were created using DAX to enhance the analytical capabilities.

4 Dashboard

4.1 Dashboard Design

The dashboard was designed to provide a comprehensive view of the patient data, highlighting key insights and trends As shown in Figure 4.1 and 4.2 , the data indicates... . The design focused on clarity, ease of navigation, and interactivity. Key elements included:

- **Overview Section:** Displayed high-level metrics such as total patient counts and average numbers.
- **Trend Analysis:** Line charts showing trends over time for different case types and specialties.
- **Case Volume Analysis:** Bar charts and pie charts illustrating case volumes across specialties, case types, and age groups.
- **Time Period Analysis:** Visualizations showing average and median times across different categories.
- **Case Mix Analysis:** Charts showing the distribution of case types within specialties and age groups.

4.2 Main Visualizations

The main visualizations included in the dashboard are detailed below:

4.2.1 Trend Analysis over Time

- **Archive_Date and Total:** Line chart showing the overall trend of cases over time.
- **Archive_Date, Case_Type, and Total:** Line chart comparing trends for different case types over time.
- **Archive_Date, Speciality_Name, and Total:** Line chart analyzing trends for specific specialties over time.

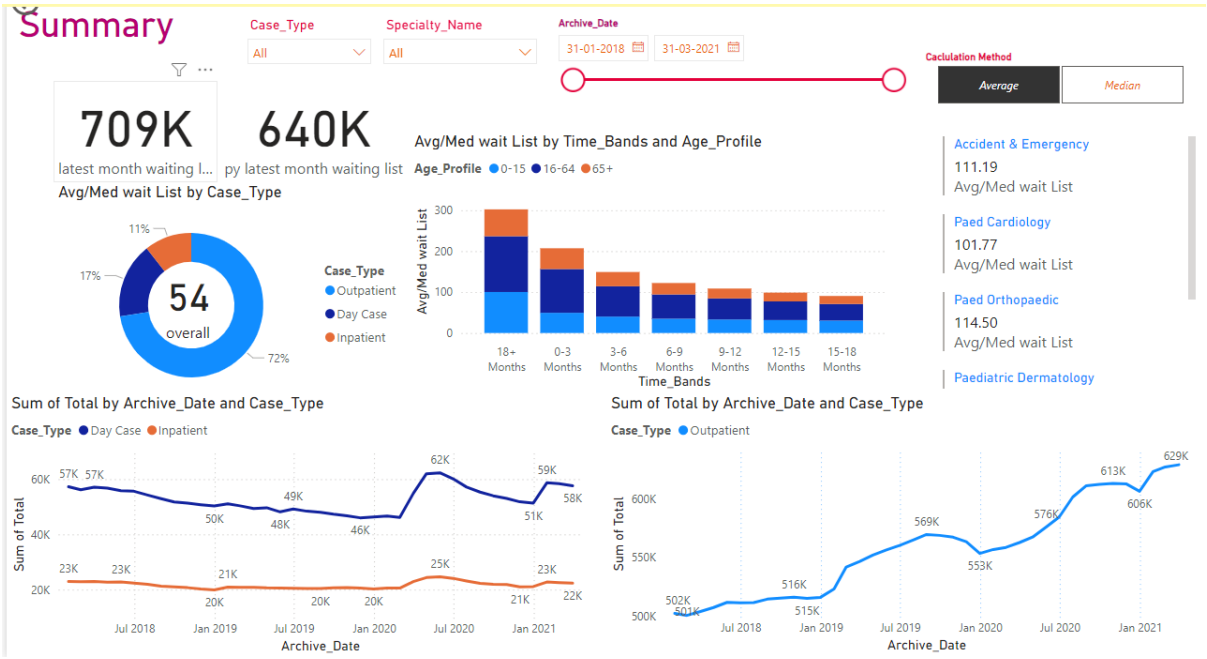

Figure 4.1: Overview of DashBoard

4.2.2 Case Volume Analysis

- **Case_Type and Total:** Bar chart showing the total number of cases for each case type.
- **Speciality_Name and Total:** Bar chart determining the case volume for different specialties.
- **Age_Profile and Total:** Pie chart analyzing case distribution across different age groups.

