Sri Lanka Institute of Information Technology

IT3021 – Data Warehousing and Business Intelligence

Continuous Assignment – 2024, Semester 1

Assignment 2

IT Number – IT22235138

Name – Thalangama A.I

1. Table of Contents

2. Data source

Data Warehouse Structure

This data warehouse is designed to store and analyze healthcare-related data efficiently. It follows a snowflake schema to support data integration and reporting across multiple dimensions.

Fact Table :

FactTreatment - Stores measurable healthcare events such as treatment cost, duration, outcomes, and links to dimension tables via foreign keys.

Dimention Tables –

DimDate - Stores date-related attributes for various treatment timestamps (start, completion, outcome).

DimPatient - Contains demographic information of patients including gender and age.

DimProvider - Includes details of healthcare providers and their affiliated hospitals.

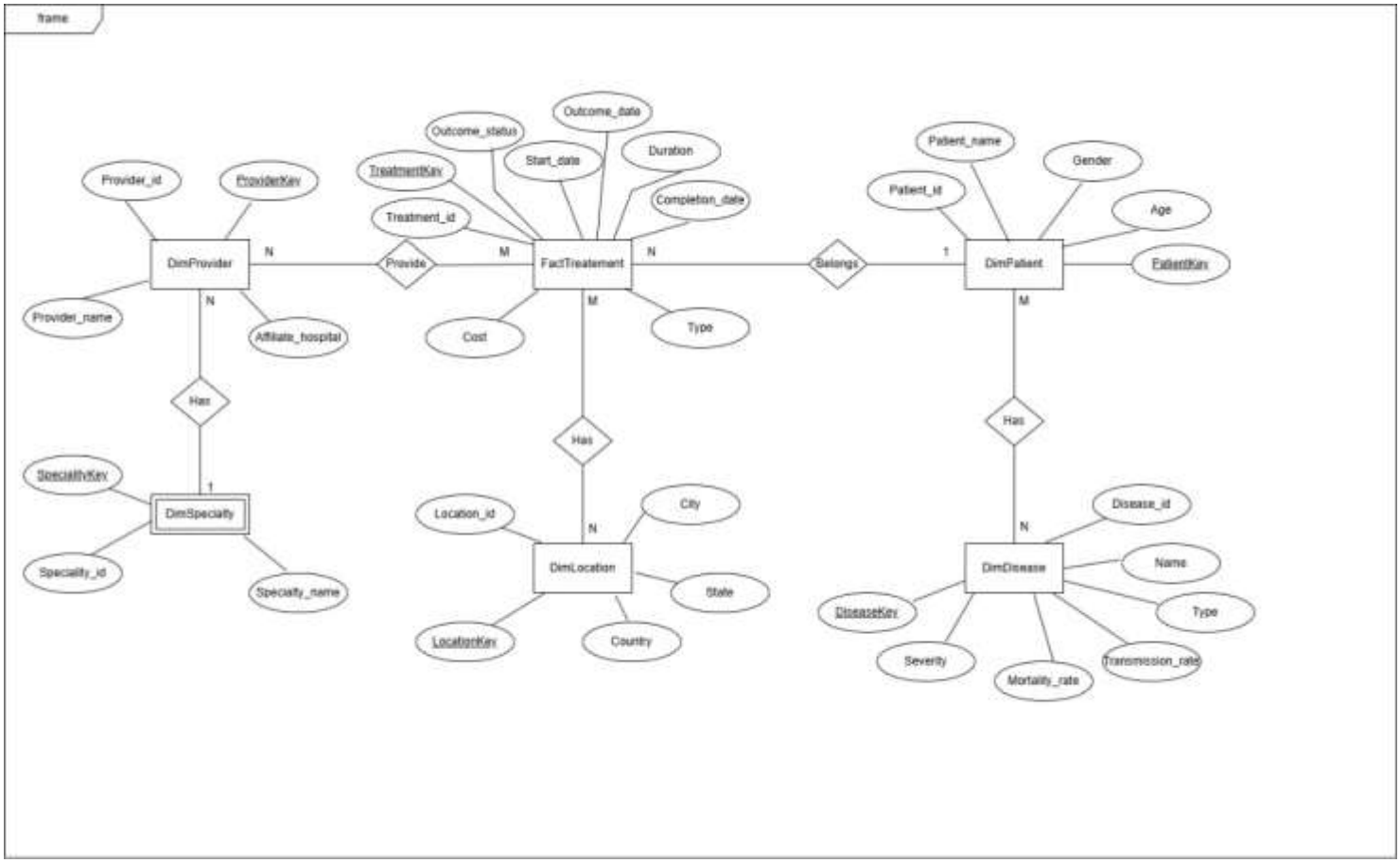
DimSpeciality - Describes provider specializations used to treat diseases.

DimDisease - Contains disease metadata including severity, type, and transmission mode.

DimLocation - Holds location information including country, state, and city where treatments occurred.

Supports Slowly Changing Dimensions (SCD) Type 2 for historical tracking of changes in dimensions such as Speciality, Provider, and Location. Includes metadata columns such as StartDate, EndDate, InsertDate, and ModifiedDate for audit and change tracking.

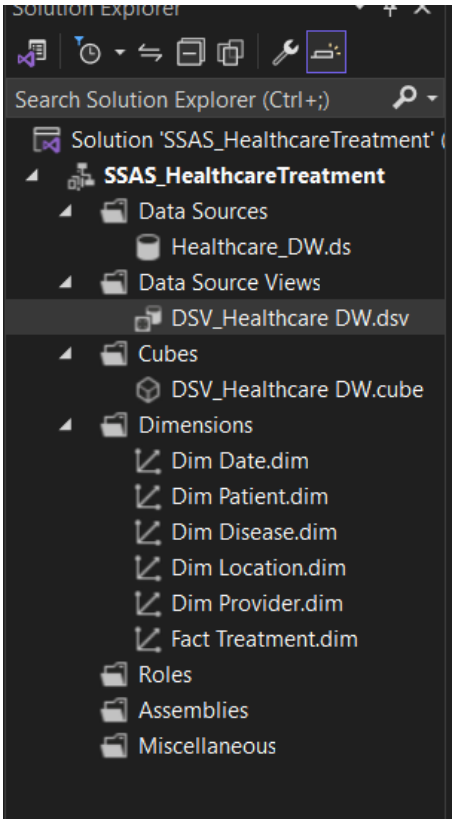
ER Diagram


frame
Provider_id
ProviderKey
DimProvider
Provider_name
Affiliate_hospital
N
N
Has
SpecialityKey
DimSpecialty
Speciality_id
Specialty_name
1
Provide
M
TreatmentKey
Treatment_id
Outcome_status
Outcome_date
Start_date
Duration
Completion_date
FactTreatement
Cost
Type
N
M
Has
Location_id
City
DimLocation
State
LocationKey
Country
N
Belongs
1
Patient_name
Patient_id
Gender
Age
DimPatient
PatientKey
M
Has
Disease_id
Name
DimDisease
Type
DiseaseKey
Severity
Mortality_rate
Transmission_rate
N

