**Objective Questions:**

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

To ensure data cleanliness in the hospital dataset with Power BI:

* Identify Missing Values: Inspected the dataset for null or inconsistent values using Power Query Editor.
* Handle Missing Data: Replaced null values, such as in the patient\_sat\_score column, with the average value to maintain data completeness.
* Ensure Consistency: Corrected any inconsistencies or erroneous data to prepare it for analysis.

This cleaning process ensures the dataset is accurate and reliable for further analysis.

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

Average Waiting Time Analysis:

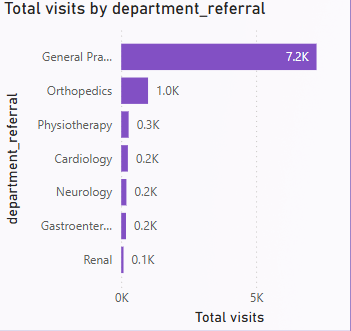
* Formula Used: DAX

Result: The average patient waiting time is 35.26 minutes.

AVG\_waiting\_time = AVERAGE('Hospital ER'[patient\_waittime])

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.

* Formula Used: Count of  Patient = COUNT('Hospital ER'[patient\_id])
* Visualization: Created a chart with:
  + X-axis: Department
  + Y-axis: Count of patients (measure).



* Result: identified the total number of visitors per department based on referrals.

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

Approach :

* To better understand patient demographics and how different age groups utilize healthcare services

Formula Used:

Created a new column using SWITCH to categorize patients by age group:  
  
Age\_Group = SWITCH(

 TRUE(),

 'Hospital ER'[patient\_age] < 18, "0-17",

 'Hospital ER'[patient\_age] >= 18 && 'Hospital ER'[patient\_age] <= 34, "18-34",

 'Hospital ER'[patient\_age] >=35 && 'Hospital ER'[patient\_age] <=49, "35-49",

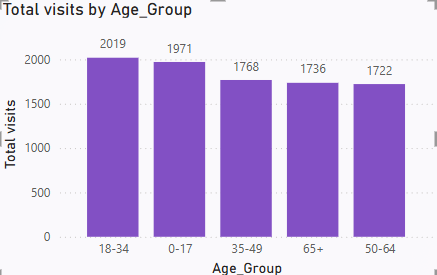
 'Hospital ER'[patient\_age] >=50 && 'Hospital ER'[patient\_age] <=64, "50-64",

 'Hospital ER'[patient\_age] >= 65, "65+",

 "Unknown"

)  
  
Visualization:

* Added the new Age Group column to the chart's X-axis and used a measure to count the number of visitors for each age group.



* Result:

You successfully created age buckets and visualized the number of visits for each demographic group.

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?**

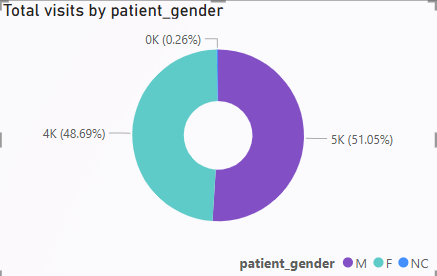
Null values were present in the patient satisfaction scores, which were subsequently filled with the average satisfaction score to ensure data completeness.

* + Identified Null Values:
    - Inspected the patient\_sat\_score column in Power Query to locate rows with missing values.
  + Calculated the Average:
    - Used Transform > Statistics > Average to calculate the column's average (4.9978298611111107)
  + Replaced Nulls with the Average:
    - Rounded the average to 5 and replaced all null values using Transform > Replace Values.

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

Analysis of Visits by Gender

* + A breakdown of patient visits reveals the following gender distribution:
    - Male: 51.05%
    - Female: 48.69%
    - Not Categorized (NC): 0.26%
  + Visualization:
    - This indicates that the number of visits is almost evenly distributed between males and females, with a slightly higher proportion of male patients. The NC category is negligible at 0.26%.



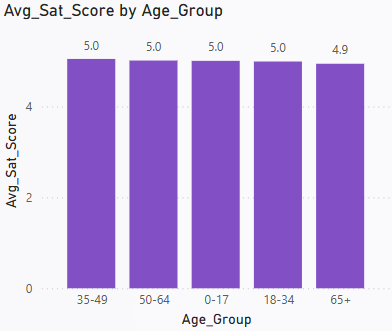
* Result:

The donut chart visually represents this distribution, highlighting the near parity in visits across genders.

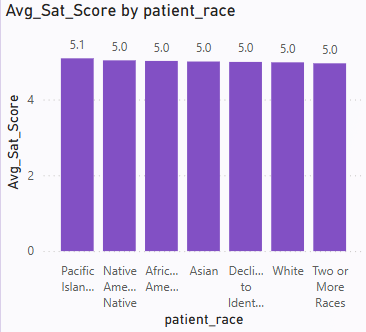
1. **Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.**

Analysis of Average Satisfaction by Demographics

* + Age Group Analysis
    1. Patient satisfaction scores across different age groups range between 4.92 and 5.05.
    2. The 36-50 age group reported the highest satisfaction, with an average score of 5.05.
    3. The 65 and above age group showed the lowest satisfaction, with an average score of 4.92.
  + Visualization:



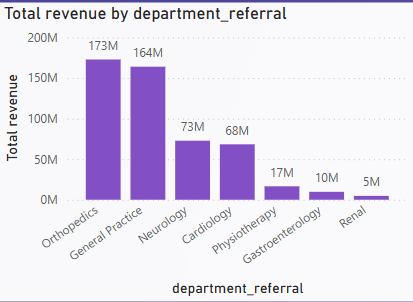
* + Insight:
    - This suggests that younger and middle-aged patients are slightly more satisfied compared to older patients, indicating potential areas for improvement in services for senior patients.
  + Race Analysis
    - Patient satisfaction scores by racial background range from 4.96 to 5.09.
    - The highest average satisfaction score of 5.09 was reported by Pacific Islanders, while other racial groups fell slightly below this mark.
    - The lowest satisfaction score of 4.96 was observed among a specific racial group, highlighting an opportunity to focus on enhancing their experience.
  + Visualization



* Insight:

These findings underline the importance of tailoring healthcare services to meet the unique needs of older patients and certain racial groups to improve overall satisfaction.

1. **The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.**
   * Department-wise Revenue Analysis
     1. To evaluate the revenue generated by each department, I created a measure: Total\_Revenue = SUM(Sheet1[Total Bill]), which calculates the total billing amount across all departments.
     2. Using this measure, I developed a column chart by plotting Department on the axis and Total\_Revenue on the values.
   * Visualization



* Insights:

Orthopedics emerged as the highest revenue-generating department, indicating its substantial contribution to the hospital’s financial performance.

