# Introduction to Data Management

**PROJECT REPORT**

(Project Semester January-April 2025)

**Patient Demographics, Conditions, and Billing Dataset: An Interactive Dashboard Approach**

Submitted by

Pranab Pasari

Registration No: 12314069 Section: K23EG

Programme: B. Tech - Computer Science and Engineering Course Code: INT217

Under the Guidance of

**Jaffar Amin Chacket (UID: 30453) Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering Lovely Professional University, Phagwara**

****

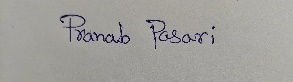
**DECLARATION**

I, Pranab Pasari, student of B.Tech Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Pranab Pasari

12314069– K23EG

B. Tech - Computer Science and Engineering Lovely Professional University, Phagwara



09/04/2025

**CERTIFICATE**

This is to certify that Pranab Pasari bearing Registration no. 12314069 has completed INT217 (Introduction to Data Management) project titled, **“****Patient Demographics, Conditions, and Billing Dataset: An Interactive Dashboard Approach”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature**

**Jaffar Amin Chacket (UID: 30453) Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University Phagwara, Punjab.

Date:

# Acknowledgement

The opportunity of attaining a course based on Data Management using Excel at Lovely Professional University was worth learning. It was a prestige for me to be part of it. During the period of my course, I received tremendous knowledge related to Microsoft Excel and Data Management.

Pre-eminently, I would like to express my deep gratitude and special thanks to my course teacher Sameeksha Khare for her theoretical knowledge and encouragement on this project and for her valuable guidance and affection for the successful completion of this project.

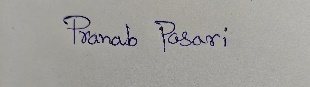
Secondly, I would like to thank Lovely Professional University for giving me an opportunity to learn this course.

Lastly, I would like to thank the almighty and my parents for their constant encouragement, moral support, personal attention, and care.

Pranab Pasari

12314069– K23EG

B. Tech - Computer Science and Engineering Lovely Professional University, Phagwara



# Contents

|  |  |  |
| --- | --- | --- |
| Sr. No. | Topic | Page No. |
| 1. | [Cover Page](#_Introduction_to_Data) | 1 |
| 2. | [Declaration](#declaration) | 2 |
| 3. | [Certificate](#certificate) | 3 |
| 4. | [Acknowledgement](#_Acknowledgement) | 4 |
| 5. | [Contents](#_Contents) | 5 |
| 6. | [Abstract](#abstract) | 6 |
| 7. | [List of Tables and Figures](#_List_of_Tables) | 7 |
| 8. | [Chapter 1 - Introduction](#Intro) | 8 |
| 9. | [Chapter 2 - Objectives](#_Chapter_2_-) | 9 |
| 10. | [Chapter 3 - Source of Dataset](#_Chapter_3_-) | 10 |
| 11. | [Chapter 4 - Dataset Preprocessing](#_Chapter_4_-) | 11-13 |
| 12. | [Chapter 5 - Analysis on dataset](#_Chapter_5_-) | 14-22 |
| 13. | [Chapter 6 - Dashboard](#_Chapter_6_-) | 23-26 |
| 14. | [Chapter 7 - Conclusion](#_Chapter_7_-) | 27 |
| 15. | [Chapter 8 - Future scope](#_Chapter_8_-) | 28-29 |
| 16. | [Chapter 9 - References](#_Chapter_9_-) | 30 |

**Abstract**

Excel is a software program created by Microsoft that uses spreadsheets to organize numbers and data with formulas and functions. Excel analysis is ubiquitous around the world and used by businesses of all sizes to perform data analysis. Excel features calculation, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications, and several other features which make Excel a perfect choice to manage and analyse data.

My project is an Excel Dashboard. The Excel Dashboard is used to display overviews of large data tracks. Excel Dashboards use dashboard elements like tables, charts, and gauges to show the overviews. The dashboards ease the decision-making process by showing the vital parts of the data in the same window.

In this report, I have shared a project where I have done data analysis of a Healthcare dataset. This report also presents my learning during my course classes.