4.2.3 Time Analysis

- **Time_Bands, Age_Profile, and Total:** Line chart analyzing average/median time periods for different age groups and time periods.
- **Case_Type and Total:** Bar chart comparing time periods for different case types.

4.2.4 Case Mix Analysis

- **Speciality_Name, Case_Type, and Total:** Bar chart analyzing the distribution of case types within each Speciality.
- **Age_Profile and Case_Type:** Pie chart examining the proportion of different case types across age groups.

4.3 Interactivity and User Experience

The dashboard was designed to be highly interactive, enabling users to explore the data in detail:

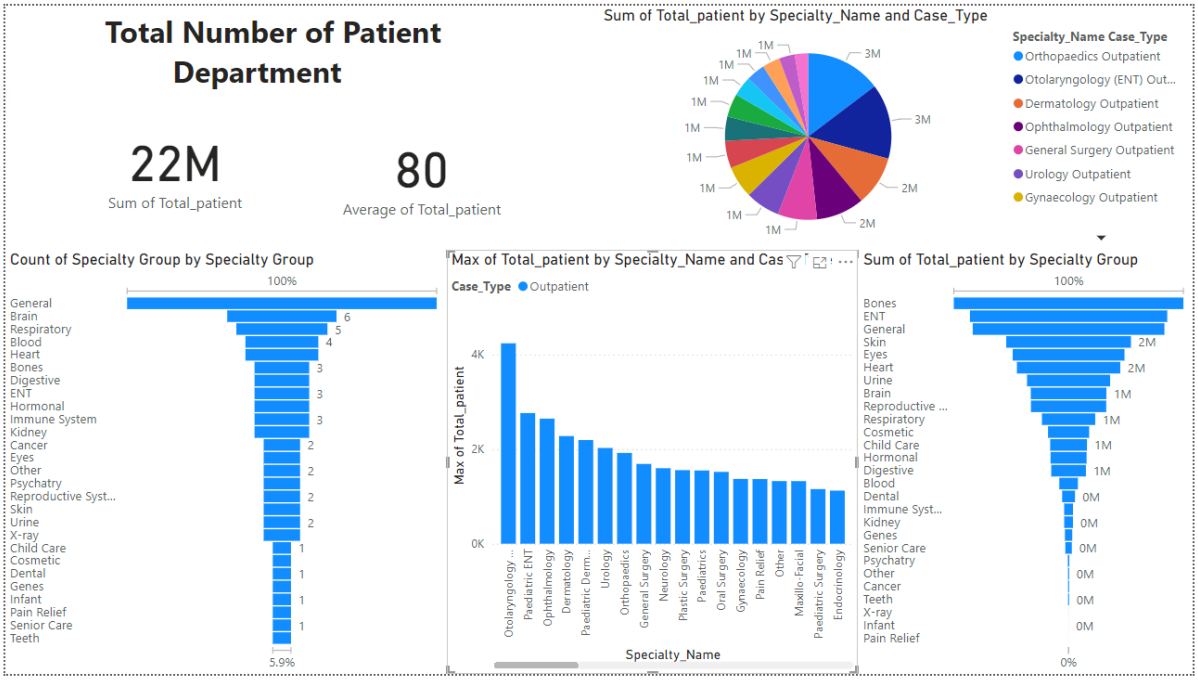

Figure 4.2: Dashboard Page 2

- **Slicers:** Used to filter data by different categories such as Speciality, case type, and age profile , the data indicates...as shown in Figure 4.3 .
- **Drill-through:** Enabled users to click on a data point and see more detailed information.
- **Responsive Design:** Ensured the dashboard was accessible and functional on different devices.

Figure 4.3: Slicers

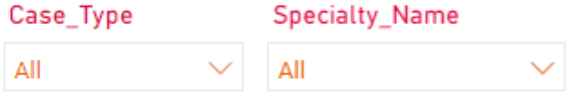

Figure 4.4: CheckBoxes

5 Analysis

5.1 Trend Analysis over Time

Trend analysis over time allows us to understand how the patient cases have evolved over different periods. By examining trends, we can identify patterns and make informed decisions.

5.1.1 Overall Trend of Cases

The overall trend of cases was analyzed using a line chart that plots **Archive_Date** against **Total**. This visualization reveals how the number of cases has changed over time, highlighting any significant increases or decreases.As shown in Figure 5.1

Analysis: The line chart shows a steady increase in outpatient cases from July 2018 to January 2021, with some minor fluctuations. This suggests a growing demand for outpatient healthcare services.