Renal recorded the lowest revenue, suggesting a need for further investigation into potential factors such as service offerings, patient inflow, or operational challenges.

* Result :

This department-wise revenue breakdown provides critical insights for decision-making, enabling management to optimize resources, improve low-performing areas, and capitalize on high-performing departments.

1. **Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.**
   * Highest Appointment Fee by Department
     1. To identify which department is charging the highest appointment fees in general.
     2. I used the LOOKUPVALUE DAX function to find the department corresponding to the maximum appointment fee. The measure created is:

Highest\_Appointment\_fee\_Department = LOOKUPVALUE('Doctor Info'[department\_referral],'Doctor Info'[Appointment Fees],max('Doctor Info'[Appointment Fees]))  
A close up of a sign

Description automatically generated

* Using this measure, I visualized the data by displaying the department charging the highest appointment fees in a card.
* Insights:

The analysis revealed that the Neurology department charges the highest appointment fee among all departments.

This insight highlights Neurology’s premium positioning, possibly due to specialized expertise, advanced equipment, or demand for its services.

* Result:

This analysis helps management better understand pricing strategies across departments and evaluate their alignment with the hospital's revenue goals and market positioning.

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.**

Month-wise Total Visits with Previous Month Comparison Table

* + To create a tabular visualization displaying Month-wise Total Visits, Previous Month's Visits, and a comparison column (Visit Increased?)

a. Create a Calendar Table

* + You created a Date Table using the CALENDAR function to cover the date range in the Hospital ER Table:
  + Date = CALENDAR(MIN('Hospital ER'[date - time]),MAX('Hospital ER'[date - time]))

Purpose: Ensures the date table spans the entire range of dates in the hospital data.

Mark as Date Table:

Mark this table as the official Date Table under Modeling > Mark as Date Table.

b. Define Year, Month, and Month Name Columns

* + To organize data, you added calculated columns in the Date Table:
  + Year Column:
    - Year = YEAR('Date'[Date])
  + Month Column:
    - Month = MONTH('Date'[Date])
  + Month Name Column:
    - MonthName = FORMAT('Date'[Date], "MMMM")
  + These columns help structure and sort data chronologically.

c. Establish Relationships

* + Create a One-to-Many relationship between the Hospital ER Table (Date Column) and the Calendar Table (Date Column).
  + Set the cross-filter direction to Single.

d. Create Measures

* You defined the following measures:

Current Month Visits: Month\_Wise\_Total\_Visits = COUNT('Hospital ER'[patient\_id])

Previous Month Visits: Pre\_Month\_Total\_visits = CALCULATE(COUNT('Hospital ER'[patient\_id]), PREVIOUSMONTH('Date'[Date]))

Visit Changetype =

VAR CurrentMonthVisits = [Total visits]

VAR PrevMonthVisits = [Previous\_Month\_Total\_visits]

RETURN

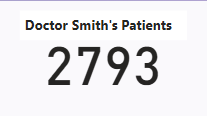
IF(CurrentMonthVisits > PrevMonthVisits, "Increased", "Decreased")

e. Tabular Visualization

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

* To calculate the total number of patients who have visited Dr. Smith, you can use the following DAX formula:

Dr Smith's Patients = CALCULATE(COUNT('Hospital ER'[patient\_id]), 'Doctor Info'[Doctor Name] = "Dr. Smith")



* Explanation:

CALCULATE changes the context of the calculation, allowing you to apply filters.

COUNT('Hospital ER'[patient\_id]) counts the number of patients.

'Doctor Info'[Doctor Name] = "Dr. Smith" filters the data to only include records where the Doctor Name is "Dr. Smith".

* Results:

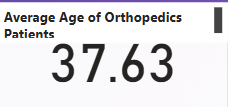
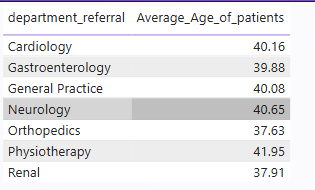
Using the DAX formula, we calculated that 5,986 patients have visited Dr. Smith.

This was achieved by filtering the Doctor info table for entries where the Doctor Name is "Dr. Smith" and counting the number of unique patient\_id entries.

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?**
   * Approach
     1. Orthopedics: Create a measure with a filter to calculate the average age for Orthopedics and display it in a Card Visual.
     2. All Departments: Create a general measure without filters, add it to a Table Visual with the department column to display averages for all departments.
   * Explanation
     1. To calculate the average age for Orthopedics, a measure is created using the following formula
     2. Average\_Age\_Othropedics\_patients = CALCULATE(AVERAGE('Hospital ER'[patient\_age]),'Hospital ER'[department\_referral] = "Orthopedics")
     3. The AVERAGE function then computes the average age, and the result is displayed in a Card Visual, as it represents a single value.
     4. For all departments, a general measure is created without any filter:

Average\_Age\_of\_patients = AVERAGE('Hospital ER'[patient\_age])

* + 1. This measure calculates the overall average age for patients. Adding this measure to a Table Visual, along with the department\_referral column, groups the data by department and shows the average age for each.
    2. Finally, the measures are formatted as whole numbers using the data format options in Power BI to ensure the results align with the realistic representation of age.
  + Visualization



* + Result

A Card Visual shows the average age for Orthopedics.

A Table Visual displays the average age for each department.

These distinct approaches ensure accurate and meaningful insights tailored to each requirement.

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

Steps Taken to Resolve Data Format Issues:

* + Combining Columns for Patient Name:
    - The patient\_first\_initial and patient\_last\_name columns were stored separately, making it inconvenient for analysis.
    - Using Power Query, I merged these columns into a single patient\_name column, ensuring proper formatting for better usability in queries and reports.
  + Extracting Dates and Creating a Calendar Table:
    - The date column included both date and time, but only the date component was needed for analysis.
    - I created a custom column to extract the date and generated a date table to enable advanced date-based filtering and trend analysis.
  + Handling Blank Spaces in Scores:
    - The patient\_sat\_scores column had blank spaces, which could disrupt calculations and analysis.
    - Initially, the data type of patient\_sat\_scores was in text format, so I converted it to a number.
    - After conversion, I replaced the blank spaces (null values) with the column's mean value, ensuring data consistency and avoiding skewed results during computations.
  + Result:
    - The dataset was transformed into a clean and consistent format, making it ready for reliable analysis and accurate insights.

