

HOSPITAL



**Power BI project by –** **Bhushan Prabhakar Dhawas**

**Portfolio Website –** ([Link](https://bhushan0016.github.io/My-Portfolio-website/))

**This word file contains all the answers and suggestions for Objective and Subjective questions.**

**Data overview**

➔ **Date**: This column contains date and time information without specifying AM or PM. The

format is **DD-MM-YYYY HH:MM**.

➔ **Patient ID**: Each patient is assigned a unique identifier, which seems to be in the format

**124-62-3289**.

➔ **Patient Gender**: This column records the gender of the patient, denoted by 'M' for male and

'F' for female.

➔ **Patient Age**: The age of the patients is listed in years.

➔ **Patient Sat Score**: It seems to represent a satisfaction score given by or for the patient.

However, the scores are single-digit, and it's not clear what the scale is.

➔ **Patient First Initial**: This column contains the first initial of the patient's first name.

➔ **Patient Last Name**: The surname of the patient is listed in this column.

➔ **Patient Race:** The racial or ethnic background of the patient is recorded here, with categories such

as 'White', 'African American', 'Asian', 'Native American/Alaska Native', and 'Two or More Races'.

➔ **Patient Admin Flag**: This column contains boolean values ('TRUE' or 'FALSE') which might indicate

whether the patient was admitted or some other administrative flag.

➔ **Patient Wait Time**: Appears to indicate the time the patient waited, possibly in minutes, before

being seen or processed.

➔ **Department Referral**: This column lists the department to which the patient was referred, with

entries such as 'General Practice', 'Orthopedics', 'Gastroenterology', or 'None' indicating no

referral.

➔ **Doctor Name:** Identifies the doctor who attended each patient.

➔ **Appointment Fees:** The cost charged for a doctor's consultation.

➔ **Total Bill:** The overall amount billed to the patient, including all services and charges.

**Objective questions**

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

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

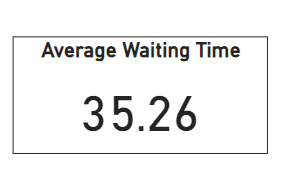
* 1. **1) Identify Missing Values:** Inspected the dataset for null or inconsistent values using Power Query Editor.
  2. **2) Handle Missing Data:** Replaced null values, such as in the patient\_sat\_score column, with the average value to maintain data completeness.
  3. **3) DAX Formula used-**
  4. patient\_new\_sat\_score= If(ISBLANK('Hospital Info'[patient\_sat\_score]),AVERAGE('Hospital Info'[patient\_sat\_score]),'Hospital Info'[patient\_sat\_score])
  5. **This cleaning process ensures the dataset is accurate and reliable for further analysis.**

**Task-2. Assess the Average Waiting Time:**

**Analyse the patient wait times to identify the average duration a patient spends before receiving care.**

**Ans –** The average waiting time is 35.26 min.

**DAX Formula used-**

****Average\_waiting\_time = AVERAGE('Hospital Info'[patient\_waittime])

**Task-3. Visits by Department Referral:**

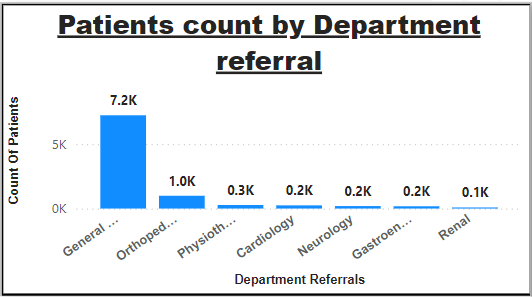
**Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.**

**Ans -**

**DAX Formula used-**

Count\_of\_patients = COUNT('Hospital Info'[patient\_id])

**X axis –** department\_referrals

**Y axis –** Count\_of\_patients

**Task-4. Patient Visits by Age Group:**

**Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.**

**Ans -**

**1)** First we will create column “**Age Group**” which divides the ages of the patients into 5 categories.

**2) Categories –**

Child - (Age<12)

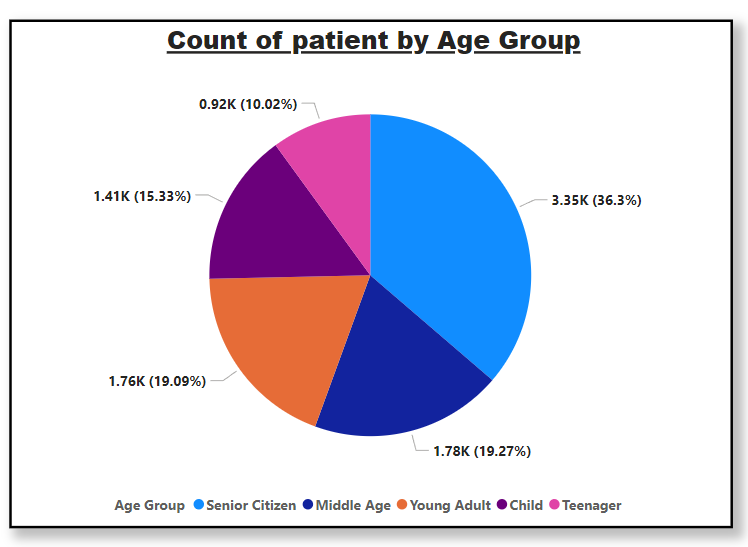
Teenager - (12<Age<=20)

Young Adult - (20<Age<=35)

Middle Age - (36<Age<=50)

Senior Citizen - (Age>50)

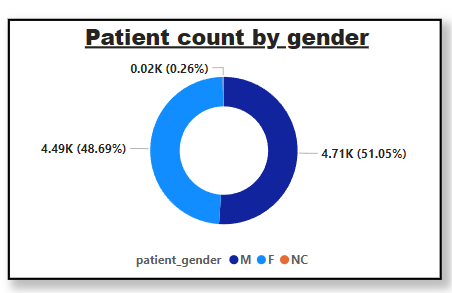
**3) DAX Formula used-**

Age Group = IF('Hospital Info'[patient\_age] <= 12, "Child",if('Hospital Info'[patient\_age] > 12 && 'Hospital Info'[patient\_age] <= 20, "Teenager",  IF('Hospital Info'[patient\_age] > 20 && 'Hospital Info'[patient\_age] <= 35, "Young Adult", IF( 'Hospital Info'[patient\_age] > 35 && 'Hospital Info'[patient\_age] <= 50, "Middle Age",  "Senior Citizen" ))))

**Task-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?**

**Ans –**

1)Yes, there are null values in the data.

* 1. 2) **Identify Missing Values:** Inspected the dataset for null or inconsistent values using Power Query Editor.
  2. 3) **Handle Missing Data:** Replaced null values, such as in the patient\_sat\_score column, with the average value to maintain data completeness.
  3. 4) **DAX Formula used-**
  4. patient\_new\_sat\_score=ROUND(if(ISBLANK('Hospital Info'[patient\_sat\_score]),AVERAGE('Hospital Info'[patient\_sat\_score]),'Hospital Info'[patient\_sat\_score]),1)
  5. **This cleaning process ensures the dataset is accurate and reliable for further analysis.**
  6. **Task-6. Is there any relation between the number of visits and Gender of the patients.**
  7. **Ans –** 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%.