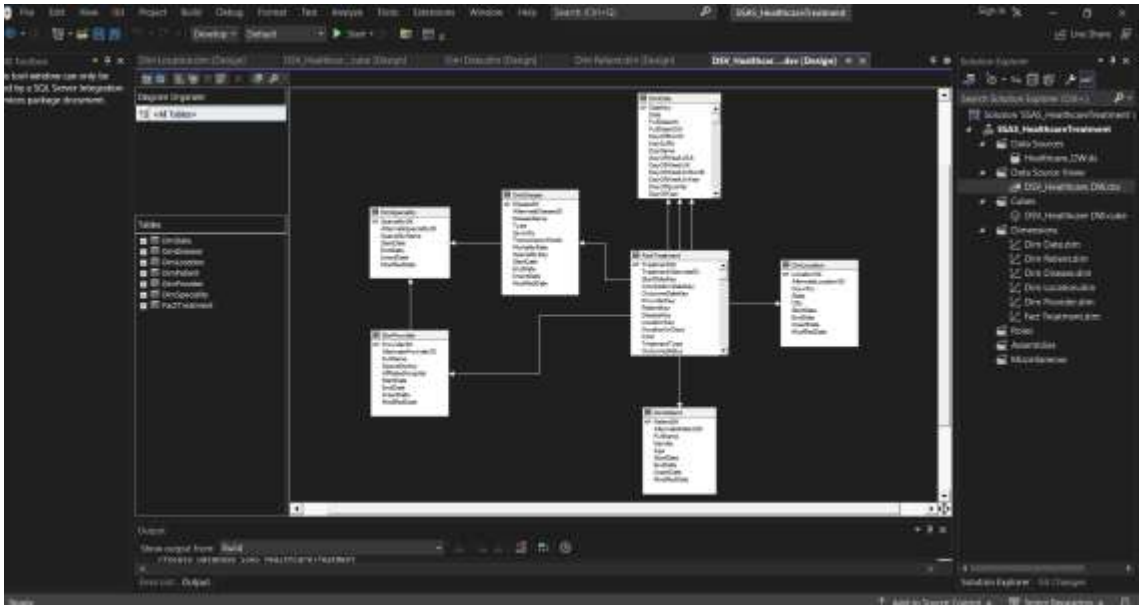
3. SSAS Cube implementation

Step 1 - Created a new Analysis Services Multidimensional and Data Mining Project named SSAS_HealthcareTreatment.

Step 2 - Connected to the SQL Server database Healthcare_DW



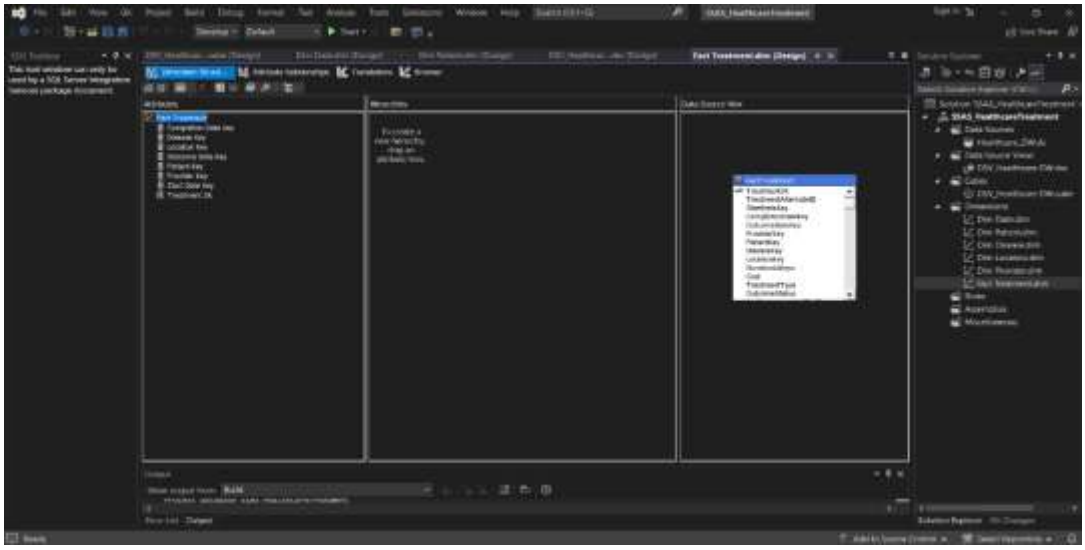
Step 3 - Created a Data Source View (DSV)



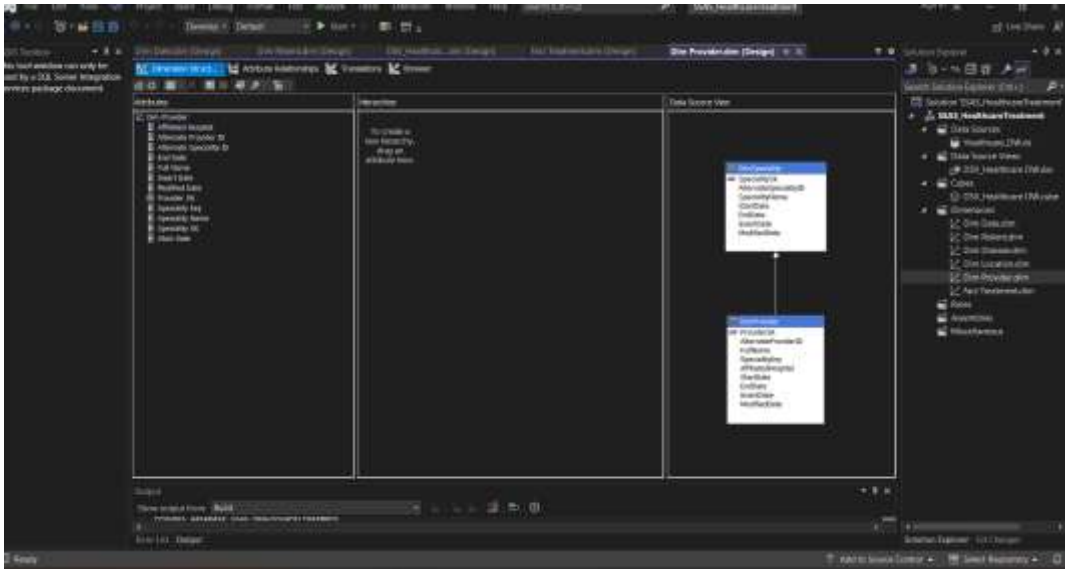
Step 4 - Created and Configured Dimensions

Created dimensions for each dimension table using the Dimension Wizard, Selected the corresponding primary key (e.g., PatientSK for DimPatient), Included relevant attributes such as Gender, Age, State, SpecialityName, etc., Ensured the correct KeyColumn (e.g., PatientSK, LocationSK) was used in each dimension.

