2.2 Introduction To Power BI

Power BI is a robust business analytics tool developed by Microsoft, offering a comprehensive suite for data visualization, interactive dashboards, and business intelligence solutions. It empowers users to effortlessly connect to various data sources, transform raw data into meaningful insights through intuitive visualizations, and share these insights across organizations.

2.2.1 Steps to Create Power BI Dashboard:

Task 1. Data Preparation Process

Prior to importing the data into Power BI for dashboard creation, a preprocessing step was undertaken using Python's pandas library. The objective was to refine the dataset to include only relevant information pertaining to Infectious Disease Screening and Specific Tests, thereby streamlining the subsequent analysis process.

Methodology

Data Extraction: The original dataset, sourced from the hospital services records, was in the form of an Excel spreadsheet encompassing diverse medical services and contain 62310 Rows and 17 Columns.

S.No	PPP ID	UHID	Patient Name	Gender	Age	Mobile Number	Visit Type	Service Name	Bill Date and Time	Sample Collection Date and Time	Result Entry Date and Time	Sign Date and Time	Referral Type	Facility	Patient Category	Patient Scheme
1				Female	57 Years		IP	Complete Blood Count (CBC)	01/12/2023 00:53:34	01/12/2023 00:55:51	01/12/2023 01:47:00	01/12/2023 01:47:26			General	MMIY
2				Female	57 Years		IP	HBsAg	01/12/2023 00:53:34	01/12/2023 00:55:52	01/12/2023 11:54:00	01/12/2023 11:54:46			General	MMIY
3				Female	57 Years		IP	Hepatitis C (HCV)	01/12/2023 00:53:34	01/12/2023 00:55:52	01/12/2023 11:54:00	01/12/2023 11:54:46			General	MMIY
4				Female	57 Years		IP	HIV Test	01/12/2023 00:53:34	01/12/2023 00:55:52	01/12/2023 14:33:00	01/12/2023 14:33:06			General	MMIY
5				Female	57 Years		IP	Lipid Profile	01/12/2023 00:53:34	01/12/2023 00:55:51	01/12/2023 06:49:00	01/12/2023 06:49:32			General	MMIY
6				Female	57 Years		IP	Liver Function Test (LFT)	01/12/2023 00:53:34	01/12/2023 00:55:51	01/12/2023 06:49:00	01/12/2023 06:49:32			General	MMIY
7				Female	57 Years		IP	Renal Function Test (RFT)	01/12/2023 00:53:34	01/12/2023 00:55:51	01/12/2023 06:49:00	01/12/2023 06:49:32			General	MMIY
8				Female	57 Years		IP	Serum Electrolytes	01/12/2023 00:53:34	01/12/2023 00:55:51					General	MMIY
9				Female	57 Years		IP	VDRL Test	01/12/2023 00:53:34	01/12/2023 00:55:52	01/12/2023 11:54:00	01/12/2023 11:54:46			General	MMIY
10				Female	57 Years		IP	Widal Test	01/12/2023 00:53:34	01/12/2023 00:55:52	01/12/2023 04:02:00	01/12/2023 04:02:36			General	MMIY
11				Female	50 Years		IP	Complete Blood Count (CBC)	01/12/2023 00:57:00	01/12/2023 01:00:56	01/12/2023 03:02:00	01/12/2023 03:02:37			General	Ayushman Bharat
12				Female	50 Years		IP	Liver Function Test (LFT)	01/12/2023 00:57:00	01/12/2023 01:00:56	01/12/2023 06:50:00	01/12/2023 06:50:16			General	Ayushman Bharat

Data Refinement: Utilizing pandas, a powerful data manipulation tool in Python, the dataset underwent a selective process. Specifically, the dataset was parsed to extract and isolate data entries relevant to Infectious Disease Screening and Specific Tests. This involved filtering techniques within pandas to extract pertinent rows and columns based on predetermined criteria.

