# 

# **Case Study**

# **Power BI Project**

# **Columbia Asia Hospital Analysis**

**Objective Questions:**

1. In analyzing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.
2. **Assess the Average Waiting Time:** Analyse the patient wait times to identify the average duration a patient spends before receiving care.
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.
4. **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.
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?
6. Is there any relation between the number of visits and the Gender of the patients?
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.
8. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.
9. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.
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.
11. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.
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?
13. Were there any data format issues in the data, and if there were/are how you handle them?
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?
15. Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)
16. Find the department where the average waiting time has decreased over three consecutive months. (SQL)
17. Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)
18. Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)
19. Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)
20. Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)
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)

**Subjective Questions**

1. What is the relation between patient wait time and satisfaction scores?
2. How do patient demographics affect the frequency of visits to different departments?
3. Is there a noticeable trend in the volume of patient visits throughout the year?
4. Which age groups report the highest and lowest satisfaction scores?
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?
6. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?
7. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.
8. Is the hospital profitable? How will you determine the profitability?
9. Any Department for which the waiting time is oddly large?
10. Come up with strategies to provide discounts to the patients.
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?
12. What do you understand by PowerBI gateway? What are its use cases?
13. How would you approach this problem, if the objective and subjective questions weren't given?
14. Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?

**Report**

The hospital has asked for a report with three tabs:

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

**Objective Questions:**

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

**Load and Inspect Data**

* Import data from all sources (Excel, CSV, etc.).
* Check row counts, column names update to standard names and check and assign the correct data types.
* Identify potential duplicates, nulls, or formatting issues.

**Handle Missing Values**

* I used the Power Query Editor in Power BI to examine the dataset for any missing (null) values or irregular data.
* During this process, I noticed that the patient\_sat\_score column contained several blank entries.
* So, to avoid the missing values I have filled the blank values with the average patient satisfaction rate (Customer Satisfaction Score).

CSAT Score = if(ISBLANK('Hospital ER'[patient\_sat\_score]), AVERAGE('Hospital ER'[patient\_sat\_score]), 'Hospital ER'[patient\_sat\_score])

This formula checks if the value is blank. If it is, it replaces it with the average of the column. If not, it keeps the original score.

**Correct Data Types**

* Ensure date fields are set to Date/DateTime.
* Columns/Attributes like numerical values such as satisfaction score, age, need to be updated as Decimal / Whole number
* IDs like (PatientID, DoctorID) were changed to Text

**Standardize Text Fields**

* IDs like (Replace inconsistent entries (e.g., "M" → "Male", "F” → Female")

**Apply and Load:**

* Columns After completing these cleaning steps, I clicked "Close & Apply" to save and load the cleaned data back into Power BI, making it ready for further analysis and visualizations on the report tabs.

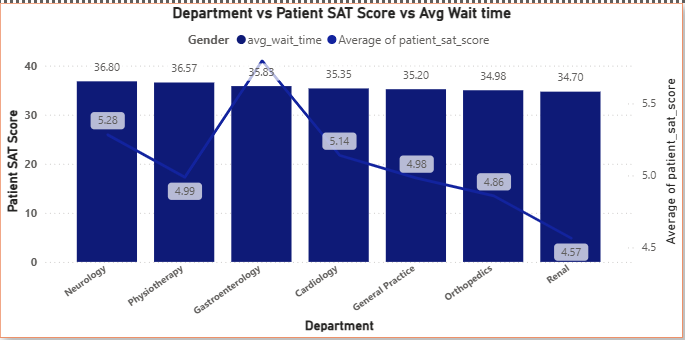
1. **Assess the Average Waiting Time:** Analyse the patient wait times to identify the average duration a patient spends before receiving care.

**Overall Average Waiting Time: 35.26 min**

avg\_wait\_time = AVERAGE('Hospital ER'[patient\_waittime])

****

With respect to each department, I have calculated the Average Waiting Time which gives an idea about the wait time for patients with respect to the consultation.



**Insights:**

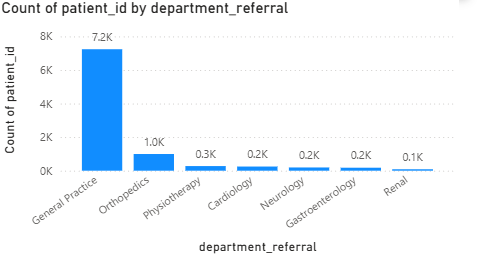
Overall Average Wait Time: The Card visual highlights the hospital’s overall baseline average wait time.

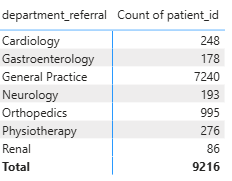
Opportunity for Improvement: As Department Wise Average Waiting Time is there, we can try to improve it by arranging the additional staff to the respective department which are having highest Average Waiting Time

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

**Steps to Calculate Total Visits by Each Department:**

1. To calculate the Total No. of Visits by Each Department: Firstly, I have calculated the total number of visits by counting the number of entries in the patient\_id column, which represents No. of patients in total.
2. Post that adding the bar chart with Department referral in x-axis and total number of patients in y-axis which makes it easy to compare visits across departments.

****

****

**Insight:**

* General Practice receives the vast majority of patient referrals (7.2K) as many of the patients will consult a doctor with normal illness which doesn’t need major attention and not many follow up checkups as the severity is less than or equal to medium level.
* Orthopedics is the distant second most common referral department (1.0K).
* All other specialized departments have significantly fewer referrals (0.1K-0.3K)
* Renal department has the lowest number of referrals at 0.1K.

**Recommendation:**

1. **Patient Demand** – Bar chart highlights departments with the highest visit volumes, showing where patient demand is greatest.
2. **Resource Allocation** – High-volume departments may require additional staff or resources to reduce wait times and maintain care quality.
3. **Referral Patterns** – Frequent referrals point to common patient pathways and departments where specialized care is sought.
4. **Service Enhancement** – As Renal department having Low-volume departments present opportunities for outreach and awareness to increase service utilization.
5. **Seasonal Trends** – Monitoring visit patterns over time reveals seasonal fluctuations, supporting proactive staffing and resource planning.
6. **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

To Categorize the age group, I have calculated in the below following format to know how many patients are falling under each category with respect to department which will give an idea about which staff is crucial and to improve based on the patient satisfaction score as well

To analyse the patient visits by each age group, I’ve categorized patient age into groups ("0 - 12", "13 - 21", "22 - 35", "36 - 59", "60+"). I used a bar chart visualization on the Patients’ Tab to display these groups along the X-axis, with total visits for each age group on the Y-axis. This layout provides a clear view of which demographics use healthcare services most frequently.

**Age\_Group\_Category** = IF('Hospital ER'[patient\_age] <= 12,  "Child",

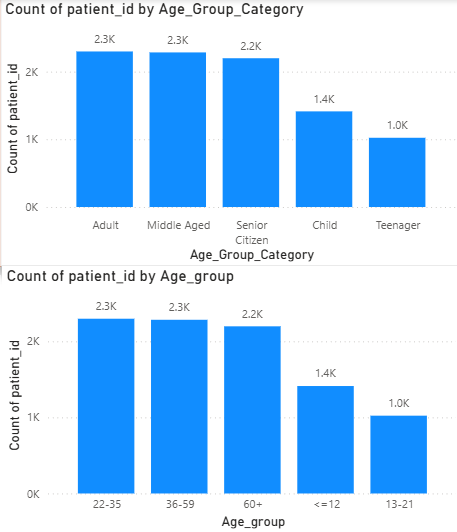
IF('Hospital ER'[patient\_age] > 12 && 'Hospital ER'[patient\_age] <=21,  "Teenager",

IF('Hospital ER'[patient\_age] > 21 && 'Hospital ER'[patient\_age] <= 40, "Adult", IF( 'Hospital ER'[patient\_age] > 40 && 'Hospital ER'[patient\_age] <= 60, "Middle Aged",  "Senior Citizen" ))))

**Age\_group** = if('Hospital ER'[Age\_Group\_Category] = "Child", "<=12", if('Hospital ER'[Age\_Group\_Category]= "Teenager", "13-21",

if('Hospital ER'[Age\_Group\_Category] = "Adult", "22-35",

if('Hospital ER'[Age\_Group\_Category] = "Middle Aged", "36-59", "60+"))))

****

**Insights:**

* **﻿**At 36.80, Neurology had the highest avg\_wait\_time and was 6.07% higher than Renal, which had the lowest avg\_wait\_time at 34.70.﻿﻿ ﻿﻿ ﻿﻿
* avg\_wait\_time and total Average of patient\_sat\_score are negatively correlated with each other.﻿﻿ ﻿﻿
* Across all 7 department\_referral, avg\_wait\_time ranged from 34.70 to 36.80 and Average of patient\_sat\_score ranged from 4.57 to 5.80.﻿﻿ ﻿﻿ ﻿﻿﻿
* ﻿﻿ **High Utilization Groups:** Adult & Middle aged were accounted for around 24.93% of total count of patients.
* **Resource Planning:** Understanding age group trends aids in planning for staffing, facilities, and resources, aligning hospital capabilities with age-specific patient demand.