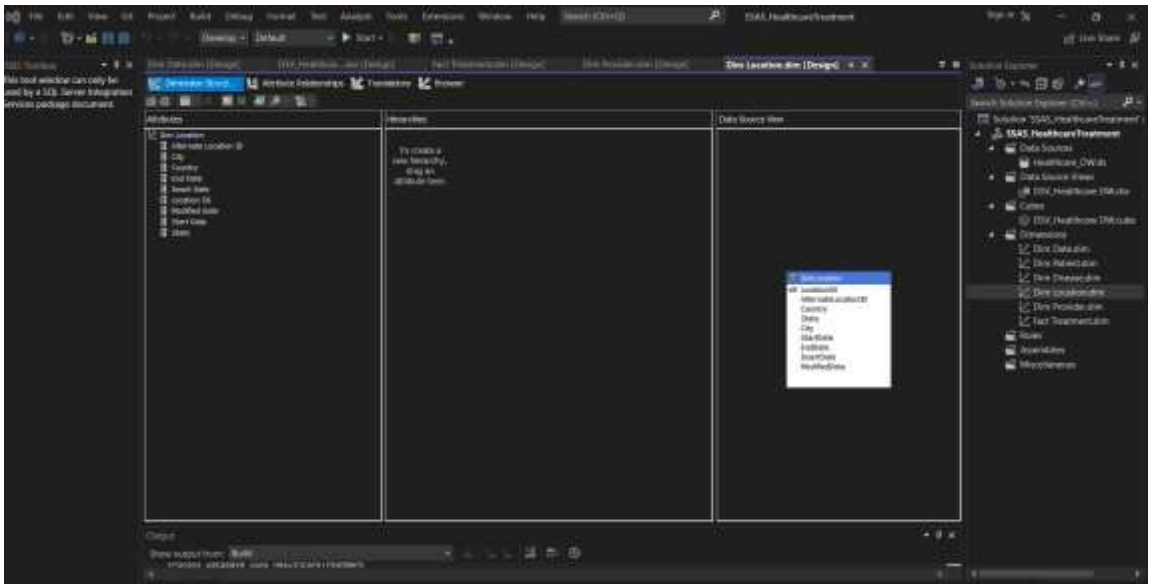
FactTreatment -



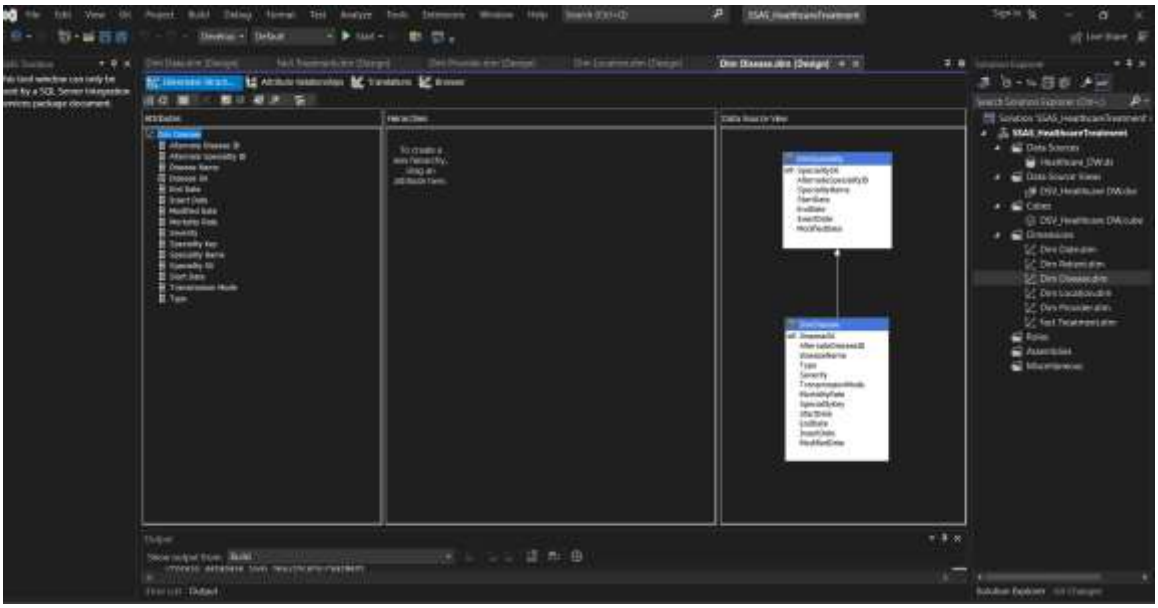
DimProvider



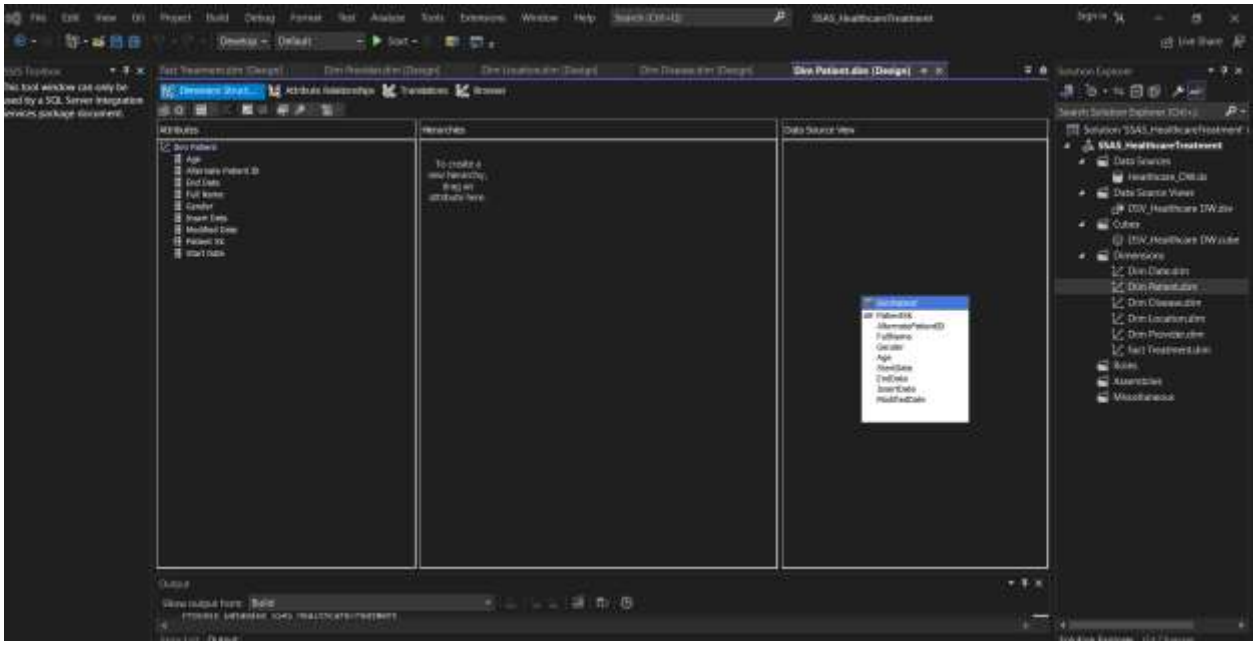
DimLocation



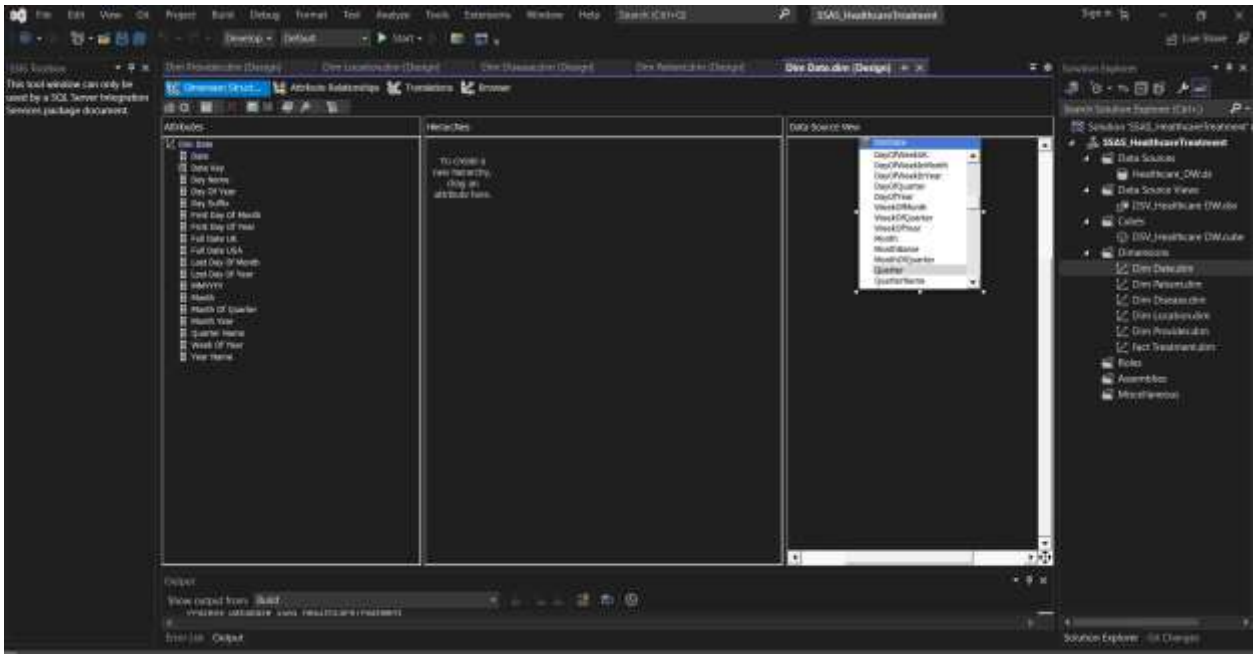
DimDisease



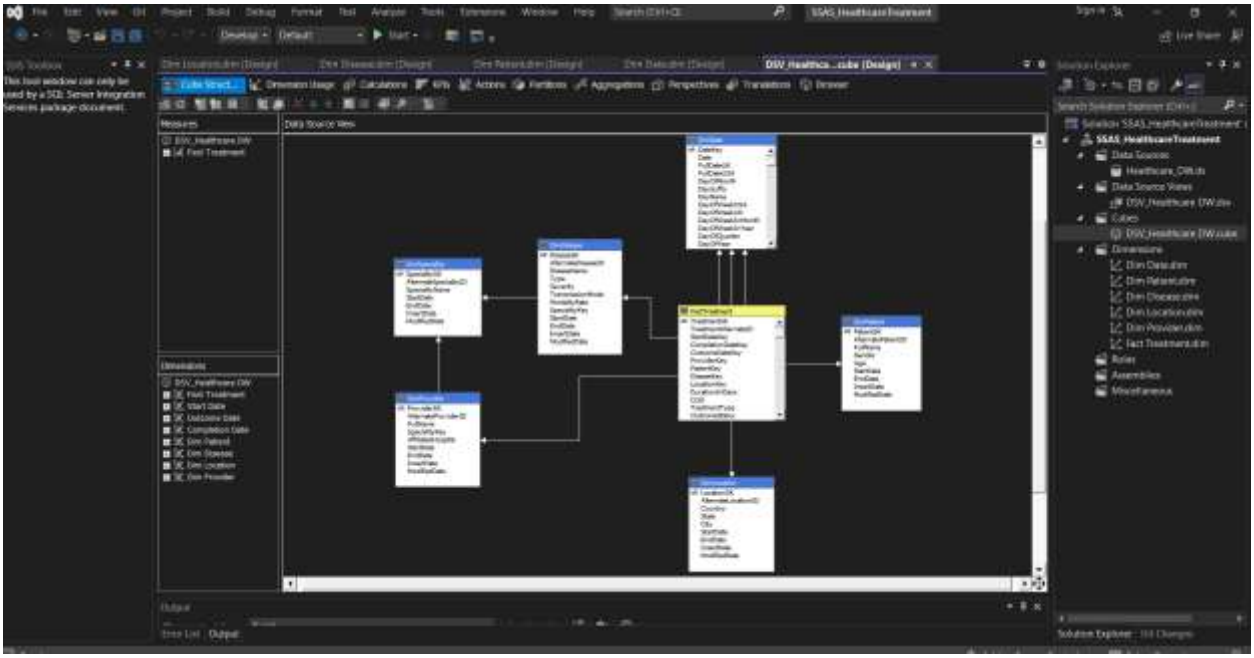
DimPatients



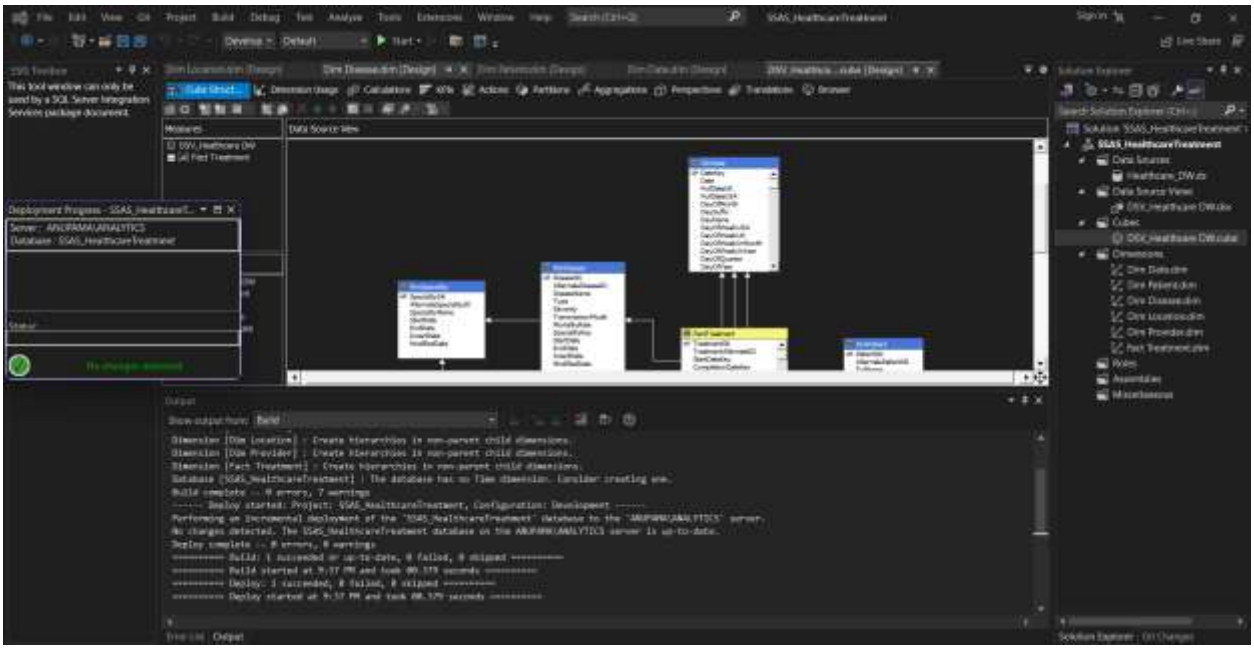
DimDate



Step 5 - Created the Cube Structure



