**Learners have to develop a Report to support the answers to the following questions and suggestions.**

**Objective Questions:**

1. In analyzing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.

**Ans:** The data is clean and no further cleaning is required. Also for all the null cells in **patient\_sat\_score** section, I just considered it as “0” rating

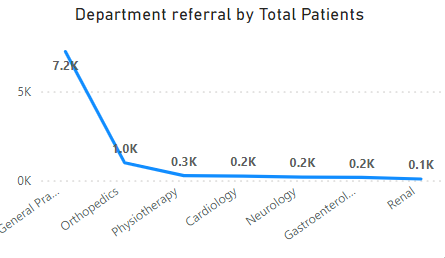
1. **Assess the Average Waiting Time:** Analyse the patient wait times to identify the average duration a patient spends before receiving care

**Ans:** We can easily get the Average Waiting Time by this simple DAX formula:  
Average wait time = AVERAGE('Hospital ER'[patient\_waittime]) and the value for it would be **35.26**

****

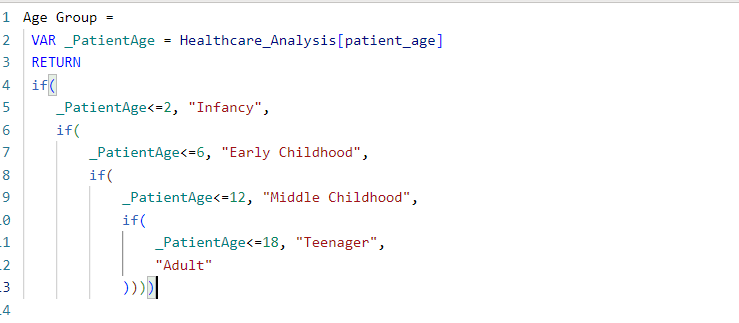
1. **Visits by Department Referral:** Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.

**Ans:** We can get the number of patients visited by each department by a table chart taking department referral and patient names as counts and the total count of patients would be **9216** and most visited department is General Practice – **7240** patients.

****

1. **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

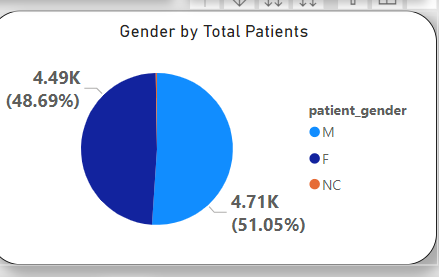
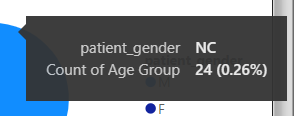
**Ans:** To get this data, we need to group the ages of all the patients in different categories to get a clearer picture of it. Here, I have grouped the people in 4 different categories as Infancy (age<= yrs), Early Childhood (age<=6yrs), Middle Childhood (age<=12 yrs), Teenager (age<=18 yrs) rest all are Adults, by using the group feature, and by this, we get to know that the **Adult** people (age>18 yrs) uses the most healthcare services i.e.,**7106** patients.  
**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data

****

1. Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for?

**Ans:** There are Null values present in the dataset. However, for the null cells in **patient\_sat\_score** section, I replaced it with 0 to get the overall average rating   
**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data  


1. Is there any relation between the number of visits and the Gender of the patients?

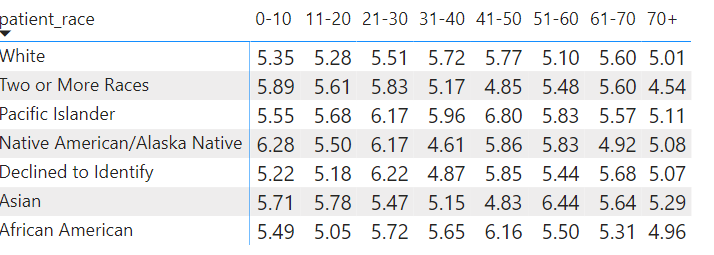
**Ans**: There is No direct relation between Number of visits with the Gender, but by further analysis, we can have found that more Males (51.05%) have visited the hospital than Female patients (48.69%), rest 0.26% have not disclosed their Gender.  
**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data  
 

1. Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.