# List of Tables and Figures

|  |  |  |
| --- | --- | --- |
| Sr. No. | Table and Figure Name | Page No. |
| 1. | [Raw Data](#f1) | 12 |
| 2. | [Clean Data](#f2) | 12 |
| 3. | [Calculated columns](#f3) | 13 |
| 4. | [Cleaned dataset sheet with columns for easy accessibility](#f4) | 15 |
| 5. | Pivot Tables to medical  coditions | 16 |
| 6. | Total Billing Overview & Trends | 17 |
| 7. | Overview of the patients Admission types | 18 |
| 8. | [Bar Chart](#f8) | 19 |
| 9. | Donut Chart | 20 |
| 10. | [Line Chart](#f10) | 20 |
| 11. | Column Chart | 21 |
| 12. | [Donut Chart](#f12) | 21 |
| 13. | [Final Dashboard](#f13) | 22 |
| 14. | [Analysing data using slicers](#f19) | 22-25 |

**Chapter 1 - Introduction**

I have created an Excel Dashboard of a Healthcare Dataset. This dashboard explains and highlights important facts, trends, and patterns related to Medical conditions and Patients different states based on gender, race, age group, and year.

The dataset used contains information regarding Healthcare. It includes details such as gender-wise data, Hispanic origin, and different age groups of people affected by diseases. The dataset also covers records from multiple years across various US states.

I have scrubbed and organized the entire dataset and performed the analysis on a clean data sheet created within the Excel workbook. I have derived and calculated important results from the dataset using various Excel features like pivot tables, slicers, and functions. I have represented the findings in the form of a dynamic dashboard using Excel's visualization tools and various charts.

# Chapter 2 - Objectives

This project on Healthcare Analysis provides the records, facts, and trends of Medical conditions and Patients different states based on gender, race, age group, and year. However, here are the few main objectives that are discussed in the dashboard:

* Track total billing amounts over time (monthly/annually).
* Insights: Identify peak cost periods, assess healthcare service usage trends.Visuals: Line chart for monthly totals, KPI cards for total & average billing.
* Medical Condition Distribution: Analyze the frequency of various medical conditions.
* Analyze the Relationship Between Medical Conditions and Test Results
* Analyze patient data by gender, age group, and blood type to understand the demographic makeup of your hospital's patient population.
* Track and visualize how patient admissions vary over time, by admission type (e.g., Emergency, Urgent, Elective), and across hospitals or doctors to improve capacity planning and resource allocation.

# Chapter 3 - Source of Dataset

The dataset is taken from Data.gov. Data.gov is the official open data portal of the United States Government that provides access to a wide range of datasets for public use. This platform allows users to find and download datasets published by government agencies.

I have selected a Healthcare dataset which contains important details of Medical conditions and Patients different states based on gender, race, age group, and year.

Here are the details of my chosen dataset:

Name -Healthcare dataset— Link - <https://drive.google.com/file/d/1tynOGr47ct1lfmU6P2jmUSOCT_xgTFVS/view?usp=sharing>

Format - CSV

# Chapter 4 - Dataset Preprocessing

In this project, the dataset preprocessing was an important step to ensure accurate and meaningful analysis. The raw dataset contained information about Medical conditions and Patients different states based on gender, race, age group, and year.

The following preprocessing steps were performed:

1. Data Cleaning:
   * Removed unnecessary columns that were not required for analysis.
   * Identified and handled missing values in important columns.
   * Replaced blank cells or missing entries with appropriate values where applicable.
2. Data Transformation:
   * Renamed column headers for easy understanding and readability.
   * Organized data into a clean and structured format suitable for analysis.
   * Filtered relevant records and removed inconsistent data entries.
3. Creating Clean Dataset Sheet:
   * After cleaning and transforming the data, a new sheet named "Clean Dataset" was created within the Excel workbook.
   * The following screenshots represent:
   * 

Figure 1: The original/raw dataset imported from Data.gov.



Figure 2: The cleaned dataset sheet created after data preprocessing.

* + This clean dataset sheet was used as the primary source for performing analysis and creating visualizations in the dashboard.

