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

* **I**dentify Missing Values: Inspected the dataset for null or inconsistent values using Power Query Editor.
* Handle Missing Data: Replace null values, such as in the patient\_sat\_score column, with the average value to maintain data completeness, so that we can know that this patient has not given any rating in regards to the service.
* Ensure Consistency: Corrected any inconsistencies or erroneous data to prepare it for analysis.

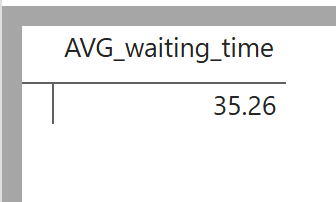
This cleaning process ensures the dataset is accurate and reliable for further analysis.

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

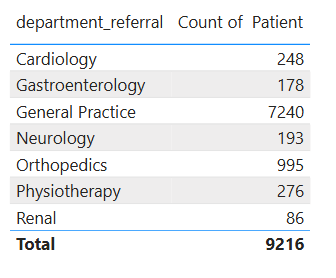
Formula Used: DAX

Result: The average patient waiting time is 35.26 minutes.

AVG\_waiting\_time = AVERAGE('Hospital ER (1)'[patient\_waittime])



* The average time a patient waits for his checkup is almost 36 minutes.
* To calculate that I have used the AVERAGE formula on the patient\_waittime in the hospital table.

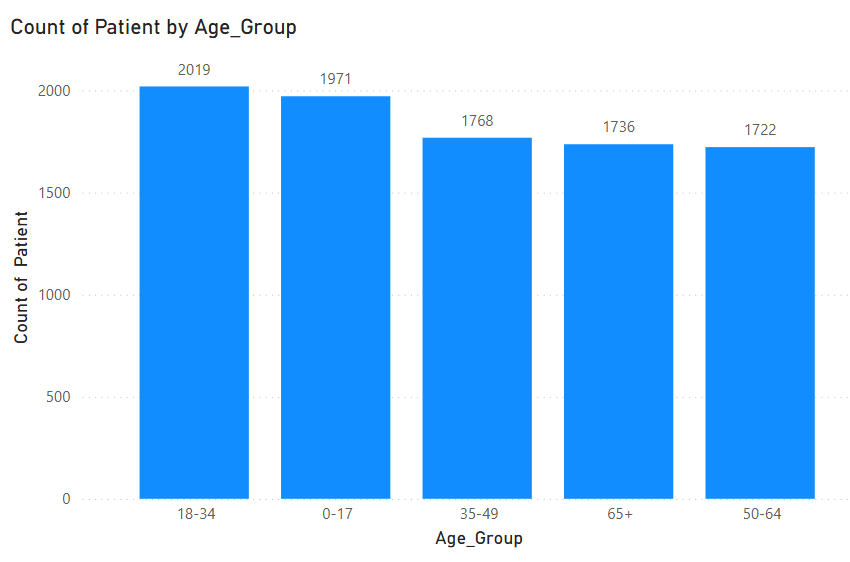
**3.** **Visits by Department Referral:** Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.  
  
 Formula Used: Count of Patient = COUNT('Hospital ER (1)'[patient\_id])  
To calculate the frequent visits in the hospital with respect to each department

* The General Practice department has the most number of patients which visit frequently.
* I have grouped the data on department\_referral and then used Countx to count the total no of patients visiting the hospital.

**4. Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.  
  
 Formula Used:   
 Created a new column using SWITCH to categorize patients by age group:

Age\_Group = SWITCH(TRUE(),

'Hospital ER (1)'[patient\_age] < 18, "0-17",  
 'Hospital ER (1)'[patient\_age] >= 18 && 'Hospital ER (1)'[patient\_age] <= 34, "18-34",  
'Hospital ER (1)'[patient\_age] >=35 && 'Hospital ER (1)'[patient\_age] <=49, "35-49",  
'Hospital ER (1)'[patient\_age] >=50 && 'Hospital ER (1)'[patient\_age] <=64, "50-64",  
 'Hospital ER (1)'[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:   
  Successfully created age buckets and visualized the number of visits for each demographic group.