**Ans**: •Pacific Islander are the most satisfied with an Average satisfaction score of 6.80 and they fall in the Age bucket of 41-50

•Two or More Races category has the least satisfaction score of 4.54 and they fall in the Age bucket of 70+

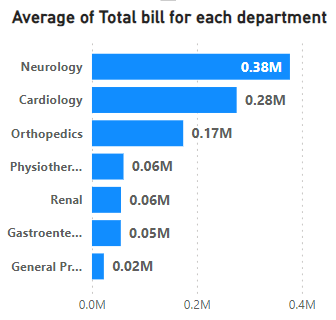
**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



1. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.

Ans: If we take department referral and Average of Total bill in a Column chart, we can see how much revenue is generated by each department.

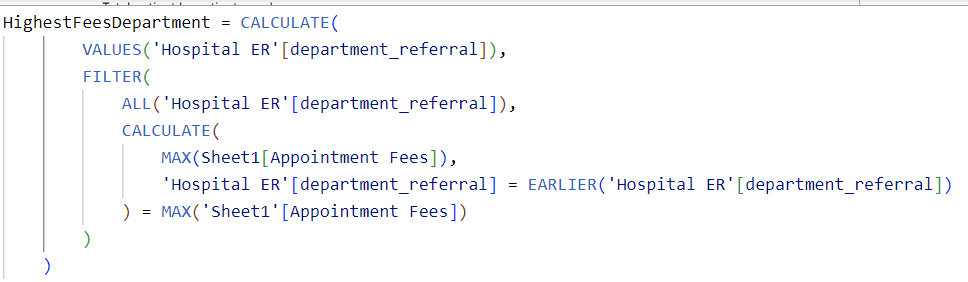
Reference: PowerBI File>Healthcare Analysis.pbix>Doctors Data



1. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.

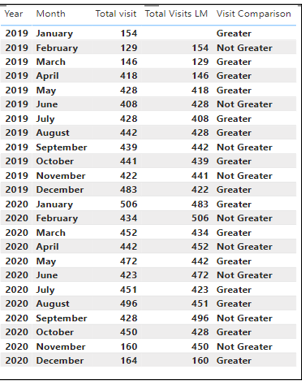
**Ans:** We can calculate the name of the department charging highest for appointment fees which is Neurology Department by using the below mentioned formula.

Reference: PowerBI File>Healthcare Analysis.pbix>Doctors Data



1. Create a tabular visualization in the Report view which consists of Month-wise total visits in the hospital. Add a third column in the table that consists of the previous month’s total visits for each month’s row. Also, include a column that states whether the visits in a month are greater than that of the previous month's visits.

Ans: We can show it in a table visual by taking year and month in columns and total visits, total visit last month by using PREVIOUSMONTH formula and lastly putting an IF DAX to mention which value is greater than the previous one or not.



To get the impression on total visit last month, we can use:

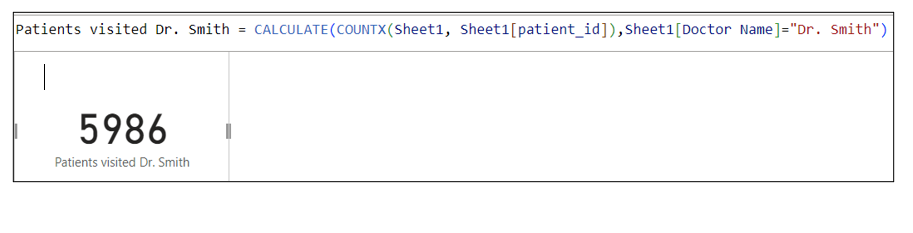
Total Visits LM = CALCULATE (COUNT ('Hospital ER'[patient\_race]), PREVIOUSMONTH ('Hospital ER'[date]), ALLEXCEPT ('Hospital ER','Hospital ER'[date]))

To get the formula for the prompt “greater” and “Not greater”, we can use:

Visit Comparison = IF([Total patient]>[Total Visits LM],"Greater","Not Greater")

1. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.

**Ans:** We can find out the number of patients visited Dr.Smith just by using a calculate function i.e., 5986, and here is the formula



1. Calculate the average age of the patients who visit the Orthopedics department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?

Ans:

* 1. Average age visiting All departments is 39.86.
  2. **Average** Average age visiting Orthopaedics department is 38.66.