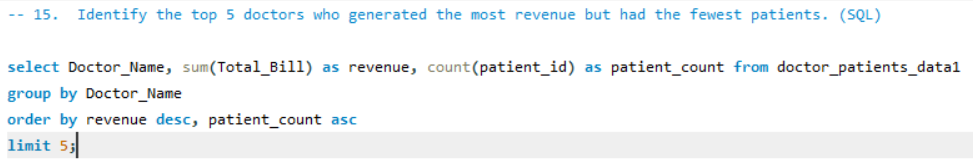
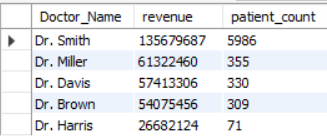
1. **When we add a column in Power Query what’s the code that comes in M language in the formula bar? What do you know about M-query?**
   * When you add a column in Power Query, the M language automatically generates a formula in the formula bar. For example, if you create a column to concatenate the first and last names of patients, the following code is generated:

= Table.AddColumn(#"Removed Duplicates", "patient\_name", each [patient\_first\_inital] & " " & [patient\_last\_name])

* + Explanation of the Formula
    1. Table.AddColumn: This function adds a new column to the existing table.
    2. #"Removed Duplicates": Refers to the previous step in the query where duplicates were removed.
    3. "patient\_name": Specifies the name of the new column being created.
    4. each [patient\_first\_inital] & " " & [patient\_last\_name]: Combines the values in the patient\_first\_inital and patient\_last\_name columns with a space in between, row by row.
  + What is M-Query?
    1. M-Query is a formula language used in Power Query to manipulate, transform, and shape data. It is case-sensitive and functions in a step-by-step manner where each step is recorded as code. Key characteristics of M-Query:
    2. Designed for data transformation.
    3. Operates in a functional programming style.
    4. Is highly readable and editable through the Advanced Editor.
    5. By leveraging M-Query, you can customize and automate complex data transformation tasks efficiently.

1. Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)

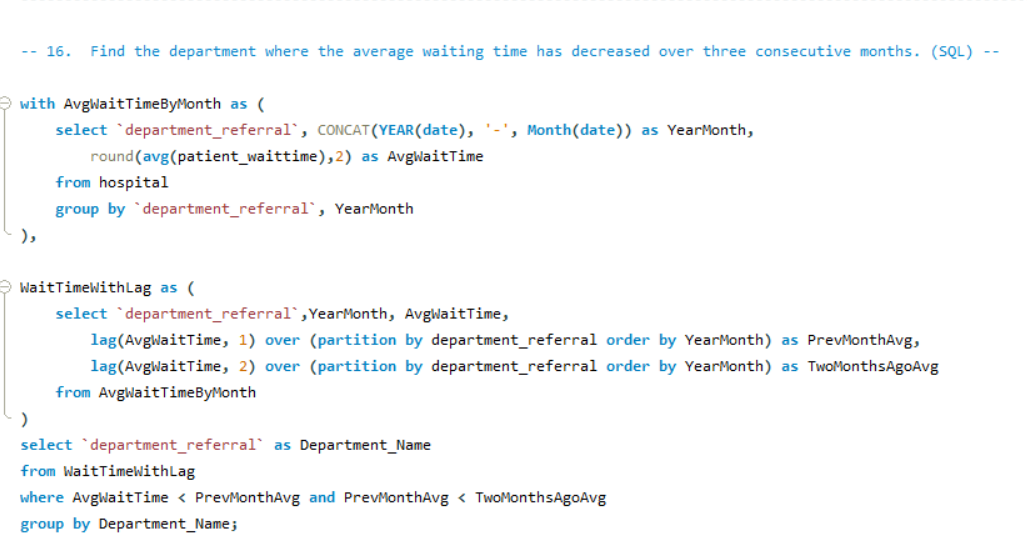
* The top five doctors with the highest revenue despite having the fewest patients are identified by analyzing their total revenue and unique patient count. This highlights doctors who specialize in high-value services, making them key drivers of revenue efficiency within the organization.

Query:   
  
  
Output:  


1. Find the department where the average waiting time has decreased over three consecutive months. (SQL)

* Departments that showed a consistent decrease in average waiting times over three consecutive months were identified. This reflects improved efficiency and better patient flow management, leading to enhanced service quality.

Query:

  
  
Output:

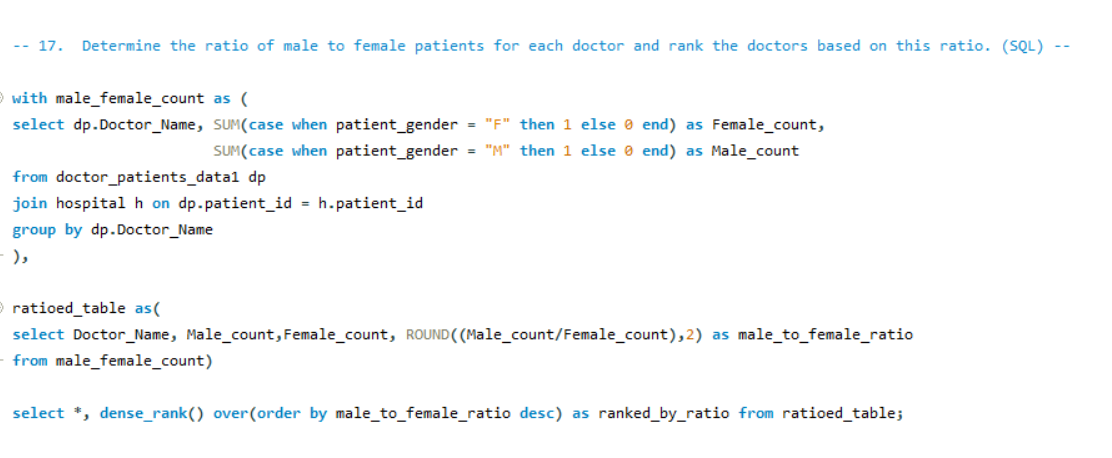
A screenshot of a computer

Description automatically generated

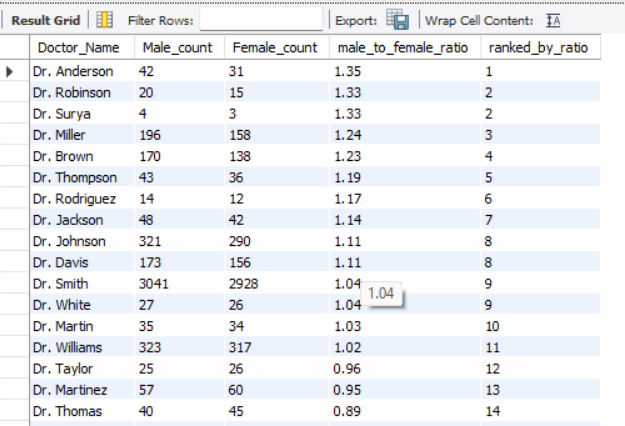
1. Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)

* The male-to-female patient ratio for each doctor is calculated and used to rank them in descending order. This analysis offers insights into patient demographics, helping to identify trends or potential biases in patient distribution.