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?

Yes, there are Null Values that exists in the data

**Null Values Detected:** The dataset contains missing (null) entries.

**Identification of Missing Data:** Power Query Editor was used to review and detect null or inconsistent records.

**Treatment of Missing Data:** Null values (e.g., in the patient\_sat\_score column) were replaced with the column average to maintain data integrity and completeness.

**DAX Formula Applied:**

CSAT Score = if(ISBLANK('Hospital ER'[patient\_sat\_score]), AVERAGE('Hospital ER'[patient\_sat\_score]), 'Hospital ER'[patient\_sat\_score])

This approach ensures all fields are complete, assuming that missing satisfaction scores likely represent a neutral or non-critical response from patients as I have considered the average sat score for the missing values in sat score of the patients. This straightforward method is especially useful when missing data points are minimal, allowing for uninterrupted insights while keeping the dataset fully intact.

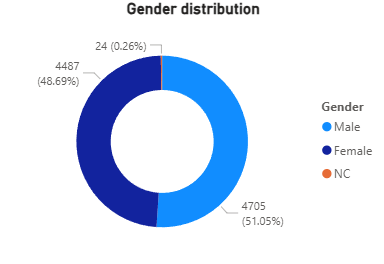
1. Is there any relation between the number of visits and the Gender of the patients?

It shows that visits are almost equally shared between male and female patients, with males making up a slightly larger share. The "NC" group is negligible, representing just 0.26% of the total.

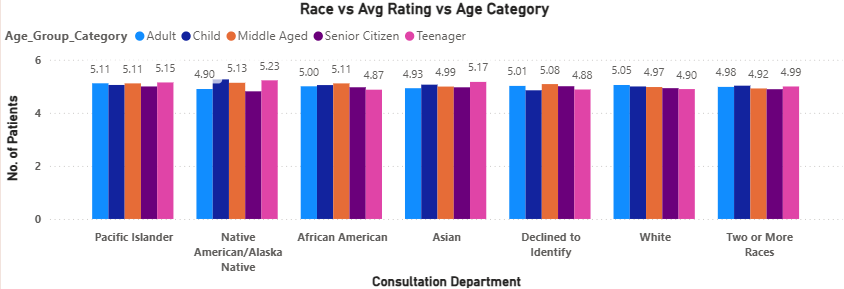
The Gender Distribution chart shows the proportion of patients by gender who visited the hospital:

* **Male Patients:** 4,705 visits (≈ 51.05%)
* **Female Patients:** 4,487 visits (≈ 48.69%)
* **Not Classified (NC):** 24 visits (≈ 0.26%)
* There is **no significant imbalance** between male and female patient visits:
* The distribution is **almost equal**, with males slightly higher (≈2.36% difference).
* This suggests that **gender does not strongly influence hospital visits**—both male and female patients access services at nearly the same rate.
* The NC category is negligible and does not impact overall analysis.

**Conclusion:**  
There is **no strong relation** between patient gender and the number of visits, since both genders contribute almost equally to hospital visits.

****

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.

****

For Analysing the relationship between patient satisfaction scores, age groups, and racial backgrounds, I have used a Line & Clustered Column Chart Titled Race vs Avg Rating vs Age Category which is in page 2.

The chart visualizes the average satisfaction scores for different age groups and racial backgrounds, using the average satisfaction score measure. Patient race was placed in the x-axis field; Age group Category was placed in Column Legend field while the average satisfaction score was shown in the y-axis field. Data labels were enabled for easy interpretation of satisfaction scores for each demographic category

**Insights:**

**Patient Race Distribution:** White (27.9%) and African American (21.2%) patients form the largest segments, together accounting for nearly 50% of total visits.

**Impact on Satisfaction:** Since these groups represent the majority, their satisfaction scores heavily influence overall patient experience.

**Asian & Multi-Race Patients:** Together make up ~28% of visits; analyzing their satisfaction can reveal cultural or service alignment gaps.

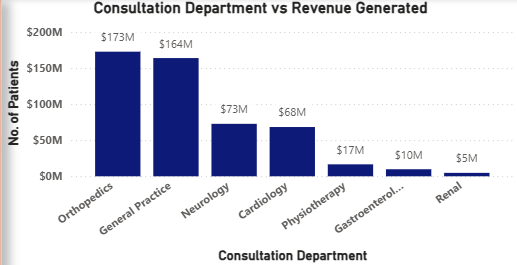
**Smaller Groups (Pacific Islander, Native American, Declined to Identify):** While smaller in number, tracking satisfaction here ensures inclusiveness and equity in care delivery.

**Actionable Insight:** Combining race with age group analysis helps identify specific clusters (e.g., older Asian patients, younger African American patients) with lower satisfaction levels. It helps us to identify the patients with respect to the Age Group in a particular regions which defines the core of different categories.

**Strategic Focus:** Improving satisfaction among the largest demographic groups boosts overall averages, while targeted outreach to smaller groups strengthens equity and trust which improves the smaller group reach and average satisfaction score.

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

**Revenue by Each Department:**

****

This is the total revenue generated by each department from patients including the Admission Fees

**Top Revenue Generators:**

Orthopedics ($173M) and General Practice ($164M) contribute the highest revenue, accounting for the majority of earnings.

**Mid-Level Departments:**

Neurology ($73M) and Cardiology ($68M) generate moderate revenue, showing consistent patient demand.

**Low Revenue Contributors:**

Physiotherapy ($17M), Gastroenterology ($10M), and Renal ($5M) generate relatively lower revenue despite patient visits.

**Key Insight:**

Revenue is heavily concentrated in Orthopedics and General Practice; other departments show potential but require strategic growth initiatives.

**Strategic Growth Opportunities:**

For departments with lower revenue, this visualization points out areas that could benefit from enhanced services, marketing, or additional support to improve their financial performance.

**Action Point:**

Explore service expansion, resource allocation, or marketing for mid/low-revenue departments to balance overall hospital revenue streams.

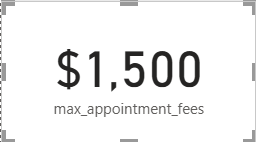
1. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.

To get the highest appointment fees being charged across all the departments.

I have used the **MAX()** function to extract the appointment fees which is highest across all departments from the Doctor Patients Table.

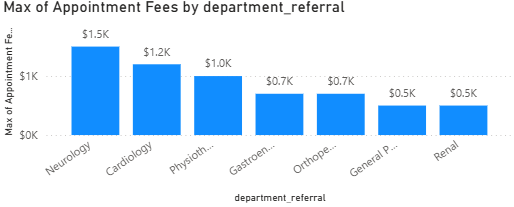


The highest admission fee is **$1500** for **Neurology**

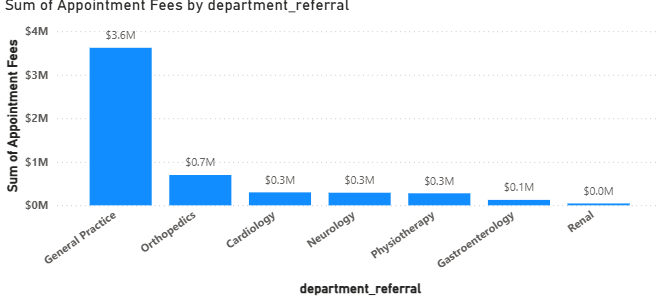


To know which department is having the highest admission fee I have used the bar chart to see which department is charging how much per consultation per patient.