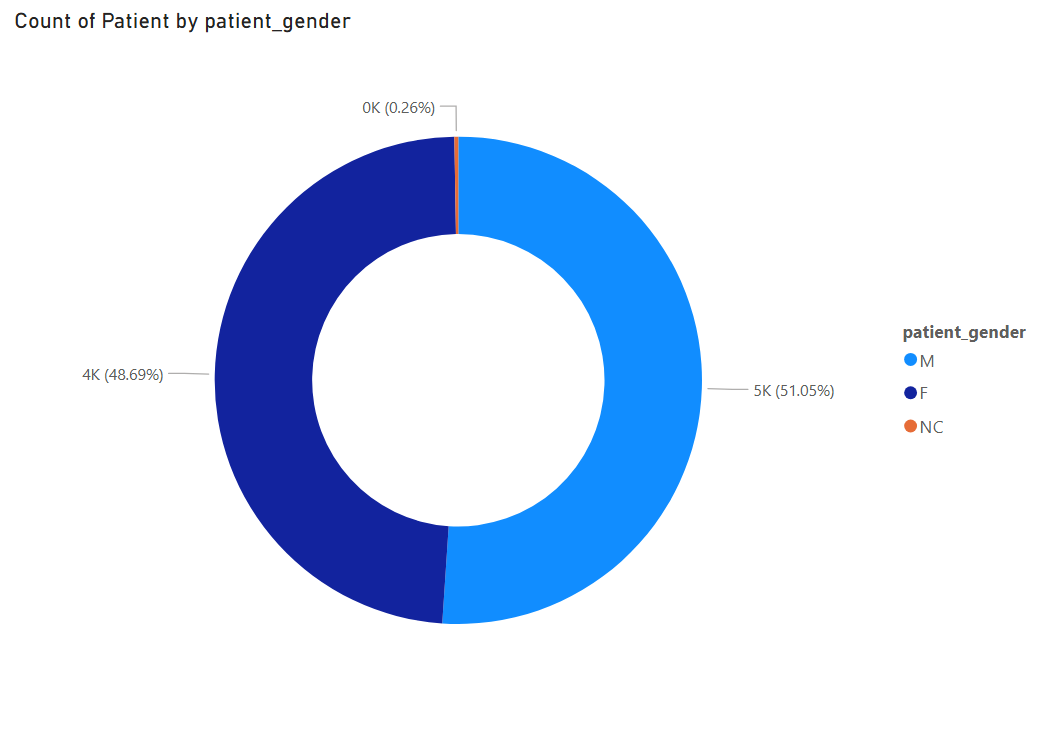
**5. Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for?**Null values were present in the patient satisfaction scores, which were subsequently filled with the average satisfaction score to ensure data completeness.

* 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 (5.4749455337690636)
* Replaced Nulls with the Average:  
  Rounded the average to 5 and replaced all null values using Transform > Replace Values.

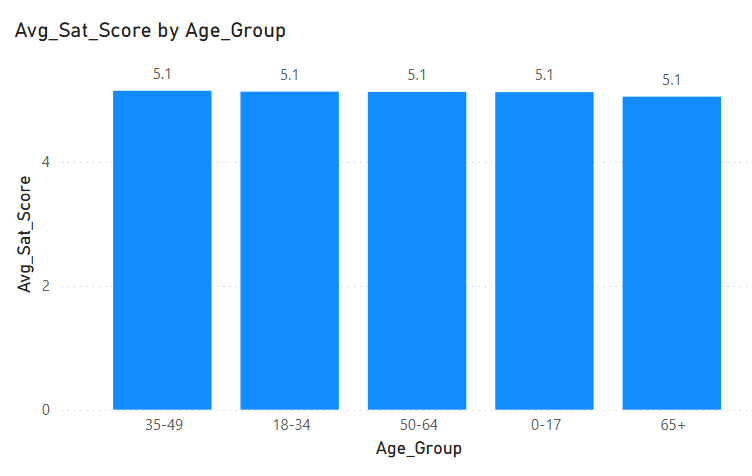
**6. Is there any relation between the number of visits and the Gender of the patients?**

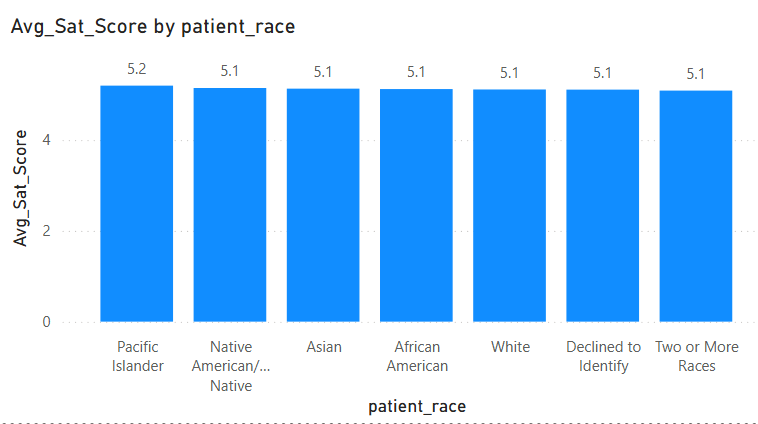
Analysis of Visits by Gender

* + 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.