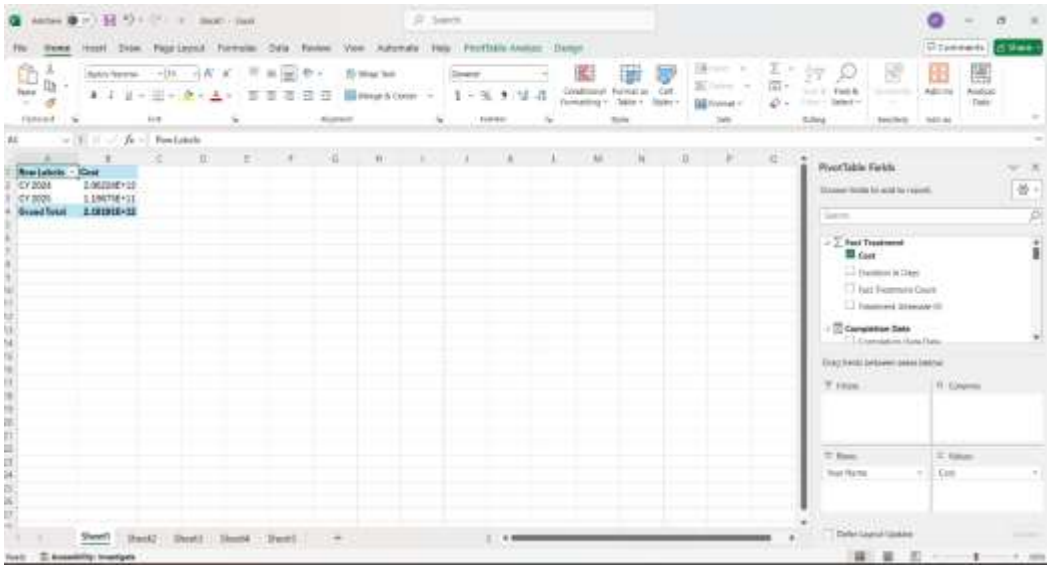
Step 6 - Deployed and Processed the Cube



4. Demonstration of OLAP operations

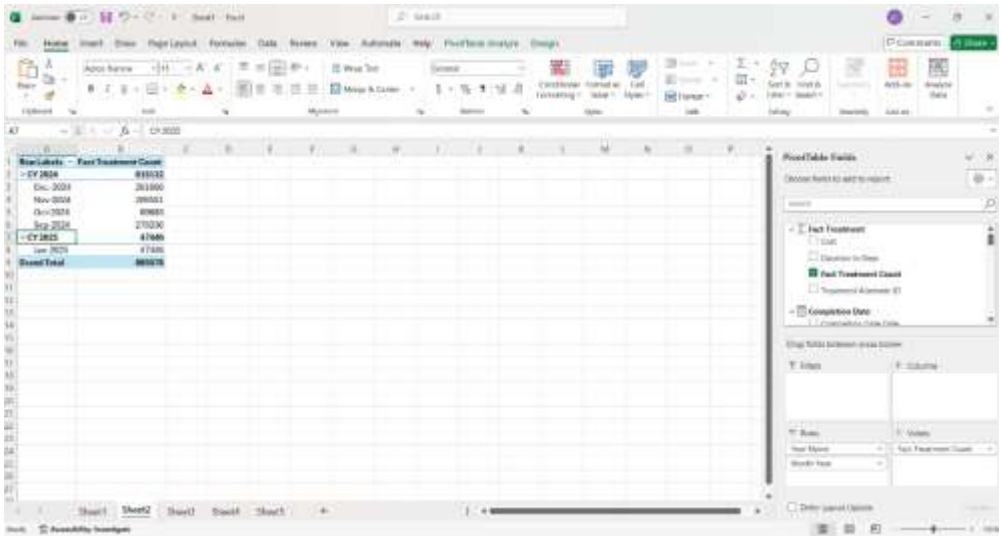
4.1 Roll-Up Operation

To create the roll-up visualization in Excel, a PivotTable was generated from the source data by dragging the "Year Name" field into the Rows area and the "Cost" field into the Values area of the PivotTable Fields pane. This setup aggregates the cost data by each year (CY 2024 and CY 2025), and also provides a Grand Total, effectively summarizing or "rolling up" the detailed cost data into higher-level yearly totals. This roll-up approach allows for quick analysis of cost trends and totals across multiple years within a single, easy-to-read table.