Data Subset Creation: Following the extraction phase, a refined subset of the dataset was generated, containing exclusively the required data for analysis in the context of the dashboard.

```python
import pandas as pd
# Read the Excel file
df = pd.read_excel("your_file.xlsx")

# Define the lists of tests for Group 1 and Group 2
group_1_tests = ["HBsAg", "Hepatitis C (HCV)", "HIV Test", "VDRL Test", "Widal Test",
                 "Dengue Serology", "Malaria", "Chikungunya", "Stool Routine",
                 "Pleural Fluid AFB", "Pleural Fluid Culture", "Pleural Fluid Protein",
                 "Pleural Fluid Routine", "Stool M/E", "Stool for Occult Blood",
                 "Blood Culture and Sensitivity", "Bleeding Time (BT)",
                 "Clotting Time (CT) - Activated", "Pap Smear", "Pus Culture and Sensitivity",
                 "AFB Sputum", "Stool Culture and Sensitivity", "Pleural Fluid Sensitivity",
                 "Blood - Serological Test", "HVS Culture and Sensitivity", "AFB Culture",
                 "Ascitic Fluid AFB", "Ascitic Fluid Protein", "Ascitic Fluid Routine",
                 "Ascitic Fluid Culture", "Cerebrospinal Fluid (CSF) - Culture",
                 "Cerebrospinal Fluid (CSF) - Sensitivity", "Cerebrospinal Fluid (CSF) - AFB",
                 "H1N1 (Swine Flu) Test", "CA 19.9", "Hepatitis A"]

# Filter data for Group 1
group_1_data = df[df["Service Name"].isin(group_1_tests)]

# Filter data for Group 2
group_2_data = df[~df["Service Name"].isin(group_1_tests)]

# Write Group 1 data to Excel
group_1_data.to_excel("group_1_data.xlsx", index=False)

# Write Group 2 data to Excel
group_2_data.to_excel("group_2_data.xlsx", index=False)
```

Outcome

Through the application of pandas, the initial dataset underwent a transformation, resulting in a tailored subset specifically tailored to the focus areas of Infectious Disease Screening and Specific Tests. This streamlined approach ensured that the subsequent analysis within Power BI was conducted on a dataset precisely aligned with the objectives of the project, enhancing the efficiency and relevance of the dashboard insights.

After splitting data only for Infectious Disease Screening and Specific Tests it contains 11248 Rows.

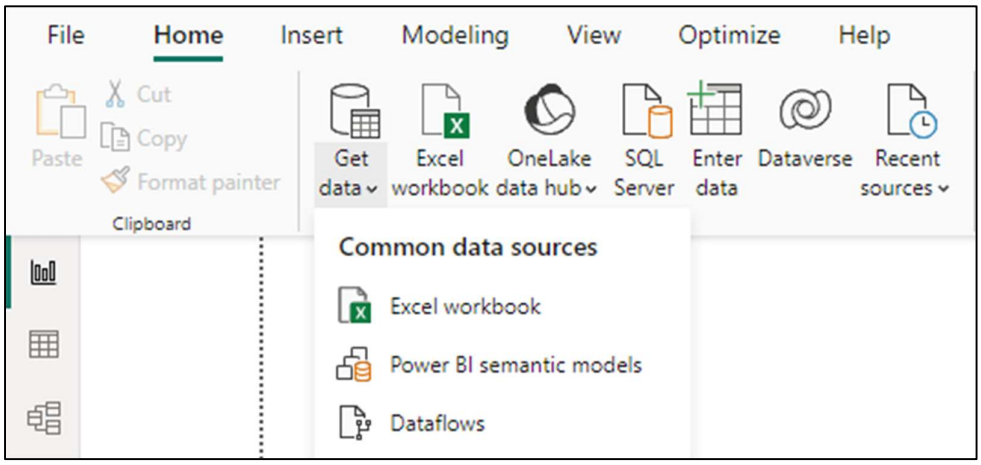
Infectious Disease Screening and Specific Tests are : ["HBsAg", "Hepatitis C (HCV)", "HIV Test", "VDRL Test", "Widal Test", "Dengue Serology", "Malaria", "Chikungunya", "Stool Routine", "Pleural Fluid AFB", "Pleural Fluid Culture", "Pleural Fluid Protein, "Pleural Fluid Routine", "Stool M/E", "Stool for Occult Blood", "Blood Culture and Sensitivity", "Bleeding Time (BT)", "Clotting Time (CT) - Activated", "Pap Smear", "Pus Culture and Sensitivity", "AFB Sputum", "Stool Culture and Sensitivity", "Pleural Fluid Sensitivity", "Blood - Serological Test", "HVS Culture and Sensitivity", "AFB Culture", "Ascitic Fluid AFB", "Ascitic Fluid Protein", "Ascitic Fluid Routine", "Ascitic Fluid Culture", "Cerebrospinal Fluid (CSF) - Culture", "Cerebrospinal Fluid (CSF) - Sensitivity", "Cerebrospinal Fluid (CSF) - AFB", "H1N1 (Swine Flu) Test", "CA 19.9", "Hepatitis A"]

Task 2. Data Import into Power BI

Data import into Power BI refers to the process of bringing external data sources into the Power BI environment for analysis, visualization, and reporting. Steps for Data Import into Power BI:

I. Open Power BI Desktop: Launch the Power BI Desktop application on your computer.

II. Click on "Get Data": In the Home tab of Power BI Desktop, click on the "Get Data" button located in the ribbon at the top.