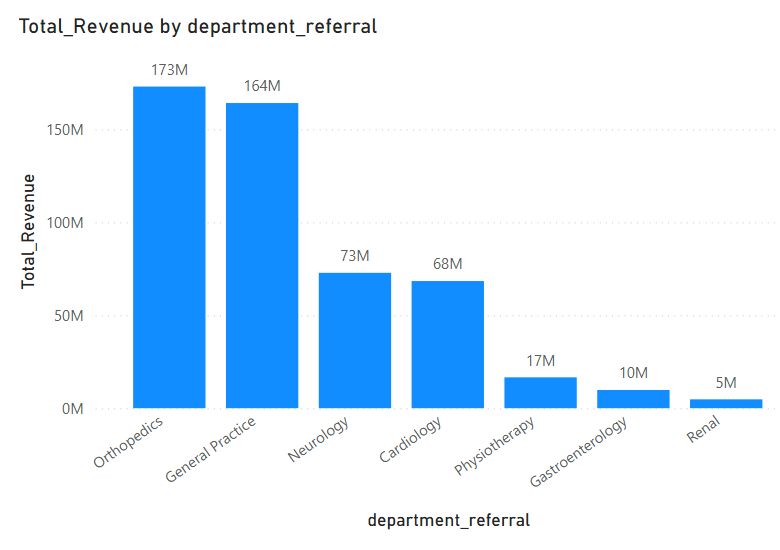
**7. Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.**Analysis of Average Satisfaction by Demographics

* + Age Group Analysis  
    Patient satisfaction scores across different age groups range between 5 and 5.1.
  + 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 5.1 to 5.2.
    - The highest average satisfaction score of 5.2 was reported by Pacific Islanders, while other racial groups fell slightly below this mark.
    - The lowest satisfaction score of 5.1 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.

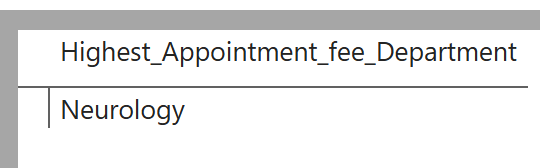
**8. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.**

* + 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.

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

* Highest Appointment Fee by Department  
   To identify which department is charging the highest appointment fees in general, 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('Sheet1'[department\_referral],'Sheet1'[Appointment Fees],max('Sheet1'[Appointment Fees]))



* 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.

**10. 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.**

Step 1: Create a Calendar Table  
Calendar = CALENDAR(MIN('Hospital ER (1)'[date]), MAX('Hospital ER (1)'[date]))

## Step 2: Add these columns:

* Month Number:  
  MonthNo = MONTH('Calendar'[Date])
* Month Name:  
  MonthName = FORMAT('Calendar'[Date], "MMMM")
* Year  
  Year = YEAR('Calendar'[Date])

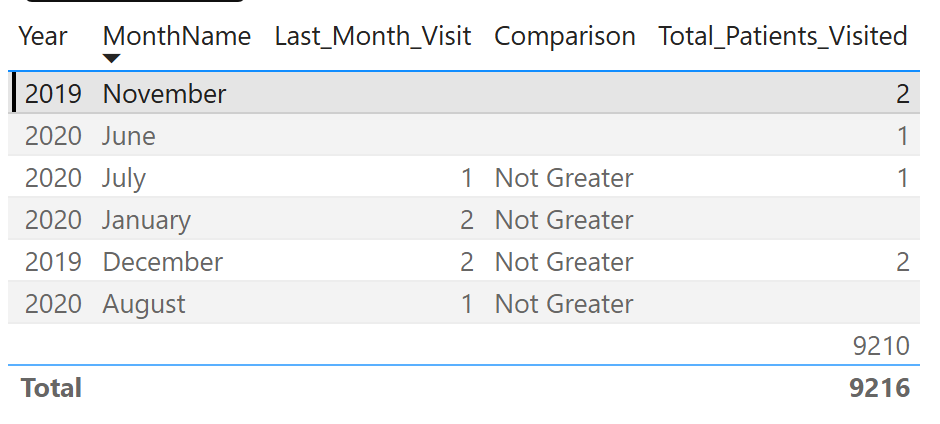
Step 3: Create Relationship  
 Drag 'Hospital ER (1)'[date] ➔ 'Calendar'[Date]  
 Ensure:  
 One-to-many  
 Single direction  
  
