**Objective Questions:**

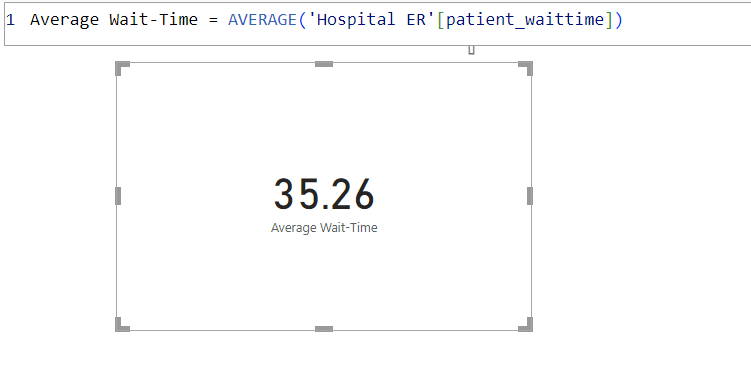
1. *In analyzing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.*

**Ans1.** No data inconsistencies or missing values were detected in either table. Otherwise, they would’ve been cleaned using methods such as replacing the null values with defaults (e.g. mean or median, etc). For primary key columns, the better remedy would be to delete rows with missing values altogether.

As for inconsistencies, the ‘Transform’ tab permits automatic type or upper/lower-case conversion for the relevant cell value to be formatted in accordance with the rest of the column.

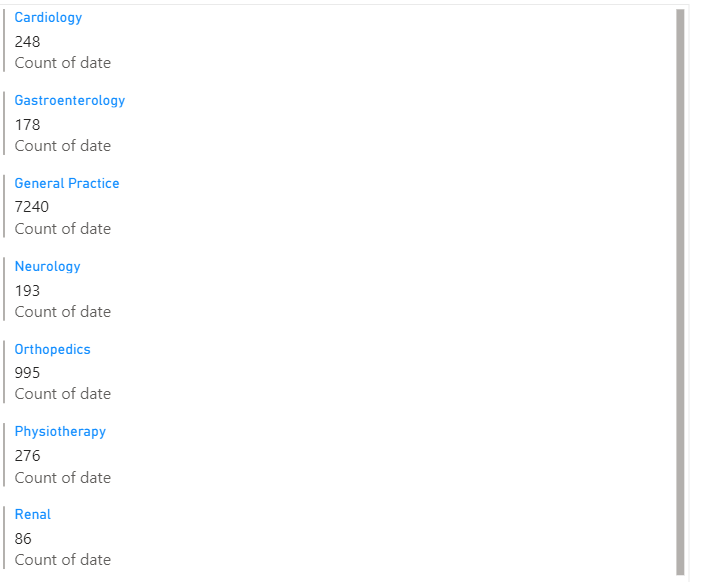
1. ***Assess the Average Waiting Time:*** *Analyse the patient wait times to identify the average duration a patient spends before receiving care.*

**Ans2.** Loading the CSV file containing patients’ wait-times into PowerBI, we use a measure to calculate the average, then a ‘card’ visualization to view the result:



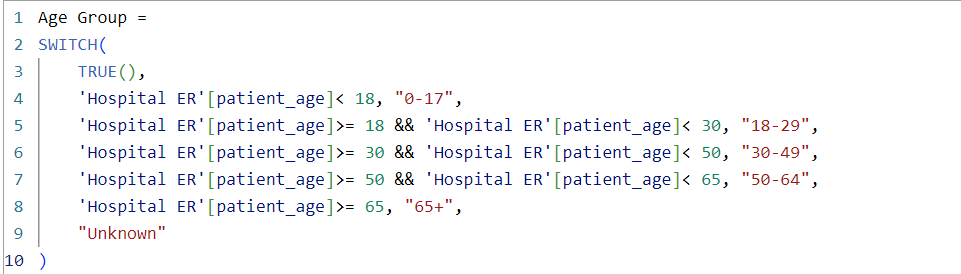
1. ***Visits by Department Referral:*** *Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.*

**Ans3.** Using a multi-row card of ‘department\_referral’ against ‘date’, we find the General Practice department to clearly be the most visited:

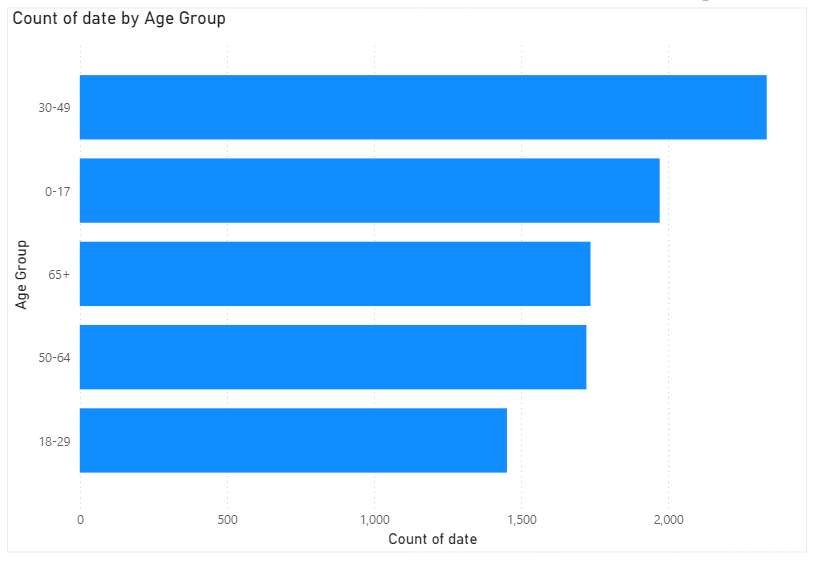
****

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

**Ans4.** To begin segregating visits by age groups, we must first define a calculated column that categorizes patients into these groups, using the formula:



This done, we generate a bar chart against the ‘date\_visited’ column:

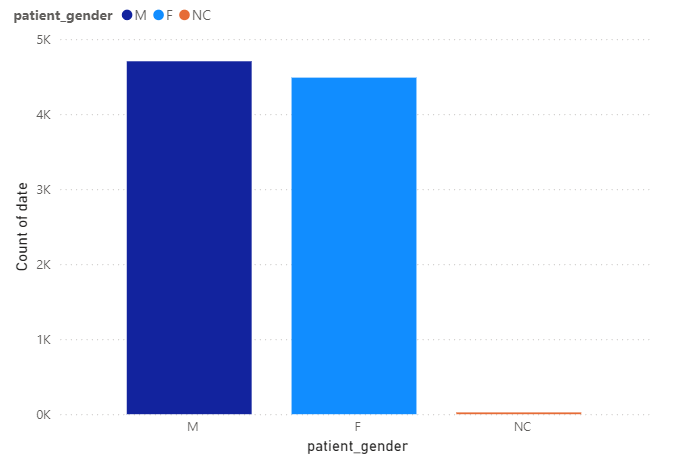
****

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?*

**Ans5.** Yes, there are Null values specifically within the column ‘patient\_sat\_score’. In this case, the chosen method of handling them was to replace by a score of 0. This was achieved in the Power Query Editor using the Replace Value feature.