III. Select Data Source: Choose the appropriate data source from the list of available options (e.g., Excel, SQL Server, CSV file, etc.). Click on the desired data source to initiate the import process.

IV. Connect to Data Source: Depending on the selected data source, you may need to provide connection details such as file location, server name, database credentials, etc. Follow the prompts to establish a connection to the chosen data source.

V. Transform and Load Data (Optional): Optionally, you can perform data transformation tasks using Power Query Editor to clean, reshape, and manipulate the imported data before loading it into Power BI. Once the desired transformations are applied, click on "Load" to import the transformed data into Power BI.

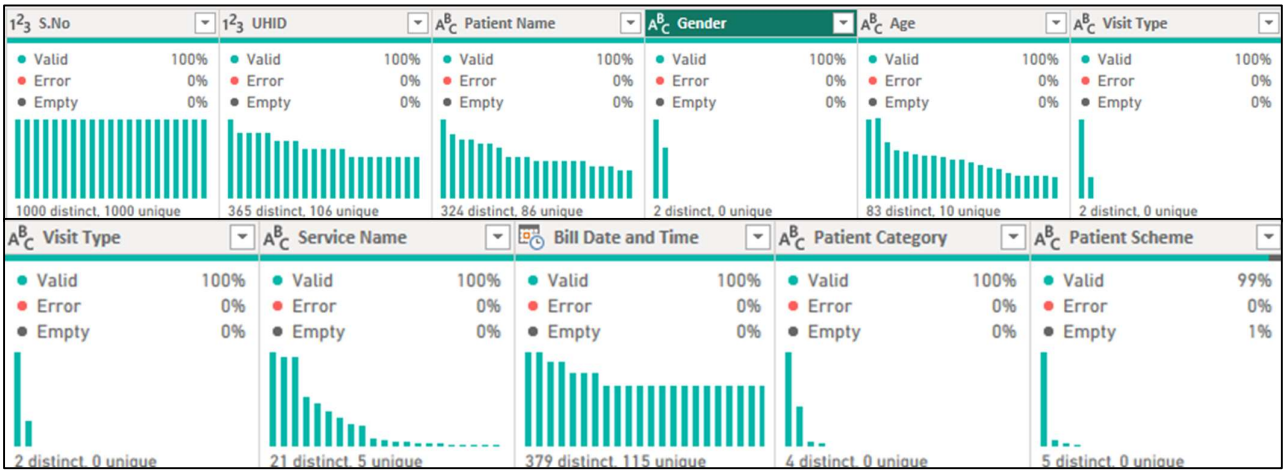
Task 3. Data Transformation

Following the import of the refined dataset into Power BI, the data underwent a series of transformation steps to enhance its quality and usability for dashboard analysis. The transformation process involved the following steps:

I. Exclusion of Unnecessary Columns:

Columns deemed irrelevant to the analysis or containing redundant information were excluded from the dataset to streamline the data structure and focus on pertinent variables.

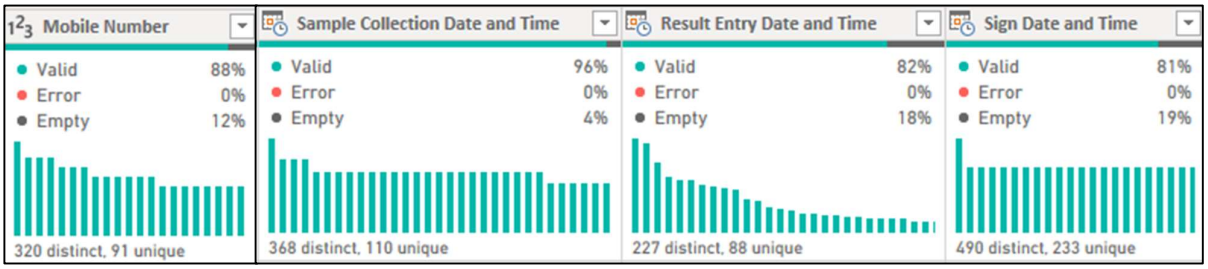
I removed several columns from the dataset that were deemed irrelevant for the dashboard analysis or contained redundant information. These columns were not conducive to the specific focus of this analysis and were therefore excluded to streamline the data structure and emphasize pertinent variables. As a result, the refined dataset now contains only the columns essential for the intended analysis, ensuring clarity and efficiency in data interpretation within



the dashboard.

Now the data sheet contains only 10 columns.