Query:



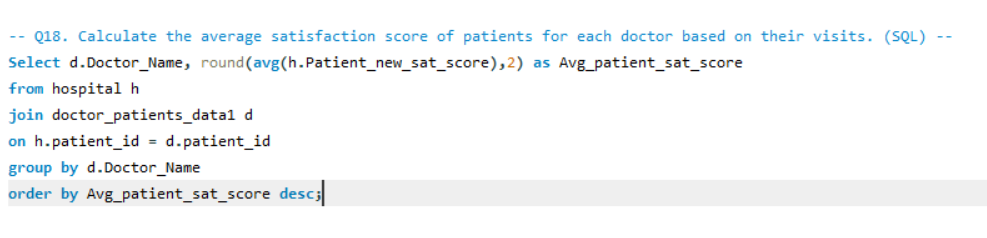
Output:



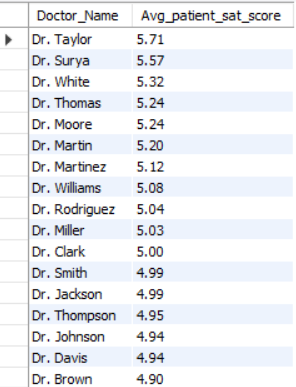
1. Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)

* The average patient satisfaction score for each doctor is calculated, with missing scores replaced by a default value of 5. This provides a comprehensive measure of patient experience, ranking doctors based on the highest satisfaction scores.

Query:



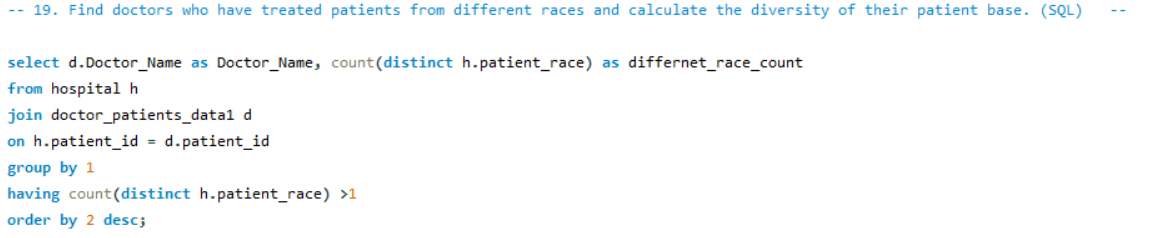
Output:



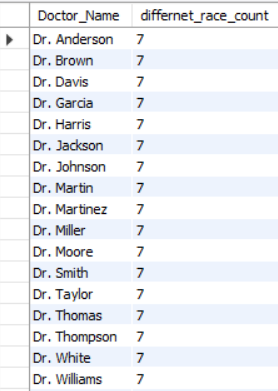
1. Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)

* Doctors who have treated patients from multiple races are identified, with the diversity of their patient base measured by the number of distinct races they have treated. This highlights doctors serving a more diverse patient demographic.

Query:



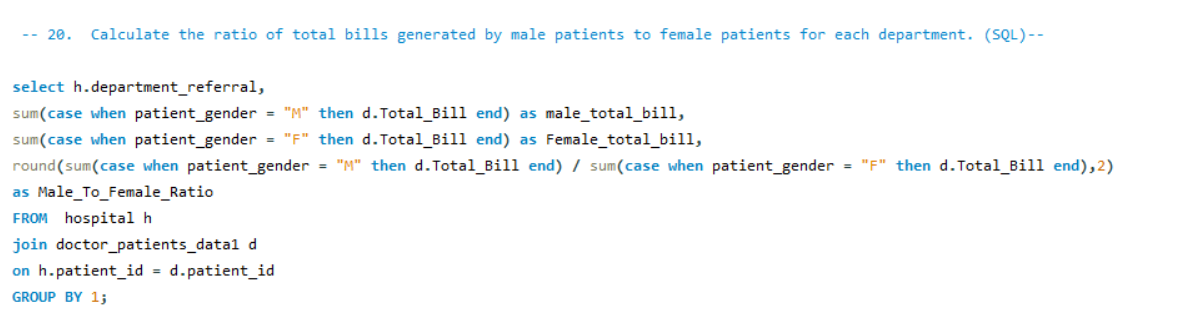
Output:



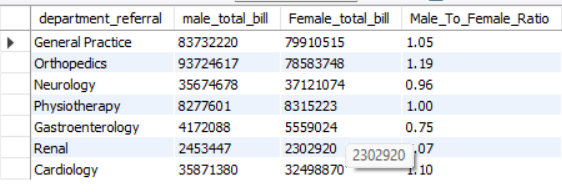
1. Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)

* The ratio of total bills generated by male to female patients is calculated for each department. This analysis offers insights into the financial contribution of each gender, helping to identify spending trends across departments.

Query:



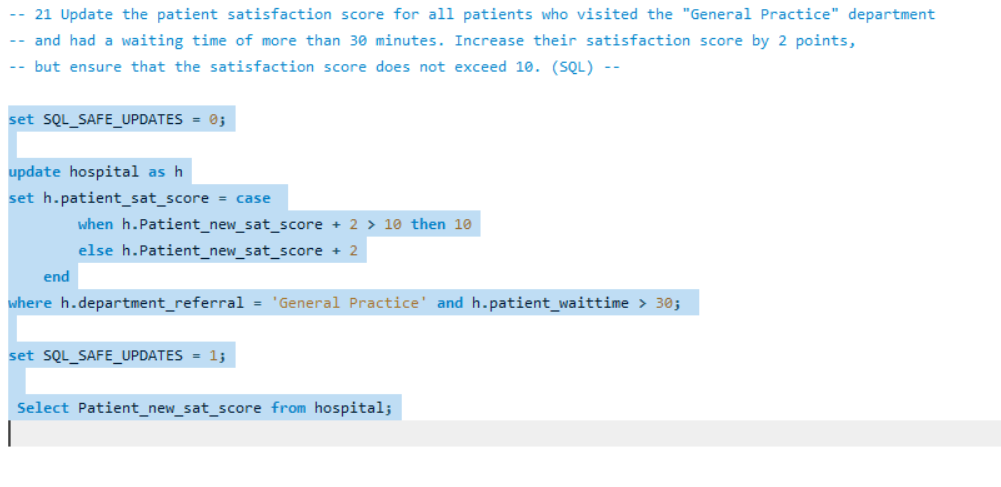
Output:



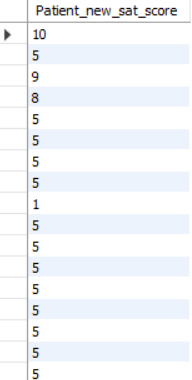
1. Update the patient satisfaction score for all patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. Increase their satisfaction score by 2 points, but ensure that the satisfaction score does not exceed 10. (SQL)

* Patient satisfaction scores in the "General Practice" department are increased by 2 points for those who experienced a waiting time of over 30 minutes. If the adjusted score exceeds 10, it is capped at 10. This ensures fair adjustments while maintaining the maximum allowable score.