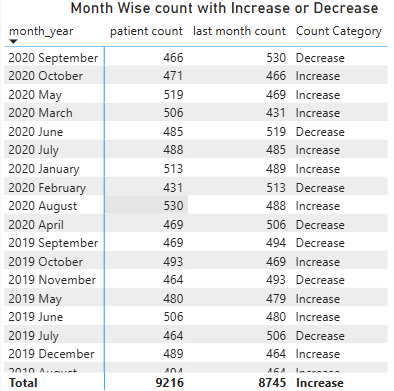
The Second highest admission fee is around $1200 for Cardiology



Though the highest admission is for Neurology the highest revenue is generated by General department which generated the revenue around $3.6M



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.



To analyse month-wise hospital visits, I created a tabular visualization that includes total visits, previous month's visits, and a comparison indicator for each month. This approach provided a detailed view of monthly trends and allowed easy tracking of changes in patient volume.

* Firstly, I have created the total patient count and assigned the year-month in Matrix and added the above created measure of patient count.
* Next, I have created the measure for the last year month patient visit count using this DAX
* last month count = CALCULATE(COUNT(Doctor\_Patients[patient\_id]), DATEADD(Dates[date], -1, MONTH))
* For category which will show increase or decrease I have used this DAX formula

**Category** = switch(true(), [Patient count] > [last month count], "Increase",

[Patient count] <[last month count], "Decrease",

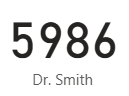
"No Change")

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 using the Calculate I have used measure which have mentioned below

Dr. Smith = CALCULATE(Doctor\_Patients[patient count], FILTER(Doctor\_Patients , Doctor\_Patients[Doctor Name] = "Dr. Smith"))





By using CALCULATE in combination with a row-level iteration function, I filtered the dataset to include only rows where **Doctor Name** equals **Dr. Smith**. Counting the patients based on filtered rows resulted in a total of **5,986 patients** who visited Dr. Smith.

This calculation offers meaningful insight into Dr. Smith’s patient load, helping analyse workload and understand how patients are distributed across doctors.

**Fact:** Dr. Smith has generated the highest revenue and treated the largest number of patients among all doctors

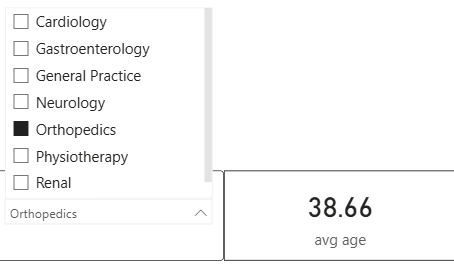
1. Calculate the average age of the patients who visit the Orthopedics department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?

**Average Age of Patients Visiting Orthopedics Department**

At First, I have calculated the overall average age of patients by creating a general measure to calculate the average age across all departments. I used the formula:



Later I have created a slicer with dropdown which consists of all departments and then when management want to check the other departments average age of the patients the card will be updated dynamically based on the department selected



This measure returned the average age as 38.66 for the Orthopedics department, this allowed me to see the Orthopedics-specific average without creating a new, separate measure.

This approach provided flexibility, as the same measure can be used to dynamically view the average age for any department based on the selected filter, rather than creating individual measures for each department.

**Insights:**

* Using a single measure with slicers simplifies the report and reduces the redundancy in calculations instead of creating multiple calculations..
* Creating and updating the slicers allows to view dynamically updates and to check the quick comparisons across departments, which makes easy to analyse average age with respect to each department.

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

The Hospital Info table had some data formatting issues, particularly in the patient satisfaction scores and date columns. The date field was stored as text, so I reformatted it as a proper date type to enable accurate time-based analysis.

For the patient satisfaction scores:

* + - 1. Date and Time Separation
      2. Missing Values in patient\_sat\_score
      3. Text Standardization
* The column which contains blank values may probably impact the calculations.
* Since it was originally stored as text, I first converted it to a numeric type.
* Then I have replaced the blanks (nulls) with the column’s average score which means patient satisfaction score to ensure data quality and avoid calculation errors.
* These steps resulted in a clean, consistent dataset, ready for accurate reporting and meaningful 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 new column in Power Query, the M language automatically generates code in the formula bar to define the transformation.

For instance, adding a custom column might produce:

= Table.AddColumn(PreviousStep, "New Column", each [Column1] + [Column2])

This command creates a new column named **"New Column"** by performing a calculation using values from the existing columns **Column1** and **Column2**.

**Understanding M-Query:**

M-Query is the powerful language behind Power Query, designed to support efficient data transformation and manipulation. Key aspects of M include:

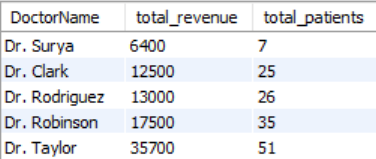
* **Data Transformation**: M allows a wide range of transformations, like filtering, grouping, and aggregating data, making it versatile for various data preparation tasks.
* **Step-by-Step Process**: Power Query records each transformation as a step, which can be viewed in the Applied Steps pane, making it easy to track and manage changes.
* **Readable Syntax:** M uses a formula-like syntax that is straightforward, enabling clear and precise adjustments in data processing.

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

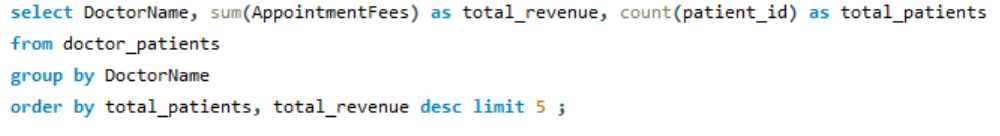
To find the top 5 doctors who has generated the most revenue but had the fewest patients were as below

To achieve this, I used an SQL query to efficiently aggregate and filter the data. The main steps included grouping by doctor, counting unique patients, and summing their total revenue. The results were then ordered by **patient count (ascending)** and **total revenue (descending)**.

This method highlights doctors who generate high revenue even with a smaller patient base.

****

Please find the below SQL query for above questionnaire:



select DoctorName, sum(AppointmentFees) as total\_revenue, count(patient\_id) as total\_patients

from doctor\_patients

group by DoctorName

order by total\_patients, total\_revenue desc limit 5 ;

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

To find the department where the average waiting time has been reduced over the 3 consecutive months

I have used SQL query to find the previous 3 months average waiting time by using **LAG()** function and the query as below.

with cte as (select department\_referral as department, date\_format(date, "%Y-%m") as year\_month\_format, avg(patient\_waittime) as average\_waiting\_time from hospital\_data

group by department\_referral, date\_format(date, "%Y-%m")),

fir\_prev as ( select department, year\_month\_format, average\_waiting\_time,

Lag(average\_waiting\_time) over(partition by department order by year\_month\_format) as first\_prev from cte),

sec\_prev as ( select department, year\_month\_format, average\_waiting\_time, first\_prev, Lag(first\_prev) over(partition by department order by year\_month\_format) as second\_prev from fir\_prev),

three\_prev as ( select department, year\_month\_format, average\_waiting\_time, first\_prev, second\_prev,

Lag(second\_prev) over(partition by department order by year\_month\_format) as third\_prev from sec\_prev)

select \* from three\_prev where first\_prev > second\_prev and second\_prev>third\_prev order by year\_month\_format;

**Monthly Average Waiting Time Calculation:**  
The process began by computing the average patient waiting time for each department on a monthly basis. This was done in the monthly\_avg\_wait CTE, which aggregated patient\_waittime values grouped by department and the month extracted from the date field.