**Task-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.**

**Ans – Analysis of Average Satisfaction by Demographics :**

**The chart titled "Average Satisfaction Score by Age Group and Race" presents an analysis of satisfaction scores based on age groups and racial backgrounds.**

**Below is a detailed breakdown:**

**1) X-Axis:** The age groups included in the analysis are:

* Middle Age
* Child
* Young Adult
* Senior Citizen
* Teenager

**2) Y-Axis:** Displays the average satisfaction scores on a scale of 0 to 6.

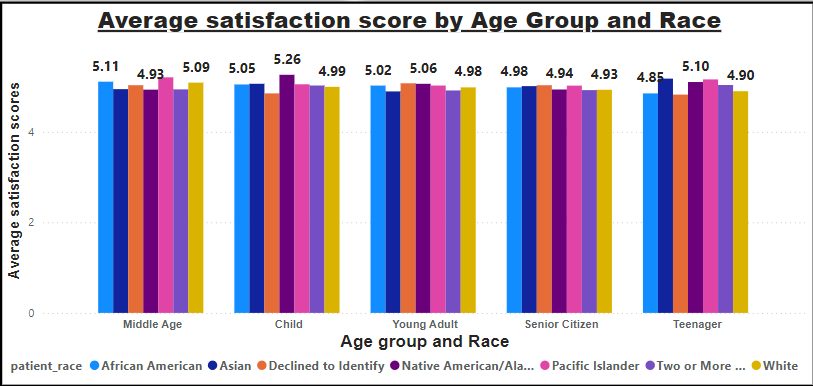
**3) Categories (Legend):** Represents different racial backgrounds:

* African American (Blue)
* Asian (Pink)
* Declined to Identify (Purple)
* Native American/Alaska Native (Orange)
* Pacific Islander (Yellow)
* Two or More Races (Dark Red)
* White (Brown)

**4) Observations by Age Group:**

* Middle Age: Average satisfaction scores range from 4.93 to 5.11
* Child: Scores range from 4.99 to 5.05
* Young Adult: Scores are consistent, around 5.98 -5.06
* Senior Citizen: Scores are mostly 4.94, with minimal variation.
* Teenager: Slightly lower scores, ranging from 4.85 to 5.10

**5) Overall Trend:** Satisfaction scores remain relatively consistent across racial backgrounds, with slight variations by age group. The Child and Teenager groups exhibit the most noticeable differences, with lower scores observed among teenagers.

****

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

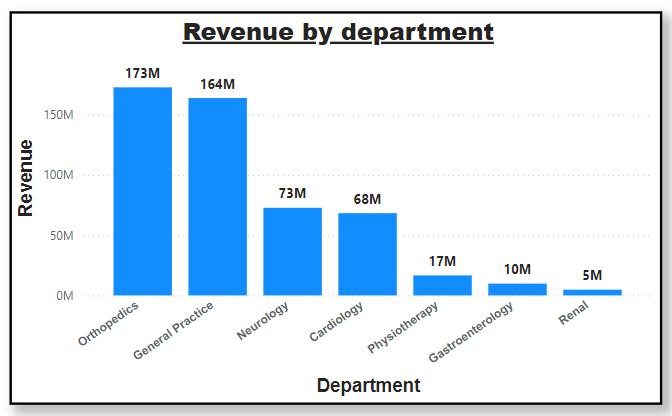
**Ans-**

**X axis –** department\_referrals

**Y axis –** Sum of Total Bill

**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.

****

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

**Ans-Neurology department is charging the highest appointment fees in general.**

1. **DAX Formula used-**

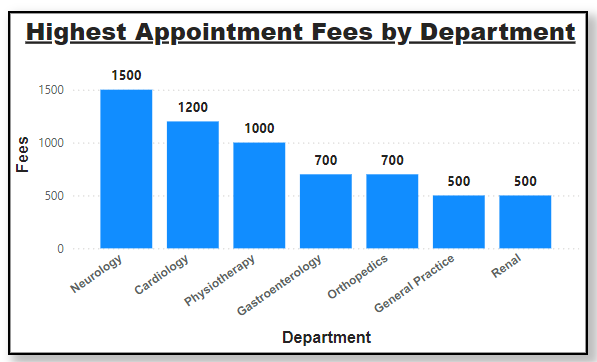
Highest Appointment Fees by Department = CALCULATE(MAX('Doctor Info'[Appointment Fees]),

  ALLEXCEPT('Doctor Info','Doctor Info'[department\_referral]))

1. **Explanation-**

* MAX(Sheet1[Appointment Fees]): Retrieves the maximum appointment fee.
* ALLEXCEPT(Sheet1, Sheet1[department\_referral]): Ensures the calculation respects the grouping by department\_referral, clearing other filters on the table except for the department context.
* CALCULATE: Applies the filter logic dynamically to calculate the maximum value within each department.



****

**Task-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.**

**Ans –**

**To show Monthly wise total visits in hospital I use table visualizations and add 4 columns in it , viz.-**

* **1st column:** Month\_no

**Dax formula used –**

**Month\_no =** MONTH('Hospital Info'[date])

* **2nd column:** Month\_name

**Dax formula used –**

**Month\_name =** FORMAT('Hospital Info'[date],"MMMM")

* **3rd column:** Monthly\_patient\_visit (Measure)

**Dax formula used –**

**Monthly\_patient\_visit =** COUNT('Hospital Info'[patient\_id])

* **4th column:** Previous\_month\_visit (Measure)

**Dax formula used –**

**Previous\_month\_visist** = CALCULATE(COUNT('Doctor Info'[patient\_id]),PREVIOUSMONTH(Calender[Date]))

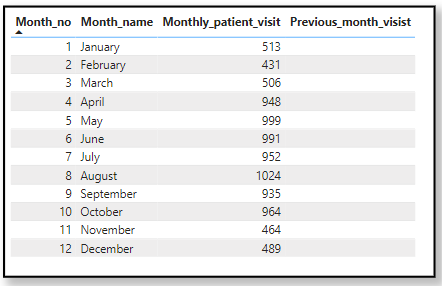
**The Previous\_month\_visit column is showing blank when it is added to the table. So it is not possible to calculate the visit are increased or not.**

**But I have added the approach and DAX formula for it.**

* **5th column:**  Visit\_increased

**Dax formula used –**

**Visit\_increased =** If([Monthly\_patient\_visit] > [Previous\_month\_visist], “Yes” , “No” )



**Task-11. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.**

**Ans-**

* **DAX Formula used-**

Dr\_Smith\_patients = CALCULATE(COUNTA('Hospital Info'[patient\_id]),'Doctor Info'[Doctor Name]="Dr. Smith")

* **Explanation:**
* CALCULATE changes the context of the calculation, allowing you to apply filters.
* COUNT('Hospital Info'[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".

****

**Task-12. Calculate the average age of the patients who visit the Orthopedic department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?**

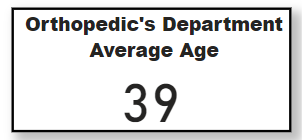
**Ans-**

**Orthopedics:** A measure is created with a filter to calculate the average age for Orthopedics and display it in a Card Visual.

**1)Explanation :**

* CALCULATE changes the context of the calculation, allowing you to apply filters.
* AVERAGE('Hospital Info'[patient\_age] calculate the average of patients age.
* 'Hospital Info'[department\_referral] = "Orthopedics" filters the data to only include records where the Department Name is “Orthopedics”.

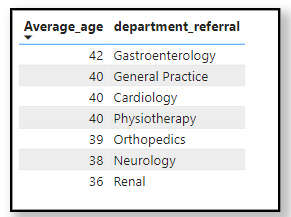
**2)DAX Formula used-**

****Orthopedic\_avg\_age = CALCULATE(AVERAGE('Hospital Info'[patient\_age]),'Hospital Info'[department\_referral] = "Orthopedics")

**All Departments:** A general measure is created without filters, add it to a Table Visual with the department column to display averages for all departments.

**DAX Formula used-**

Average\_age = AVERAGE('Hospital Info'[patient\_age])

****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.

**Task-13. Were there any data format issues in the data, and if there were/are how you handle them?**

**Ans –**

1)Yes, there was data format issues for the patient\_sat\_scores column and date column in the “Hospital Info” table.

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

3)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.

**The dataset was transformed into a clean and consistent format, making it ready for reliable analysis and accurate insights.**

**Task-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?**

**Ans –**

* 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 you add a column in Power Query UI.