Query:



Output:



**Subjective Questions**

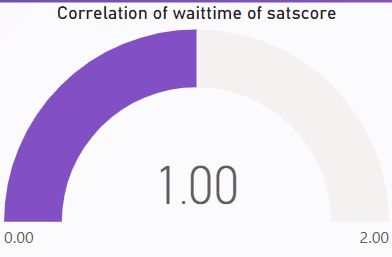
1. What is the relation between patient wait time and satisfaction scores?

**Approach:**

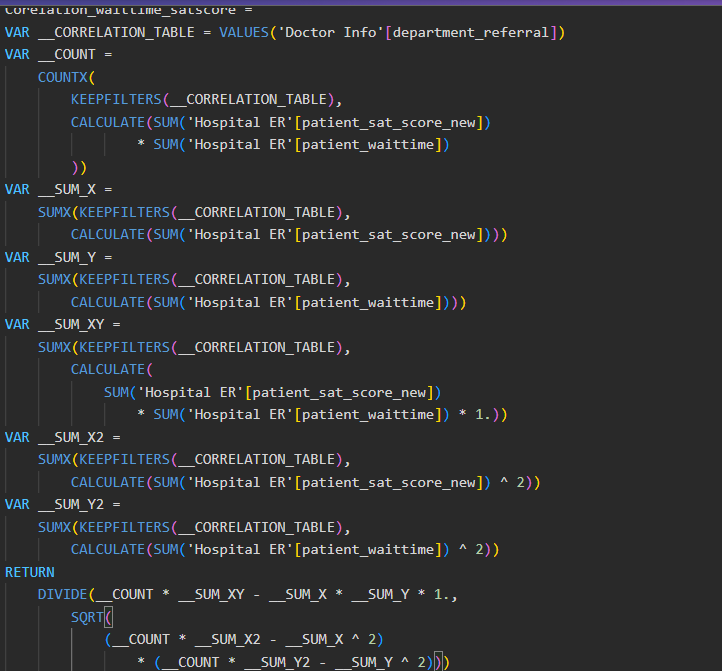
* A quick measure was created in Power BI by assigning department\_referral as the category, patient\_sat\_score as Measure1, and patient\_waittime as Measure2.
* The DAX formula was automatically generated to calculate the correlation between these two measures.
* To visualize the correlation value, a Gauge Chart was used with a range set from 0 to 2, where the final correlation value observed was 1.

**Explanation:**

* Correlation analysis in this context assesses the strength and direction of the relationship between patient wait times and satisfaction scores.
* A correlation value of 1 indicates a moderate positive relationship, implying that as wait times increase, satisfaction scores generally decrease, though not in a strictly linear manner.
* This suggests that while wait times impact patient satisfaction, other factors—such as staff behavior, quality of care, or facility conditions—may also play a role.
* The Gauge Chart visually represents this correlation, making it easier to interpret the score and understand its significance within the given range.
* Visualization:



**Insight:**

* The analysis identified a moderate positive correlation of 1, indicating that while patient wait times influence satisfaction scores, the effect is not dominant.
* Departments that minimize wait times or effectively manage patient expectations may achieve higher satisfaction levels, but factors like service quality and communication likely play a significant role in shaping patient perceptions.
* **DAX Formula  
  **

**Result:**

* This analysis underscores the importance of balancing wait times with other quality-of-service factors to enhance overall patient satisfaction.

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

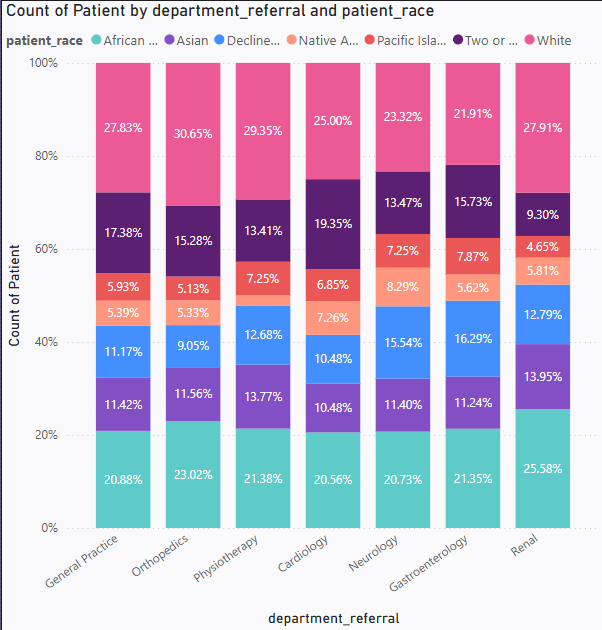
**Approach**

* Organize the data with columns for **Department, Patient Count,** and key demographics (**Age Group, Race, Gender**).
* Construct a **Stacked Column Chart**, setting **Department** on the x-axis and **Patient Count** on the y-axis.
* Implement **Field Parameters** to enable dynamic filtering, allowing users to switch between demographic variables (**Age Group, Race, Gender**) for deeper insights.

**Explanation**

* This analysis aims to understand how patient demographics affect visit frequency across different departments. The **Stacked Column Chart** provides a visual breakdown of how various demographic groups contribute to the total patient count per department.
* **Field Parameters** enable dynamic toggling between demographic categories (**Age Group, Race, Gender**), allowing for a detailed analysis of their impact on department visit frequency.
* This approach helps identify whether specific departments are preferred by certain demographic groups and reveals potential patterns or disparities in departmental utilization based on patient characteristics.

**Visualizations**

****

**Insight**

* Demographics do not influence the number of visitors to various departments, as visit frequency primarily depends on the patient's health condition and the hospital's location. The demographic composition of different departments remains largely unaffected.

**Result**

* The **Stacked Column Chart** confirms that demographic factors have minimal impact on overall visit counts across departments.