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

**Ans6.** Plotting the ‘patient\_gender’ column against the ‘date’ on a column chart reveals Male patients to have made roughly 5% more visits than Female ones, while Non-Confirmed (NC) patients comprise an extremely miniscule fraction:

****

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.*

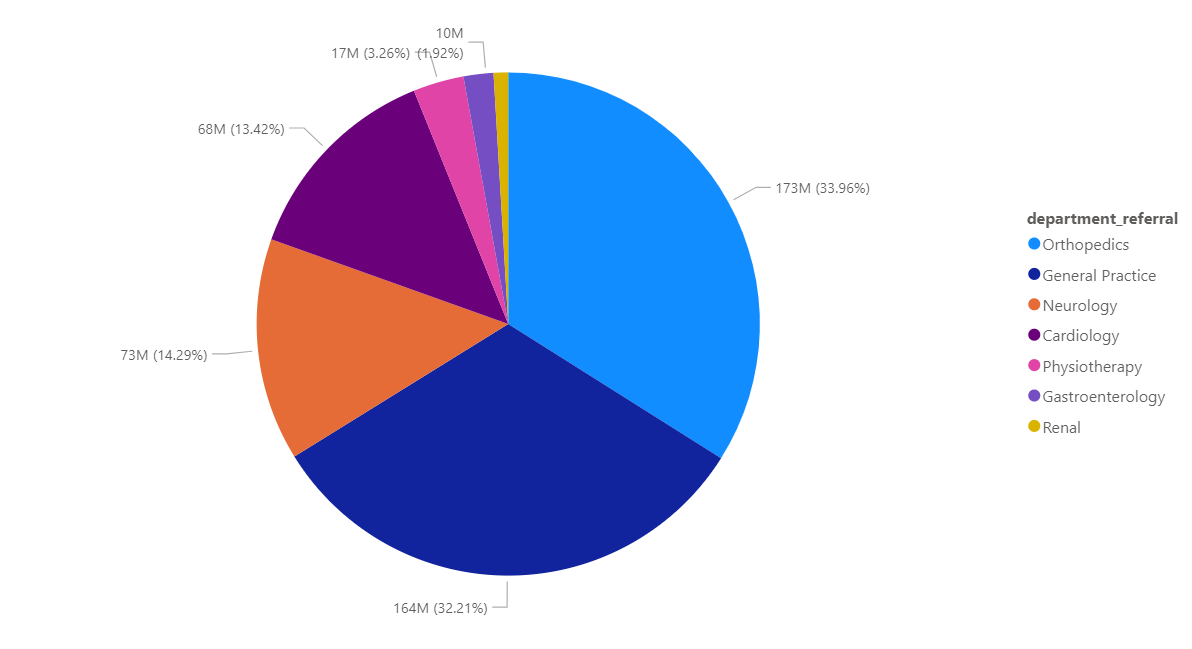
**Ans7.** To determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds, we can use the three-dimensional matrix visualization, taking averages of satisfaction scores for every age group, race and pairing of these columns:



**Ans7[contd].** We find that, among younger patients [ages 0-29], Native Americans/Alaska Natives and Asians offer the most favorable ratings, while among those over 49, Pacific Islanders and Asians do so. The least satisfaction among the races comes from the African Americans, Biracials and Whites.

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

**Ans8.** A pie chart for the ‘Total Bill’ column against ‘department\_referral’ columns from the second dataset shows us that the General Practice and Orthopedics departments generate the most revenues:

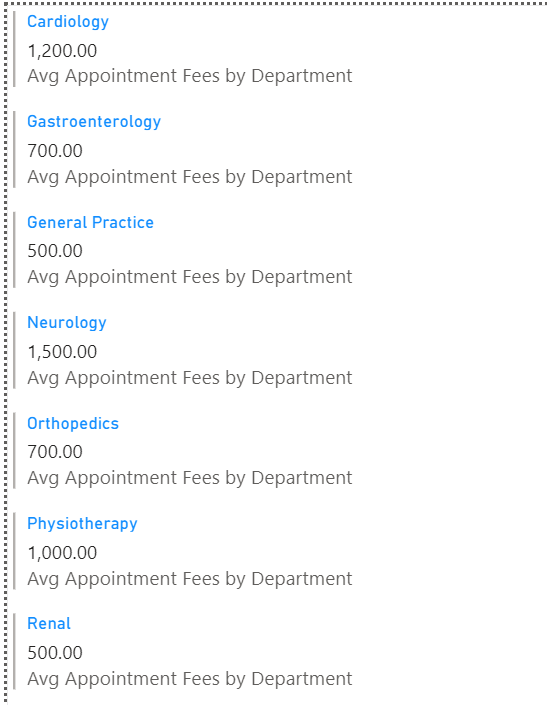


1. *Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.*

**Ans9.** The following aggregation DAX function is used to calculate the average appoint fee each department charges:

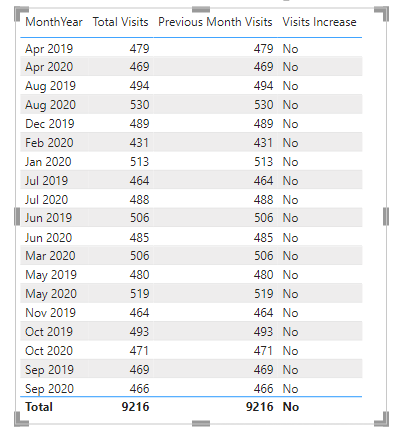
Avg Appointment Fees by Department=AVERAGEX(VALUES('Sheet1'[department\_referral]), CALCULATE(AVERAGE('Sheet1'[Appointment Fees])))