II. Removal of Null Values:



Null values, if present, were identified and subsequently removed from the dataset to ensure data integrity and accuracy.

1. Null values were primarily found in columns deemed unnecessary for the analysis, leading to their removal.

2. However, in the 'Patient Scheme' column, null values persisted but accounted for only around 1% of the total data.
3. Due to their minimal prevalence, null values in the 'Patient Scheme' column were retained.
4. Removing rows with null values risked losing crucial information, thus justifying their retention.
5. Retaining these null values in the 'Patient Scheme' column was deemed acceptable, considering their insignificant impact on the analytical outcomes.

III. Elimination of Duplicate Entries:

Duplicate entries within the dataset were identified and eliminated to prevent redundancy and maintain data consistency.

1. Notably, duplicates were observed in the age section, such as '2 Month' and '2 Months'.

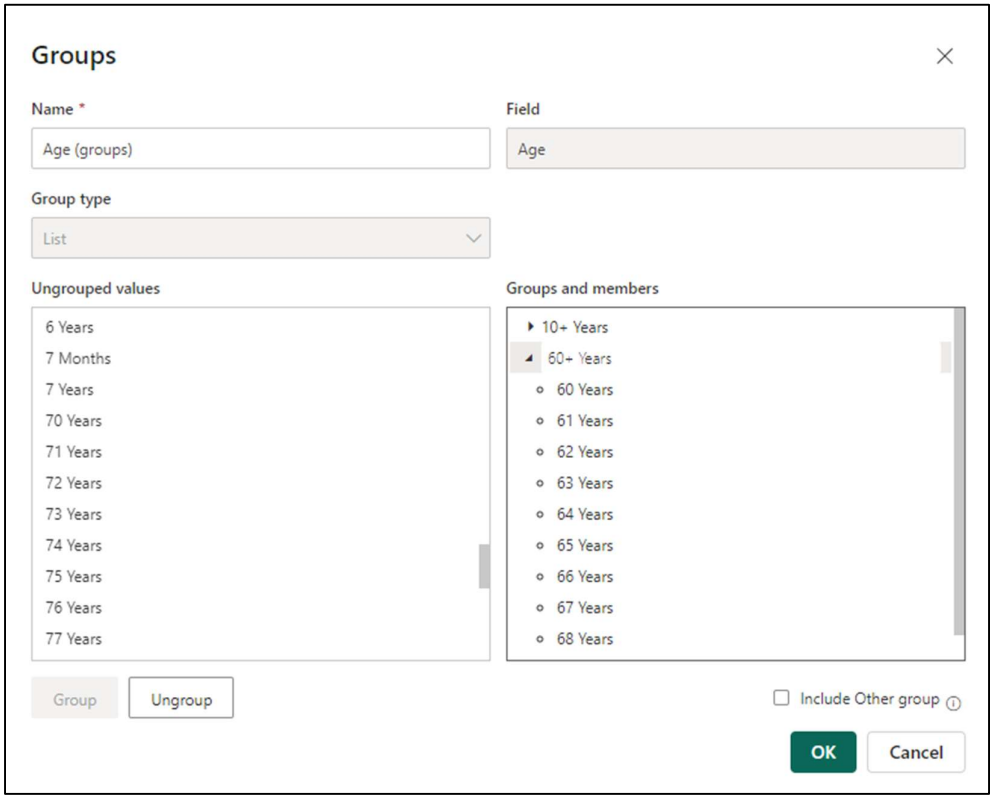
2. These inconsistencies were rectified through a systematic review and deduplication process.
3. By addressing such duplications, the dataset's integrity and consistency were upheld, ensuring accurate analysis outcomes.

IV. Creation of Age Groups:

A categorical column representing age groups was generated to categorize individuals into broader age ranges for analytical purposes. This involved grouping similar age intervals (e.g., 10-19 years, 20-29 years) into broader categories (e.g., 10+, 20+ years) to simplify analysis and interpretation.

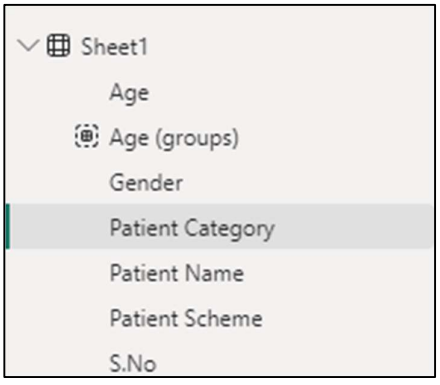
1. Definition of Age Groups: Define the age groups based on the requirements of the analysis. For example, age groups could be defined as 0-9 years, 10-19 years, 20-29 years, and so on.