**Comparison Across Consecutive Months:**  
Next, for a previous comparison CTE was created to evaluate waiting times over three consecutive months for each department. The LAG() function was used to retrieve the average waiting times of the preceding two months, allowing for direct month-over-month comparisons directly.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| department | year\_month\_format | average\_waiting\_time | first\_prev | second\_prev | third\_prev |
| Cardiology | 2019-07 | 46.3333 | 36.25 | 34.5833 | 30.1667 |
| Neurology | 2019-07 | 42.2222 | 42.8667 | 38.9091 | 35.7273 |
| Cardiology | 2019-08 | 36.2941 | 46.3333 | 36.25 | 34.5833 |
| Orthopedics | 2019-08 | 34.7551 | 36.537 | 34 | 30.9143 |
| Cardiology | 2019-12 | 32.7222 | 41.8462 | 30 | 27.3846 |
| Neurology | 2020-02 | 41 | 37.6667 | 37.6364 | 31.5 |
| General Practice | 2020-03 | 35.7871 | 36.4575 | 35.9204 | 34.4794 |
| Neurology | 2020-03 | 39.2 | 41 | 37.6667 | 37.6364 |
| Orthopedics | 2020-05 | 36.3725 | 36.6364 | 36.2881 | 35.4348 |
| Physiotherapy | 2020-06 | 35.8182 | 37.6 | 37.4667 | 34.3571 |
| Neurology | 2020-07 | 34.7143 | 35.1667 | 35.125 | 31.4 |
| Neurology | 2020-10 | 38 | 44.9 | 34.9286 | 34.7143 |
| Physiotherapy | 2020-10 | 30.4118 | 39.3333 | 36.7778 | 29.0833 |

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

The below SQL query calculates the ratio of male to female patients for each doctor and ranks them accordingly, I have used the dense rank to not to skip the rankings in case of any ties

with gender\_count as (select DoctorName,   
sum(case when patient\_gender = "M" then 1 else 0 end) as male\_count,

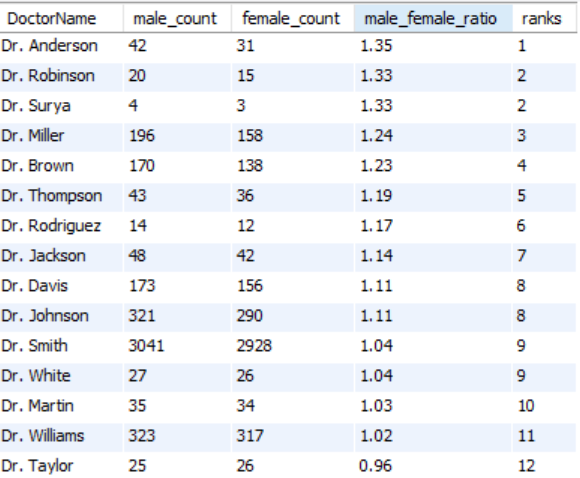
sum(case when patient\_gender = "F" then 1 else 0 end) as female\_count   
from doctor\_patients d inner join hospital\_data h on h.patient\_id = d.patient\_id

group by DoctorName)

select DoctorName, male\_count, female\_count, round(male\_count/female\_count,2) as male\_female\_ratio,

dense\_rank() over(order by round(male\_count/female\_count,2) desc) as ranks

from gender\_count order by ranks;



**Insights:**

**Gender Distribution:** The analysis provides visibility into the patient demographics served by each doctor, highlighting potential imbalances or gaps.

**Resource Allocation:** These ratios help hospital management optimize resource distribution to ensure equitable care across all patient groups.

**Tailored Healthcare Services:** The findings support the design of targeted programs and outreach initiatives to better meet the needs of both male and female patients.

**Data-Driven Strategy:** By leveraging these insights, hospital leaders can adopt inclusive strategies that enhance patient care and overall service quality.

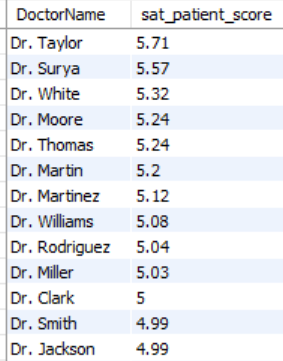
1. Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)

select DoctorName, round(avg(case when patient\_sat\_score = "" then 5 else patient\_sat\_score end ),2) as sat\_patient\_score

from hospital\_data h inner join doctor\_patients d on

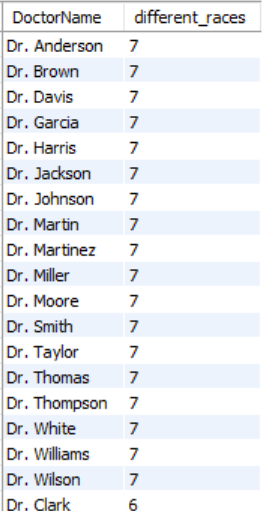
d.patient\_id = h.patient\_id group by DoctorName

order by sat\_patient\_score desc;

****

This SQL query computes the average patient satisfaction score for each doctor. Missing scores are replaced with a default value of 5. It joins patient satisfaction records from the hospital table with doctor details from the doctor\_patients table, groups the results by doctor, and sorts them in descending order so that doctors with the highest satisfaction appear first.

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

****

The query identifies each doctor and counts the number of distinct races among their patients. This count, shown as “different\_races,” represents the diversity of the doctor’s patient base. A higher number indicates that the doctor has treated patients from a wider variety of racial backgrounds.

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

select d.department\_referral as department,

sum(case when patient\_gender = "M" then TotalBill end) as male\_total\_bill,

sum(case when patient\_gender = "F" then TotalBill end) as female\_total\_bill,

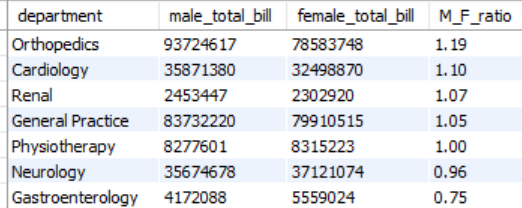
round((sum(case when patient\_gender = "M" then TotalBill end) / sum(case when patient\_gender = "F" then TotalBill end)),2) as M\_F\_ratio

from hospital\_data h inner join doctor\_patients d on

d.patient\_id = h.patient\_id

group by d.department\_referral

order by M\_F\_ratio desc;



The query calculates the total medical bills paid by male and female patients for each department. It then computes the ratio of male-to-female billing (M\_F\_ratio) to measure gender-based revenue contribution. The results are grouped by department and ordered in descending ratio, showing which departments generate a higher share of bills from male patients compared to female patients.

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)

update hospital\_data

set patient\_sat\_score = least(patient\_sat\_score+2,10)

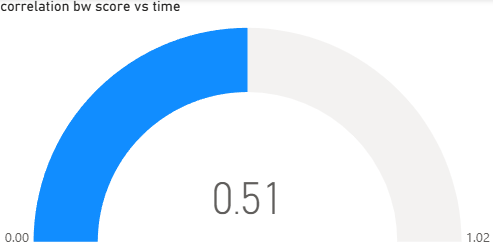
where department\_referral = 'General Practice'

and patient\_waittime>30;

The UPDATE query adjusts patient satisfaction scores for individuals who visited the General Practice department and experienced waiting times longer than 30 minutes. Specifically, it increases their satisfaction score by 2 points, with a safeguard to ensure the score never exceeds the maximum threshold of 10. This logic is designed to simulate the effect of a post-visit intervention or service recovery effort, where patients who faced longer waits are given additional consideration to reflect an improvement in their overall experience. By applying this rule, the dataset models a more realistic scenario in which hospitals actively address patient dissatisfaction and attempt to boost patient perceptions through targeted interventions.

**Subjective Questions**

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



The Interpretation of Correlation = 0.51

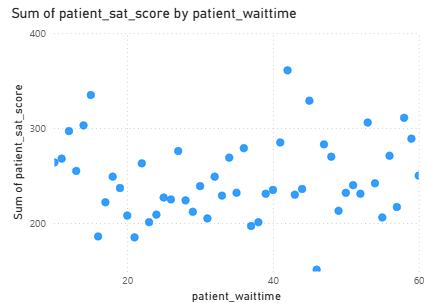
* A correlation of 0.51 indicates a moderate positive relationship between the two variables whereas if wait time increase the satisfaction score will be coming down as no patient want to wait too much and it will impact it in the satisfaction score and gradually may impact the revenue as they will be checking for the alternate hospital for the treatments.
* I used **Quick Measure** in Power BI, which helps create common calculations like averages or correlations without writing DAX manually.
* I applied the **Correlation Coefficient** option to see how patient wait time

(Y-axis) relates to satisfaction scores (X-axis), grouped by department referral.