It is then visualized using a multi-row card:

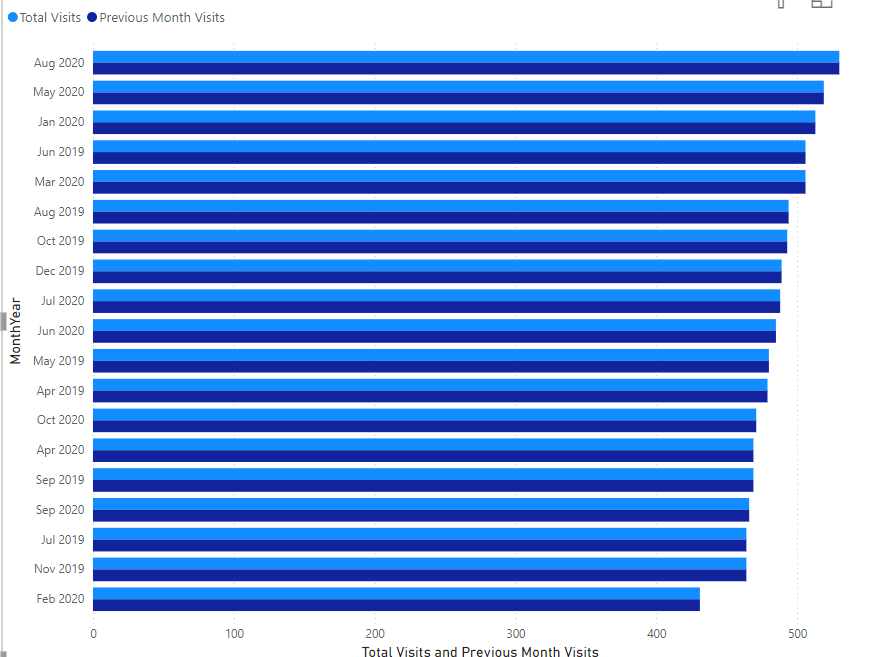


As seen above, the highest fee, generally, is charged by the Neurology department.

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.*

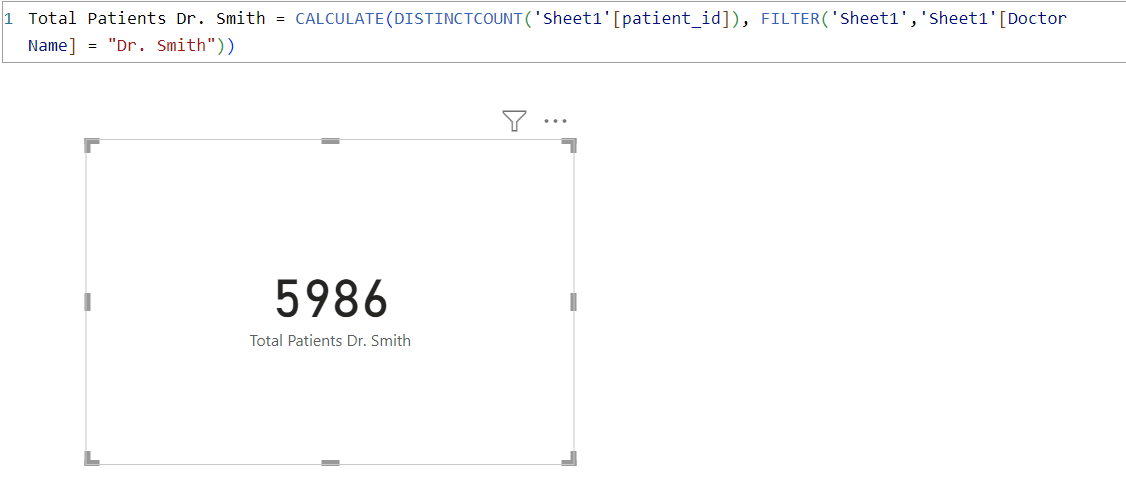
**Ans10.**

The following is a bar-chart visualization of the above table for easier comprehension of data:



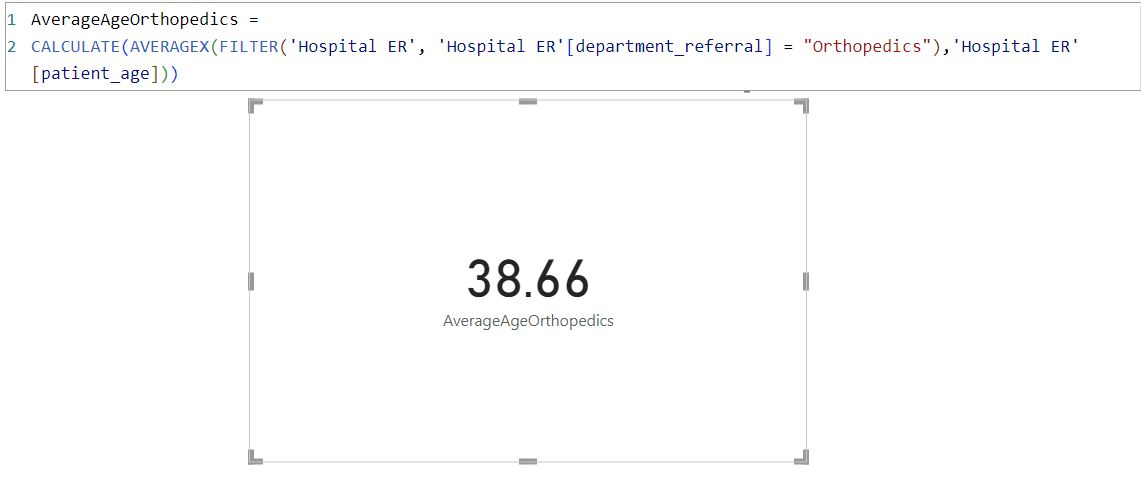
1. *Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.*

**Ans11.** We use the following row-iteration DAX function to calculate Dr. Smith’s total patients:



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?*

**Ans12.** We use the following row-iteration DAX function to calculate the average age of the patients who visit the Orthopedics department:



Yes, a different approach would have been necessary in order to calculate the average age of patients from all departments in general, namely, averaging the entire patient\_age column without filtering for department.