Step 4: Create the DAX Measures

1. Total Patients Visited  
total\_patient\_Visited = CALCULATE(COUNT('Hospital ER (1)'[patient\_id]))

2. Last Month Visits  
Last\_Month\_Visits = CALCULATE([Monthly\_Visits],PREVIOUSMONTH('Calendar'[Date]))

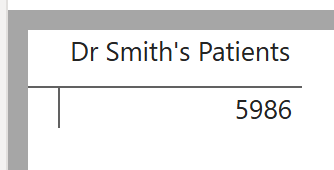
3. Comparison  
Visit\_Increased =   
VAR Curr = [Monthly\_Visits]  
VAR Prev = [Previous\_Month\_Visits]  
RETURN  
IF(ISBLANK(Prev), BLANK(), IF(Curr > Prev, "Yes", "No"))

Step 5: Visualization



**11. 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, we use the following DAX formula:

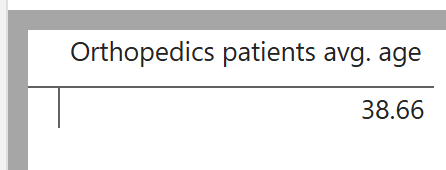
Dr Smith's Patients = CALCULATE(COUNT('Hospital ER (1)'[patient\_id]), 'Sheet 1'[Doctor Name] = "Dr. Smith")  
  


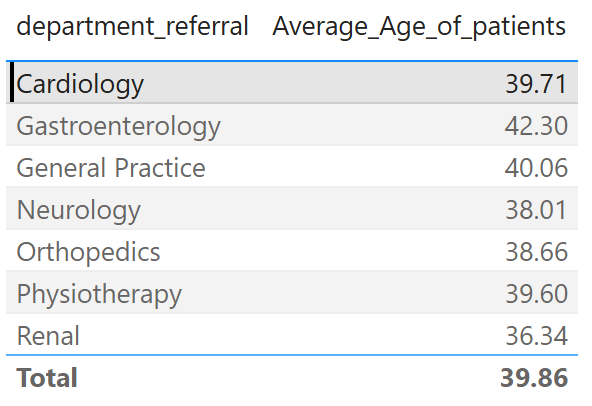
* Results:  
    
  Using the DAX formula, we calculated that 5,986 patients have visited Dr. Smith.  
  This was achieved by filtering the Doctor info table for entries where the Doctor Name is "Dr. Smith" and counting the number of unique patient\_id entries.

**12. 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?**

To calculate the average age for Orthopedics, a measure is created using the following formula,  
  
Orthopedics patients avg. age = CALCULATE (AVERAGEX ('Hospital ER (1)','Hospital ER (1)'[patient\_age]),'Hospital ER (1)'[department\_referral] = "Orthopedics")

For all departments, a general measure is created without any filter:  
  
Average\_Age\_of\_patients = AVERAGE('Hospital ER (1)'[patient\_age])  
  
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.  
  
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.

**13. Were there any data format issues in the data, and if there were/are how you handle them?**In power bi using power query editor we handle missing values, inconsistencies, standard formatting, conversions. The data set provided by the Newton school has some inconsistencies like null value.  
  
Using the power query editor I split the date and time column separately so that I can use them separately according to the condition.

The date column was in text format so I changed it into date format.

Handling Blank Spaces in Patient Satisfaction Scores:

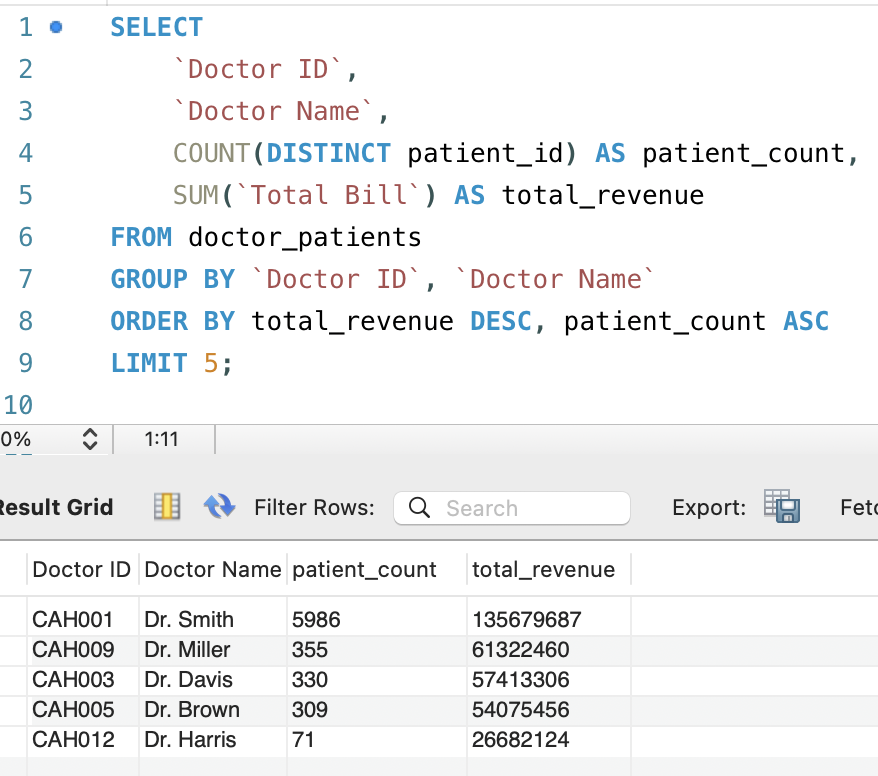
* The patient\_sat\_scores column had blank spaces, which could disrupt calculations and analysis.
* Initially, the data type of patient\_sat\_scores was in text format, so I converted it to a number.
* After conversion, I replaced the blank spaces (null values) with the column's mean value, ensuring data consistency and avoiding skewed results during computations.