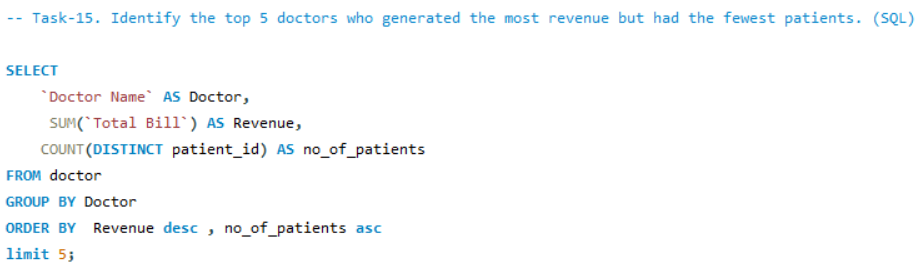
**= Table.AddColumn(#"Replaced Value", "Full Name", each Text.Combine({[patient\_first\_inital], " ", [patient\_last\_name]}), type text)**

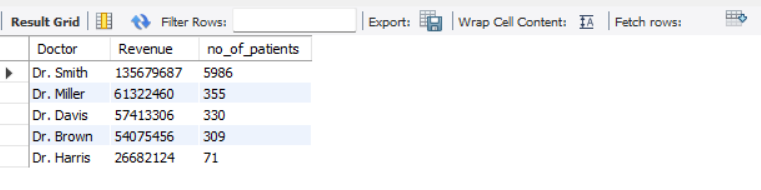
* In the above M-query we are merging the “patient\_first\_initial” and the “patients\_last\_name” column to create a new column called “Full Name” using the “Column from Examples” feature of Power Query UI.

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

**Ans –**

* The query calculates the total revenue generated by each doctor and the number of unique patients they have treated.
* It groups the data by doctor to aggregate the revenue and patient count for each one.
* Doctors are ranked by revenue in descending order, with ties resolved by ascending patient count.
* Only the top 5 doctors based on this ranking are displayed.
* The output provides insights into the performance of doctors in terms of earnings and patient volume.

****

**Result -**

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

**Ans-**

* **Calculate Monthly Average Wait Times:**

The first CTE (AvgWaitTimeByMonth) calculates the average patient wait time (AvgWaitTime) for each department (department\_referral) on a monthly basis by grouping data by department and date.

* **Add Lagged Wait Time Values:**

The second CTE (WaitTimeWithLag) uses the LAG function to fetch the wait times from the previous three months for each department, creating columns PrevMonthAvg, TwoMonthsAgoAvg, and ThreeMonthsAgoAvg.

* **Identify Departments with Decreasing Trends:**

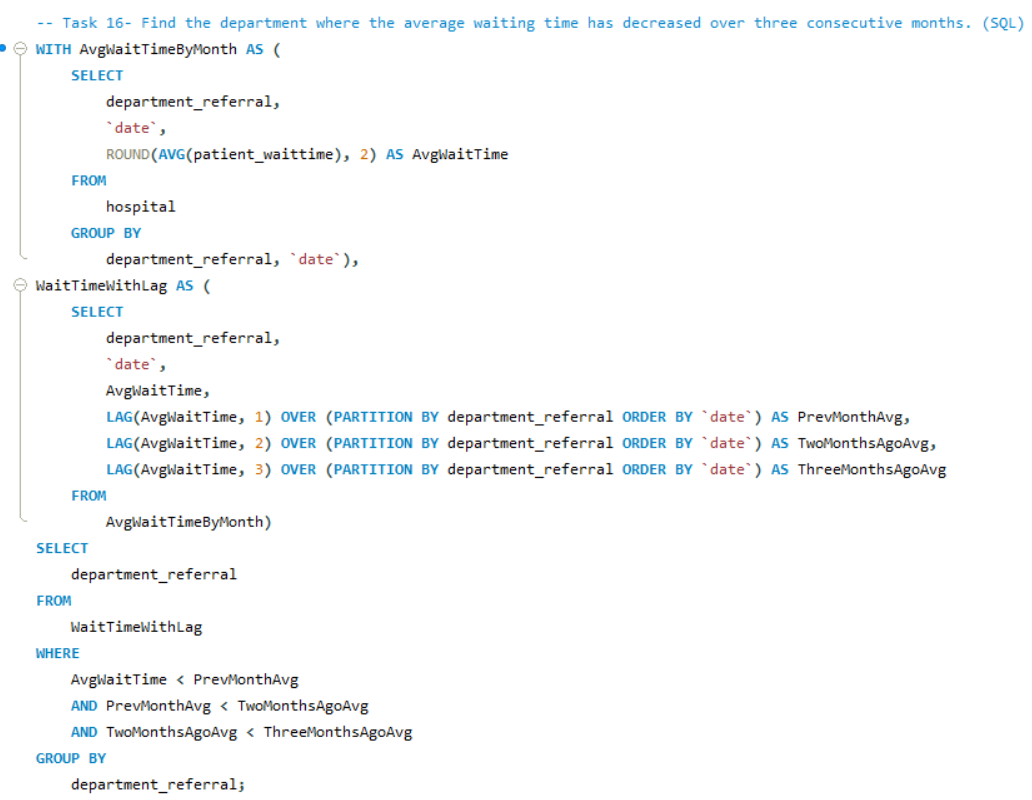
The main query filters departments where the current month's wait time is less than the previous month's, the previous month's is less than two months ago, and two months ago is less than three months ago. This ensures a consistent decrease for three consecutive months.

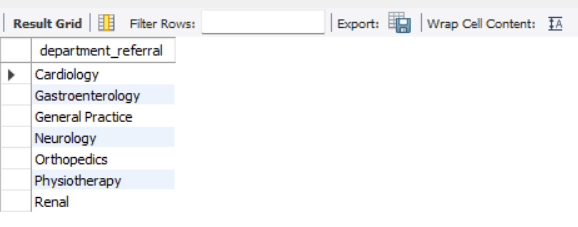
* **Group by Department:**

To avoid duplicate department entries in the output, the query groups the results by department\_referral.

* **Output Departments:**

The final result lists departments showing a consistent downward trend in average wait times over the last three months.

****

** Result -**

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

**Ans –**

* **Count Male and Female Patients per Doctor:**

The first CTE (patientCount) calculates the number of male and female patients for each doctor by checking the patient\_gender column and grouping the data by Doctor Name.

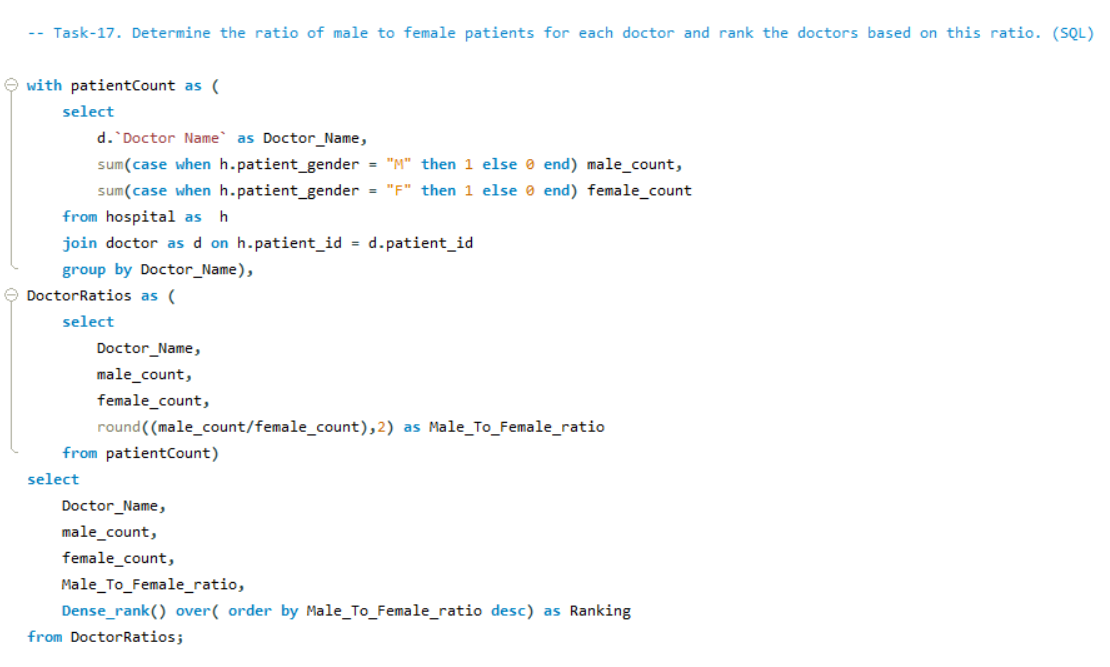
* **Calculate Male-to-Female Ratio:**

The second CTE (DoctorRatios) uses the counts from patientCount to calculate the male-to-female ratio for each doctor and rounds it to two decimal places.