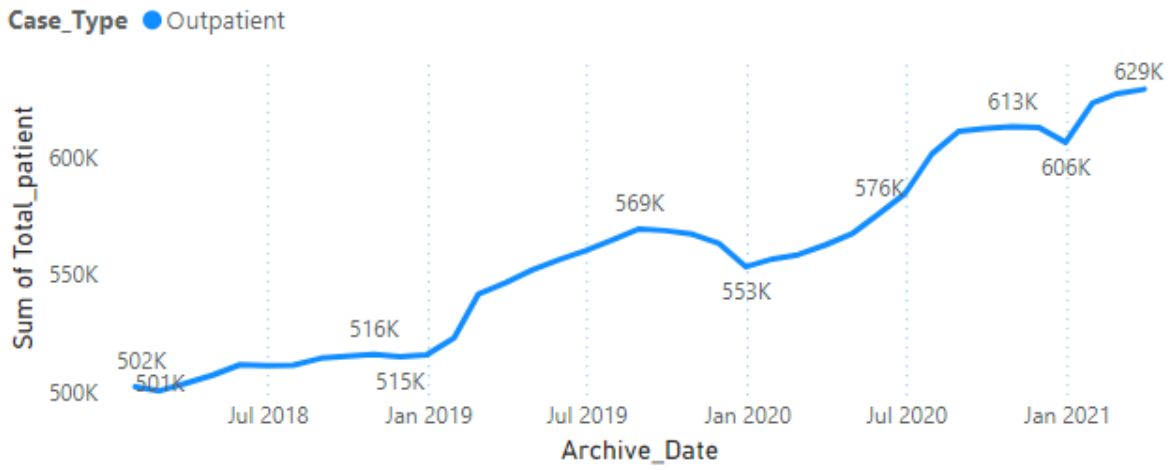

Figure 5.1: No. of Outpatient over Time

5.1.2 Comparison of Case Types

To compare trends across different case types, a multi-line chart was used, plotting **Archive_Date**, **Case_Type**, and **Total**. This chart helps in identifying whether certain case types are more prevalent during specific periods and how their trends differ from one another.As shown in Figure 5.2

Analysis : The line chart compares day cases and inpatient cases. It shows an increase in day cases and a relatively stable trend for inpatient cases, suggesting a potential shift from inpatient to day-case procedures.

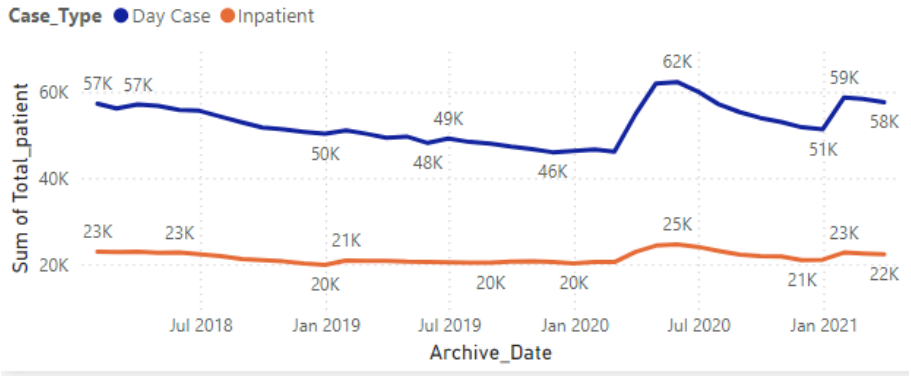

Figure 5.2: No. of Day Case and Inpatient over Time

5.1.3 Speciality Trends

Speciality trends were analyzed by plotting **Archive_Date**, **Speciality_Name**, and **Total**. This analysis highlights trends specific to each Speciality, providing insights into how different medical specialties have evolved over time.As shown in Figure 5.3

Key Findings

- Orthopedics and General Surgery are the two largest specialties, each accounting for approximately 3 million patients.
- Otolaryngology (ENT), Dermatology, and Ophthalmology each have around 1 million patients.
- The remaining specialties (Urology, Gynecology) have slightly fewer patients.