2. Grouping Similar Age Intervals: Group similar age intervals together to create broader categories. For instance, combine age ranges such as 10-19 years, 20-29 years, etc., into



broader categories such as 10+, 20+, etc.

3. Creation of New Column: Create a new column in the dataset to represent the age groups.



4. Mapping Age Values to Age Groups: Map each individual's age to the corresponding age group based on the defined criteria. For example, an individual aged 15 would be mapped to the '10+' age group.

V. Aggregation of Hospital Services:

Similar types of hospital services were grouped together to facilitate a more cohesive analysis. This involved categorizing related services under broader categories to provide a comprehensive overview of healthcare offerings.

1. Analysis of Hospital Services: Analyze the hospital services to identify similar types that can be grouped together. This may involve reviewing service descriptions, codes, or other relevant information.

Hepatitis Tests:
- HBsAg
- Hepatitis C (HCV)
- Hepatitis A

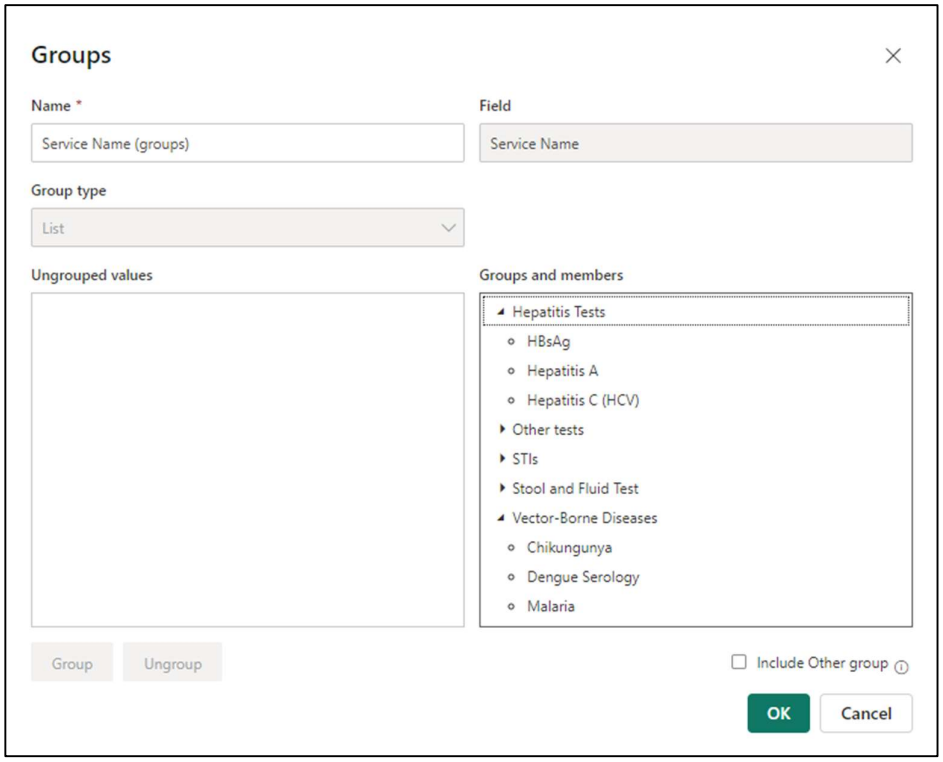
HIV and Sexually Transmitted Infections (STIs):
- HIV Test
- VDRL Test

Vector-Borne Diseases:
- Dengue Serology
- Malaria
- Chikungunya

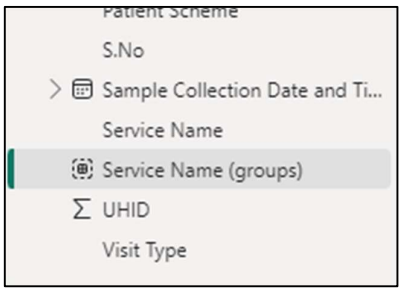
Other Tests:
- Bleeding Time (BT)
- Clotting Time (CT) - Activated
- Blood - Serological Test
- HVS Culture and Sensitivity
- AFB Culture
- H1N1 (Swine Flu) Test
- CA 19.9

Stool and Fluid Tests:
- Stool Routine
- Stool M/E
- Stool for Occult Blood
- Stool Culture and Sensitivity
- Pleural Fluid AFB
- Pleural Fluid Culture
- Pleural Fluid Protein
- Pleural Fluid Routine
- Pleural Fluid Sensitivity
- Blood Culture and Sensitivity
- Pap Smear
- Pus Culture and Sensitivity
- AFB Sputum
- Ascitic Fluid AFB
- Ascitic Fluid Protein
- Ascitic Fluid Routine
- Ascitic Fluid Culture
- Cerebrospinal Fluid (CSF) - Culture
- Cerebrospinal Fluid (CSF) - Sensitivity
- Cerebrospinal Fluid (CSF) - AFB