1. Data Preparation:
   * Created calculated columns for better analysis (e.g., calculating total suicide rates, average rates, etc.).
   * This screenshot represent:

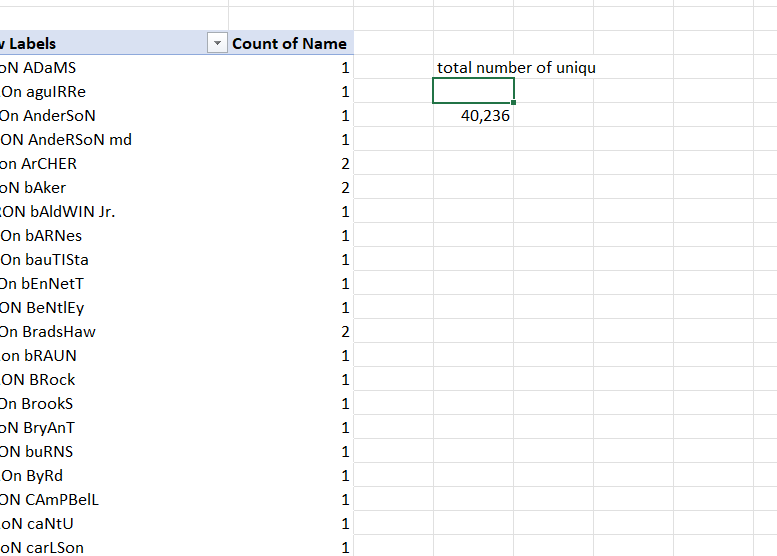


Figure 3: The calculated columns for better analysis

* + Used sorting and filtering techniques to analyse specific data points.

This preprocessing phase ensured that the dataset was accurate, consistent, and ready for further analysis using Excel tools like Pivot Tables, Charts, Slicers, and Conditional Formatting.

# Chapter 5 - Analysis on dataset

* + General Description:

This analysis focuses on the "Medical Condition" field in the dataset, which represents the primary diagnosis or health issue associated with each patient record. Common values include conditions such as **Cancer**, **Diabetes**, **Obesity**, and others.

By aggregating and categorizing this data, we aim to identify which health conditions are most prevalent in the patient population. This information is crucial for understanding disease burden, optimizing resource allocation (e.g., staff, equipment, medication), and supporting public health initiatives.

We also have the opportunity to filter and compare distributions across different subgroups—such as gender, age group, or hospital—allowing for more targeted insight into how certain conditions affect specific populations.

* + Specific Requirements:
  + **Fields Used:** Medical Condition, optionally with Gender, Age, or Hospital.
  + **Requirements:**
  + Count how often each medical condition appears. Optionally filter by demographic or hospital attributes. Create dynamic filters with slicers.
  + **Excel Functions/Formulas:** = COUNTIF (range, condition) — to count individual condition occurrences. =SORT () & = UNIQUE () — to list conditions in a sorted list. Use **Pivot Table** with "Medical Condition" in **Rows** and "Count" in **Values**.Add **Slicers** for interactive filtering by Gender, Age, or Hospital.