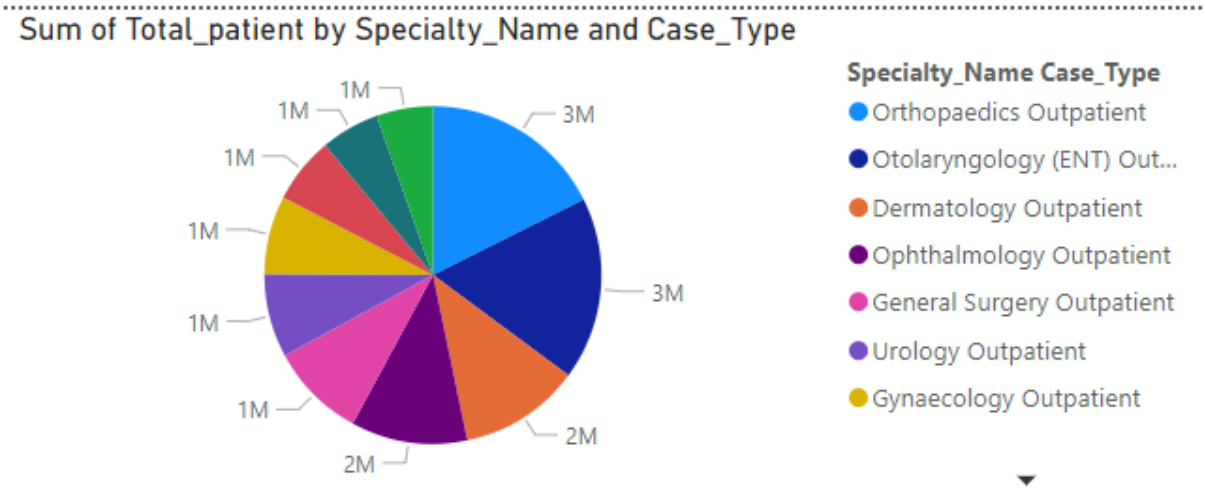

Figure 5.3: Pie chart for different Dept. by sum of Total Patients(Top 10)

5.2 Case Volume Analysis

Case volume analysis focuses on understanding the distribution of patient cases across different categories.

5.2.1 Total Cases by Type

The total number of cases by type was visualized using a bar chart, which plots **Case_Type** against **Total**.This chart helps in understanding the proportion of inpatient and outpatient cases and identifies which type is more common.As shown in Figure 5.4

Analysis:

- Outpatient cases account for the largest portion of the wait list, representing 72%.
- Day cases make up 17% of the wait list.
- Inpatient cases have the smallest share, comprising 11% of the wait list.

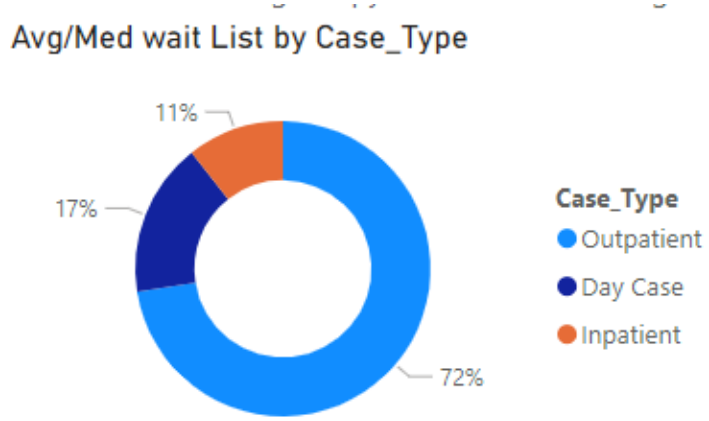

Figure 5.4: Total Cases by Types

5.2.2 Cases by Speciality

A bar chart was used to represent the case volume for different specialties, plotting **Speciality_Name** against **Total**. This analysis helps in determining which specialties handle the most cases and highlights any specialties with particularly high or low volumes.As shown in Figure 5.5

Analysis : Otolaryngology (ENT) has the highest number of outpatient cases, followed by Ophthalmology and Dermatology.