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

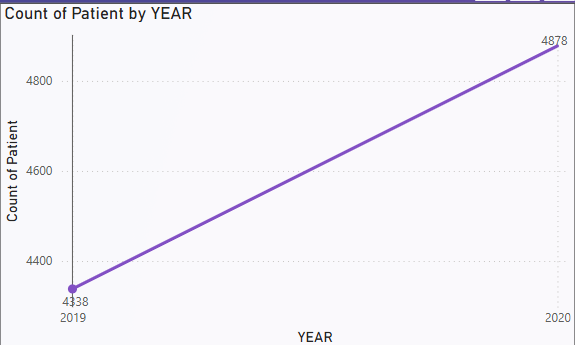
**Approach**

* Utilize a **Line Chart** to illustrate the trend in patient visit volume over time.
* Place **Year and Month Name** (in a hierarchy) on the x-axis to enable **drill-up and drill-down** functionality, allowing trend analysis at different time granularities.
* Use the **Count\_of\_Patients** measure on the y-axis to track patient visits per month.

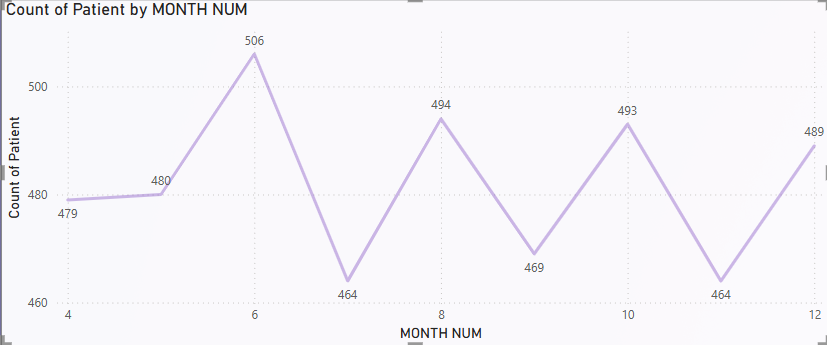
**Explanation**

* This analysis aims to uncover patterns in patient visit volumes throughout the year. A **Line Chart** with a **Year-Month hierarchy** helps visualize both annual and monthly fluctuations.
* The **Count\_of\_Patients** measure highlights variations in patient visit frequency over time.
* The **drill-up and drill-down** features allow for both a broad (yearly) and detailed (monthly) view, helping identify **seasonal trends** or **irregular fluctuations** in patient visits.

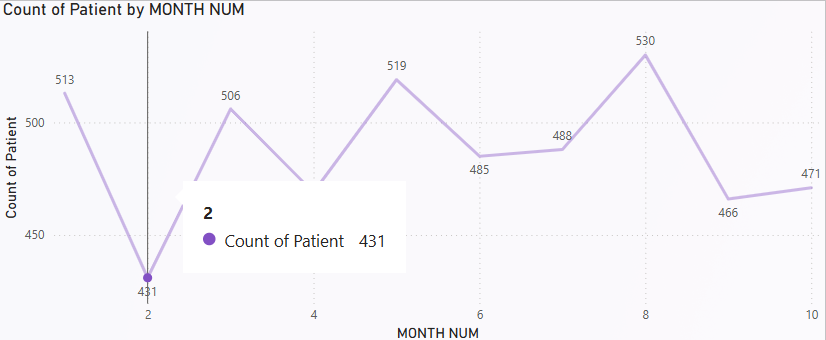
**Visualization**



Year 2019



Year 2020



**Insight**

* In **2019**, the total patient visits were **4,338**, increasing to **4,878** in **2020**, reflecting a **7.5% rise** in visits.
* This growth suggests potential shifts in **healthcare demand**, possibly influenced by factors such as **greater healthcare awareness, new medical services,** or **external events like the pandemic**.

**Result**

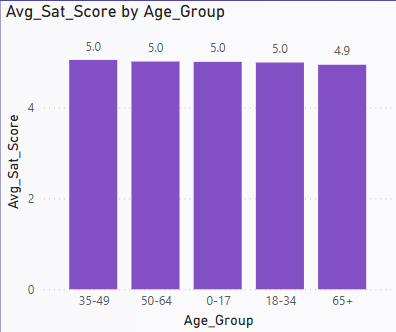
* The analysis highlights a **notable upward trend** in patient visits in **2020 compared to 2019**, indicating changing healthcare dynamics.

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

**Age Group Analysis**

* Patient **satisfaction scores** across different age groups range between **4.92 and 5.05**.
* The **36-50 age group** reported the **highest satisfaction**, with an **average score of 5.05**.
* The **65 and above age group** had the **lowest satisfaction**, with an **average score of 4.92**.

**Visualization**



**Insight**

* Younger and middle-aged patients exhibit **slightly higher satisfaction** levels compared to older patients.
* This indicates potential areas for **service improvements** tailored to the needs of senior patients, such as **better accessibility, personalized care, or enhanced communication.**

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?

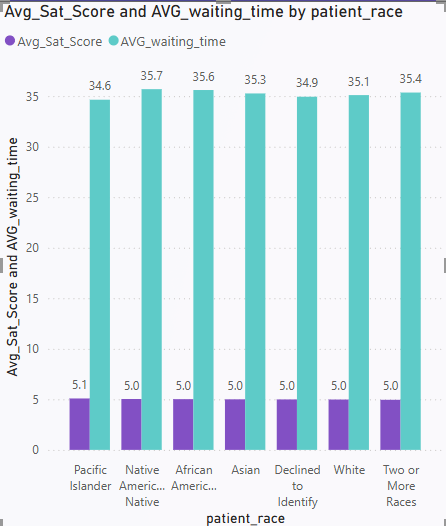
**Approach**

* Create a **Clustered Column Chart** with a **Field Parameter** on the **X-axis**, enabling the ability to toggle between **race** and **gender** for comparison.
* On the **Y-axis**, plot both **average waiting time** and **average satisfaction score** for each group (race and gender).
* This chart facilitates the comparison of **patient experiences** across different **race and gender groups**, specifically in terms of waiting time and satisfaction scores.

**Explanation**

* A **Clustered Column Chart** is used to investigate any potential evidence of **racial or gender-based discrimination** within the hospital.
* The **Field Parameter** on the **X-axis** allows for dynamic switching between **race** and **gender** categories, while the **Y-axis** displays **average waiting times** and **average satisfaction scores**.
* By comparing these metrics, we can assess whether certain demographic groups face **longer wait times** or **lower satisfaction**, potentially highlighting areas for improvement or discrimination.

**Visualization**

****

**Insight**

* Analyzing the **average waiting times** and **patient satisfaction scores** indicates **no evidence** of racial or gender-based discrimination within the hospital.
* The data reflects **consistent and equitable treatment** across different **race and gender** groups, with **no significant disparities** in either **waiting times** or **satisfaction levels**.