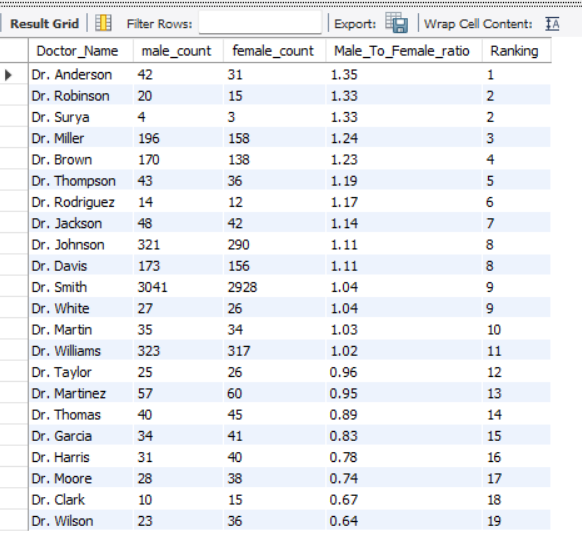
* **Rank Doctors by Gender Ratio:**

The main query assigns a dense rank to doctors based on their male-to-female ratio in descending order, where doctors with the same ratio share the same rank.

* **Final Output:**

****The query outputs the doctor’s name, male and female patient counts, the male-to-female ratio, and their ranking by the ratio.

**Result -**

****

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

**Ans –**

* **Calculate Average Satisfaction Score:**

For each doctor, the query calculates the average patient satisfaction score (patient\_sat\_score). If the score is missing (""), it defaults to a score of 5.

* **Join Hospital and Doctor Data:**

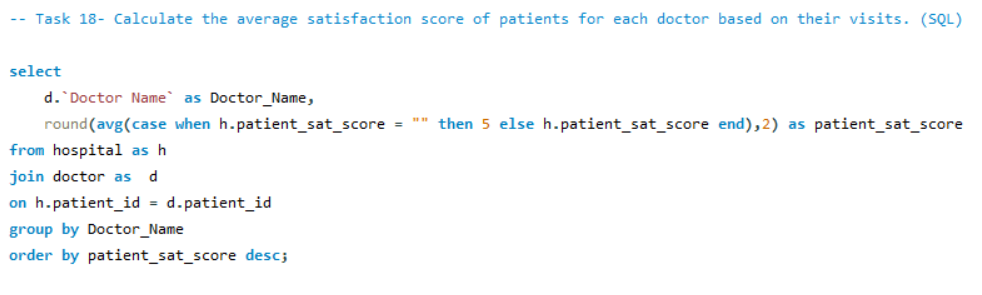
The hospital and doctor tables are joined on the patient\_id column to associate patient satisfaction scores with their respective doctors.

* **Group by Doctor:**

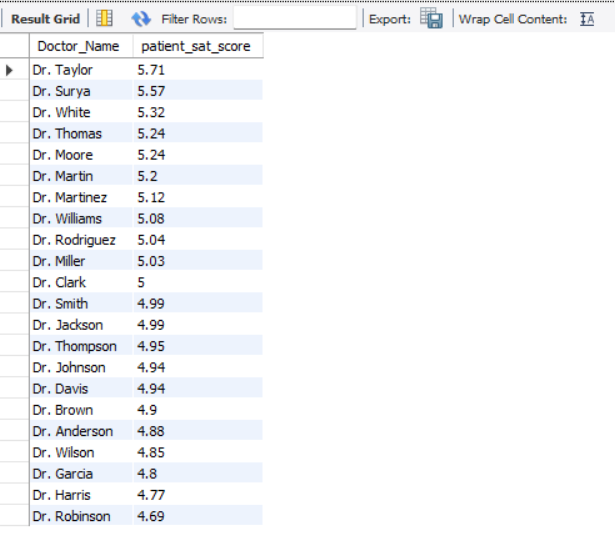
The query groups the data by Doctor Name to calculate the average satisfaction score for each doctor.

* **Order by Satisfaction Score:**

The results are sorted in descending order of satisfaction scores, displaying doctors with the highest average scores first.

****

**Result -**

****

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

**Ans-**

* **Count Different Patient Races per Doctor:**

The query calculates the number of unique patient races (different\_race\_count) treated by each doctor using COUNT(DISTINCT h.patient\_race).

* **Join Hospital and Doctor Data:**

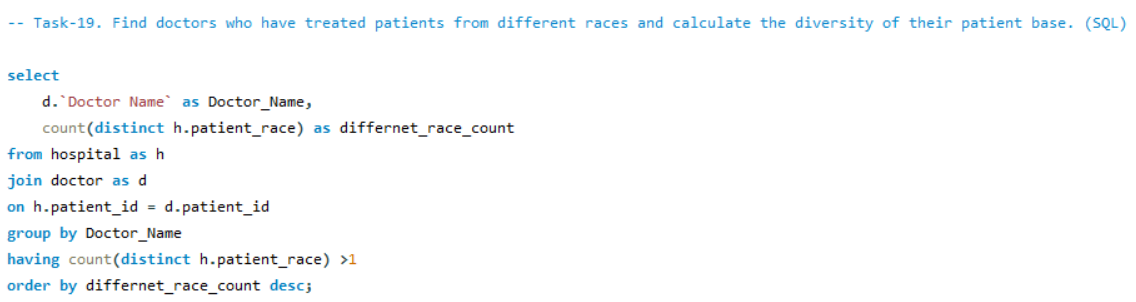
The hospital and doctor tables are joined on patient\_id to link patients' race data to their respective doctors.

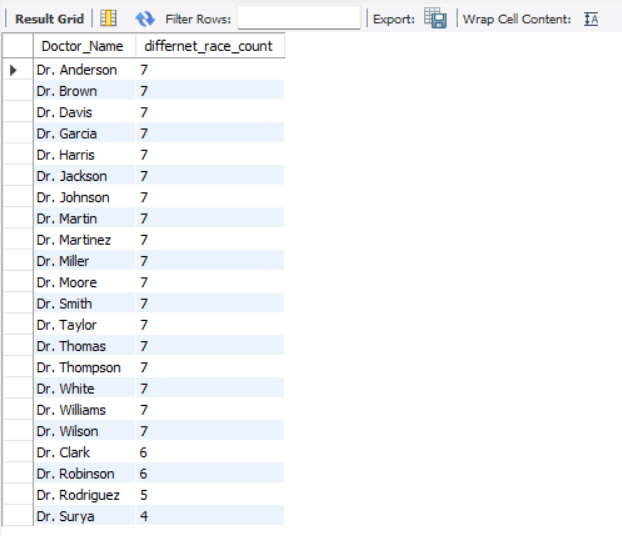
* **Filter Doctors with Diverse Patient Races:**

The HAVING clause filters the results to include only those doctors who have treated patients of more than one race (different\_race\_count > 1).

* **Order by Diversity Count:**

The results are sorted in descending order of different\_race\_count, showing doctors with the most racially diverse patient groups first.

****

**Result –**

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

**Ans –**

* **Calculate Total Bills by Gender for Each Department:**

The query computes the total billing amount for male (Male\_total\_bill) and female (Female\_total\_bill) patients in each department using conditional aggregation with CASE statements.