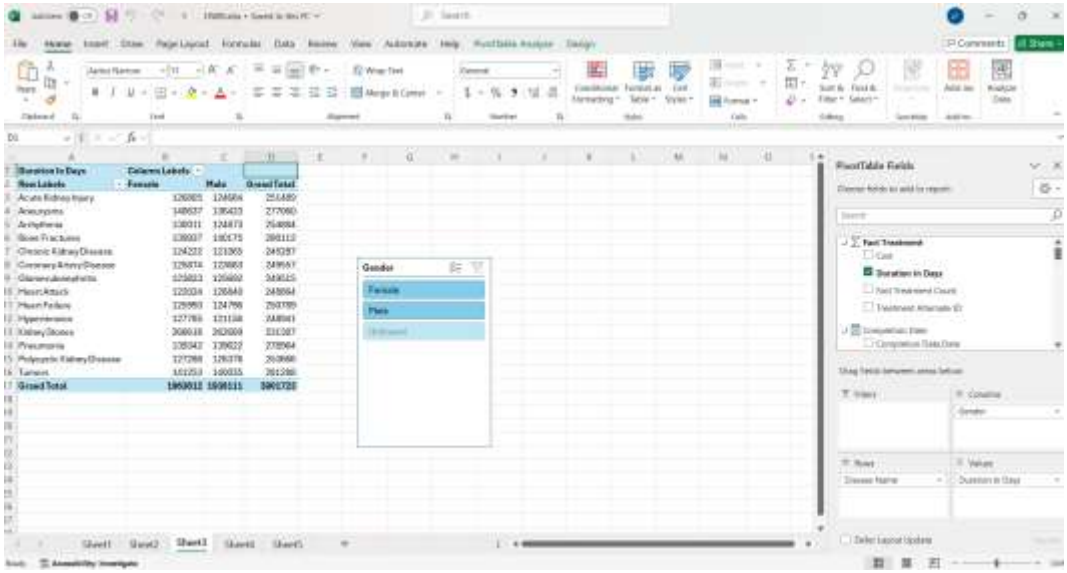
4.2 Drill-Down Operation

To perform a drill-down in Excel using a PivotTable, the data was first summarized by year by placing the "Year Name" field in the Rows area and "Fact Treatment Count" in the Values area. To explore the data in more detail, the "Month Year" field was added below "Year Name" in the Rows area. This allowed the PivotTable to display not only the total treatment count for each year but also to break down (drill down) those yearly totals into monthly values. For example, under "CY 2024," you can see the counts for individual months such as December, November, October, and September. This drill-down approach lets users analyze summary data at a higher level and then expand it to see more granular details directly within the PivotTable



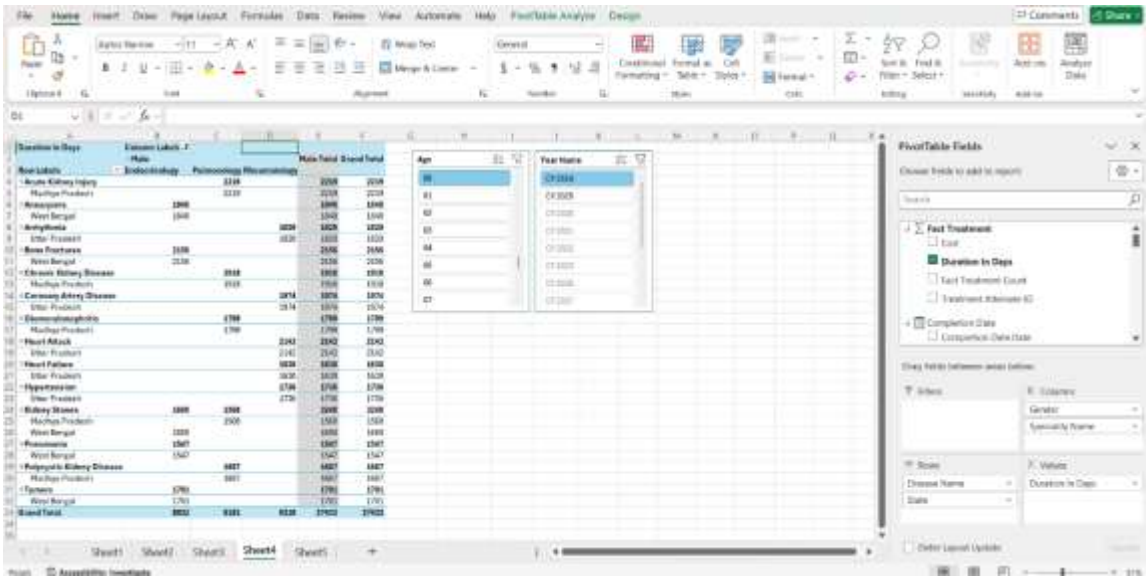
4.3 Slice

To create the slicer visualization in Excel, first a PivotTable was created from the source data by placing the categorical field (Disease Name) in the Rows area, the filter field (Gender) in the Columns area, and the numeric field (Duration in Days) in the Values area to summarize the data. Then, by selecting the PivotTable and navigating to the PivotTable Analyze tab, the Insert Slicer option was used to add a slicer for the Gender field. This slicer provides interactive buttons representing each gender category, allowing users to filter the PivotTable dynamically by clicking on the desired gender(s). The slicer was then positioned and optionally formatted for better usability and visual appeal, enabling quick and intuitive filtering of the summarized data based on gender.



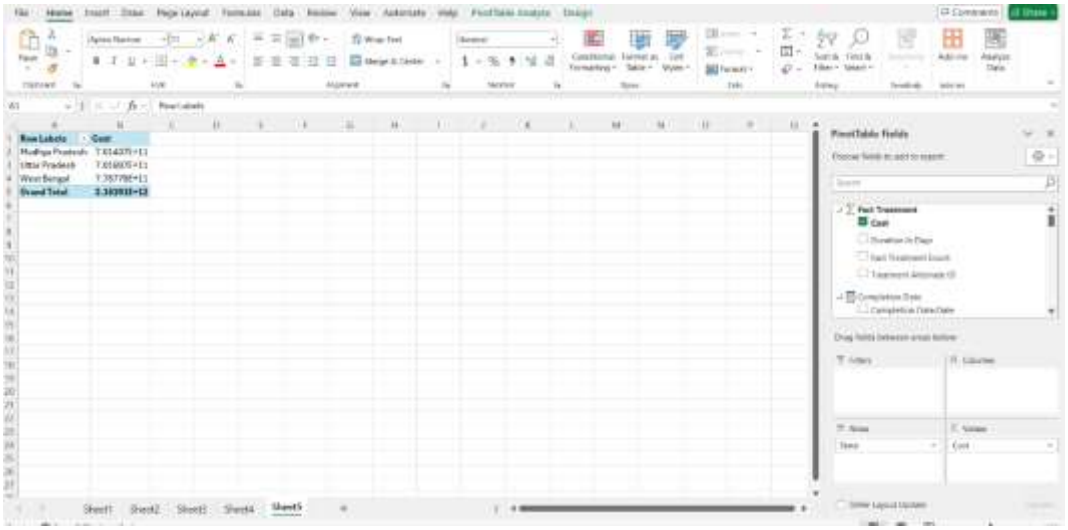
4.4 Dice

To create the dice visualization shown in the Excel, a PivotTable was first generated from the dataset by placing Disease Name and State in the Rows area, Speciality Name and Gender in the Columns area, and Duration In Days in the Values area to summarize the data. Slicers were then added for the Age and Year Name fields by selecting the PivotTable, navigating to the PivotTable Analyze tab, and using the Insert Slicer option. These slicers allow users to interactively filter the PivotTable by specific age groups and years, instantly updating the summarized data to reflect the selected criteria. The slicers were positioned next to the PivotTable for easy access, enabling dynamic multi-dimensional analysis-commonly referred to as "dice" in data analysis-by allowing users to view the data from different perspectives based on age and year.