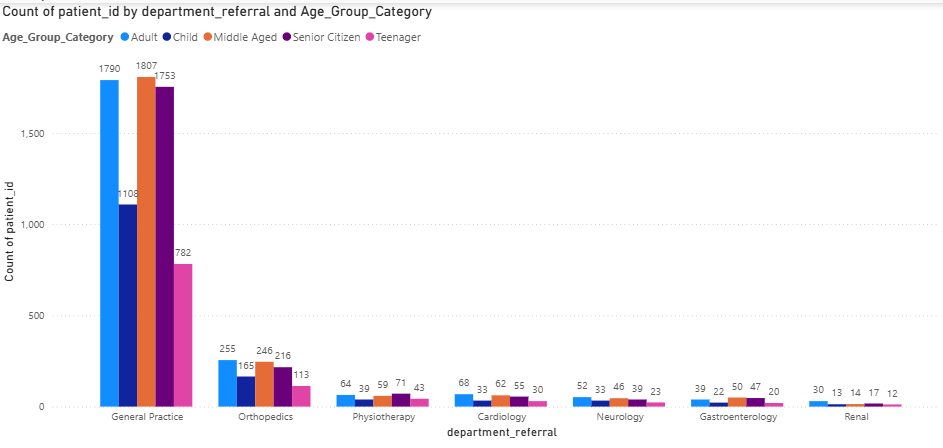
* Power BI calculated the result as **0.51**, showing a **moderate positive relationship** between wait time and satisfaction.
* Since we expect wait time to reduce satisfaction, this positive correlation suggests the data might not be perfectly linear (e.g., satisfaction may sometimes rise slightly with wait time in certain cases, or other factors are influencing the trend).
* In practical terms:
  + 0.0 → No relationship
  + 0.1 – 0.3 → Weak relationship
  + 0.3 – 0.7 → Moderate relationship
  + 0.7 – 1.0 → Strong relationship

Thus, 0.51 means there is a noticeable, moderate connection, but it’s not strong enough to be the only factor affecting satisfaction. Other factors (like staff behaviour, treatment quality, or communication) also the significant roles.

As we aim to reduce the wait time make sure the pre-check-ups are being done for the patients in the wait time and can be avoided during the actual consultation which may reduce the time duration for actual appointment which will reduce the patient wait time indirectly as consultation can be done faster than expected.



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



**Observations and Insights**

* General Practice leads across all age groups, with the highest visits from Senior Citizens (2,645) and Young Adults (1,373).
* Middle Age (1,418) and Children (1,103) also show strong reliance on General Practice.
* Teenagers contribute fewer visits overall but still account for 696 visits to General Practice.
* **Specialty departments record lower volumes**:
  + Orthopedics attracts Middle Age (165), Senior Citizens (344), and Children (107).
  + Physiotherapy is mostly used by Senior Citizens (99) and Middle Age (48).
  + Cardiology demand is concentrated among Senior Citizens (95) and Young Adults (52).
  + Neurology is more frequent among Senior Citizens (67) and Children (33).
  + Gastroenterology and Renal visits remain low across all groups, though Senior Citizens still account for the highest usage.

**Conclusion**

* Senior Citizens represent the highest demand across departments, particularly for General Practice and Cardiology.
* Children primarily rely on General Practice and Neurology, reflecting pediatric and neurological needs.
* Young Adults and Middle Age groups focus on General Practice, with moderate use of Orthopedics and Physiotherapy.
* Teenagers show the lowest overall visits, mainly to General Practice.
* Specialty services like Renal and Gastroenterology remain underutilized across all demographics.

**Key Takeaway**

Patient age groups play a major significant role in shaping the healthcare service demand as age will impact the health.

General Practice serves as the common entry point for the consultation before going to any other department, while older patients rely more heavily on specialty departments compared to younger groups majorly because of their age.

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



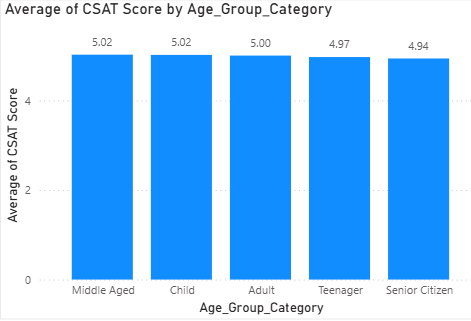
**Insights and Observations**

* Slow Start: Patient visits begin at a lower level in January (513) and hit the lowest point in February (431).
* Steady Growth: Numbers rise consistently from March (506) to April (948), continuing upward through the summer and early fall.
* Peak Season: The highest activity occurs in August (1,024), with strong volumes maintained from April to October (roughly 950–1,024).
* Minor Fluctuations: July (952) and September (935) show slight dips, though overall volumes remain high.
* Year-End Decline: After October (964), visits drop sharply — November falls to 464 and December stays low at 489.

**Conclusion**

* Patient volumes follow a clear seasonal pattern: lowest in the beginning and end of the year, and highest during spring through early fall.
* The busy period spans April to October, likely influenced by seasonal conditions, health awareness campaigns, or planned checkups.
* The sharp decline in November–December may be linked to holidays, adverse weather, or patients deferring non-urgent care.

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

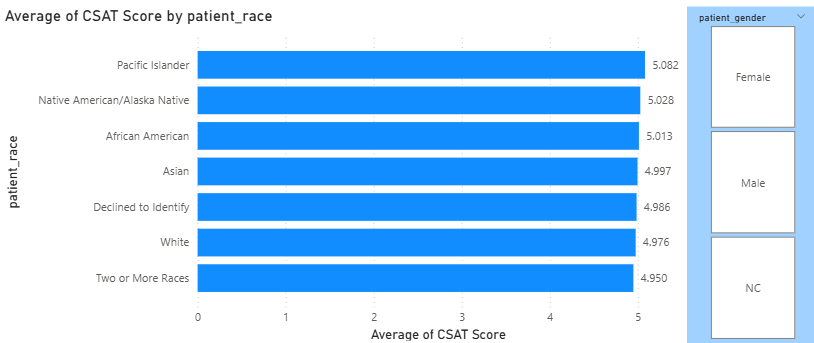


**Insight Summary: Customer Satisfaction by Age Group**

* Overall Trend: Satisfaction scores are stable across all age groups, ranging narrowly from 4.94 to 5.02.
* Top Performing Segments:
  + Middle-Aged and Child groups report the highest satisfaction (5.02).
* Areas of Concern:
  + Senior Citizens (4.94) and Teenagers (4.97) reflect slightly lower satisfaction levels compared to other groups.
* Interpretation:
  + Middle-aged and child customers appear highly engaged and satisfied.
  + Senior citizens and teenagers may face unmet expectations or experience gaps, signaling opportunities for service improvement or tailored communication.

Recommendation: Focus on enhancing experiences for teenagers and senior citizens through personalized campaigns, improved accessibility, and targeted service enhancements to balance satisfaction across all demographics.

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?



**Identifying the Racial or Gender-Based Discrimination in the Hospital**

1. Approach to Verification
   * To check if the claim of discrimination is valid, we examine CSAT (Customer Satisfaction) scores across race and gender groups.
   * If there is a large gap in satisfaction scores between races or genders, it may indicate bias or unequal treatment.
   * If scores are consistent across groups, it suggests no evidence of systemic discrimination.
2. Insights from the Chart
   * Race-based comparison:
     + Pacific Islander (5.08) and Native American/Alaska Native (5.03) report slightly higher satisfaction.
     + Two or More Races (4.95) and White (4.97) report slightly lower satisfaction.
     + The difference across all races is minimal (0.13 points), indicating no major disparity.
   * Gender-based comparison (via slicer):
     + Filtering by Male, Female, or NC shows scores remain stable, with no evidence of systematic discrimination by gender.
3. Conclusion
   * The data does not support claims of racial or gender discrimination, since the variation in satisfaction scores is very small.
   * Continuous monitoring is recommended to ensure fairness and maintain equity in patient care.

Recommendation: While no strong evidence of discrimination is visible, the hospital should maintain regular equity audits, encourage feedback across demographics, and ensure transparent reporting to build trust.

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

To make the discount allocation more practical, we need to define a minimum spending threshold. Only when a patient’s total expenses cross that threshold should they become eligible for discounts.

* The overall patient spend is approximately 509.3M.
* The average spend per patient is around 55.26K.

However, using the overall average spend as the benchmark for discount eligibility is not feasible, since not every patient spends that much.

To address this, I considered two feasible checkpoints:

* Normal Discount: Set at 20% of the average spend (~15K). This creates a reasonable entry point for patients to qualify.
* High Discount: Set at 90% of the average spend (~50K). This ensures only significantly high spenders qualify.