* **Join Hospital and Doctor Data:**

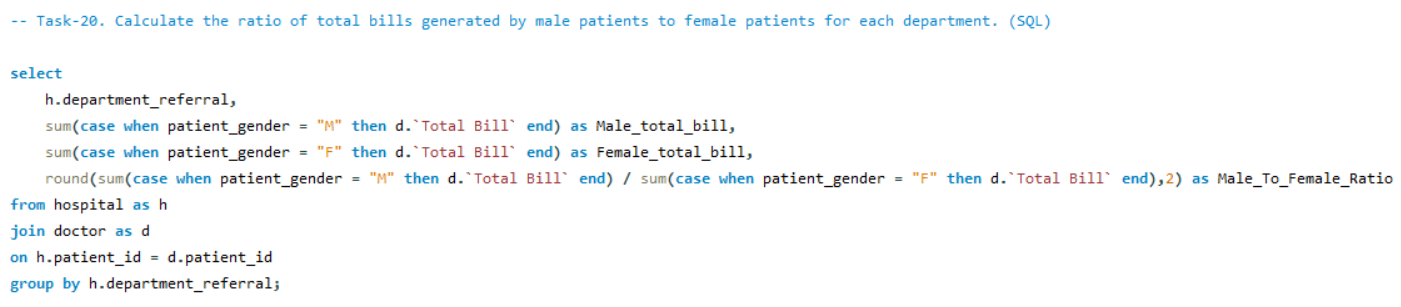
The hospital and doctor tables are joined on patient\_id to link billing data from doctors to their respective departments.

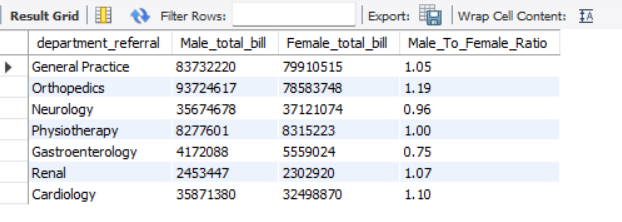
* **Compute Male-to-Female Billing Ratio:**

The query calculates the ratio of total billing for male patients to female patients (Male\_To\_Female\_Ratio) and rounds it to two decimal places.

* **Group by Department:**

The data is grouped by department\_referral to generate these statistics for each department individually.

****

**Result –**

**Task-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)**

**Ans –**

* **Disable Safe Updates:**

SET SQL\_SAFE\_UPDATES = 0 ensures that the update query can execute without requiring a key or full table conditions.

* **Update Patient Satisfaction Score:**

The UPDATE query increases patient\_sat\_score by 2 for patients who meet the conditions:

They visited the 'General Practice' department (h.department\_referral = 'General Practice').

Their waiting time exceeded 30 minutes (h.patient\_waittime > 30).

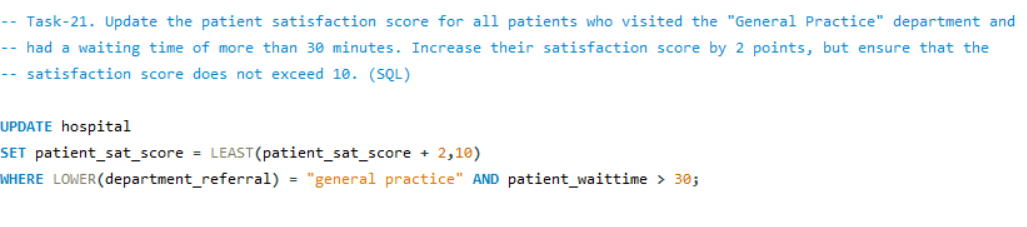
If the new score exceeds 10, it is capped at 10 using a CASE statement.

* **Conditions in WHERE:**

Ensures only patients in the 'General Practice' department with a wait time greater than 30 minutes are affected.

* **Re-enable Safe Updates:**

SET SQL\_SAFE\_UPDATES = 1 restores safe update mode to prevent accidental changes in subsequent queries.

****

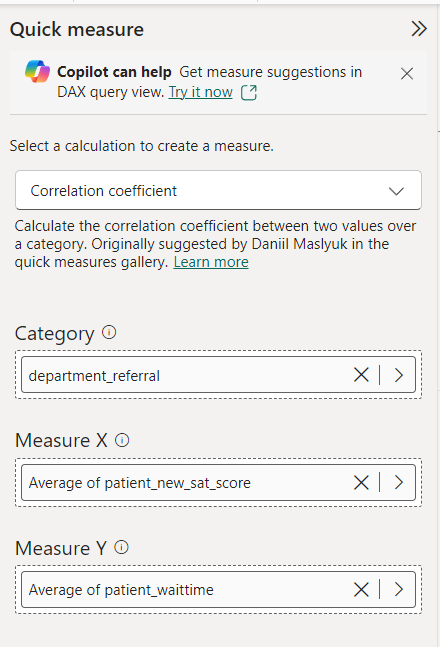
**SUBJECtive questions**

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

**Ans -**

* Here I use **Quick Measure** :

It’s a Power BI feature that helps create commonly used calculations (like correlation, averages, etc.) without needing to write DAX formulas.

* I selected "Correlation Coefficient" to measure the relationship between patient wait time (Measure Y) and satisfaction score (Measure X), grouped by department referral (Category).
* It used a statistical formula to find the correlation coefficient, which is 0.49, showing a moderate positive relationship between the two measures.
* ****It automates complex calculations, saves time, and makes it easy to analyze relationships between data variables without DAX knowledge.

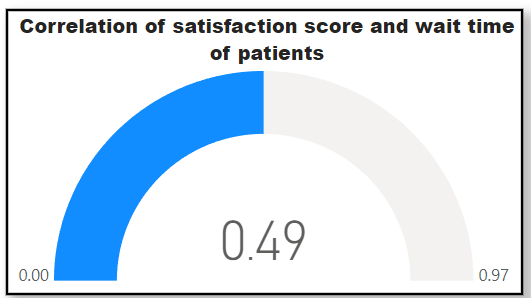
**Explanation –**

1. **Correlation Coefficient Value**:

* A correlation coefficient ranges from **-1 to +1**:
  + - **+1**: Strong positive correlation (as one variable increases, the other also increases).
    - **0**: No correlation (no relationship between variables).
    - **-1**: Strong negative correlation (as one variable increases, the other decreases).
* The value **0.49** indicates a **moderate positive correlation**.

1. **Interpretation**:
   * The Gauge Chart effectively illustrates the result, making it easier to interpret the correlation score and its significance within the range.

* This result suggests that as **patient wait time increases**, there is a slight tendency for **satisfaction scores** to increase as well, but the relationship is not very strong.
* This may suggest that wait times are not the only factor affecting patients' satisfaction; other factors may have a substantial impact.

****

1. **Possible Implications**:

* **Counterintuitive Insight**: Given the common association between greater wait times and decreased satisfaction, a minor positive connection may appear surprising.
* **This could mean:**
  + - Longer wait times may result in better services or care for patients, which raises satisfaction.
    - In addition to wait time, other factors like communication or service quality may have a greater impact on customer satisfaction.
    - **Data Context**: The specific dynamics of the department or survey design might play a role.