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

**Ans13.** No data format issues were found in the dataset. Otherwise, to check and handle data format issues, the following procedure would’ve been used.

1. Identify Data Format Issues:

* Ensuring all date columns are in a consistent format.
* Checking that numeric fields contain only numbers and are not mixed with text.
* String fields with unexpected characters or spaces.
* Columns with null or empty values.
* Verifying that each column is stored with the correct data type (e.g., integer, float, date, string).

2. Check for Data Format Issues:

* Inspecting data in the "Data View" to visually check for inconsistencies.
* Using DAX functions like ISNUMBER, ISDATE, ISTEXT, and ISBLANK to create calculated columns or measures that flag potential issues.
* Using Power Query to filter or identify rows with format issues, then applying transformation steps to detect and correct data types.

3. Handle Data Format Issues:

* + Transforming data in Power Query by the "Replace Errors" or "Change Type" options to convert data to the desired format.
  + Creating custom columns to clean or transform data, using functions like FORMAT, VALUE, DATEVALUE, etc.

Alternatively, we could automate data format error-handling by taking steps such as:

* Setting up Data Quality rules/conditional formatting to flag issues.
* Creating summary reports/dashboards that highlight data quality issues, allowing proactive monitoring.

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?*

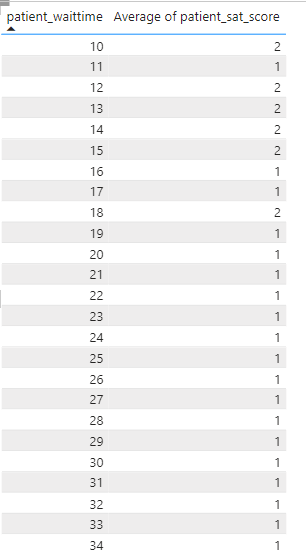
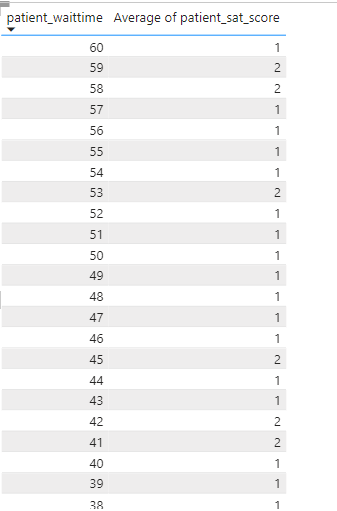
**Ans14.** On adding a column in Power Query, the code that appears in the M language formula bar starts with the function name that represents the action being performed. The specific code depends on the type of column added (e.g. a custom column, an index column, etc.).

An M-query is a query written in the M language, which is a data transformation language used in Microsoft Power Query, Power BI, Excel, and other Microsoft data manipulation tools. M (also known as Power Query Formula Language) is designed specifically for data manipulation and transformation, allowing users to import, transform, and combine data from different sources. It is a functional language and treats computations as the evaluation of mathematical functions without changing the state and data. Each query is a series of function applications, building upon the previous one. M supports various data types like text, numbers, lists, records (similar to rows in a table), tables, and more.

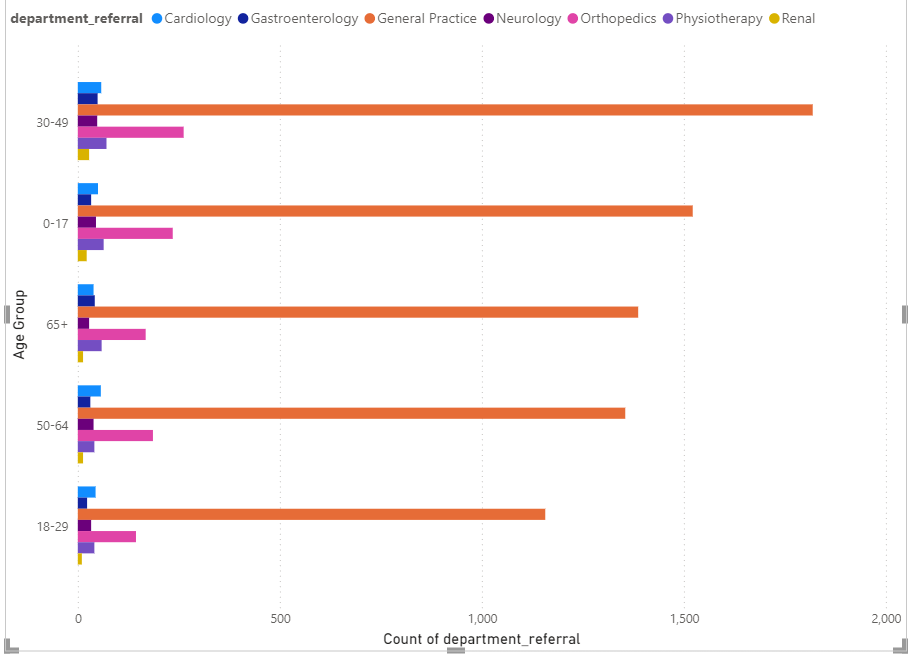
**Subjective Questions:**

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

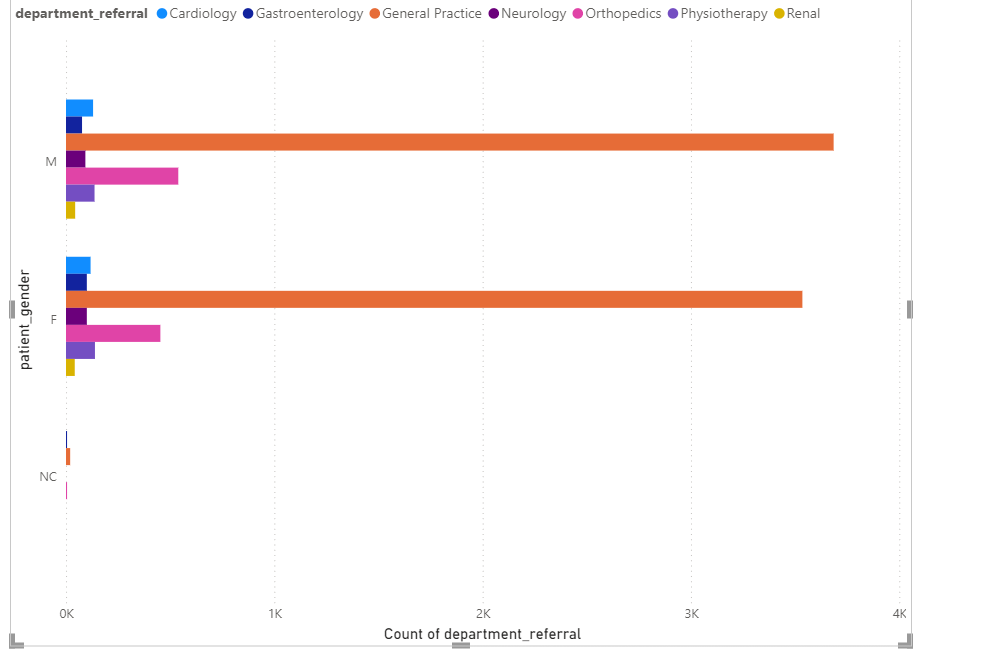
**Ans1.**  Tabulating both these attributes, it is seen that a waiting time shorter than 20 minutes yields, on average, a higher satisfaction score:



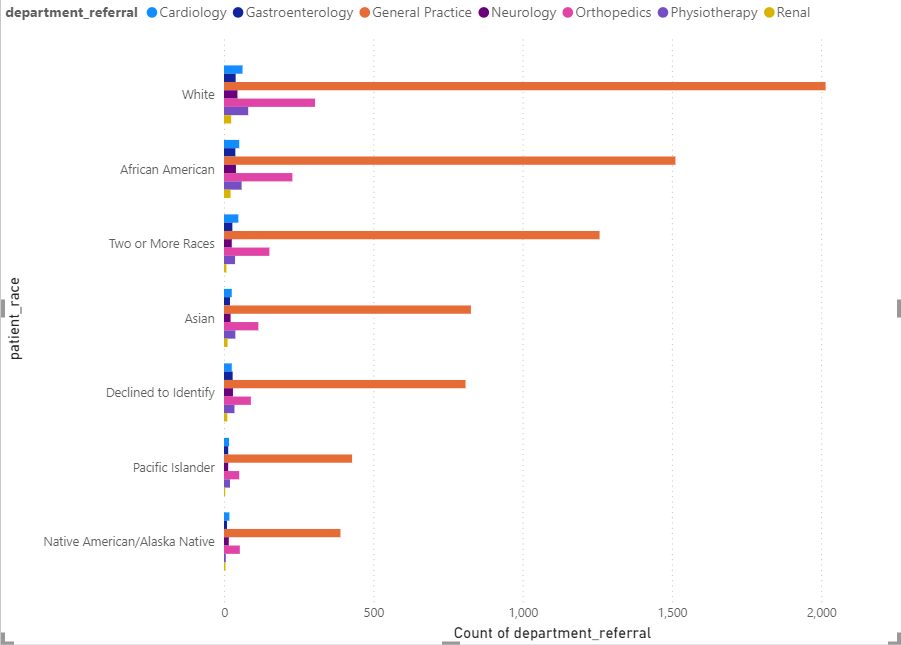
1. *How do patient demographics affect the frequency of visits to different departments?*

**Ans2.** The three given data on demographics (age, gender and race) are plotted against the visits on a clustered bar chart and prioritized in the aforementioned order:

**Ans2[contd].** This is drilled down to check by gender:



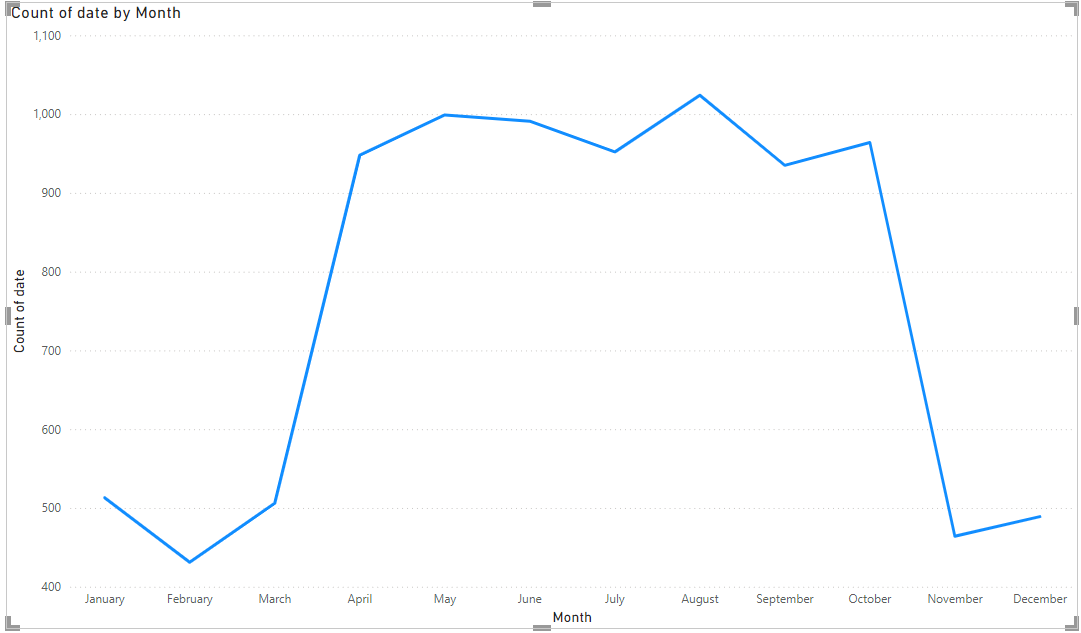
And again for race:



From each of the above, we gather the following insights: General Practice and Orthopedics remain the two most visited departments regardless of age, gender and race. Senior citizens make slightly more visits to Gastroenterology, Males make more to Cardiology and all race categories besides Whites, Afro-Americans and Multiracials make less-frequent visits in general.