The DAX logic used is:

**Discount eligibility** = SWITCH(TRUE(),

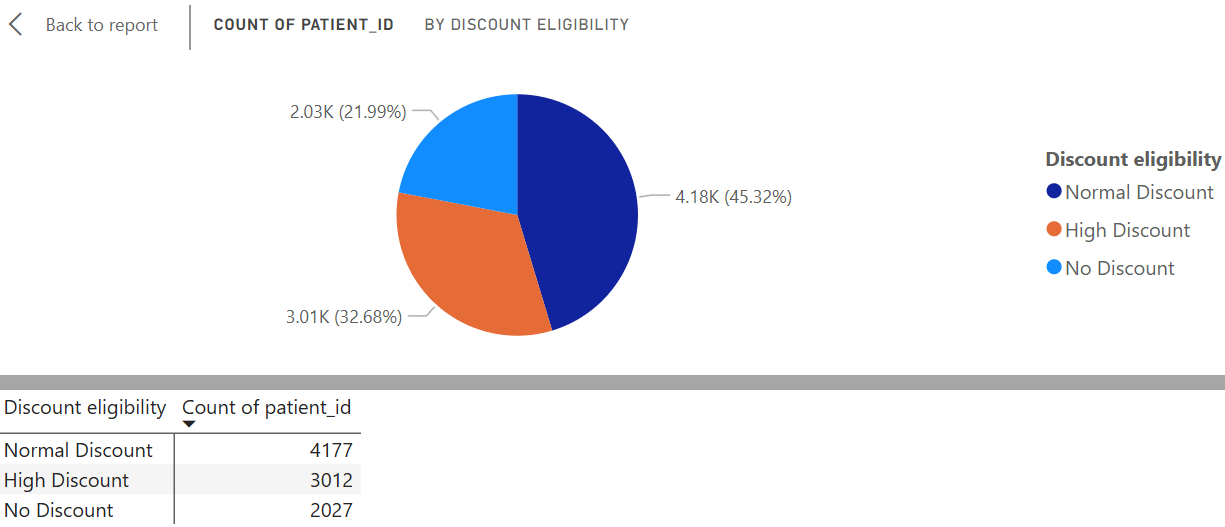
Doctor\_Patients[Total Bill] >=35000, "High Discount",

Doctor\_Patients[Total Bill] >= 15000, "Normal Discount",

"No Discount")

**Insights:**

* Around 22% of patients qualify for the High Discount, as their total bill exceeds 35K.
* About 45% of patients fall into the Normal Discount category, with total bills between 15K and 35K.



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

For this Firstly we need to sure to review some of the important factors such as Satisfaction Rate, the greater Revenue generation department, the greatest number of patients for the which department, the greater wait time for the department.

I have created below charts to compare each department’s patient visits, average bill per visit, total number of patients per each department, number of doctors per each department and average wait time, helping guide staffing decisions.

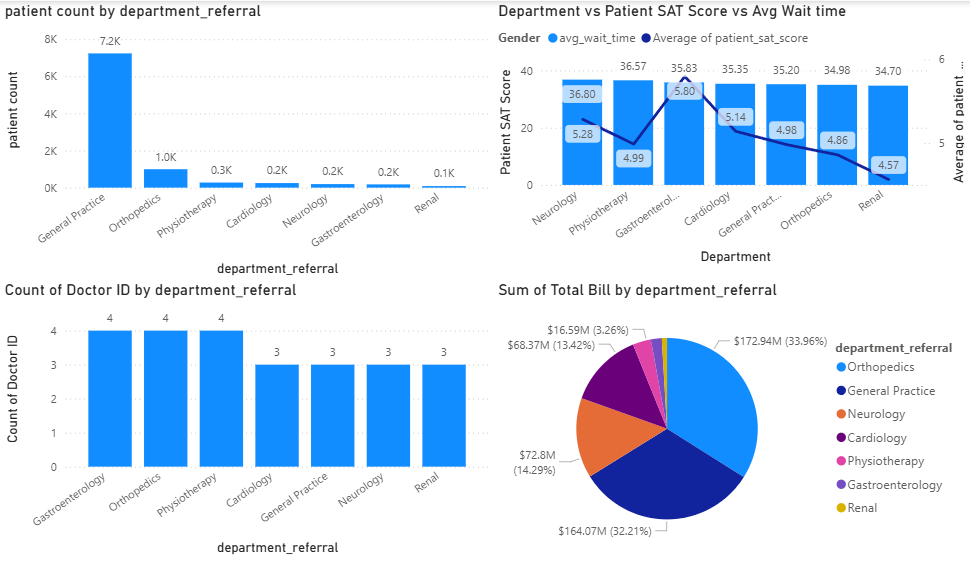
* Average Bill: Shows revenue contribution per patient.
* Total Visits: Highlights departments with high demand.
* Doctors Per department: displays the count of doctors assigned to each department
* Wait Time: Identifies service bottlenecks.

**Insights:**

* General Practice & Neurology: High patient demand.
* Neurology & Cardiology: Higher revenue generation.
* Neurology & Orthopedics: Longer wait times.

**Recommendation:**

* Neurology & Cardiology: Higher revenue generatiAs Renal department is having less patients and revenue is less, we can shift the doctors in their free time to check on the General Practice as it may help the in reducing the wait time for the patients
* Hire additional doctors in General Practice, Neurology, and Orthopedics to handle patient volume, reduce wait times, and support revenue growth.

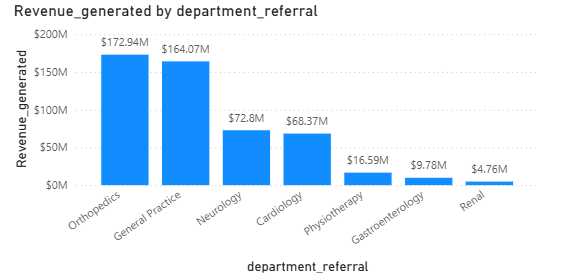


1. Is the hospital profitable? How will you determine the profitability?

Yeah, the hospital is in profitable state.

To determine hospital profitability, we need to compare **revenues vs. expenses**:

1. **Revenue Analysis**
   * Patient billing (admission, total overall).
   * Average revenue per patient visit.
   * Department-wise contribution (which specialties generate most revenue). ~ From Orthopedics department the greater revenue was generated

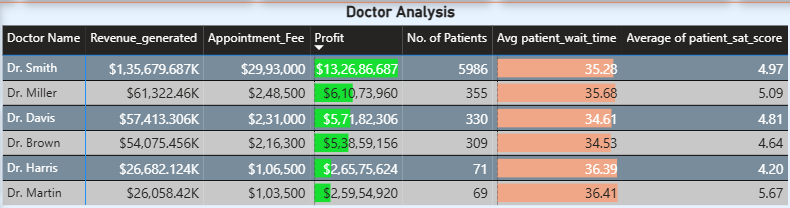


1. **Cost Analysis**

As we don’t have the details of salaries for the below, we cannot determine the overall expenses of the cost analysis

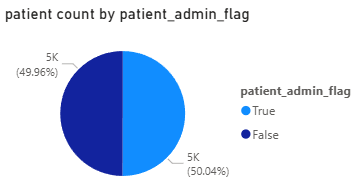
* + Fixed costs (salaries of doctors, nurses, admin staff, rent, equipment depreciation).
  + Variable costs (medicines, consumables, utilities, operational supplies).

1. **Profitability Metrics**
   * Net Profit = Total Revenue – Total Expenses
   * Profit Margin % = (Net Profit ÷ Revenue) × 100

****

1. **Other Indicators**
   * High patient retention & revisit rates = stable revenue.

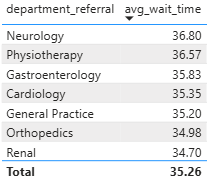
To check the patient retention & revisit I have taken the consideration of patient admit flag which is marked as yes (retention or revisit patients)



**Conclusion**:  
The hospital is profitable if **total revenues consistently exceed total costs** with a healthy profit margin, while also maintaining quality care and patient satisfaction. Undoubtedly, we can say that hospital is in profitable state.