2. Definition of Service Categories: Define broader categories or groups based on the identified similarities among hospital services. For example, services related to diagnostic tests, medical procedures, consultations, etc., can be grouped into distinct categories.
3. Grouping Similar Services: Group similar hospital services together under the defined categories. This may involve manually grouping services or using automated techniques based on predefined criteria.
4. Creation of New Column: Create a new column in the dataset to represent the aggregated service categories.



5. Mapping Services to Categories: Map each hospital service to the corresponding aggregated category based on the defined criteria. For example, services such as blood tests, urine tests, and imaging tests may be mapped to the 'Diagnostic Tests' category.



Outcome

Through the implementation of these data transformation steps, the dataset underwent a refinement process, resulting in a cleaner, more structured dataset optimized for dashboard analysis within Power BI. By addressing issues such as null values, duplicates, unnecessary columns, and the creation of meaningful categorical variables, the transformed dataset provides a solid foundation for deriving valuable insights through the subsequent dashboard development phase. Total Columns 10 + 2 group Columns = 12 Columns.

Task 4. Creating Data Visualizations

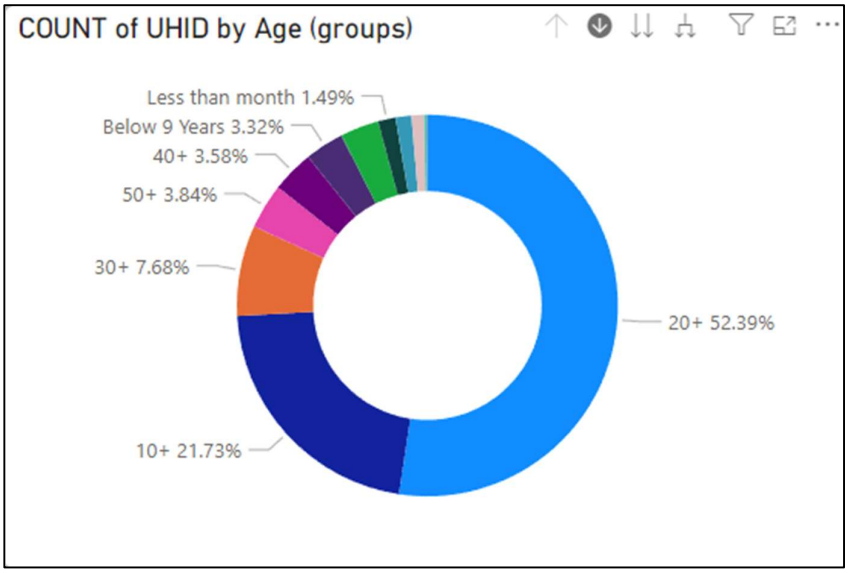
With the refined and transformed dataset now prepared, the next phase involves the creation of data visualizations using Power BI. Data visualization serves as a powerful tool for conveying insights and trends hidden within the dataset, enabling stakeholders to make informed decisions. The process of creating data visualizations encompasses the following steps:

Selection of Visualization Types: Determine the most suitable types of visualizations to represent the dataset effectively. This may include bar charts, line graphs, pie charts, scatter plots, and others, depending on the nature of the data and the insights to be conveyed.

1. Donut Chart:

A donut chart is a circular visualization with a hole in the center, similar to a pie chart. It is used to display categorical data and show the proportion of each category relative to the whole.