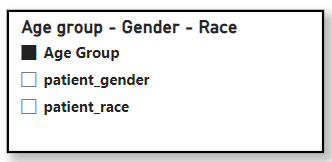
**This analysis highlights the need to balance wait times and other quality-of-service factors to improve overall patient satisfaction.**

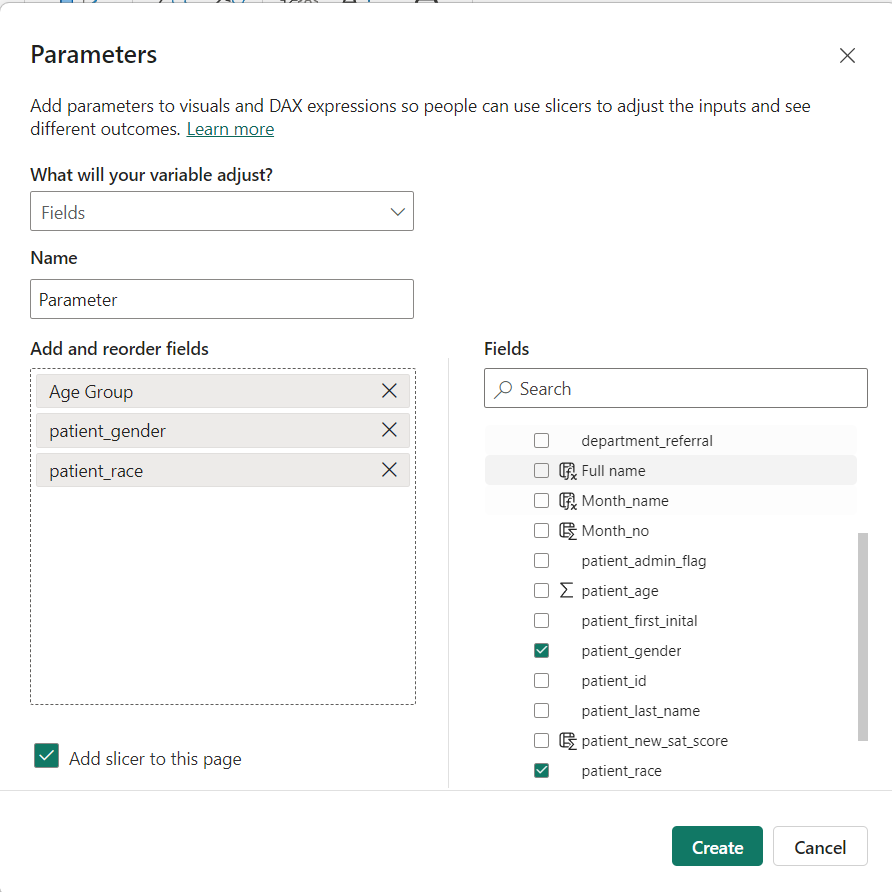
**Task-2. How do patient demographics affect the frequency of visits to different departments?**

**Ans –**

**Approach :**

* Prepare the data with columns for Department, Patient Count, and demographics (Age Group, Race, Gender).
* Create a Stacked Column Chart with Department on the x-axis and Patient Count on the y-axis and demographics on the legend as it will show the numbers.
* Use Field Parameters to switch between demographic variables (Age Group, Race, Gender) to dynamically filter the data.



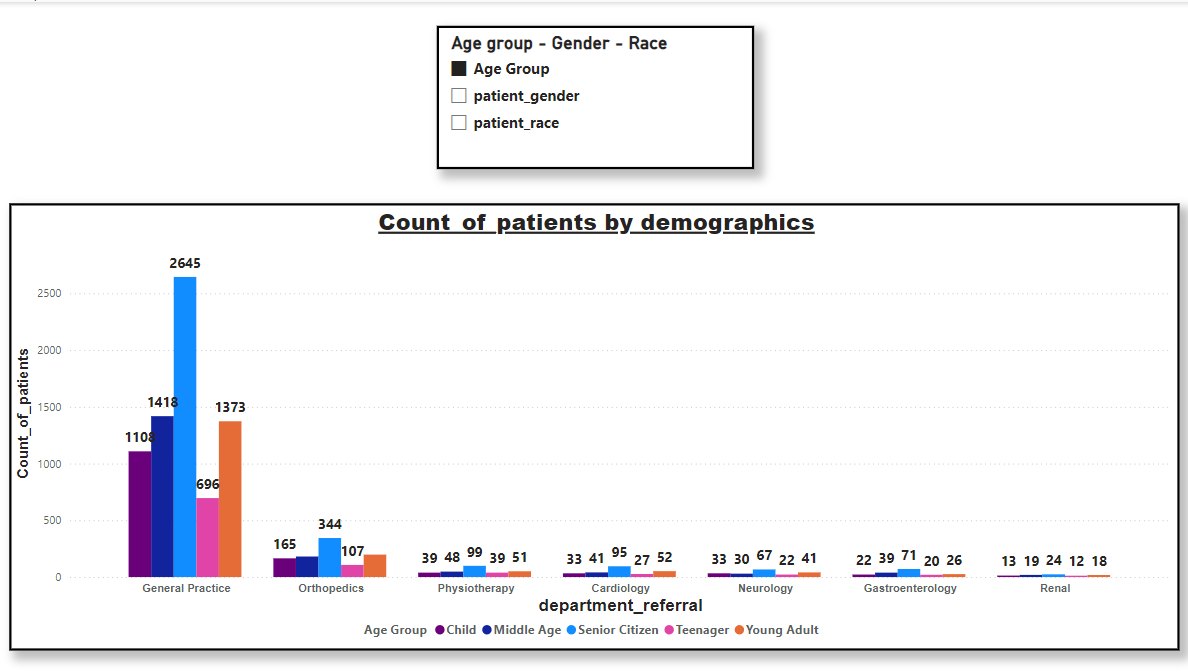


**Explanation :**

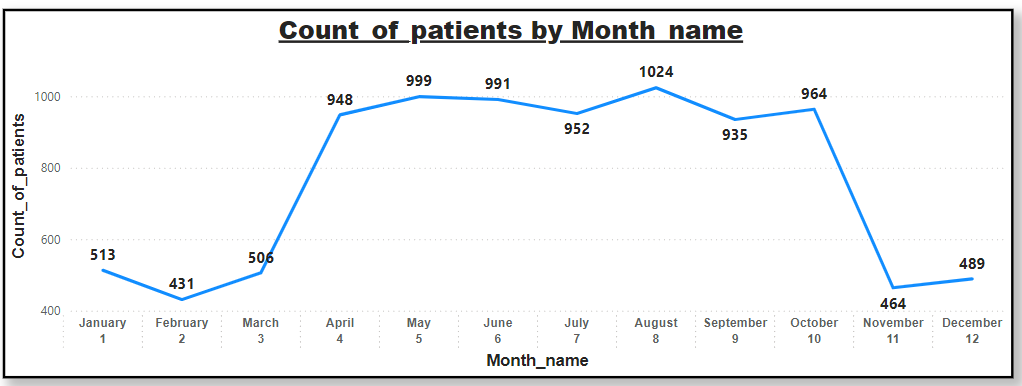
* The goal of this analysis is to explore how patient demographics influence the frequency of visits to various departments. By using a Stacked Column Chart, we can see how different demographic groups contribute to the overall patient count per department.
* Field Parameters allow us to toggle between demographic categories (such as Age Group, Race, and Gender) to analyze the impact of each on department visit frequency.
* This helps in understanding whether certain departments are preferred by specific demographic groups and whether there are patterns or imbalances in departmental utilization based on patient characteristics.

**Insight :**

* Demographics do not affect the number of visitors to other departments because the number of visits depends entirely on the patient’s health and the location of the hospital. The demographics of different departments are unaffected directly.

****

**Task-3. Is there a noticeable trend in the volume of patient visits throughout the year?**

**Ans –**

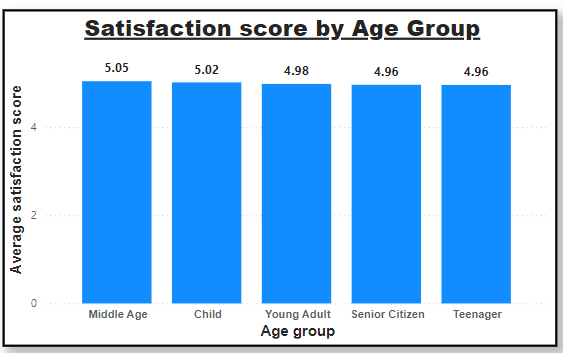
**Observations –**

* We can see that August was the month with the most visit and February was the month with the least.
* There is a high volume of patients all through April to October and then a sharp decline in November.
* Start of the year (January – March) and end of year (November – December) has low volume of patients.