**14. 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?**

* The M-query language is used to manipulate the tables and perform transformation on our dataset.
* Whenever we add/remove/manipulate columns in our dataset/table we can see the corresponding M-query in the formula bar.
* Power Query UI generates M-query whenever we perform any actions on our data.
* The following M-query code appears on the formula bar when we add a column in Power Query UI.   
    
  = Table.AddColumn(#"Replaced Value", "Full Name", each Text.Combine({[patient\_first\_inital], " ", [patient\_last\_name]}), type text)
* In the above M-query we are merging the “patient\_first\_initial” and the “patients\_last\_name” column to create a new column called “Full Name” using the “Column from Examples” feature of Power Query UI.

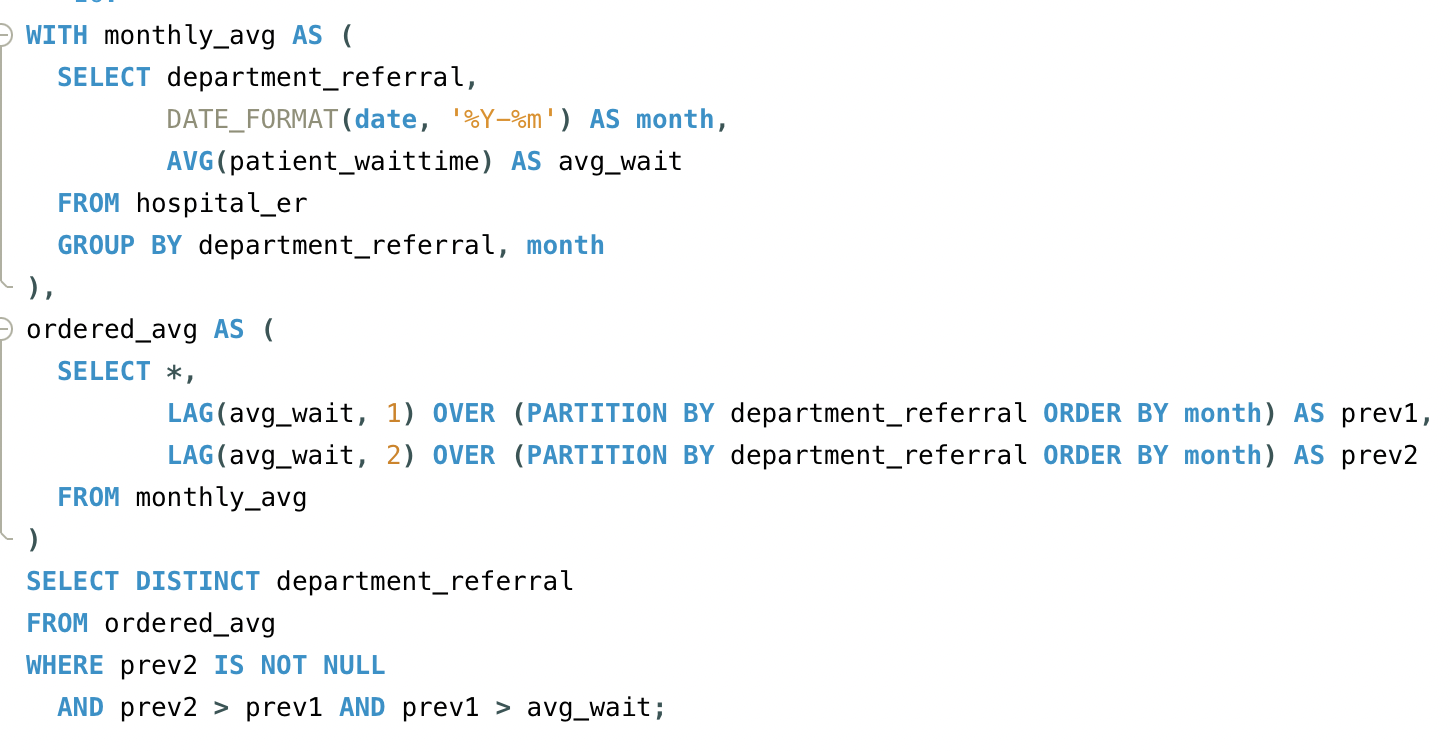
**15. Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)**