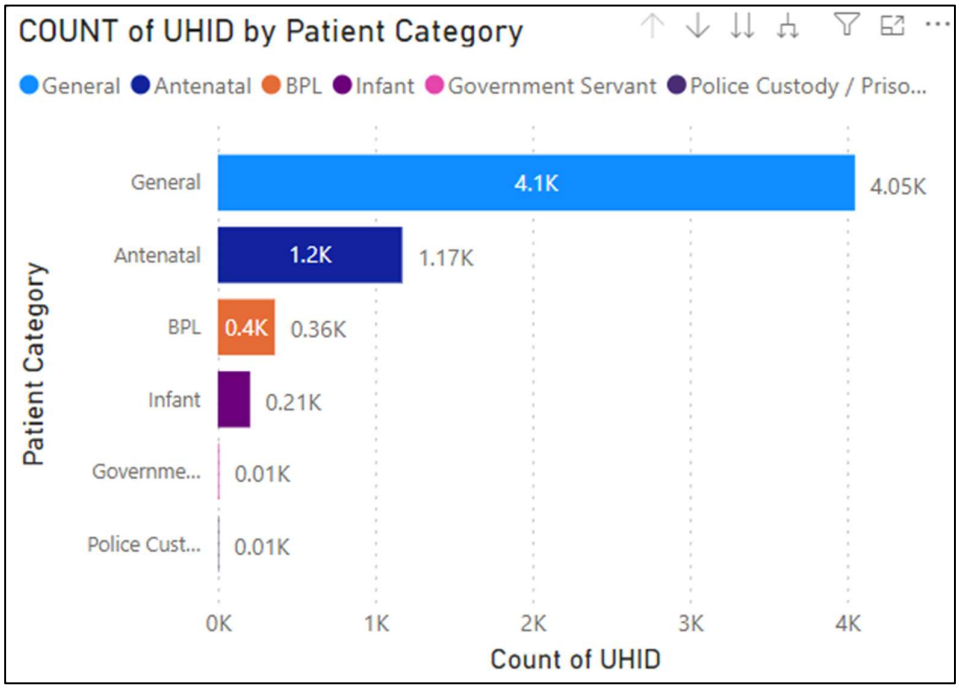
In this dashboard, the donut chart is utilized to represent age groups, where the age is broken down



into groups. Each age group is represented by a segment in the donut, and the count of UHID (Unique Hospital Identification) is displayed for each segment. Additionally, the visit type (such as inpatient or outpatient) is represented as the legend, providing further insight into the distribution of visits within each age group.

2. Stacked Bar Chart:

A stacked bar chart is a bar chart that displays multiple data series stacked on top of each other. It is useful for comparing the total sizes across categories and understanding the composition of each category.

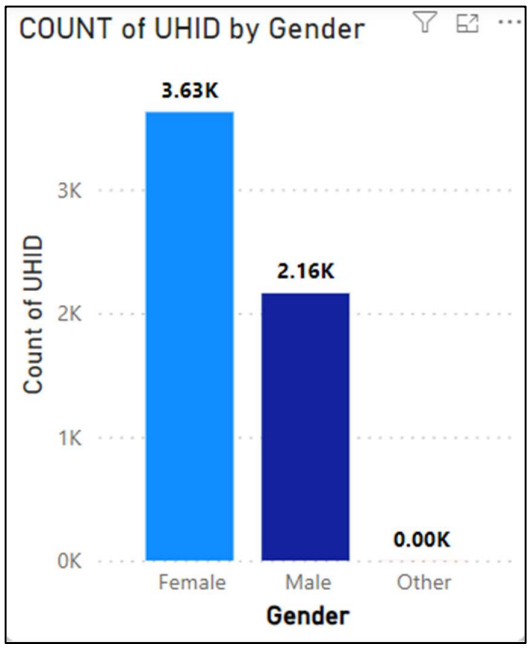


In this dashboard, the stacked bar chart is used to visualize patient categories stacked by patient scheme. Each bar represents a patient category, and the height of the bar represents the total count of UHID for that category. Within each bar, different colors represent the breakdown of patient schemes, allowing stakeholders to understand the distribution of patients within each category.

3. Stacked Column Chart:

Similar to the stacked bar chart, a stacked column chart displays data series stacked on top of each other, but in a vertical orientation.

In this dashboard, the stacked column chart is employed to illustrate the distribution of genders. Each column represents a gender (such as male or female), and the height of the column represents the count of UHID for that gender. The columns are stacked on top of each other to show the total count of UHID across genders.



4. Matrix:

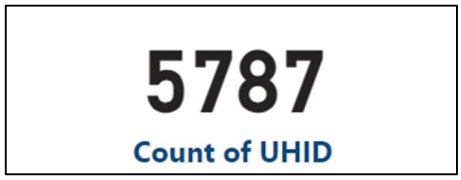
A matrix is a tabular visualization that displays data in rows and columns, similar to a pivot table. It is useful for comparing multiple categories simultaneously and providing a detailed breakdown of data.