4.5 Pivot Operation

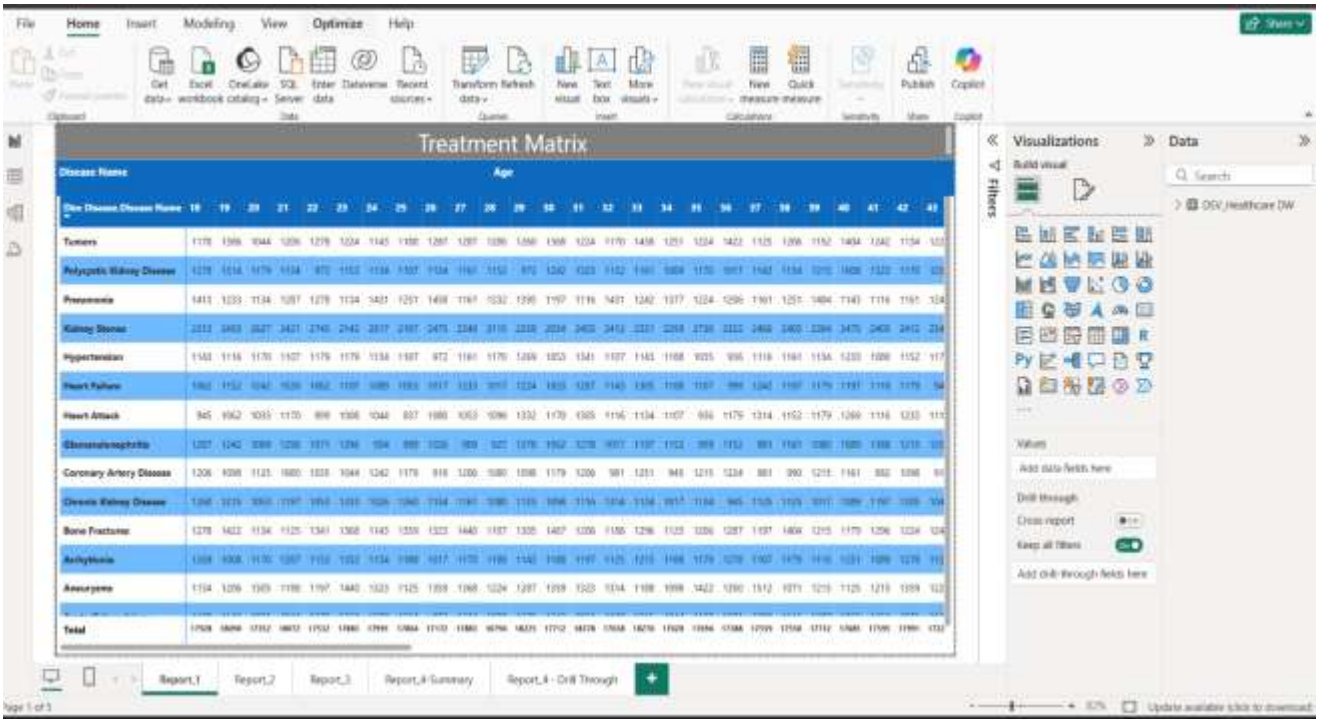
PivotTable was created to summarize and analyze cost data across different states and medical specialties. The "State" field is placed in the Rows area, listing each state (Madhya Pradesh, Uttar Pradesh, and West Bengal) vertically. The "Speciality Name" field is placed in the Columns area, creating separate columns for each specialty (Endocrinology, Pulmonology, and Rheumatology), along with a Grand Total column. The "Cost" field is placed in the Values area, displaying the total cost for each combination of state and specialty, as well as overall totals for each row and column. This pivot operation enables users to quickly compare and aggregate cost data across multiple dimensions, making it easier to identify patterns and insights within the dataset.