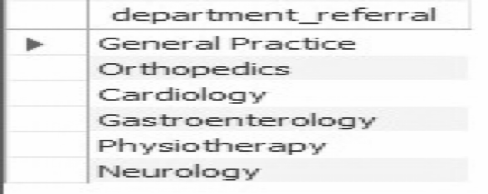
The top five doctors with the highest revenue despite having the fewest patients are identified by analyzing their total revenue and unique patient count. This highlights doctors who specialize in high-value services, making them key drivers of revenue efficiency within the organization.



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

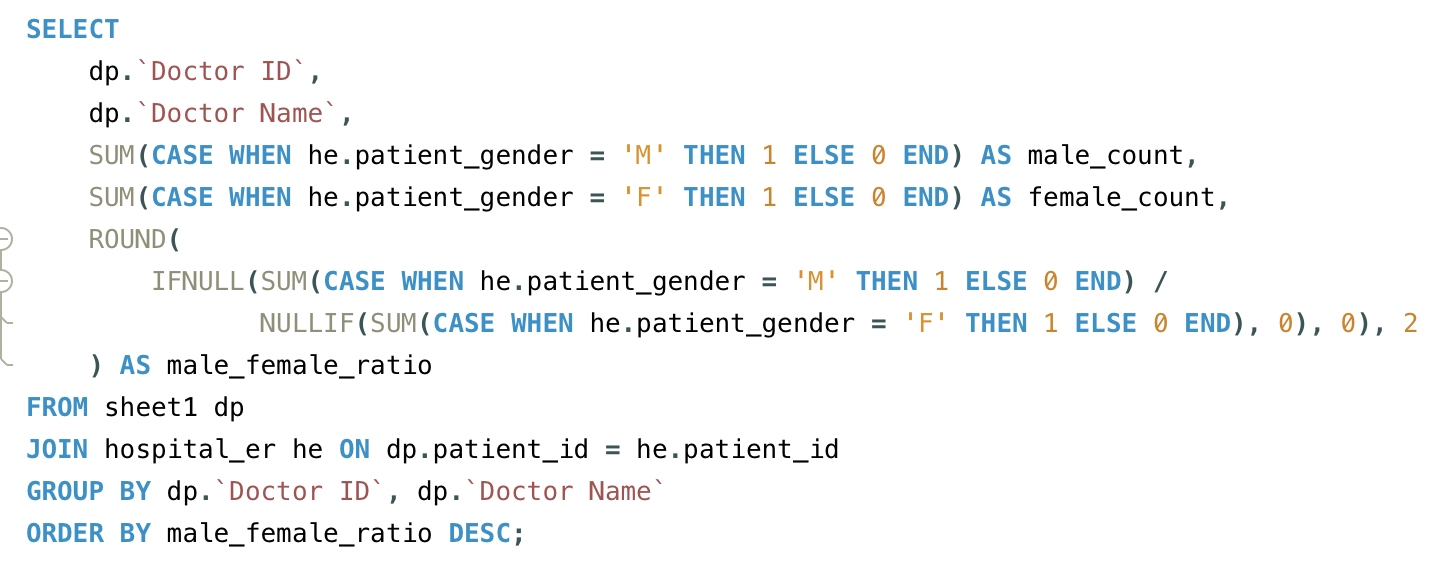
Departments that showed a consistent decrease in average waiting times over three consecutive months were identified. This reflects improved efficiency and better patient flow management, leading to enhanced service quality