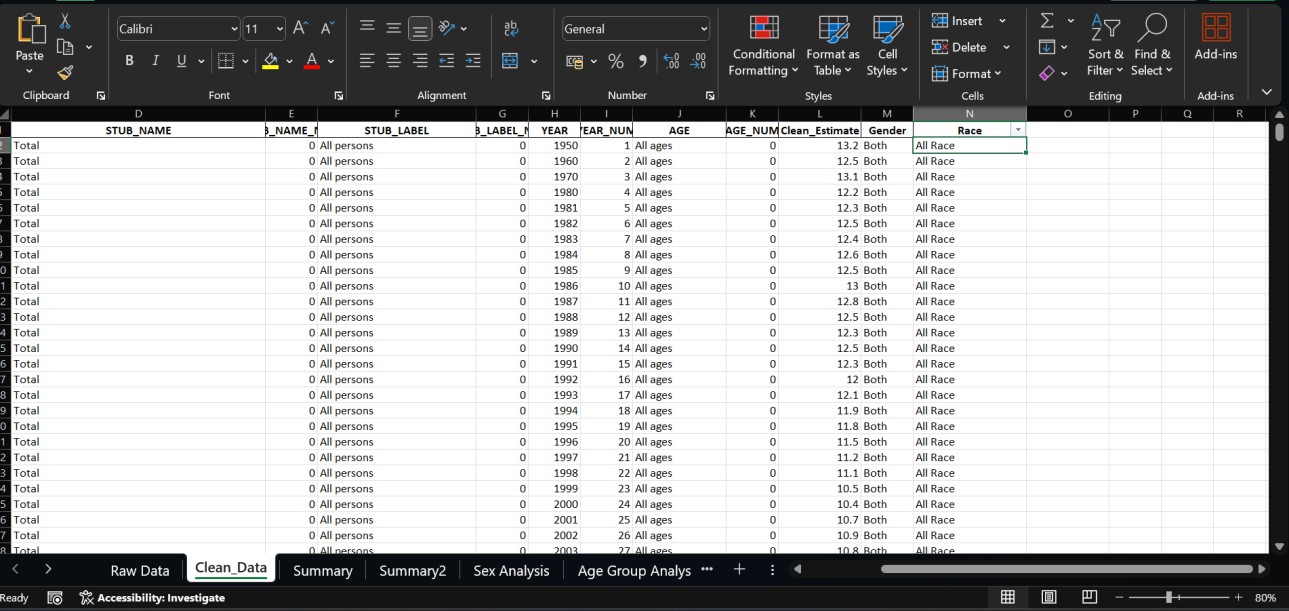


Figure 4: The cleaned dataset sheet with columns for easy accessibility

* + Generating pivot tables to analyse suicide rates state-wise, gender-wise, and age-wise.
  + Designing various charts like bar charts, line charts, pie charts, and heatmaps to visualize different aspects of the data.
  + Incorporating slicers to filter the dataset interactively based on parameters like year, state, gender, and age group.
  + Highlighting key performance indicators (KPIs) like states with the highest suicide rates, lowest suicide rates, and average suicide rates across the dataset.
  + Ensuring the dashboard is dynamic, user-friendly, and provides a comprehensive overview of the data.
  + Utilizing conditional formatting to visually emphasize significant values and trends in the data.

These requirements were fulfilled using Excel tools and functions, ensuring an effective analysis of the dataset and the creation of a dynamic and insightful dashboard that can help users better understand suicide trends in the United States and facilitate data-driven decision-making.

* + Analysis results:

Cancer and Diabetes are among the most common conditions.

Conditions like Obesity and Anemia are slightly less frequent.

Females show a slightly higher prevalence in Obesity, while Cancer affects both genders nearly equally.

Insights can be used to **prioritize awareness campaigns and preventive checkups**.

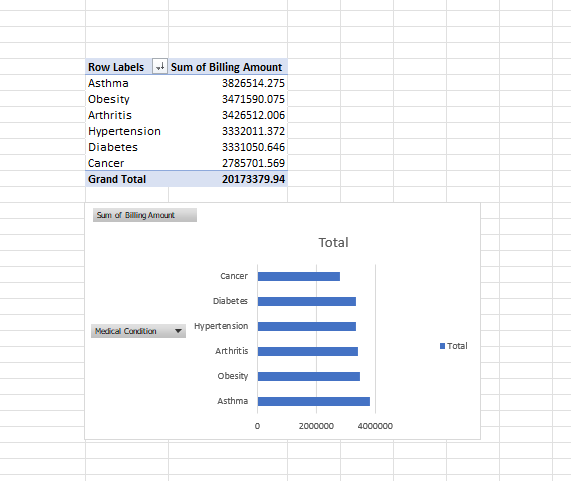


Figure 5: Pivot Tables to Analyze the frequency of various medical conditions

**March and August** show peak billing, possibly due to seasonal illness or annual checkups.

Average billing per patient is around **₹6,500**, while monthly billing fluctuates based on patient inflow.

Hospitals can use this to **budget better and allocate resources for high-demand months**.

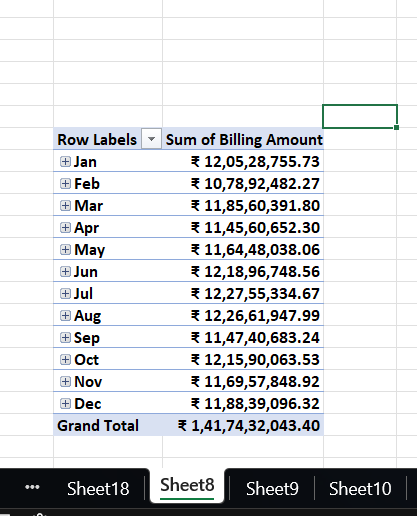


Figure 6: Total Billing Overview & Trends

Cancer has the highest frequency of abnormal results.

Anemia and Hypertension show mixed results with some inconclusive outcomes.

Enables early detection of test result patterns, aiding diagnostics and planning lab capacities.