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

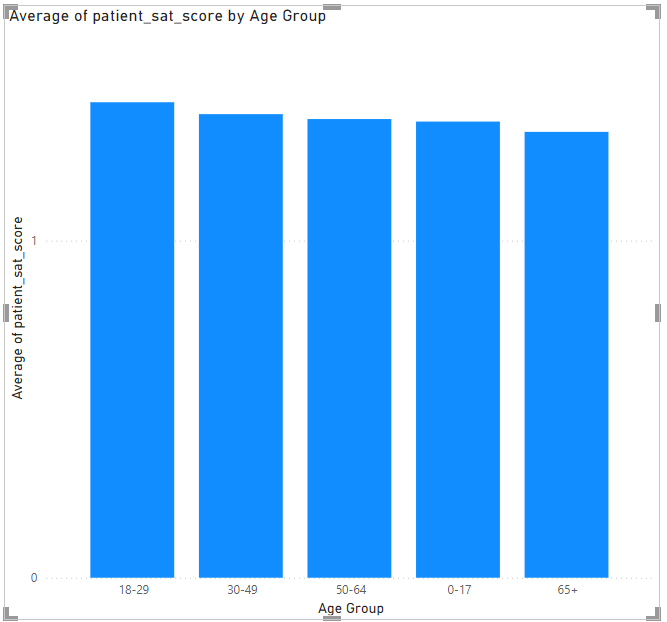
**Ans3.** On graphing a line chart with Month on the x-axis and the aggregated Count of Dates on the y-axis:



We find that the busiest periods for the hospital begin from March onwards and continue relatively consistently till late October (minus a decline in July, possibly attributable to Independence Day holidays), peaking in mid-August.

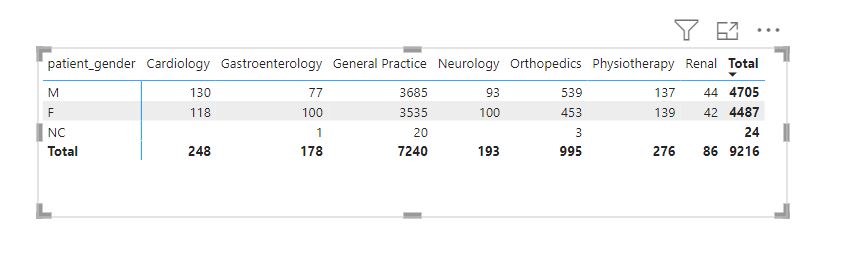
1. *Which age groups report the highest and lowest satisfaction scores?*

**Ans4.** The 18-29 reports the highest average satisfaction score, while the 65+ group reports the lowest:

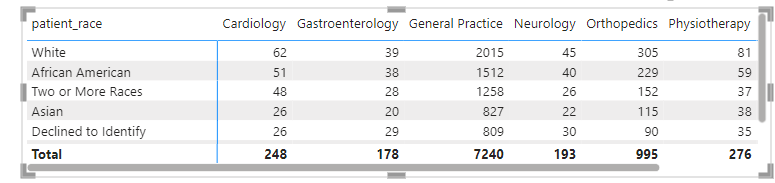


1. *Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?*

**Ans5.** To identify whether there is racial or gender-based discrimination in the hospital, we carry out a structured analysis of the patient data, focusing on comparing the treatment, frequency of visits, and outcomes across different racial and gender groups.

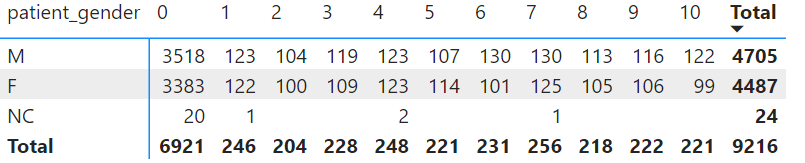


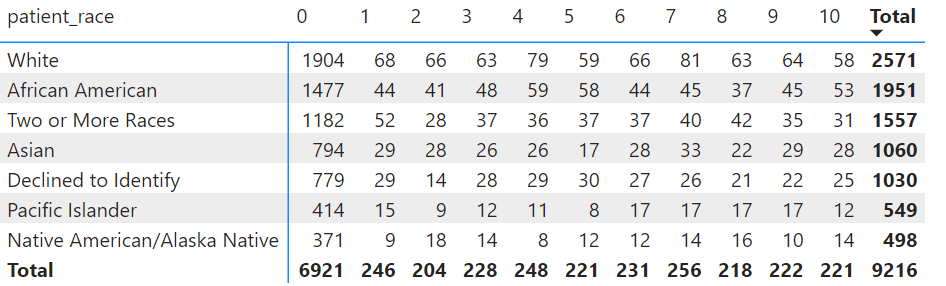
From the gender breakdown above, we find that males make more visits to cardiology, general practice, orthopedics and renal, while females make more to gastroenterology, neurology and physiotherapy, and that males in general comprise slightly more visits.



From the racial breakdown above, we find that Whites make up the most visits to all departments, and Asians/DTIs (overall) the least.

Additionally, we tabulate satisfaction scores across gender and race for more thorough verification:



****

**Ans5[contd].** Gender-Based Analysis**:** The total number of patients for each gender is nearly even: 4,705 males (M) and 4,487 females (F). Both males and females have a consistent spread of satisfaction scores across all levels. The highest number of patients for both males and females is at satisfaction score 0, which could indicate a dissatisfaction trend.

Race-Based Analysis: The largest racial group is White patients (2,571), followed by African American (1,951) and those identifying as Two or More Races (1,557). White patients have a broad distribution of satisfaction scores, but a significant number (804) rated their satisfaction at 0, which is the highest for this group. African American patients also show a high number of 0 satisfaction scores (474), suggesting dissatisfaction, but they have slightly more evenly spread satisfaction scores across the spectrum compared to White patients.

Conclusion: There aren’t notable differences between males and females in satisfaction scores. This is not the case for racial differences, particularly with a high number of dissatisfaction scores in certain racial groups.

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

**Ans6.** When assigning offers or discounts to patients, the hospital management should consider multiple factors to ensure that the discounts are both fair and effective in improving patient satisfaction and engagement.

1. Based on Satisfaction Scores: Patients who have reported low satisfaction scores should be prioritized for discounts. Offering discounts to these patients can help improve their perception of the hospital and encourage them to return, potentially boosting their satisfaction in the future.

2. Based on Demographic Factors: Elderly patients, who often require more frequent visits and treatments, could be offered discounts to reduce their financial burden. Alternatively, young patients might be targeted with discounts to encourage early preventive care.