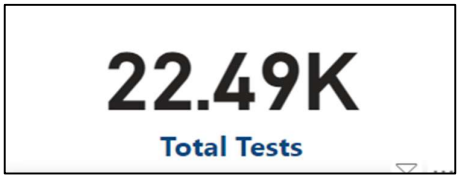
Service Name (groups)	Count of UHID
Vector-Boner Diseases	**4064**
Widal Test	1614
Malaria	1232
Dengue Serology	868
Chikungunya	350
Stool Tests	**252**
Other Tests	**1974**
HIV & STIs	**7024**
VDRL Test	2512
HIV Test	4512
Hepatitis Tests	**8268**
Hepatitis C (HCV)	4148
Hepatitis A	2
HBsAg	4118
Fluid Tests	**912**
Total	**22494**

In this dashboard, the matrix is used to present a detailed breakdown of service names grouped by service name (group). Each cell in the matrix represents the count of UHID for a specific combination of service name (group) and service name.

5. Cards:

Cards are single-value visualizations that display a specific metric or KPI (Key Performance Indicator). They are useful for highlighting key metrics or summarizing important information.





In this dashboard, two cards are included: one card displays the distinct count of UHID, providing an overview of unique hospital identifications in the dataset, and the other card showcases the total number of tests conducted, summarizing the overall test count.

6. Slicers:

Slicers are interactive visualizations that allow users to filter data dynamically by selecting specific values or ranges.

In this dashboard, two slicers are incorporated: one slicer is used to filter between inpatient and outpatient data, allowing stakeholders to focus on a specific type of visit, and the other slicer enables users to select a date range for analysis, facilitating temporal analysis of the data.

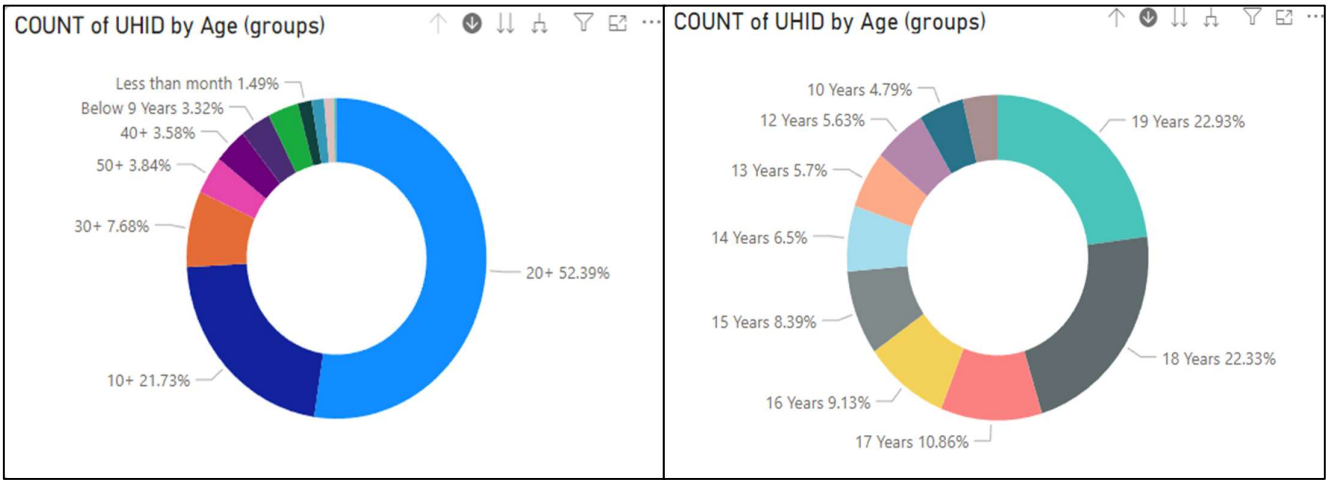
These visualization choices were made to provide a comprehensive overview of the dataset, allowing stakeholders to gain insights into age distribution, patient categories, gender distribution, service utilization, and overall test counts. The combination of visualizations and interactive slicers facilitates data exploration and analysis, empowering stakeholders to make informed decisions based on the data presented.

Task 5. Enhancing Data Exploration with Drill-Down Functionality

In each chart, data drill-down functionality enhances the user experience by allowing them to explore the dataset at different levels of granularity and gain deeper insights into the data. This interactivity empowers users to investigate specific trends or anomalies and make informed decisions based on the data presented.

I. Drill into Donut Chart:

In the donut chart, users can drill down from the overall age group to more granular data by clicking on a specific segment representing an age group. For example, if a user clicks on the segment representing the "18-30" age group, they can drill down to see more detailed information about patients in that age group, such as the distribution of visit types (e.g., inpatient or outpatient) within that age group.



Age (Group) -> Age: Data Drill into 10+

II. Drill into Stacked Bar Chart:

In the stacked bar chart, users can drill down from the overall patient category to explore the composition of each category by patient scheme. By clicking on a specific bar representing a patient category, users can drill down to view the distribution of patients within that category, broken down by different patient schemes.