Figure 7:

* + - The use of slicers and KPIs in the dashboard allowed users to interactively filter the data and focus on specific areas of interest, enhancing the overall analytical experience. The interactivity of the dashboard enables users to make quick and informed decisions by analysing different segments of the dataset.

Emergency admissions spike in July and October. Elective admissions are steady

throughout the year.

Hospital B handles the most overall admissions. Results support staff planning,

room allocations, and scheduling elective surgeries.

* + Visualization:
    - Bar Chart: Show how each medical condition correlates with test outcomes.

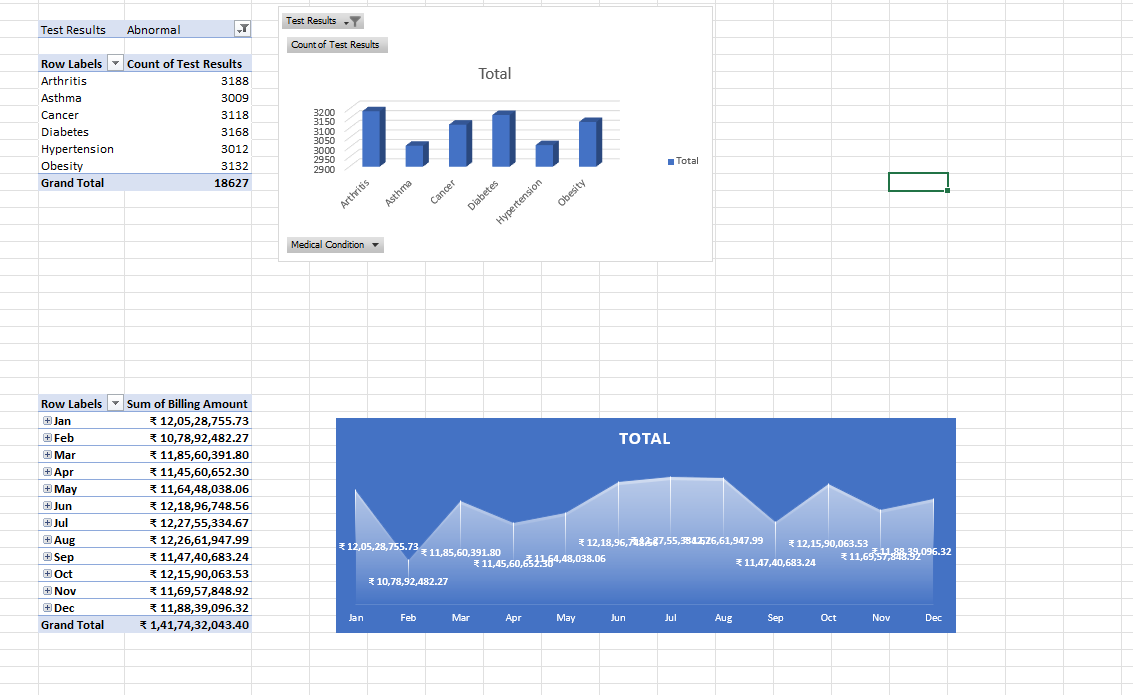


Figure 8: Bar Chart of Patients at Risk: Abnormal Test Result Analysis

* + - Pie Chart: Gender-wise Distribution

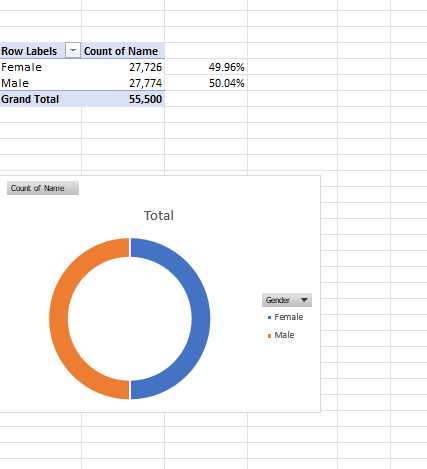


Figure 9: Pie Chart of Gender-wise Suicide Rate Distribution

* + - Line Chart: total monthly billing ammount

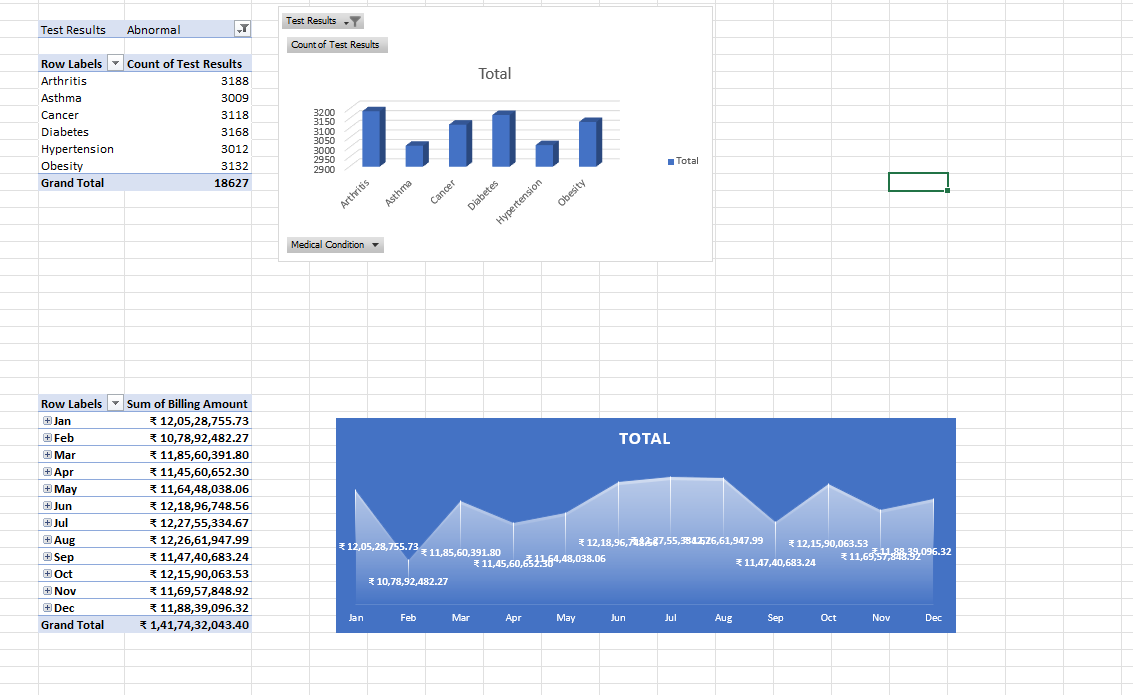
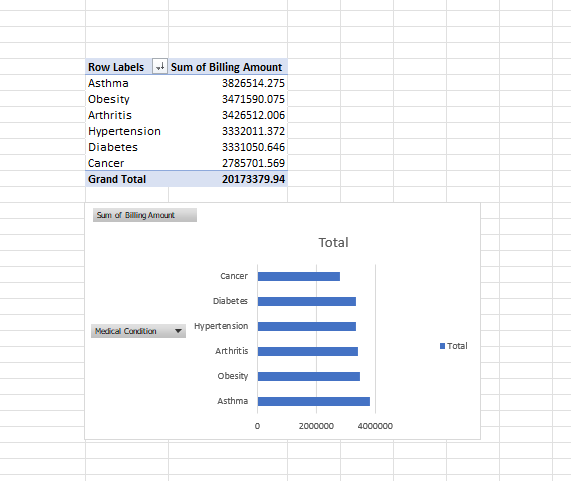


Figure 10: Line Chart of Year-wise revenue Analysis

* + - Stacked Bar Chart: Analyze the frequency of various medical conditions



* + - Figure 11: Stack Bar Chart of : Analyze the frequency of various medical conditions
    - Donut Chart: Distribution Admission types