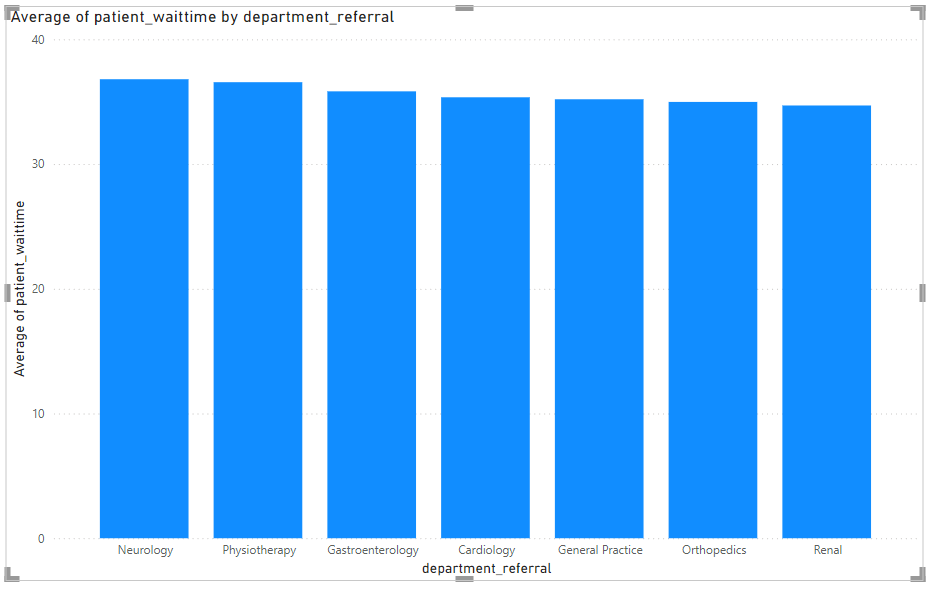
3. Based on Frequency of Visits: Patients who visit frequently could be rewarded with discounts as a form of loyalty incentive. E.g., after a certain number of visits, a patient might receive a discount on their next visit.

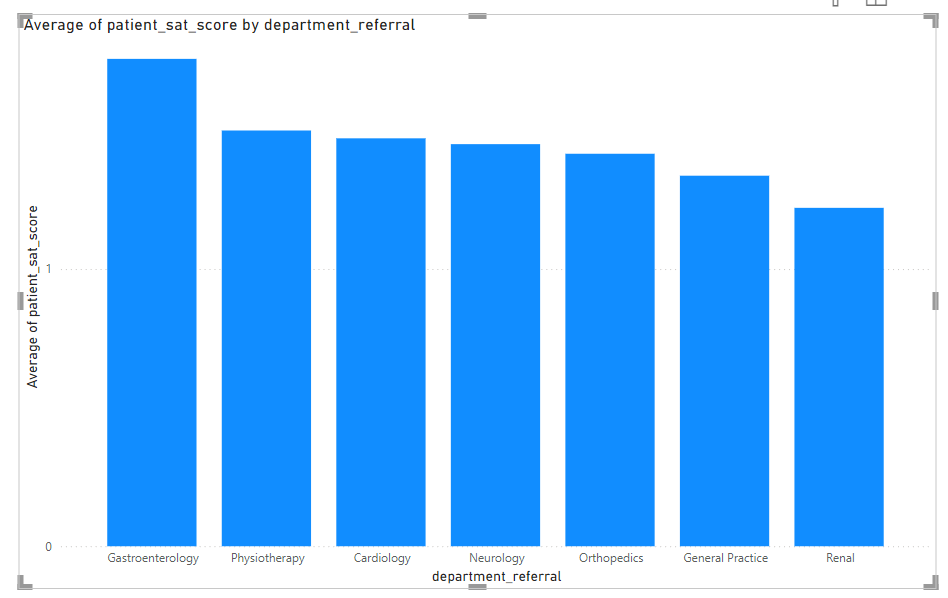
4. Target High-Cost Departments: Patients who frequently visit departments associated with high costs (e.g., oncology, surgery) might benefit from discounts to reduce financial strain.

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

**Ans7.** To provide well-informed suggestions on which departments the hospital should prioritize for hiring new doctors, let's consider the following key factors:

* 1. Departments that consistently handle a large number of patients or have long wait times due to high demand should be prioritized. Hiring in these areas can help reduce patient wait times, improve service delivery, and prevent burnout among existing staff. Here, they are Neurology and Physiotherapy.



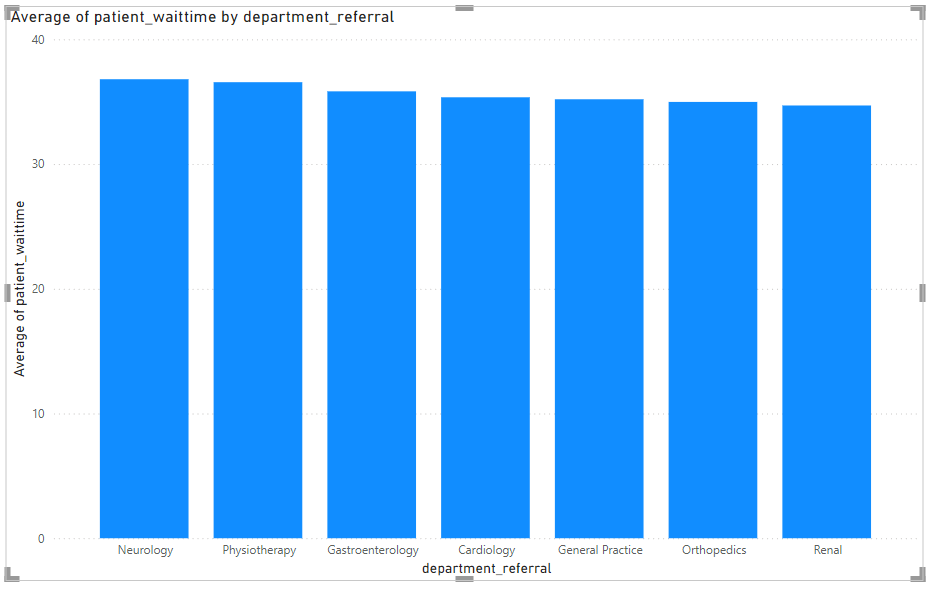
2. If certain departments have lower patient satisfaction scores, this could indicate a need for more staff to improve the quality of care. Hiring additional doctors in these departments could help address patient concerns and improve overall satisfaction.