**Age- Orthopedics = CALCULATE(AVERAGE('Hospital ER'[patient\_age]),'Hospital ER'[department\_referral]= "Orthopaedics")**

**Average Age- All Departments = CALCULATE(AVERAGE('Hospital ER'[patient\_age]))**

Note: Approach to calculate the metric is same. However, Filter will not be used if we have to calculate Average age of patients visiting All departments

1. Were there any data format issues in the data, and if there were/are how you handle them?

Ans: Approach: Data cleaning performed in Power Query Editor:

* + Data type conversion (implicit or explicit)
  + Handling missing values (replacing, filling, or removing)
  + Standardizing formats (e.g., date format)
  + Removing special characters
* Specific issue addressed: Inconsistent date formatting:
  + Splitting date from time
  + Combining them into a single, consistent format

Based on the provided “Hospital ER” dataset, below changes are carried out:

* The date format appears inconsistent. Some dates are in "dd-mm-yyyy" format (e.g., "20-03-2020"), while others are in "mm-dd-yyyy" format (e.g., "12-11-2019"). It's important to have a consistent date format for easier processing and analysis. Hence, the Date is converted to “dd-mm-yyyy” format.
* Patient\_First\_Initial and Patient\_Last\_Name are combined to Full\_Name.

1. When we add a column in Power Query what’s the code that comes in M language in the formula bar? What do you know about M-query?

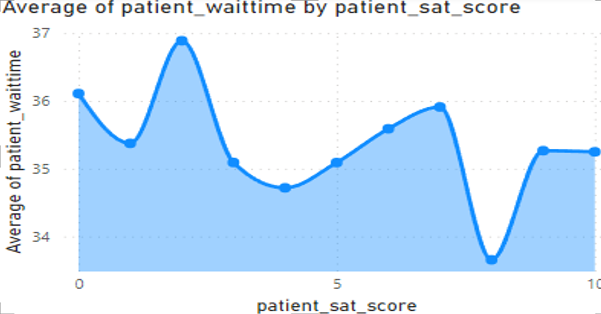
Ans: When we add a column in Power Query, the code that appears in the formula bar is written in the M language. M is the formula language used in Power Query, and it's designed for data transformation tasks.

The code appears there is: =Table.AddColumn(#"PreviousStep", "NewColumnName", each [Column1] + [Column2]

**Subjective Questions**

1. What is the relation between patient wait time and satisfaction scores?

* Ans: Relationship between patient wait time and satisfaction score: Negative correlation.
* Visualization: Chart of Average Patient Wait Time vs. Satisfaction Score
* Conclusion: Satisfaction score decreases as wait time increases.

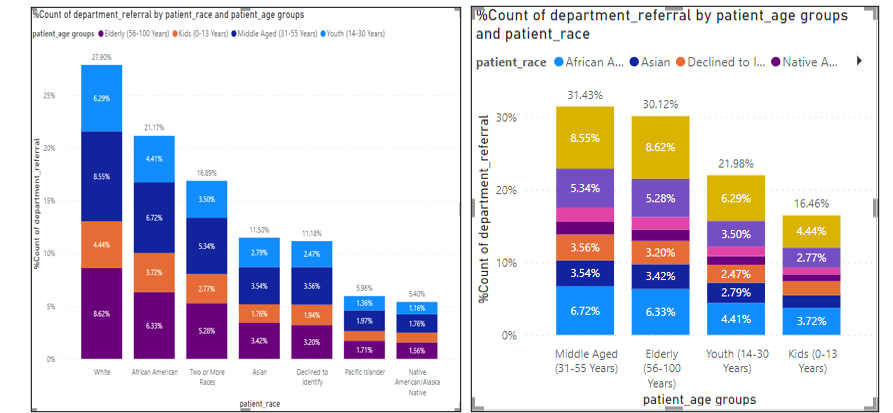


Here we can clearly see the Sat score is highest when the waiting time is lower, and lowest when the patients had to wait more.

1. How do patient demographics affect the frequency of visits to different departments?

Ans: Approach: We can answer this question by making a chart between patient race in X axis and % of Department count in Y axis putting age groups as legends

**Reference**: PowerBI File>Healthcare Analysis.pbix

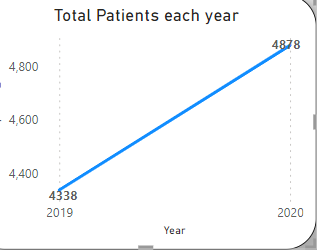


* **Race:**
  + Most referred: White (over 25% of total referrals)
  + Least referred: Native American (under 6% of total referrals)
* **Age Group:**
  + Highest referred: Middle Aged (31-55 yrs)
  + Lowest referred: Kids (0-13 yrs)

1. Is there a noticeable trend in the volume of patient visits throughout the year?

Ans: There is a noticeable trend going up from past year to the current year 2020 which we can see through the line chart year wise volume.