5. PowerBI Reports

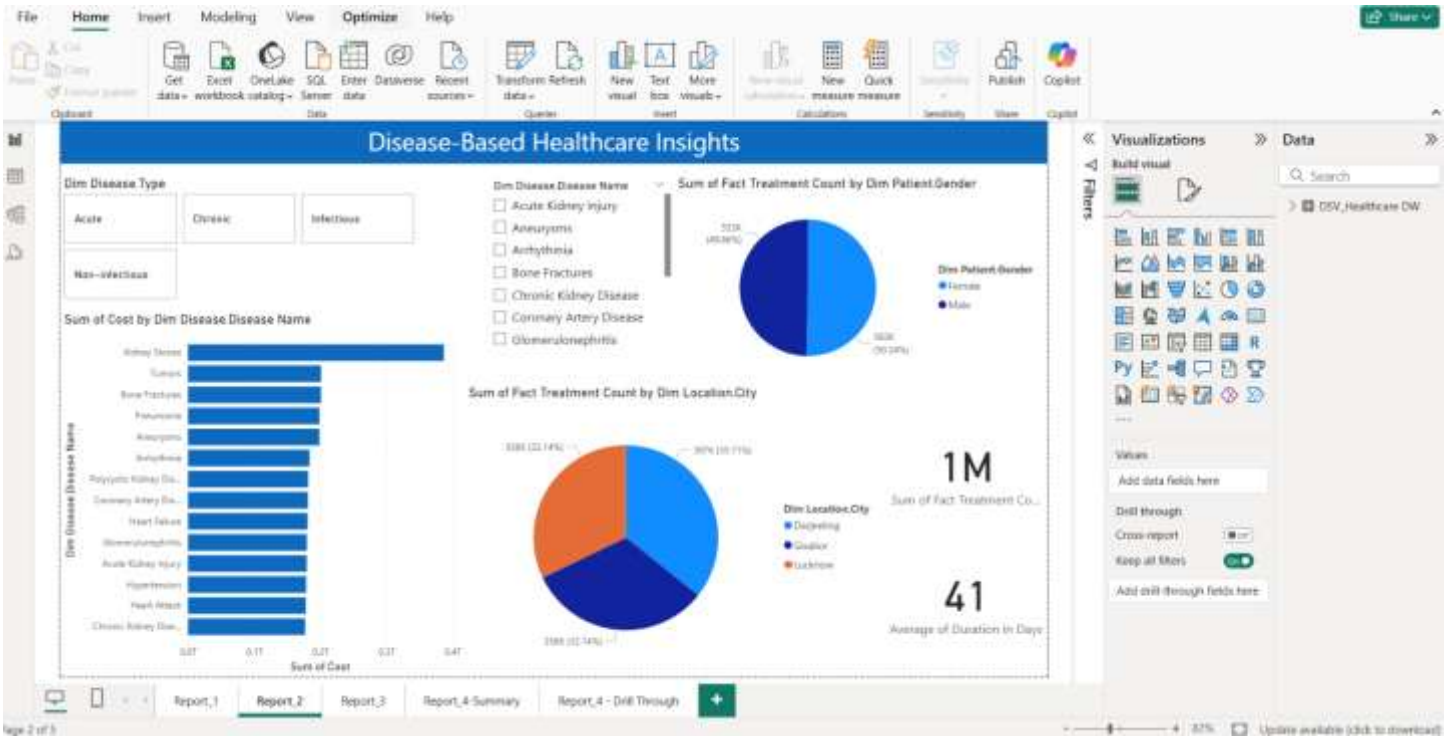
5.1 Report 1

To create this matrix report, the process began with data preparation by connecting to the healthcare cube (DSV_Healthcare DW) and importing relevant tables containing disease names, patient counts, and age information. For the visual design, a matrix visual was selected from the Visualizations pane and configured with disease names as rows and age as columns, displaying the corresponding patient counts in each cell. Conditional formatting was applied to enhance readability, using alternating row colors for better visual separation. The report was further refined by adding a clear title ("Treatment Matrix"), adjusting column widths, and ensuring the layout was clean and easy to interpret. The result is an interactive matrix that allows users to quickly analyze the distribution of various diseases across different age groups.



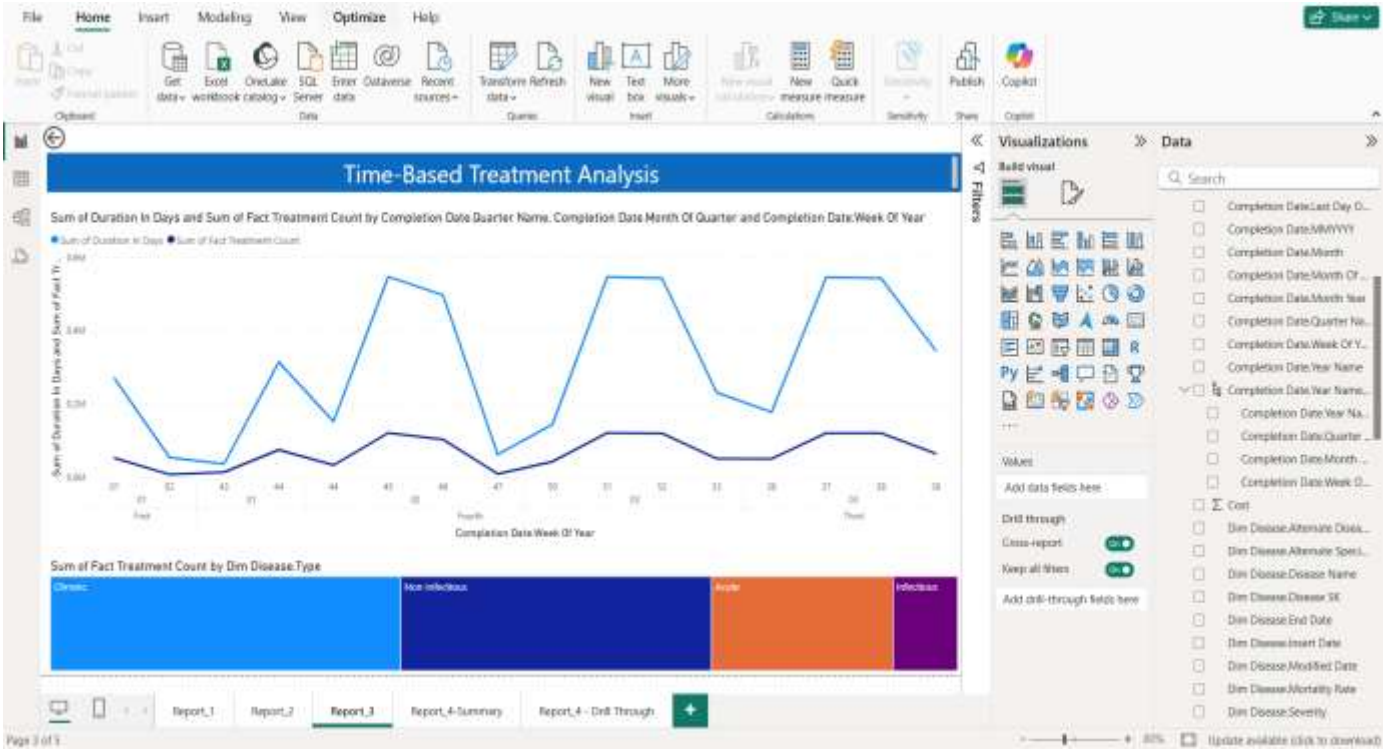
5.2 Report 2

The first slicer allows users to select a disease type (Acute, Chronic, Infectious, Non-infectious), which dynamically filters the available options in the second slicer (disease names), ensuring users only see relevant diseases based on their initial selection. For the visual design, a combination of bar charts, pie charts, and card visuals were used to present insights such as treatment costs by disease, treatment counts by gender and city, and key metrics like total treatments and average duration. Each visual was carefully formatted for clarity, with interactive filters and slicers allowing users to drill down into specific segments of the data. The end result is a visually engaging and highly interactive report that empowers users to explore disease-based healthcare insights efficiently.



5.3 Report 3

Hierarchies were then created within the date dimension, allowing users to drill down from year to quarter to month directly within the visuals. For the visual design, a line chart was used to display the sum of treatment duration and treatment count over time, with the x-axis configured to support drill-down through the time hierarchy. Additionally, a stacked bar chart was included at the bottom to show the distribution of treatment counts by disease type, offering further insights. The report's layout was carefully formatted for clarity, with a prominent title and color-coded visuals to distinguish between different disease categories. This interactive design empowers users to explore treatment trends at various time granularities, uncovering patterns and insights across years, quarters, and months.



5.4 Report 4

The first report page was designed with a bar chart visual displaying disease type on the x-axis and the sum of costs on the y-axis. Drill-through functionality was then configured by creating a second page as the drill-through target, where the drill-through field was set to disease type to filter data based on the user's selection on the first page. The detailed second page includes a table showing patient full name, disease name, treatment count, sum of cost, and sum of duration in days, alongside pie charts illustrating specialty names with sum of treatment count and sum of cost by providers' affiliated hospitals. Additionally, two KPI visuals display the total sum of cost and treatment count for quick reference. This setup allows users to right-click a disease type in the bar chart on the first page and navigate seamlessly to a detailed page focused on that disease, providing a comprehensive view of related patient and treatment information with interactive filtering applied automatically.