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

**Ans8.** To calculate profit(or loss), we would need to first derive the hospital’s total revenue via a measure for SUM of the ‘Total Bill’ column. Then, another measure for the SUM of the expenses column would be subtracted from the first in order to confirm if and how large the hospital’s profit was.

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

**Ans9.** No department has an unexpectedly high waiting time, as inferred from the following:



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

**Ans10.** 1. Offer discounts based on the patient’s income level. Patients with lower incomes receive higher discounts.

2. Offer discounts to patients who frequently visit the hospital or have multiple appointments.

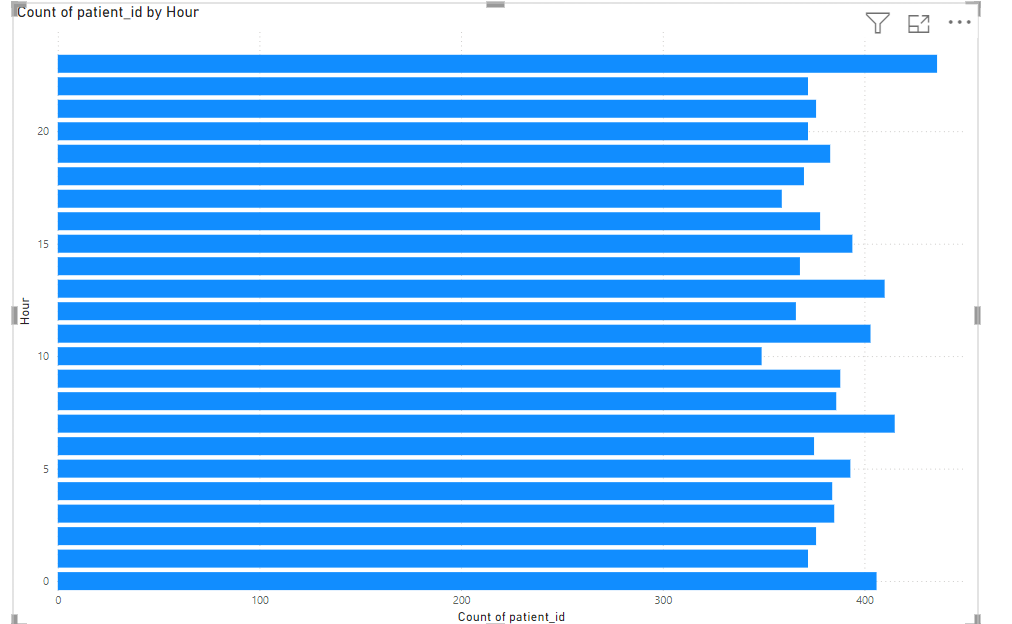
3. Offer special discounts to elderly patients as well as for pediatric care or young adult services.

4. Patients with chronic conditions that require regular treatment could be eligible for discounts.

5. Offer special discounts during holiday seasons or specific health awareness months.

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 will this 2 shift policy be helpful for the hospital?*

**Ans11.** We begin by plotting a bar chart for hour of visit [derived column using the Hour = HOUR('Hospital ER'[date]) expression] against count of Patient ID:



From the chart, it is observed that visits begin increasing by 6AM and peak from midday to 5PM, then declining.

Proposed shift timings are thus:

1. Morning Shift: 6 AM to 2 PM: This shift covers the time when patient visits start increasing and goes through the morning.
2. Afternoon/Evening Shift: 2 PM to 10 PM: This shift covers the afternoon peak and extends into the evening, just before visits start to decline.

Ideally, an equal number of doctors should be apportioned to each shift. Since the Hospital has 22 doctors on payroll, each shift could receive 11 each.

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

**Ans12.** Power BI Gateway is a software that acts as a bridge, allowing secure data transfer between Power BI services in the cloud and on-premises data sources. This is particularly useful when you want to keep your data on-premises for security reasons but still leverage Power BI's cloud-based services for data analysis and reporting. A few essential features are:

1. The Gateway allows us to connect to various on-premises data sources like SQL Server, Oracle, SAP, and files stored in network folders.
2. Enables the scheduling of automatic data refreshes, ensuring that Power BI reports and dashboards always display the latest information from on-premises data sources.
3. Power BI Gateway supports DirectQuery, which allows running live queries against on-premises data sources, ensuring real-time data availability in reports.
4. Central management of all gateways in the Power BI service, including adding data sources, monitoring gateway health, and configuring gateway settings.

Power BI Gateway’s use-cases include:

* When an organization has sensitive data stored on-premises that needs to be kept secure while still using Power BI's cloud services.
* When we need to schedule automatic refreshes for reports that connect to on-premises data.
* When we require real-time data reporting using DirectQuery.

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

**Ans13.** Without the guiding objective/subjective questions, I would have attempted to gather my own observations by experimenting with differing permutations of the dataset’s given metrics, visualizing each to better understand potential trends or patterns pertaining to a certain demographic, department or time of year.

Next, I would decide key objectives by identifying core metrics that would be valuable to the hospital management. Also essential would be to consider what hospital management, doctors and patients might want to know.

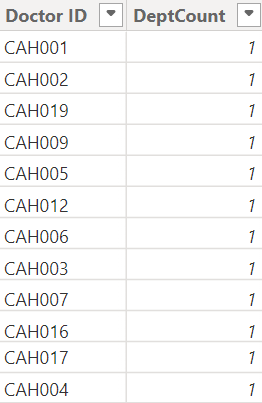
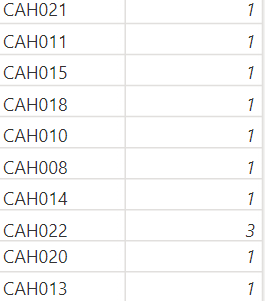
After this, I would conduct more advanced analysis, such as segmentation analysis to understand the behavior and outcomes of different patient groups, profitability analysis to compare revenues and costs across departments & satisfaction analysis to pinpoint areas needing improvement.

The insights gained from these analyses may then be integrated towards actionable recommendations for hospital management.

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

**Ans14.** To investigate the relationship between these attributes, we begin by deriving a new table for distinct departments per doctor, using the following DAX expression:

DoctorDeptCount = SUMMARIZE(Sheet1, Sheet1[Doctor ID], "DeptCount", DISTINCTCOUNT(Sheet1[department\_referral]))



As seen in the visualization, one of these doctors, namely CAH022, links to multiple departments. Thus, not every row has a DeptCount of one and the relationship is one-to-many.