**Task-4. Which age groups report the highest and lowest satisfaction scores?**

**Ans –**

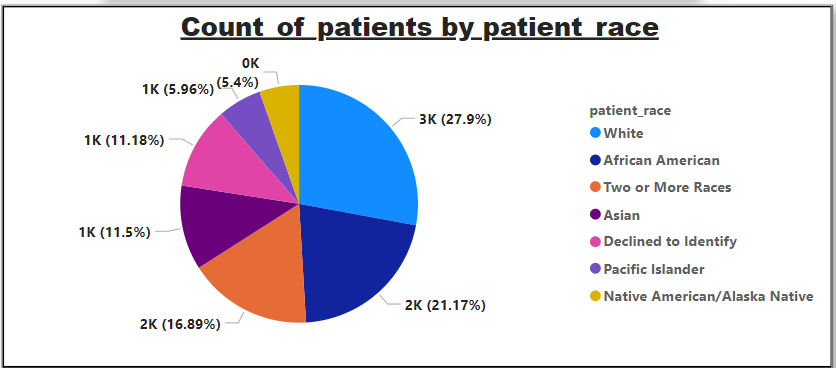
1. **Highest Satisfaction Score**: The **Middle Age** (36<Age<=50) group reports the highest satisfaction score of **5.05**, slightly above other groups.
2. **Lowest Satisfaction Score**: The **Senior Citizen** (Age>50) and **Teenager** (12<Age<=20) group share the same lowest score of **4.96**, indicating no significant variation among these groups.
3. **Observation**: The difference between the highest and lowest scores is minimal , meaning satisfaction levels are generally consistent across all age groups.
4. **Key Insight**: Middle-aged individuals might have slightly better experiences or expectations met compared to others, but overall satisfaction is stable across the board.



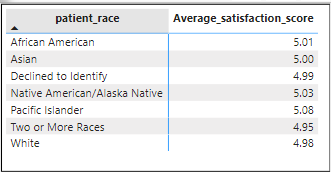
**Task-5. Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?**

**Ans - We can deny this claim multiple ways through the visualization of our patients data.**

**1)Count of patients by race.**

* We can clearly see that white patients only account for around 27.9% of our total patient base.
* The other approx. 70% of our patient base is non-white.
* We have patients from various walks of life such as African American (21.7%), Mixed Race (16.89%), Asian (11.5%), Pacific Islander (5.6%) and Native American/ Alaska Native with 5.4%.
* ****The remaining patients who declined to identify their race are around 11.18% which still if we assume all are White, still leave a significant gap between white patient base and others.

**2)Average patient satisfaction score by race.**



* As we can observer from the above matrix that Pacific Islander highly rate the hospital with an average satisfaction score of 5.09.
* Followed by Native America/Alaskan Native with a score of 5.03 and African American with the score of 5.02.
* Asians, Whites and Mixed race people rate the hospital in a very similar way with 5.0, 4.98 and 4.96 average satisfaction scores respectively.
* And, people who refused to disclose their race have an average patient satisfaction score of 4.99.
* All of the above metrics can put any claims of the hospital being racially discriminating false.

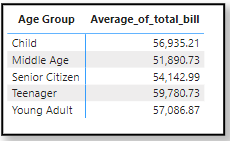
**The above analysis reveals that the hospital does not exhibit any discriminatory practices.**

**Task-6. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?**

**Ans –**

**1st Approach –**

**Giving discounts on the basis of Age Group :**

****

* We can provide discounts to age groups who spent the most at the hospital.
* We can see that Teenagers (Aged between 12 to 20) are the ones who spent the highest on average in the hospital.
* Amounting to over $59,780.73 on average they are hospital’s biggest customer.
* Secondly, if we attract more patients from the teenager age group, we might be able to create relationships with them and make a life long customers out of them.
* For all the above reasons I think the hospital should give patients discounts on the basis of their age and more specifically to the teenagers.

**2nd Approach –**

**Giving discount whose Total Bill is more than 10,000 :**

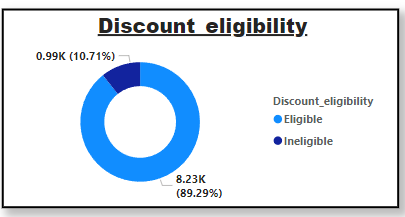
1. **Dax Formula used –**

Discount\_eligibility = if(calculate(SUM('Doctor Info'[Total Bill]),ALLEXCEPT('Doctor Info','Doctor Info'[patient\_id])) > 10000 ,"Eligible","Ineligible")

1. **Explanation –**

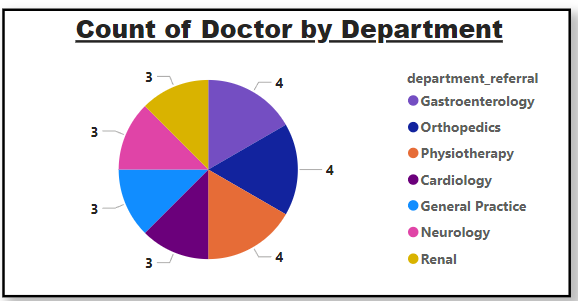
* **Purpose:** This DAX formula calculates whether a patient is eligible for a discount based on the total bill amount across all their visits. If their total exceeds ₹10,000, they are marked as "Eligible"; otherwise, "Ineligible".
* **Breakdown:**
* SUM('Doctor Info'[Total Bill]): Adds up the total bill amounts for the data being analyzed.
* ALLEXCEPT('Doctor Info', 'Doctor Info'[patient\_id]): Removes filters on all columns except patient\_id, ensuring the calculation considers all bill entries for the same patient, even if other filters are applied.
* **Logic:** If the calculated total bill for a patient (using the above logic) is greater than ₹10,000, the formula assigns "Eligible". Otherwise, it assigns "Ineligible".
* This formula is commonly used in scenarios where patient-level aggregations are needed to determine eligibility based on cumulative criteria, such as loyalty rewards or discounts.

1. **The hospital management intended to offer discounts to high-spending. I used a DAX formula to categorize patients based on their total bill, creating a field to reflect whether a patient qualifies for the discount.**

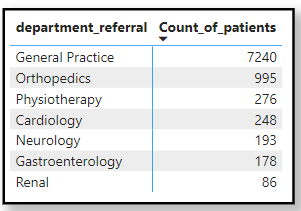


For all the above reasons I think the hospital should give patients discounts on the basis of their age and more specifically to the teenagers and also who spent more than 10,000.

**Task-7. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.**

**Ans –**

* The above pie chart shows current departments and the number of doctors per department.
* We can see that most of the hospital’s department have 3 or 4 doctors.
* Further analyze the number of visits each department get.



**Observation –**

* General Practices gets the highest amount of visits with 7,240.
* Even Orthopedics which comes in at 2nd place with 995 visits is no-where near that.

**Suggestion-**

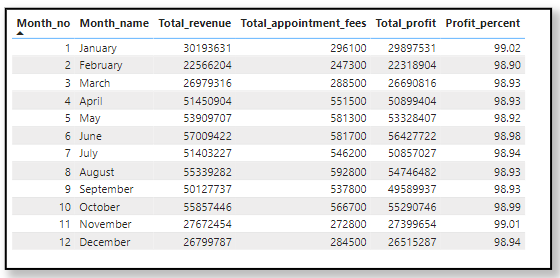
* My suggestion would be to hire all 3 new doctors to the General Practice department as it has only 3 doctors currently.

**Task-8. Is the hospital profitable? How will you determine the profitability?**

**Ans –**

**Approach -**

* To determine the hospital's profitability, I calculated the total revenue, total appointment fees, total profit, and profit percentage using DAX measures.
* These metrics helped assess whether the hospital is generating a profit or not.



**Explanation –**

* **Total Appointment Fee:** Calculated by summing the appointment fees for each patient.
* **Total Revenue:** Sum of the total bill amounts for all patients.
* **Total Profit:** Derived by subtracting the total appointment fee from the total revenue.
* **Profit Percentage:** Calculated by dividing the total profit by the total revenue and multiplying by 100 to get the percentage.