5.2.3 Age Group Distribution

The distribution of cases across different age groups was visualized using a pie chart that plots **Age_Profile** against **Total**. This chart provides insights into which age groups have the highest and lowest case volumes.As shown in Figure 5.6
AnaLysis :

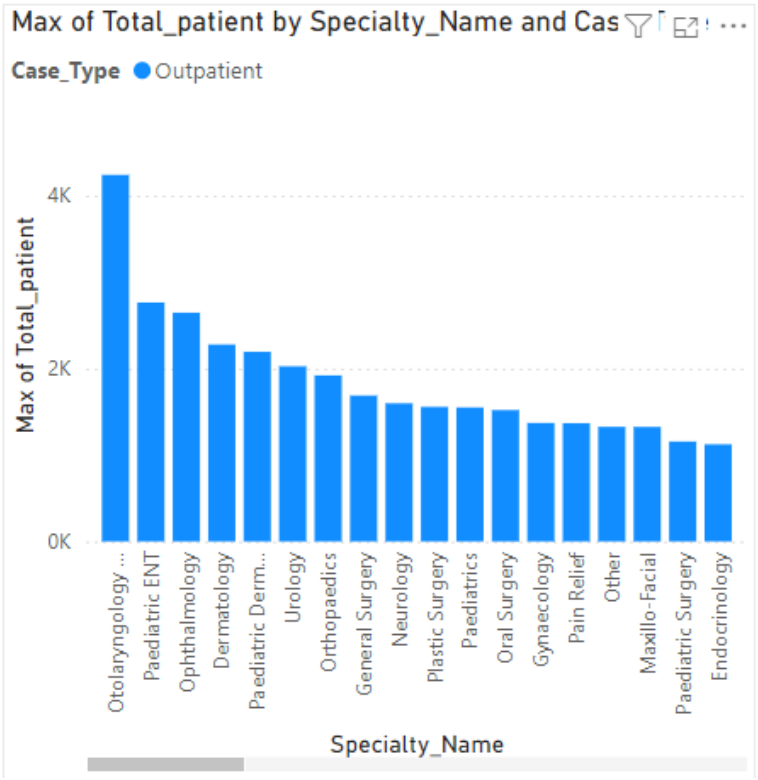

Figure 5.5: Cases By Speciality

- Wait times increase with age: Across all time bands, the average/median wait list is generally longer for older age groups (65+) compared to younger age groups (0-15 and 16-54).
- Wait times increase with time: Within each age group, the average/median wait list tends to increase as the time band progresses.
- Age group 16-54 has the longest wait times: In most time bands, the 16-54 age group has the longest average/median wait list.

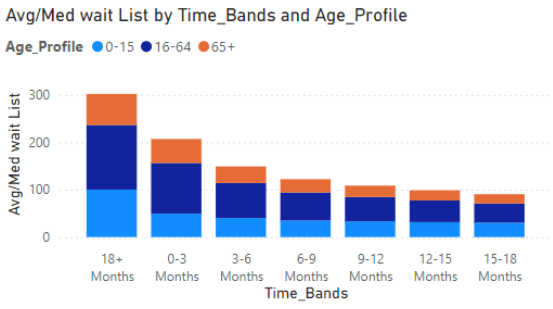

Figure 5.6: Age Group Distribution

5.3 Time period Analysis

Time period analysis is crucial for understanding the efficiency and performance of healthcare services.

5.3.1 Average/Median Time Periods

The average and median time periods were analyzed by plotting **Time_Bands**, **Age_Profile**, and **Total** using a line chart. This analysis helps in identifying how time periods vary across different age groups and time periods.We can Toggle (Figure 4.1)between Average and Median by the given Tile at top of the Dashboard.As shown in Figure 5.2 and 5.1.

5.4 Case Mix Analysis

Case mix analysis examines the distribution of case types within different specialties and age groups.

5.4.1 Distribution within Specialties

The distribution of case types within each Speciality was visualized using a funnel chart that plots **Speciality_Name**, **Case_Type**, and **Total**. This chart helps in understanding the variety of cases handled by each Speciality.As shown in Figure 5.7

Analysis :

- Bones is the specialty with the highest number of patients, followed by ENT and General.
- Eyes, Heart, and Skin also have a significant number of patients.