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data

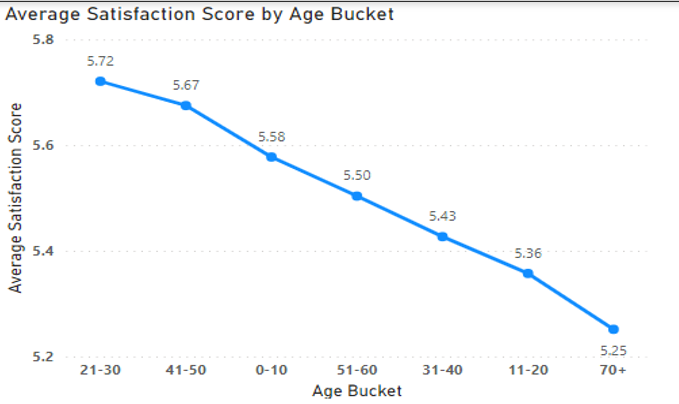


Here we can clearly see the patient volume rise from 2019 to 2020 which is around ~6%

1. Which age groups report the highest and lowest satisfaction scores?

Ans: Simply by asserting a chart for Patient’s age group vs satisfaction scores, we can easily see which age group has rated the **highest** sat scores, and in this case, it is the **Youth (21-30 yrs)**, and **lowest** would be the **Elderly** people **(70+yrs)**

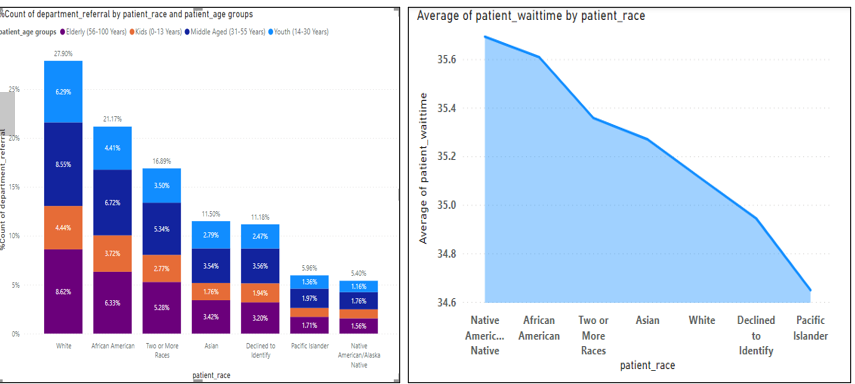
**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



1. Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?

Ans: We can answer this question by making a chart between patient race in X axis and % of Department count in Y axis putting age groups as legends:

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



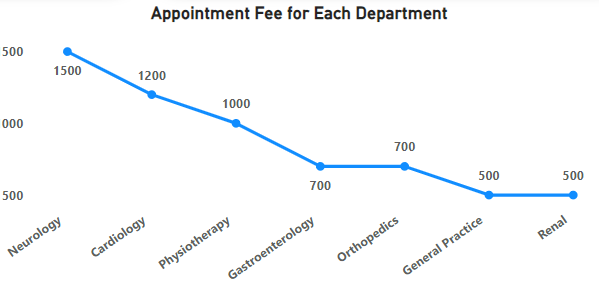
Here from this chart, we can clearly see the most referred Race is White and its more than 25% of total referrals, whereas the least referred Race is Native American, lesser than 6% of total refers.

Also in the 2nd chart, we can clearly see the Native people had to wait longer than white people. **This concludes the fact that the claim on Gender discrimination is justified** but in very minimal scale.

1. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?

Ans: As we can observe from the chart that the Neurology and Cardiology department charges the highest compared to others, the management can try and provide discounts to the patients visiting Neurology and Cardiology.

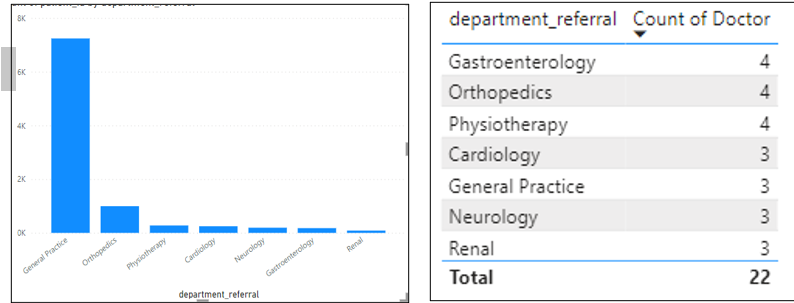
**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



1. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.