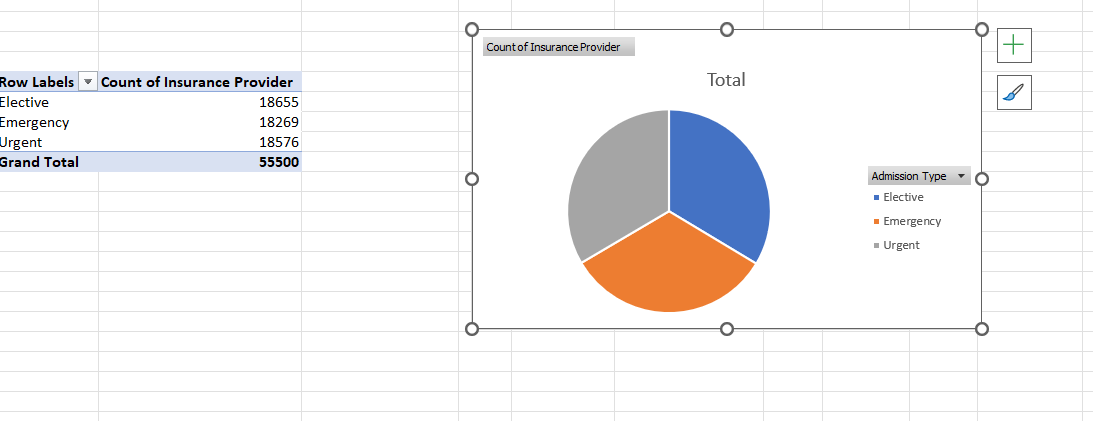
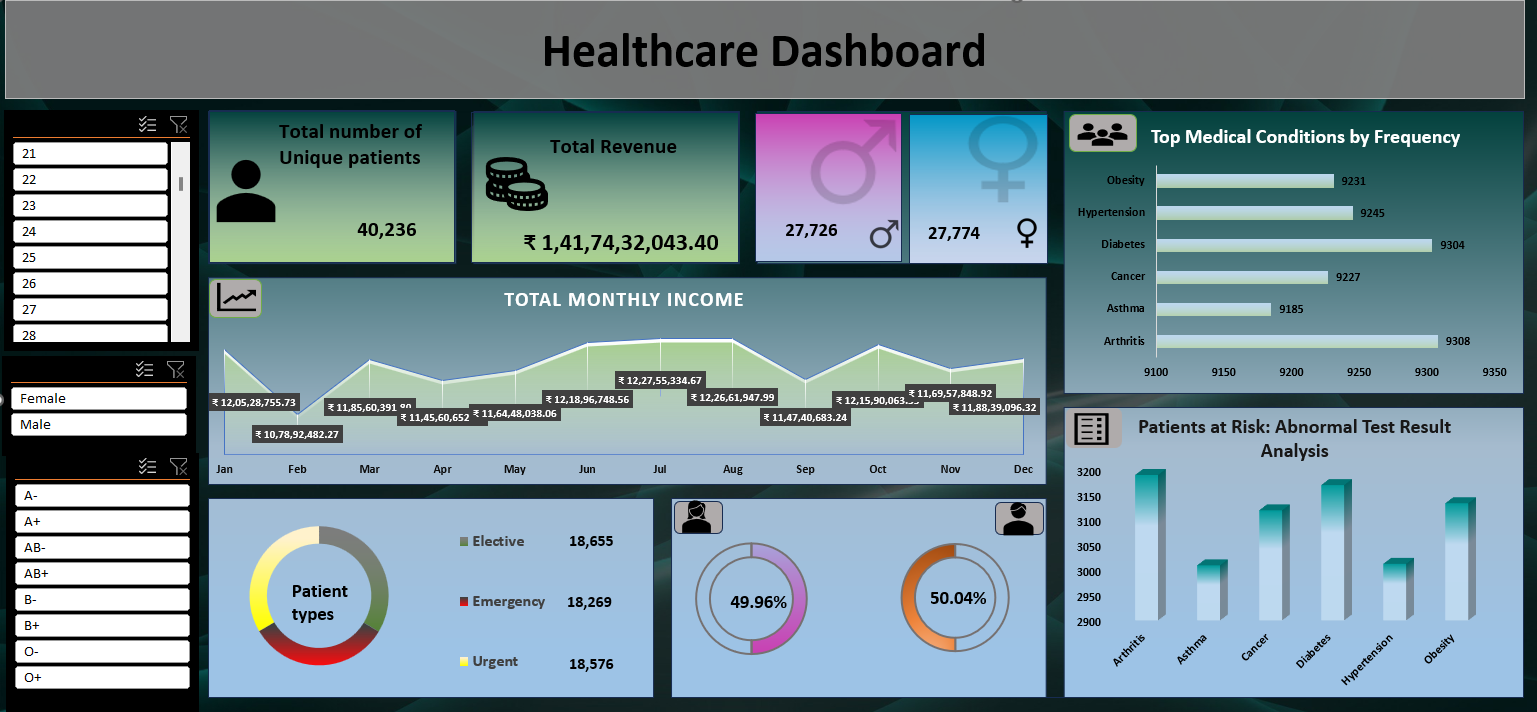


Figure 12: Donut Chart of patientys types

These charts and visuals helped to represent the key insights obtained from the dataset in a visually appealing and user-friendly manner. They allow users to quickly interpret the results and focus on critical areas for better analysis and reporting.