**Result**

* The analysis confirms that **waiting times and satisfaction scores** remain **uniform** across **race and gender groups**, suggesting that the hospital maintains **fair and non-discriminatory** practices.

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

**Approach:**

* A **DAX formula** was applied to determine **discount eligibility** based on the following conditions:
  + **Total Bill ≥ 10,000**
  + **Satisfaction Score ≥ 6**
* A **DiscountEligibility** column was created to classify patients as either **"Eligible"** or **"Not Eligible."**
* A **Donut Chart** was used to visualize the proportion of **eligible** vs. **non-eligible** patients.

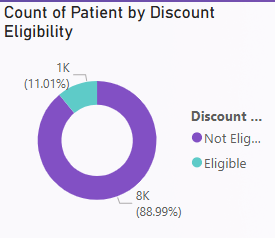
**Explanation:**

* The hospital management aimed to offer **discounts** to patients who were **high spenders** and had **high satisfaction scores**.
* A **DAX formula** was used to categorize patients based on their **total bill** and **satisfaction score**, ensuring that only those meeting both criteria qualified for the discount.

The formula used is:

Discount Eligibility = IF(AND('Doctor Info'[Total Bill]>=10000,'Doctor Info'[Sat\_score]>=6),"Eligible","Not Eligible")

**Visualisation:**

****

**Insight:**

* Out of **9,216** total patients, only **11.01%** qualified for the discount, while the remaining **88.99%** did not.

**Result:**

* The **Donut Chart** visually highlights that a **small proportion (11.01%)** of patients meet the discount criteria, whereas the **majority (88.89%)** do not qualify.

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

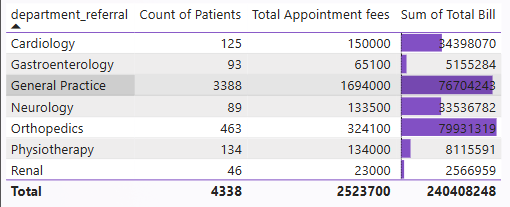
**Approach:**

* Analyzed hospital **departments** using key metrics:
  + **Count of Patients**
  + **Sum of Appointment Fees**
  + **Sum of Total Bill**
* Identified departments needing **2-3 additional doctors** based on high patient volumes and revenue.
* Used a **table visualization** to compare metrics across departments.

**Explanation:**

* Departments with the highest demand and revenue were prioritized.
* **Orthopedics** had the **highest** patient count (**995**), **appointment fees (696,500)**, and **total bill (172,939,773)**, indicating a need for more doctors.
* **General Practice** had the **largest patient count (7,240)** and significant revenue, making it a key department for expansion.
* **Neurology** had **fewer patients (193)** but generated **high appointment fees and total bills**, suggesting a **high-value specialty** needing additional support.

**Visualization:**



**Insight:**

* **Orthopedics** and **General Practice** have the **highest patient load and revenue**, making them the **top priorities** for new hires.
* **Neurology**, despite having **fewer patients**, still generates **high revenue**, indicating **strong demand** for specialized care.

**Result:**

* **Hiring additional doctors** in **Orthopedics, General Practice, and Neurology** is recommended to manage patient demand and maintain service quality.

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

**Approach:**

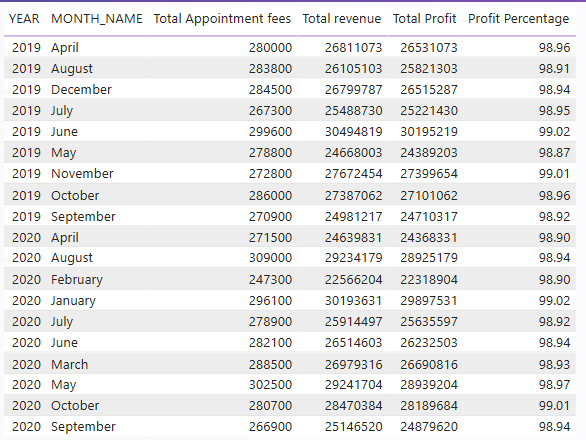
* Created a **visual table** displaying **Year** and **Month** as columns.
* Used **DAX measures** to calculate key profitability metrics:
  + **Total Revenue** (Sum of total bill amounts)
  + **Total Appointment Fees** (Sum of all appointment fees)
  + **Total Profit** (Revenue - Appointment Fees)
  + **Profit Percentage** ((Total Profit / Total Revenue) \* 100)
* Assessed whether the hospital is generating a **profit** based on these calculations.

**Explanation:**

* **Total Appointment Fee:** Sum of all patient appointment fees.
* **Total Revenue:** Sum of total bill amounts.
* **Total Profit:** Revenue minus appointment fees.
* **Profit Percentage:** (Total Profit / Total Revenue) \* 100, showing how much of the revenue translates into profit.
* The **profit margin** remained consistently **high**, around **98.96% and 98.87%**, indicating **strong financial performance**.

**Insight:**

* The hospital maintains a **high profit margin** exceeding **90%**, confirming its **strong profitability**.
* **Revenue generation** is **substantial**, while **costs (appointment fees) remain low**, contributing to the **high margins**.



**Result:**

* With a **consistent profit percentage above 98%**, the analysis confirms that the **hospital is highly profitable** and operates with a **very high margin**.

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

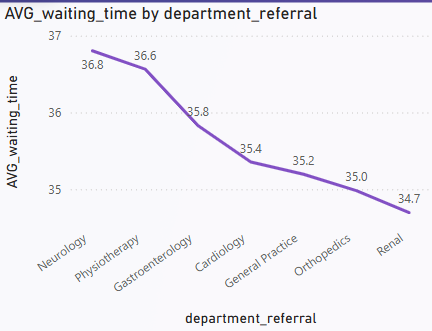
**Approach:**

* **Created a Line Chart** with **Department** on the **x-axis** and **Average Waiting Time** on the **y-axis** to visualize waiting time variations across departments.
* This allowed for a **clear comparison** and identification of departments with **unusually long waiting times**.

**Explanation:**

* The **line chart** revealed **differences in waiting times** between departments.
* **Neurology** had the **highest average waiting time** at **36.80 minutes**, standing out **significantly** from other departments.

**Visualization:**

****

**Insight:**

* The **high waiting time** in the **Neurology department** may point to potential **staff shortages**, **high patient demand**, or **scheduling inefficiencies** that need addressing.

**Result:**

* **Neurology** was identified as having the **longest average waiting time** at **36.80 minutes**, which is notably higher than other departments, highlighting an area for improvement.