Ans: As per my analysis, departments which have the highest patient visits and lowest doctors in comparison to others should hire the new doctors, we can clearly check it in the below mentioned chart.

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



So, as per the charts, the Highest visited department is General practice, but the count of doctor is lower than other departments, where we can hire more doctors to compensate and ask them to work shift-wise.

1. Is the hospital profitable? How will you determine the profitability?

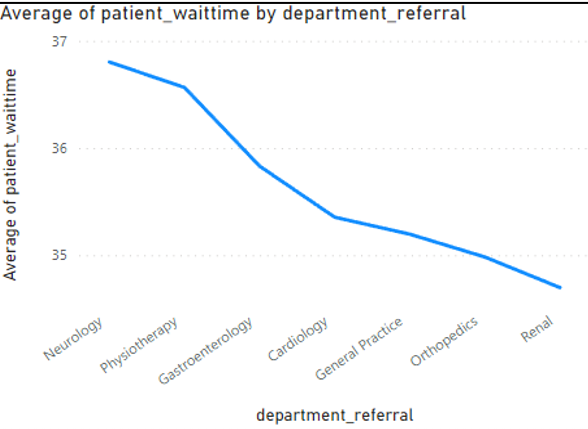
Ans: The hospital is totally profitable, in order to determine, if we take total appointment fees as total expenses, and subtract it with total revenue generated, we get the Total Profit of $503.96 Million

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



1. Any Department for which the waiting time is oddly large?

Ans: If we draw a Line chart between departments and average waiting time, we can see, the Neurology and Physiotherapy departments’ waiting time is little bit higher than the other departments in the hospital.



1. Come up with strategies to provide discounts to the patients.

Ans: **Senior Citizen Discounts**:

Utilizing demographic information allows for precise targeting of discounts based on age groups that comes under senior citizen above 60 and provide them some discounts.

**Package Deals**:

Creating bundled packages encourages patients to utilize more services while saving money, thus increasing overall revenue, and promoting comprehensive healthcare.

**Promotional Discounts**:

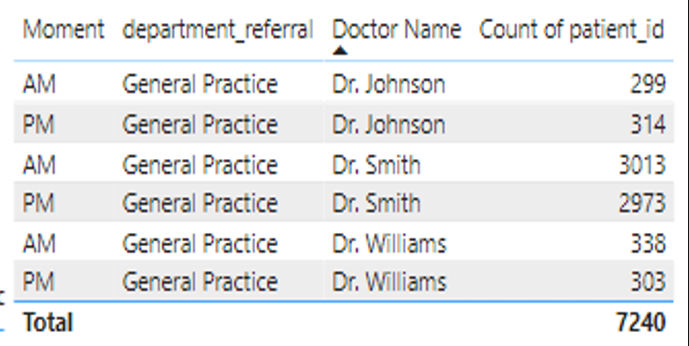
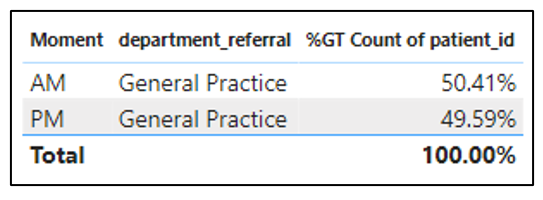
Running promotional campaigns with limited-time discounts on select services or procedures can attract new patients and incentivize existing patients to utilize additional services, contributing to revenue growth and

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data  
patient retention.

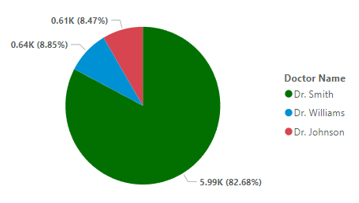
1. Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also will this 2 shift policy be helpful for the hospital?