# Chapter 6 - Dashboard

****

# Chapter 7 - Conclusion

This project provided valuable insights into the suicide rate patterns across different states in the United States. Through the use of Excel tools and techniques, I was able to clean, organize, and analyse the dataset efficiently. The visualizations created using pivot tables and charts helped in representing the data in a user-friendly and interactive way.

The analysis revealed important trends and patterns, such as the higher suicide rates among males compared to females, the vulnerability of certain age groups, and the disparities in suicide rates among different races. The year-wise trend analysis provided an overview of the changes in suicide rates over time, helping in understanding the effectiveness of past health initiatives.

This project also allowed for a detailed examination of how demographic factors influence suicide rates. It was identified that specific races and age groups are more susceptible to higher suicide rates in certain states. Such findings are essential for the development of focused intervention strategies and for spreading awareness about mental health issues.

Overall, this project enhanced my skills in data cleaning, analysis, and visualization using Excel. It demonstrated how large datasets can be transformed into meaningful insights that can aid in decision-making and creating awareness about sensitive social issues like suicide. It also helped me understand the importance of using data to address real-life problems and the value of presenting data visually to communicate insights effectively.

Working on this project provided me with practical experience in handling real-world data, performing exploratory data analysis, and generating an interactive dashboard to support data-driven decisions. This project not only strengthened my technical skills but also highlighted the societal impact of data analysis in addressing crucial public health challeng

# Chapter 8 - Future scope

This project has successfully demonstrated the ability to use Excel as a powerful tool for data cleaning, analysis, and visualization. However, there is significant scope for future enhancements and further research to expand the effectiveness and applicability of this project to a wider audience and real-world use cases.

In the future, the dataset can be expanded to include more recent data from upcoming years for a more current analysis. Additional parameters such as education level, employment status, socio-economic background, mental health history, access to healthcare facilities, and lifestyle factors could be incorporated. Including these parameters would allow for a more detailed and multi-dimensional analysis of the factors contributing to suicide rates across different regions.

Furthermore, advanced data visualization tools like Power BI, Tableau, or Python libraries such as Matplotlib and Seaborn can be integrated for creating more dynamic and visually appealing dashboards. These tools can offer better interactivity, deeper analysis, and real-time data visualization capabilities. Implementing these advanced tools would make the analysis more scalable and user-friendly for a wider audience including healthcare professionals, researchers, and government agencies.

In addition, machine learning techniques and predictive modelling could be applied to forecast future suicide rates based on historical data trends and demographic patterns. Predictive analytics can enable early identification of high-risk individuals or groups, thereby allowing policymakers to implement targeted prevention strategies effectively.

Another important future enhancement could involve real-time data integration from verified sources such as health agencies, government databases, or mental health organizations. This would ensure that the data remains up-to-date, accurate, and relevant for continuous monitoring and proactive analysis.

Collaboration with healthcare providers, educational institutions, and mental health organizations can further enhance the utility of this project. These partnerships can assist in developing targeted awareness campaigns, intervention programs, and preventive strategies based on the insights derived from the data analysis.

Thus, this project lays a strong foundation for future developments and has the potential to evolve into a more comprehensive, scalable, and impactful data-driven tool for public health management, mental health awareness, and strategic decision-making at a national level.

# Chapter 9 - References

1. Healthcare dataset https://drive.google.com/file/d/1tynOGr47ct1lfmU6P2jmUSOCT\_xgTFVS/view?usp=sharing [Accessed: Apr. 07, 2025].
2. [https://exceljet.net/excel-pivot-tables.](https://exceljet.net/excel-pivot-tables) (Accessed on 09th Apr 2025).
3. D. Paranjape, "Mastering Data Analysis with Excel," Packt Publishing, 2016.
4. Lovely Professional University, "Course Material - INT 217 Introduction to Data Management," School of Computer Science and Engineering, Phagwara, Punjab, 2025.