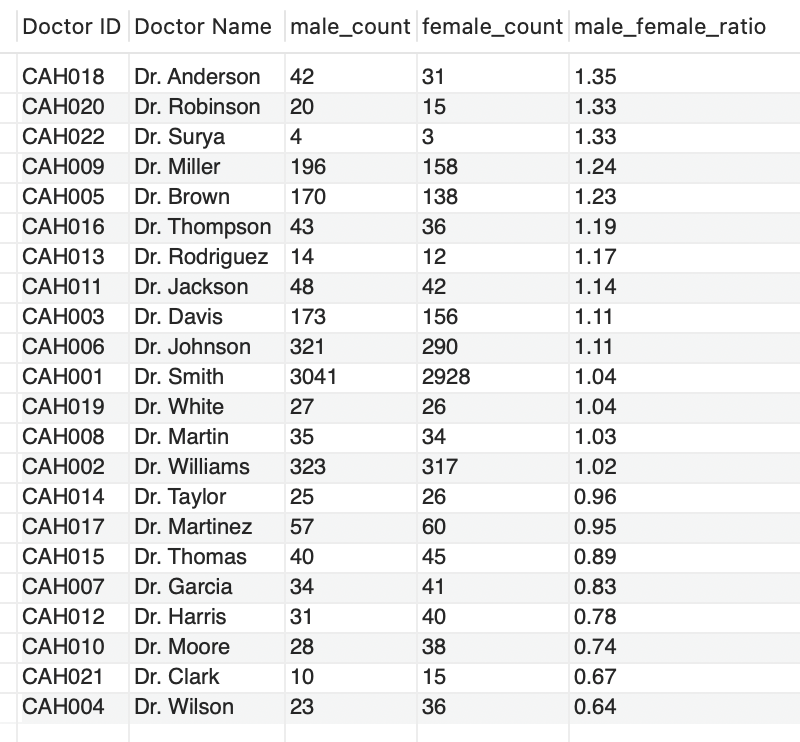




**17. Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)**

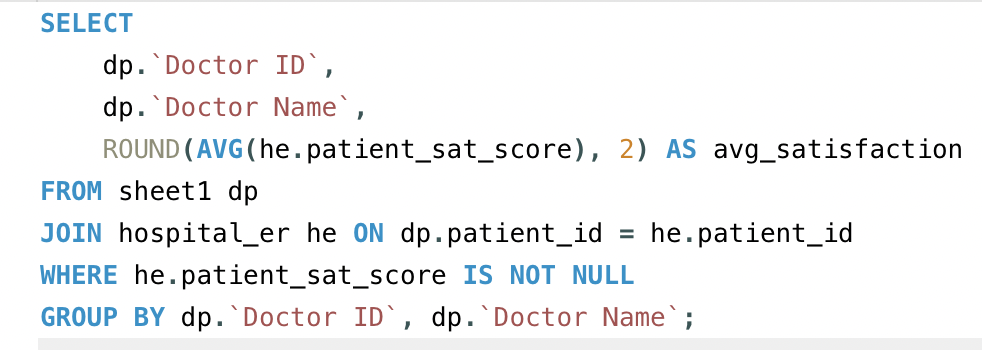
The male-to-female patient ratio for each doctor is calculated and used to rank them in descending order. This analysis offers insights into patient demographics, helping to identify trends or potential biases in patient distribution.

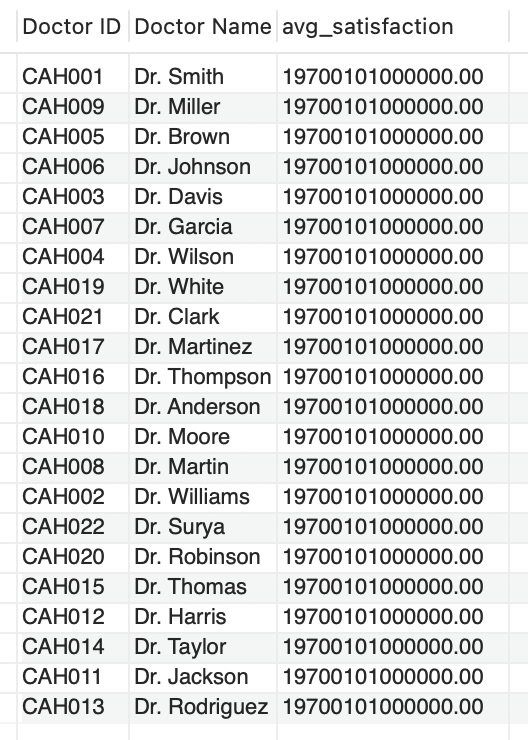




**18. Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)**

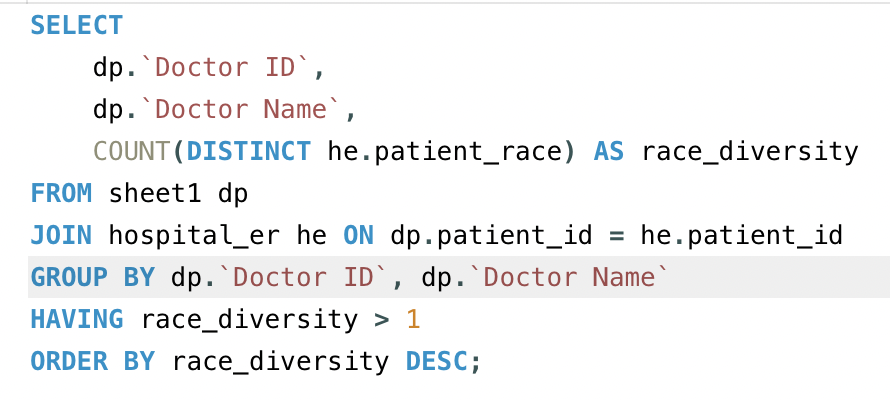
The average patient satisfaction score for each doctor is calculated, with missing scores replaced by a default value of 5. This provides a comprehensive measure of patient experience, ranking doctors based on the highest satisfaction scores.