**Observation -**

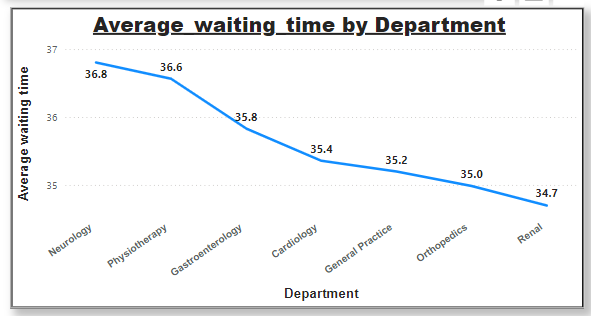
* The Profit Percentage values were consistent, showing around 98.96% and 98.87%, indicating a very high profit margin.
* The profit margin remains consistently high, well above 90%, which indicates the hospital is highly profitable.
* Based on the profit percentage consistently being above 98%, it is clear that the hospital is highly profitable.

**Task-9. Any Department for which the waiting time is oddly large?**

**Ans –**

**Approach -**

To identify departments with unusually large waiting times, I created a line chart with Department on the x-axis and Average Waiting Time on the y-axis. This visualization allowed for a clear comparison of waiting times across all departments.



**Observation –**

* By analyzing the chart, I observed that the Neurology department had the highest average waiting time of 36.80 minutes, which stands out compared to other departments
* All the departments have the average wait time between 34.70 minutes to 36.80 minutes which is approximately a 2.1 min of difference between the maximum and minimum.

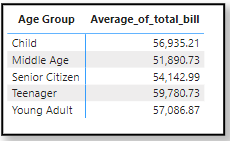
**This difference is very insignificant and hence we can conclude that there is no particular department with relatively very high wait time.**

**Task-10. Come up with strategies to provide discounts to the patients.**

**Ans -**

**1st Approach –**

**Giving discounts on the basis of Age Group :**

****

* We can provide discounts to age groups who spent the most at the hospital.
* We can see that Teenagers (Aged between 12 to 20) are the ones who spent the highest on average in the hospital.
* Amounting to over $59,780.73 on average they are hospital’s biggest customer.
* Secondly, if we attract more patients from the teenager age group, we might be able to create relationships with them and make a life long customers out of them.
* For all the above reasons I think the hospital should give patients discounts on the basis of their age and more specifically to the teenagers.

**2nd Approach –**

**Giving discount whose Total Bill is more than 10,000 :**

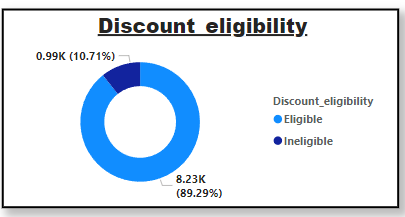
1. **Dax Formula used –**

Discount\_eligibility = if(calculate(SUM('Doctor Info'[Total Bill]),ALLEXCEPT('Doctor Info','Doctor Info'[patient\_id])) > 10000 ,"Eligible","Ineligible")

1. **Explanation –**

* **Purpose:** This DAX formula calculates whether a patient is eligible for a discount based on the total bill amount across all their visits. If their total exceeds ₹10,000, they are marked as "Eligible"; otherwise, "Ineligible".

1. **The hospital management intended to offer discounts to high-spending. I used a DAX formula to categorize patients based on their total bill, creating a field to reflect whether a patient qualifies for the discount.**

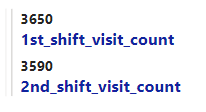


**Task-11. Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also will this 2 shift policy be helpful for the hospital?**

**Ans -**

**Approach -**

* Divide day into two shifts i.e.
* 1st Shift – 0.00 to 12.00
* 2nd Shift – 12.00 to 24.00
* Calculate the count of patient visits in those 2 shifts.
* **Dax formula used –**
* **1st\_shift\_visit\_count =** CALCULATE([Count\_of\_patients],'Hospital Info'[Hours] >=0 && 'Hospital Info'[Hours]<12 && 'Hospital Info'[department\_referral] ="General Practice")
* **2nd\_shift\_visit\_count =** CALCULATE([Count\_of\_patients],'Hospital Info'[Hours] >=12 && 'Hospital Info'[Hours]<=23 && 'Hospital Info'[department\_referral] ="General Practice")



**Observation –**

* By visualizing these measures using the multi-row card,

we can see that Shift 1 has more patients than Shift 2.

* While assigning shifts to doctors we should assign more doctors in shift 1 than in shift 2.
* Since we have 3 doctors currently in General Practice Department, we will assign 2 doctors to Shift 1 and 1 doctor to Shift 2.
* The main benefit of dividing duty of the doctors in shifts is that the workload get distributed evenly.
* This will turn leads to lesser wait time for patients and better satisfaction scores.

**Task-12. What do you understand by Power BI gateway? What are its use cases?**

**Ans -**

* **Power BI Gateway:**

It is a tool that enables secure data transfer between on-premises data sources and the Power BI cloud service, allowing you to refresh and access data stored locally in reports and dashboards.

* **Use Cases:**
* **Data Refresh:** Automatically refreshes on-premises data for up-to-date reports.
* **Secure Data Transfer:** Ensures encrypted, secure data transfer between on-premises sources and Power BI.
* **Access On-Premises Data:** Allows on-premises data (e.g., SQL Server, Excel) to be used in Power BI.
* **Hybrid Data Integration:** Combines on-premises and cloud data for unified reports.
* **Direct Query:** Enables real-time querying of live data sources in Power BI.
* **Types of Gateways**:
* **Personal Gateway:** For individual use, refreshing personal data.
* **Enterprise Gateway:** For organizations, supporting multiple users and data sources.
* **Some benefits of Power BI gateway are:**
* The connection between the data source and cloud is very secure and reliable which help in protection of sensitive data.
* Data can be accessed directly without moving it to expensive cloud storage solution hence saving money.
* Dashboards can be refreshed using Power BI gateway with a connection to the on-premise data source.
  1. **Task-13. How would you approach this problem, if the objective and subjective questions weren't given?**
  2. **Ans -**
* If the objective and subjective question were not given to me I would have approached the problem according to the goals.
* First I would have import the data into Power BI to begin the analysis process.
* Use Power Query to transform the data by filtering, merging, and shaping it to fit the required structure.
* Clean the data by addressing any inconsistencies, handling missing values, correcting data types, and ensuring the format is standardized across the dataset.
* Identify the key performance indicators (KPIs) that are most relevant to the objectives of the analysis.
* Analyze each KPI in detail, exploring the trends, relationships, and patterns to derive meaningful insights.
* To gain insights for hiring new employees, I would not just focus on doctors but also the nursing staff and the support staff.
* Having a bigger picture for each department will help in sharing workload and hiring/firing employees according to the hospital’s requirements.
* For patients discounts, my strategy would remain the same, I would still give discounts to the age group which gives us the most business.
* Along with the immediate benefit of increased cash flow, catering to a specific age group could pave way for a specialty department which in turn would attract more of the same high revenue generating group.
* Create comprehensive reports and dashboards using various visualizations such as bar charts, line graphs, and tables to present the findings.
* Add interactive slicers and filters to the reports, enabling users to dynamically explore the data based on different dimensions.
  1. **Review the final report for accuracy, clarity, and relevance, and provide actionable recommendations based on the insights gathered from the analysis.**
  2. **Task-14. Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?**
  3. **Ans -**Each doctor\_id is associated with one department.
  + Each department can have multiple doctors assigned to it.
  + This makes it a **one-to-many** relationship from **the perspective of the doctors** to the department. Alternatively, you could describe it as a **many-to one** relationship **from the department's perspective** to the doctors.
  + The relationship between doctor\_id and department is one-to-many because many doctors can belong to a single department, but each doctor is assigned to only one department.