Ans: For this question, if we draw a table for General Practice department AM and PM wise by doctor’s name, we can already see all the 3 doctors have been aligned with 2 shifts each, which is AM and PM

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data

* Currently, all three doctors (Dr. Johnson, Dr. William, and Dr. Smith) are assigned to both AM and PM shifts.
* A proposal is made to change the shift allocation:
  + Dr. Johnson and Dr. William would work PM shifts only.
  + Dr. Smith would work AM shifts only.
* This change would result in a more balanced distribution of patients across AM and PM shifts, with Dr. Smith seeing most of the AM patients.
* **Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



1. What do you understand by PowerBI gateway? What are its use cases?

Ans: A Power BI gateway is a software application that acts as a secure bridge between on-premises data sources and Microsoft cloud services like Power BI, PowerApps, and Azure Logic Apps.

**Use Cases:**

There are several scenarios where a Power BI gateway comes in handy:

1. How would you approach this problem, if the objective and subjective questions weren't given?

Ans: Here's how I would approach identifying objective and subjective questions in a hospital management system using Power BI, even without pre-defined categories:

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data

* Define the objectives and scope of the "Columbia Asia" project, outlining the key questions the Power BI solution should address.
* Perform data cleaning, transformation, and modelling to prepare the data for analysis.
* Create a data model in Power BI that accurately represents the relationships between different data entities.
* Design and develop interactive reports and dashboards in Power BI to visualize the data and answer the key questions identified during requirement gathering.
* Use a variety of Power BI visualizations, such as charts, graphs and tables, to present insights effectively.
* Incorporate interactive features like slicers, filters, and drill-downs to allow users to explore the data dynamically.

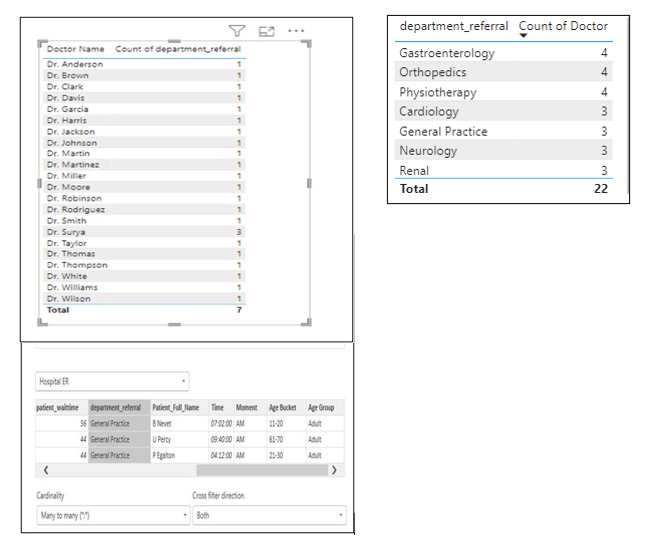
By following this approach, we can develop solution for the "Columbasia" project, enabling stakeholders to gain valuable insights from their data and make informed decisions.

1. Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?

Ans: The answer for this question should be **Many-to-many** cardinality because one department can have many doctors, also one doc can handle many departments as well, refer to the below mentioned table:

As per the table below, we can see that Dr.Surya is attending 3 different departments.

**Reference**: PowerBI File>Healthcare Analysis.pbix>Patients Data



Here we can clearly see for 1 department, there are 3-4 doctors have been allotted based on their specialities, and vice versa so the conclusion is correct and just for Many-to-Many relationship, and not One-to-one.

**Report**

The hospital has asked for a report with three tabs:

* Main Tab
* Doctors’ Tab
* Patients’ Tab
* **Using the Main tab in the report,** the hospital should be able to look at the overall metrics like the number of daily visits, revenue produced on that day, customer satisfaction, how busy are different departments on that day, and general waiting time on that day. This tab should have a slicer of date.
* **Using the Doctors’ Tab,** the Chief Of Staff at the hospital should be able to look at the individual doctor’s performance metrics like customer satisfaction, the number of patients he was visited by, the revenue he has generated, and his appointment fees. This tab should have a slicer of the Doctor's Name or ID.
* **Using the Patients’ Tab,** the Patient’s Care Chief at the hospital wants to look at a customer’s profile which would involve metrics like the most frequently visited department, their age, their race, their waiting time, number of visits, the total amount that they have paid to the hospital, etc. All the metrics using which they can address the patient very carefully in their visits. This tab should have a slicer of the Patient's Name or ID.

**Make sure that all the visualizations look decent and are placed in a proper order. Each tab has different POCs (Point Of Contact), so make sure you involve all the metrics that POC may look at in that tab along with those mentioned in the tab description.**

**After making the report on the Desktop ensure that it is hosted on PowerBI service and use the hosted link for submission of the dashboard and mentioning on the resume.**