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

Well, there is no much difference in waiting time for the departments

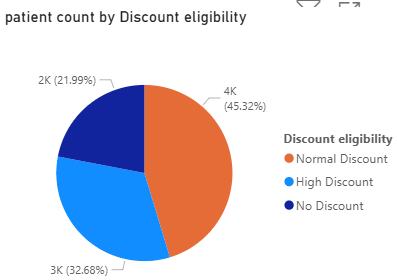


* Neurology Has the greater Wait Time: With an average of 36.8 minutes, Neurology records the highest wait time, highlighting significant patient delays.
* Noticeable Decline Across Departments: Wait times gradually decrease from Neurology down to Renal, reflecting differences in demand or operational efficiency. (depending on the patient illness and based on department it will vary)
* Renal Department is Most Efficient: At 34.7 minutes, Renal shows the lowest wait time, as lighter patient load which indicates the lesser wait time on average.
* Specialized Departments Face Longer Delays: Neurology, Physiotherapy, and Gastroenterology show the highest wait times, likely due to greater demand for their specialized care and the illness as it is time consuming.

**Conclusion:**

Neurology stands out with particularly long wait times, signalling a possible bottleneck or high patient demand that warrants closer review.

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



To strengthen patient loyalty, encourage repeat visits, and optimize revenue, a structured discount strategy can be implemented based on patient spending and engagement patterns rather than region-wise or average bill–based discounts.

Discount eligibility = SWITCH(TRUE(),

Doctor\_Patients[Total Bill] >=35000, "High Discount",

Doctor\_Patients[Total Bill] >= 15000, "Normal Discount",

"No Discount")

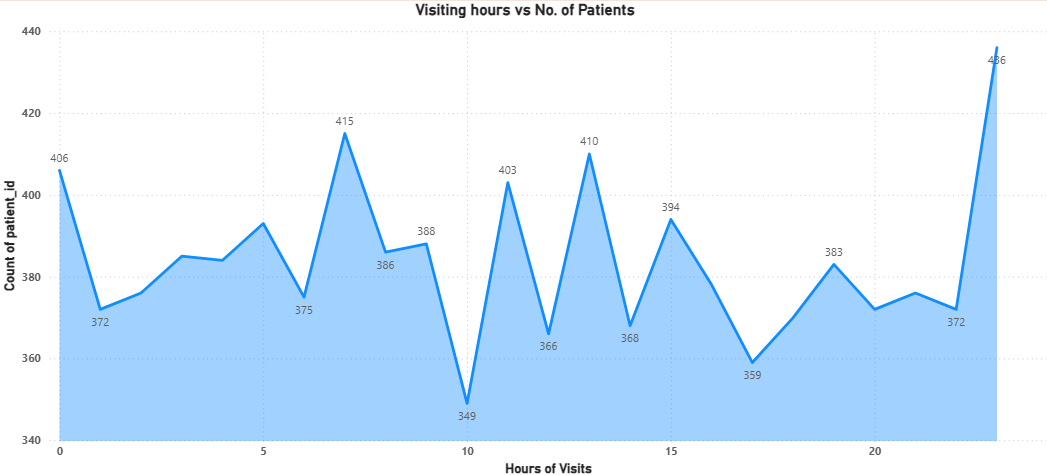
**Proposed Discount Strategy:**

* Normal Discount (7%): Patients with a total bill greater than $15,000 are eligible for a 7% discount on their total bill.
* High Discount (10%): Patients spending over $35,000 qualify for a 10% discount, rewarding high-value patients.
* Revisiting Patients: Individuals who return within 3 months can avail the same discount tier again, encouraging patient retention and boosting repeat visits.
* Annual Membership Patients: Patients opting for the hospital’s annual membership program receive complimentary annual health checkups and extended discounts for their family members, promoting holistic healthcare engagement.

**Recommendations/ Suggestions:**

* Introduce seasonal wellness packages (e.g., winter immunity or summer hydration checkups) with bundled discounts to attract patients during off-peak periods.
* Offer loyalty points or cashback on bills, redeemable for future consultations or pharmacy purchases.
* Provide special discounts for senior citizens and long-term patients to enhance trust and goodwill.

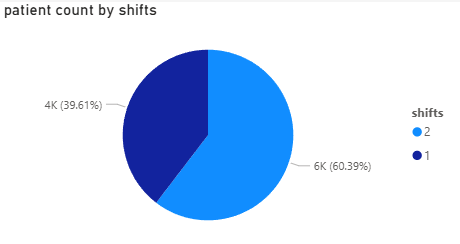
1. Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also will this 2 shift policy be helpful for the hospital?

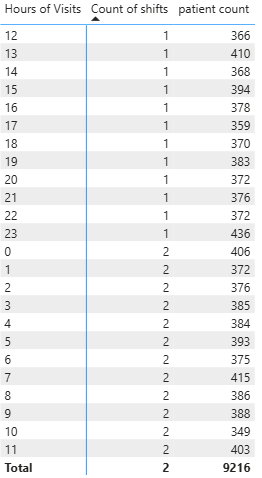


1st-shifts = CALCULATE([patient count], 'Hospital ER'[Hours of Visits]>=0 && 'Hospital ER'[Hours of Visits]<12 && 'Hospital ER'[department\_referral] = "General Practice")

**This is marked as Shift 1**

**And for Shift 2 :**shifts = if(ISBLANK('Hospital ER'[1st-shifts]), 2, 'Hospital ER'[1st-shifts])





**Step 1: Identify the Shift Timings**

To define two effective shifts for the General Practice department, you need to analyse both patient flow and doctor availability.

* Patient Data Analysis:
  + Look at hospital records (OPD visits, walk-ins, appointments) by hour of the day.
  + Identify peak hours (e.g., 9 AM–1 PM and 4 PM–8 PM).
* Shift Design Based on Demand:
  + Morning Shift: 8 AM – 2 PM (covers early OPD rush, routine checkups).
  + Evening Shift: 2 PM – 8 PM (caters to working professionals, school children, post-office visits).

(If hospital operates late, an 8 AM–2 PM and 2 PM–10 PM schedule may be even more effective.)

**Step 2: Divide Doctors into Shifts**

* Patient-to-Doctor Ratio:  
  Divide doctors proportionally based on expected patient volume per shift.  
  Example: If 60% of visits are in the morning and 40% in the evening, allocate doctors in a 60:40 ratio.
* Fair Rotation:  
  Implement a weekly or bi-weekly rotation system so doctors get balanced exposure across shifts and avoid burnout.
* Special Cases:  
  Senior doctors or specialists can be scheduled during peak hours, while junior doctors can cover lighter hours with supervision.

**Step 3: Benefits of a Two-Shift Policy**

Operational Efficiency: Continuous coverage ensures patients can get consultations at flexible times.  
Reduced Wait Times: Shorter queues improve patient satisfaction (ties back to your earlier wait-time analysis).  
Doctor Well-being: Avoids long 10–12 hour stretches, improving productivity and morale.  
Revenue Growth: Extended consultation hours attract working patients who cannot visit during standard timings.  
Patient Retention: Flexibility and shorter wait times encourage repeat visits and loyalty.

**Step 4: Monitoring & Adjustments**

* Track shift-wise patient inflow, waiting times, and revenue.
* Adjust timings if patient demand shifts (e.g., increase evening hours if demand rises).
* Use Power BI dashboards to visualize doctor performance and patient satisfaction per shift.

**Conclusion:**  
Yes, a two-shift policy for General Practice doctors is highly beneficial. It balances the workload, reduces wait times, aligns with patient availability, and improves overall hospital efficiency while enhancing both patient and doctor satisfaction.

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

**Power BI Gateway – What It Is**

A Power BI Gateway is a bridge that connects on-premises data sources (like SQL Server, Oracle, SAP, Excel files on a local network) to Microsoft Power BI Service (cloud).

Without the gateway, Power BI Service cannot directly access data stored on local servers or private networks.

It acts like a secure tunnel between on-premises data and cloud-based Power BI reports/dashboards.

**Types of Power BI Gateways**

1. Personal Mode
   * Tied to a single user.
   * Only supports Import mode (scheduled refresh).
   * Best for personal use or testing.
2. Standard (Enterprise) Mode
   * Can be used by multiple users.
   * Supports both Import and Direct Query/Live Connection.
   * Best for enterprise scenarios where multiple reports/datasets connect to on-premises systems.