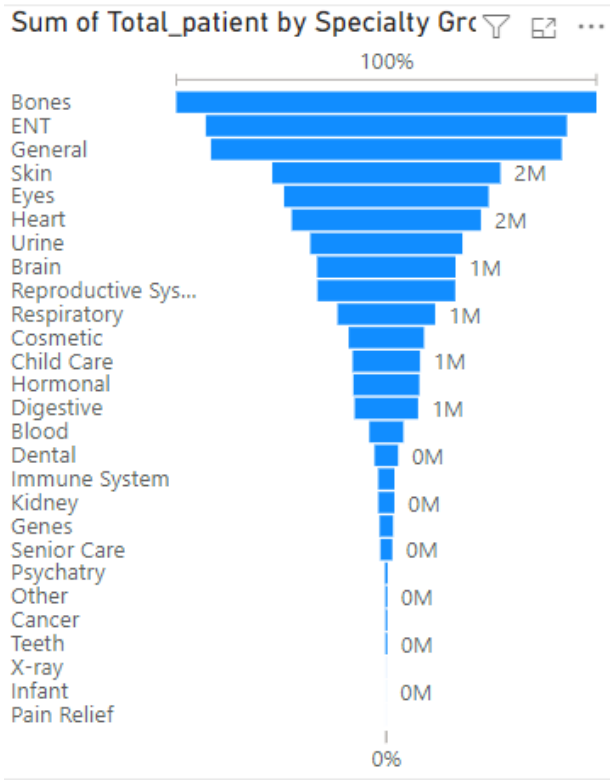

Figure 5.7: Distribution within Specialties

5.4.2 Age Group and Time in Hospital(admitted)

A pie chart examining the proportion of different case types across age groups, plotting **Age_Profile** and **Time_band**, was used to understand how long different age groups of people present in hospital .As shown in Figure 5.8
Analysi :

- Age group 16-54 is the most dominant across all time bands, with the largest slice in each segment.
- Age group 0-15 has a significant presence in the earlier time bands (0-3 months, 3-6 months) but becomes less prominent in later time bands.
- Age group 65+ is relatively smaller in the earlier time bands but becomes more prominent in the later time bands, particularly 12-15 months and 15-18 months.

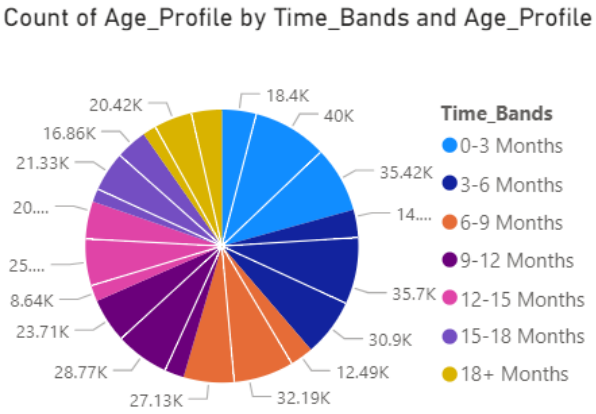

Figure 5.8: Pie chart for differnt time bands and age profile

5.5 Summary of Findings

The analysis provided several key insights:

- **Overall Trends:** The overall number of cases showed a steady increase over time, with notable peaks during specific periods.
- **Case Types:** Outpatient cases were more prevalent than inpatient cases, with distinct trends observed for each type.
- **Speciality Insights:** Certain specialties, such as emergency medicine and pediatrics, handled higher volumes of cases compared to others.
- **Age Distribution:** The majority of cases were concentrated in specific age groups, with children and the elderly showing higher case volumes.
- **Time periods:** time periods varied significantly across different case types and age groups, highlighting areas where efficiency can be improved.

Specialty Group	Sum of Total_patient
Blood	277814
Bones	387255(
Brain	1159509
Cancer	16103
Child Care	556832
Cosmetic	769681
Dental	226216
Digestive	551078
ENT	3164796
Eyes	2018385
General	3461164
Genes	109024
Heart	1799645
Hormonal	538049
Immune System	156906
Total	**24640969**

Figure 5.9: Sum of each patient among different departments

- **Case Mix:** The distribution of case types within specialties and age groups provided insights into the specialization and focus areas of different medical departments.

These findings can inform decision-making and strategic planning to improve patient care and operational efficiency in healthcare services.