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

* **Reward High Spenders:**
* Provide discounts to patients with a **total bill exceeding ₹10,000** to recognize and reward high spenders.
* **Encourage Loyalty:**
* Offer discounts to patients with a **satisfaction score of 6 or above** to foster loyalty and incentivize positive feedback.
* **Tiered Discounts:**
* Introduce **tiered discount levels** based on the total bill amount:
  + **5%** for bills between **₹10,000–₹20,000**
  + **10%** for bills above **₹20,000**
* **Compensate for Long Wait Times:**
* Provide **compensatory discounts** to patients with unusually long waiting times, helping improve their overall experience and satisfaction.
* **Seasonal or Promotional Discounts:**
* Offer **seasonal or promotional discounts** during **low-demand periods** to attract more patients and balance patient flow.

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?

**Approach:**

* **Created a table visualization** with **Doctor Name** and **Department** to display the list of doctors and their respective departments.
* Used **DAX** to assign doctors to two shifts:
  + **8:00 AM - 2:00 PM**
  + **2:00 PM - 8:00 PM**
* Applied **conditional formatting** to highlight the **General Practice department** for better visibility.

**Explanation:**

* The **DAX formula** used to assign doctors to shifts is as follows:

Shift =

IF(

MOD(RANKX(ALL('Doctor Info'),'Doctor Info'[Doctor Name], ,ASC),2) = 0,

"8:00 AM - 2:00 PM",

"2:00 PM - 8:00 PM"

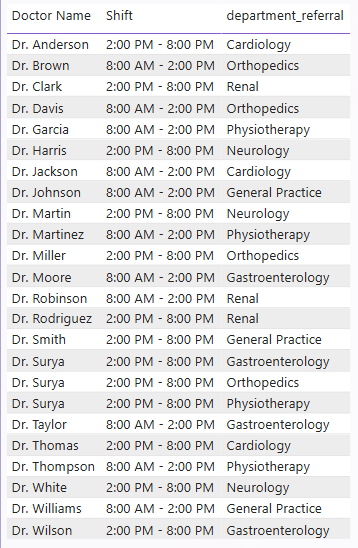
)

The formula used also **ranks doctors by name** and alternates shift assignments using the **MOD function**. This ensures a balanced distribution of doctors across shifts, based on patient volume.

In this case:

* **Dr. Johnson** and **Dr. Williams** were assigned to the **morning shift** (8:00 AM - 2:00 PM) due to **higher patient traffic**.
* **Dr. Smith** was assigned to the **evening shift** (2:00 PM - 8:00 PM) to manage **lower patient volume** during those hours.

**Visualisation:**



**Result:**

* The **two-shift policy** helps achieve an **even distribution of workload** among doctors, ensuring that patient care is more **efficient** and that **wait times are reduced**.
* This approach is particularly beneficial in the **General Practice department**, where the patient flow is typically higher, allowing for **better management** of patient demand and enhancing overall service quality.

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

**Power BI Gateway:**

A **Power BI Gateway** is a tool that enables the secure transfer of data between **on-premises data sources** (such as databases or Excel files) and the **Power BI cloud service**, allowing you to refresh and access data stored locally within your reports and dashboards.

**Use Cases:**

1. **Data Refresh:**
   * **Automatically refreshes** on-premises data to ensure reports and dashboards are always up to date.
2. **Secure Data Transfer:**
   * Ensures that data is **encrypted** and **securely transferred** between on-premises sources and Power BI, maintaining confidentiality and compliance.
3. **Access On-Premises Data:**
   * Allows you to incorporate **on-premises data** (e.g., SQL Server, Excel files) into Power BI for comprehensive reporting.
4. **Hybrid Data Integration:**
   * Combines **on-premises** and **cloud data** into unified, seamless reports for better analysis and decision-making.
5. **DirectQuery:**
   * Enables **real-time querying** of live data sources in Power BI, ensuring that the most recent data is always used in reports and visualizations.

**Types of Gateways:**

1. **Personal Gateway:**
   * Designed for **individual use**, allowing personal data refreshes without the need for multiple users or large-scale deployments.
2. **Enterprise Gateway:**
   * Aimed at **organizations** with **multiple users** and **data sources**, supporting large-scale data refresh and integration for team-based reporting.
3. How would you approach this problem, if the objective and subjective questions weren't given?

**Steps for Data Analysis in Power BI:**

1. **Import the Data:**
   * Begin by importing the dataset into **Power BI** to initiate the analysis process.
2. **Use Power Query for Data Transformation:**
   * Utilize **Power Query** to **filter**, **merge**, and **shape** the data to match the required structure, ensuring it’s ready for analysis.
3. **Clean the Data:**
   * Address any **inconsistencies**, handle **missing values**, **correct data types**, and **standardize formats** across the dataset to ensure accuracy and consistency in the analysis.
4. **Identify Key Performance Indicators (KPIs):**
   * Identify the **most relevant KPIs** based on the objectives of the analysis, focusing on metrics that will provide meaningful insights for decision-making.
5. **Analyze Each KPI:**
   * Dive deep into each KPI by exploring **trends**, **relationships**, and **patterns** to uncover insights that align with the analysis goals.
6. **Create Reports and Dashboards:**
   * Develop **comprehensive reports** and **dashboards** using various **visualizations** like **bar charts**, **line graphs**, and **tables** to effectively communicate findings.
7. **Add Interactive Slicers and Filters:**
   * Enhance the reports by adding **interactive slicers** and **filters**, allowing users to **dynamically explore** the data based on different dimensions for more personalized analysis.
8. **Review and Finalize the Report:**
   * **Review the final report** for **accuracy**, **clarity**, and **relevance**. Provide **actionable recommendations** based on the insights gathered during the analysis process to support informed decision-making.
9. Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?

* **Unique Doctor Assignment**:
* Each doctor is assigned to one department, ensuring clarity in the structure of the organization.
* **Multiple Doctors per Department**:
* A single department can have multiple doctors, which allows for specialization within that department.
* **Data Modeling**:
* In a relational database, you can set up a **foreign key** on the doctor table that references the department table, linking each doctor to their respective department.
* **One-to-Many Relationship**:
* When analyzing the relationship, think of the department as the “one” side and the doctors as the “many” side. Each department can have many doctors, but each doctor is limited to only one department.
* **Querying and Reporting**:
* In reports or queries, you can aggregate data by department (e.g., count of doctors per department) or drill down to see the details for each individual doctor.
* **Data Integrity**:
* Enforcing this relationship helps maintain data integrity, ensuring that each doctor is assigned to only one department, preventing errors in reporting or data analysis.
* **Department-Level Metrics**:
* Departments can be analyzed independently (e.g., department-wise doctor count) or combined with other metrics like patient load, performance, or financials.