**Use Cases of Power BI Gateway**

1. Scheduled Data Refresh
   * Keep Power BI reports up to date with on-prem data (e.g., nightly refresh from SQL Server).
2. Real-Time / Direct Query
   * Run live queries against on-premises databases without storing data in Power BI Service.
3. Hybrid Data Scenarios
   * Combine cloud data (Azure, Salesforce) with on-premises data (SQL, ERP) in a single dashboard.
4. Enterprise-Wide Access
   * Multiple users in an organization can securely access reports based on the same on-premises data source.
5. Governance & Security
   * Ensures data never leaves your organization without encryption; admins can manage access centrally.
6. How would you approach this problem, if the objective and subjective questions weren't given?

If the objective and subjective questions were not provided, I would take an exploratory, data-driven approach to uncover meaningful patterns, relationships, and actionable insights that align with potential business goals or hospital management needs. My process would include:

**1. Understanding the Dataset**

* Initial Review: Examine available tables (doctor\_patients\_data, hospital\_er), their columns, and data types.
* Data Profiling: Check for null values, unique counts, and distributions across key fields (e.g., patient\_gender, patient\_age, doctor\_name, department\_referral). This ensures awareness of data quality and completeness.

**2. Identifying Potential Areas of Analysis**

* Patient Demographics: Analyze age, gender, and visit patterns to understand service demand and satisfaction.
* Departmental Analysis: Evaluate patient volume, visit frequency, and revenue by department to identify strong and weak areas.
* Doctor Performance: Measure satisfaction scores, patient loads, and revenue generated per doctor.
* Temporal Trends: Study variations across days, weeks, and months to detect seasonal or time-based patterns.

**3. Developing Key Metrics & KPIs**

* Wait Time vs. Satisfaction: Test correlations between patient wait times and satisfaction scores.
* Revenue Analysis: Track revenue per department and per doctor to gauge financial performance.
* Diversity Metrics: Assess demographic diversity within each doctor’s patient base.
* Satisfaction Trends: Compare department-level and doctor-level satisfaction scores to highlight top and low performers.

**4. Visualization Strategy (Power BI)**

* Line Charts: To show satisfaction and visit trends over time.
* Bar/Column Charts: To compare department visits, demographics, and revenue.
* Scatter Plots: To highlight correlations (e.g., wait time vs. satisfaction).
* KPI Cards & Gauges: For at-a-glance metrics like revenue per department, average satisfaction, or patient volume.

**5. Insights & Recommendations**

* If long wait times correlate with poor satisfaction, recommend staff reallocation or improved scheduling.
* If certain doctors or departments underperform, propose additional training or resources.
* Use patient demographic insights to design personalized services.
* Leverage seasonal/temporal analysis for resource planning and workload balancing.

**Summary:** My approach would combine data exploration, KPI development, and Power BI storytelling to transform raw data into actionable hospital management insights.

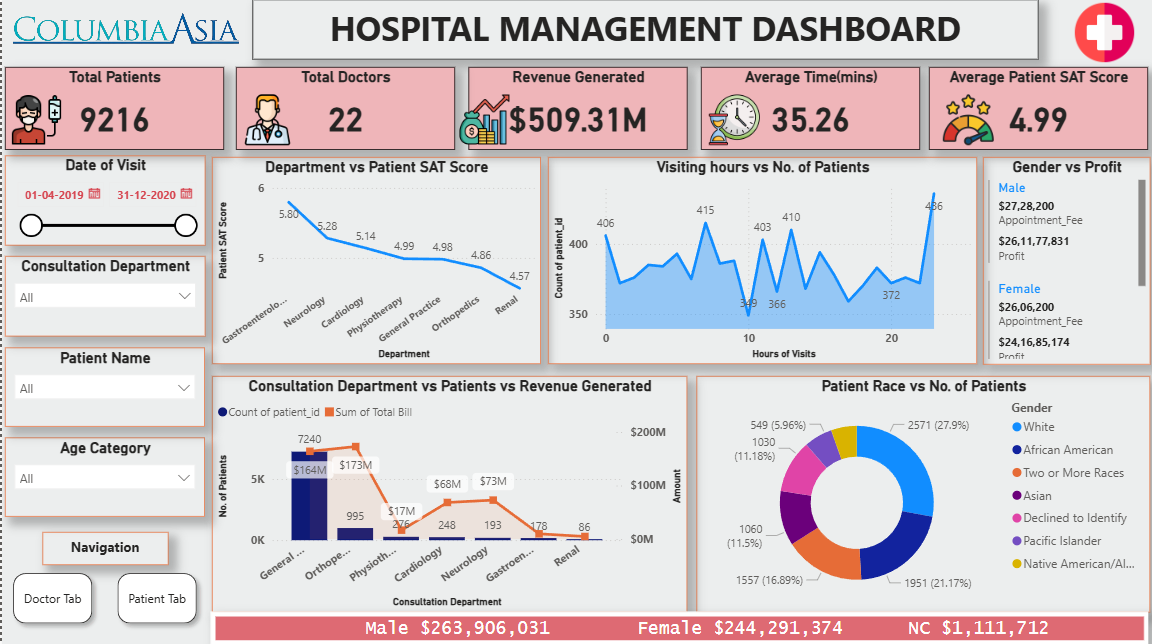
1. Can you analyse and write the type of relationship between the doctor id and department, is it one-to-one?

* Every doctor ID is connected to only one department.
* But a department can have many doctors working in it.
* That makes the relationship one-to-many: one department, many doctors.
* Or if you look at it the other way, it’s many-to-one: many doctors belong to one department.
* So, it’s not one-to-one, it’s a one-to-many relationship between doctor and department.

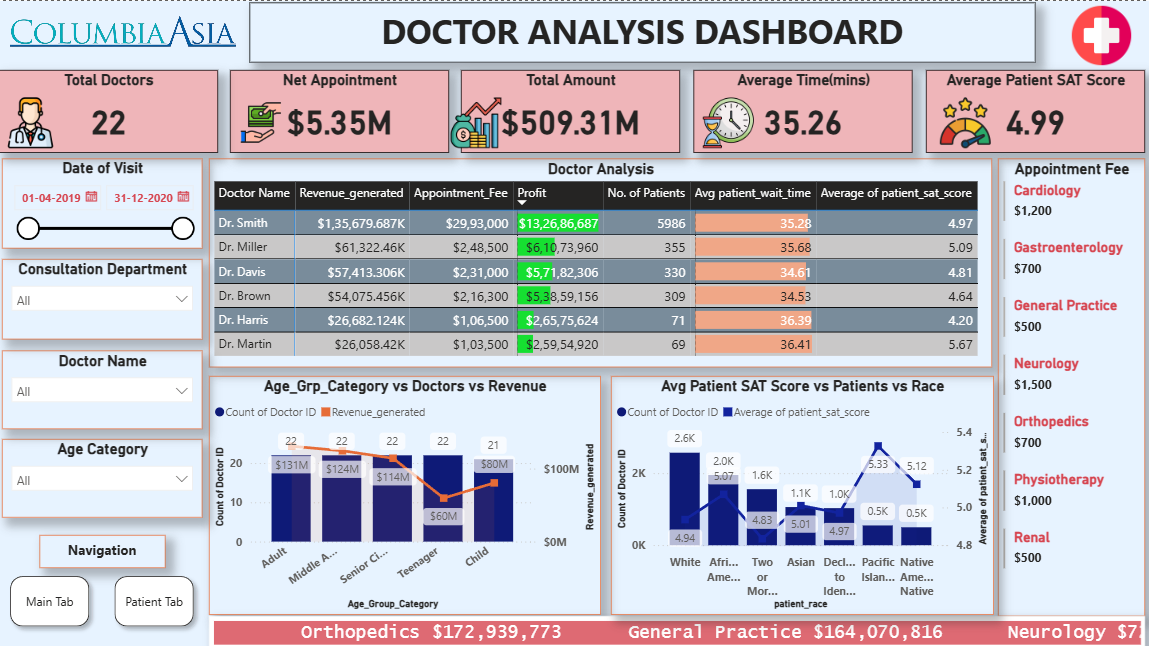
For these below dashboards I have used bookmark option which is at

top right  corner to reset the page filters in all the report pages

1. Hospital Management Dashboard



2. Doctor Analysis Dashboard



3. Patient Analysis Dashboard