Archive_Date	Day Case	Inpatient	Outpatient	**Total**
28 February 2018	56180	22859	500800	**579839**
31 January 2018	57267	22937	502482	**582686**
31 March 2018	57095	22963	504111	**584169**
30 November 2018	50760	20229	515360	**586349**
31 December 2018	50324	19880	516162	**586366**
30 April 2018	56731	22683	507507	**586921**
31 July 2018	54244	21912	511675	**587831**
30 September 2018	51725	20993	515547	**588265**
31 October 2018	51287	20714	516363	**588364**
31 August 2018	52911	21278	514585	**588774**
30 June 2018	55617	22397	511415	**589429**
31 May 2018	55795	22801	511904	**590500**
31 January 2019	51087	20940	523225	**595252**
28 February 2019	50345	20860	541899	**613104**
31 March 2019	49367	20852	546630	**616849**
31 December 2019	46339	20224	553434	**619997**
30 April 2019	49655	20640	551965	**622260**
31 January 2020	46687	20616	556770	**624073**
31 May 2019	48156	20609	556411	**625176**
29 February 2020	46142	20563	558554	**625259**
30 June 2019	49168	20503	560251	**629922**
30 November 2019	46010	20584	563410	**630004**
31 July 2019	48395	20412	564829	**633636**
31 October 2019	46771	20740	567221	**634732**
30 September 2019	47314	20671	568769	**636754**
31 August 2019	47984	20406	569498	**637888**
31 March 2020	54848	22900	562693	**640441**
30 April 2020	61974	24369	567329	**653672**
31 May 2020	62283	24663	575863	**662809**
Total	**2059882**	**845348**	**21735739**	**24640969**

Figure 5.10: For detailed analysis , and to use Drill through

6 Conclusion & Future Enhancement

6.1 Summary of Work

This report presented a comprehensive analysis of patient data using Power BI. The dataset, comprising various attributes such as **Archive_Date**, **Speciality_HIPE**, **Speciality_Name**, **Case_Type**, **Adult_Child**, **Age_Profile**, **Time_Bands**, and **Total**, was thoroughly examined to uncover valuable insights. The key components of the work included:

- Data cleaning and preparation to ensure the accuracy and reliability of the dataset.
- Exploration of Power BI features for effective data visualization.
- Development of a detailed dashboard with interactive visualizations.
- Analysis of trends, case volumes, time periods, and case mix.

The report demonstrated the power of data visualization in making informed decisions and highlighted the importance of using advanced tools like Power BI to handle complex datasets.

6.2 Key Insights

Several key insights emerged from the analysis:

- **Overall Trends:** The number of patient cases exhibited a steady increase over time, with specific peaks indicating periods of high demand.
- **Case Types:** Outpatient cases were more prevalent than inpatient cases, suggesting a higher frequency of non-critical visits.
- **Speciality Trends:** Certain specialties, such as emergency medicine and pediatrics, consistently handled higher volumes of cases, indicating their critical role in patient care.
- **Age Distribution:** Children and the elderly represented the largest patient groups, highlighting the need for targeted healthcare services for these age ranges.
- **Time periods:** time periods varied significantly across case types and age groups, pointing to potential areas for improving service efficiency.
- **Case Mix:** The diversity of case types within specialties and age groups provided insights into the specialization and focus areas of different medical departments.

These insights can guide healthcare providers in optimizing resources, improving patient care, and enhancing operational efficiency.

6.3 Future Work

While the current analysis provided valuable insights, there are several areas for future enhancement:

- **Expanded Dataset:** Incorporating additional data sources, such as patient outcomes, treatment details, and cost information, can provide a more comprehensive view of healthcare services.
- **Predictive Analytics:** Utilizing advanced machine learning algorithms to predict future trends, patient outcomes, and resource needs can enhance decision-making.
- **Real-time Data Integration:** Integrating real-time data feeds into the Power BI dashboard can provide up-to-date insights and enable more responsive actions.
- **Detailed Time period Analysis:** Conducting a more granular analysis of time periods by considering additional factors such as patient acuity and resource availability can help identify specific bottlenecks.
- **User Feedback Integration:** Collecting and incorporating feedback from healthcare providers and patients can improve the usability and relevance of the dashboard.

Implementing these enhancements will further strengthen the ability to analyze and visualize patient data, ultimately contributing to better healthcare outcomes.