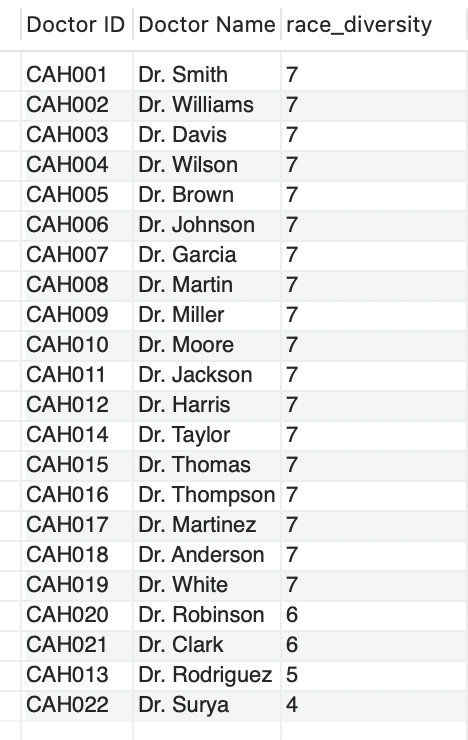




**19. Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)**

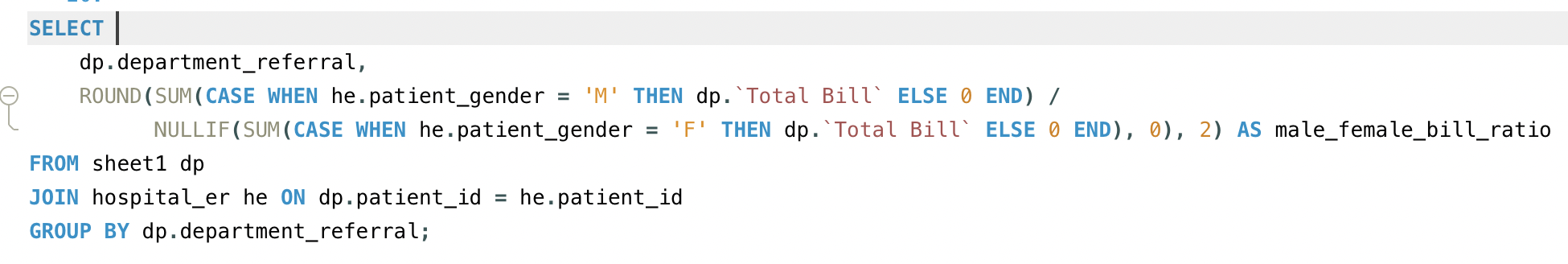
Doctors who have treated patients from multiple races are identified, with the diversity of their patient base measured by the number of distinct races they have treated. This highlights doctors serving a more diverse patient demographic.

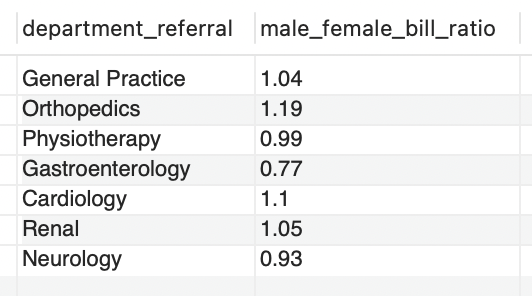




**20. Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)**

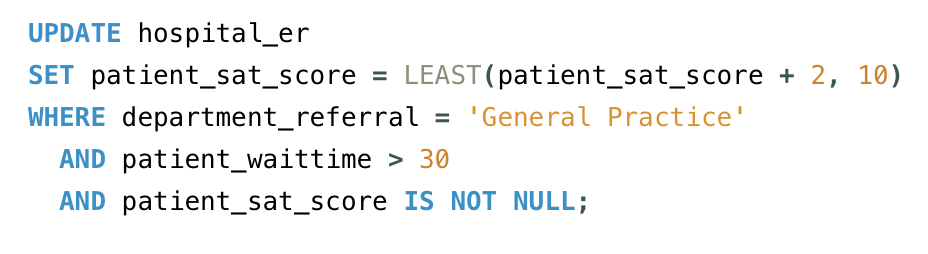
The ratio of total bills generated by male to female patients is calculated for each department. This analysis offers insights into the financial contribution of each gender, helping to identify spending trends across departments.





**21. 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.

****