**Columbia Asia Hospital Project**

* **Himanshu Goswami**

**Objective questions**

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

To ensure data cleanliness and accuracy, specifically by addressing missing values in the patient\_sat\_score column.

1. Methodology:

The Power Query Editor was used to identify all null or blank entries. A data imputation strategy was chosen to replace these missing values with the overall average patient satisfaction score, thus preserving the dataset's integrity without deleting rows.

1. Finding:

A new, complete column patient\_new\_sat\_score was created, making the data reliable for further analysis.

1. DAX Formula Used:

patient\_new\_sat\_score =

IF(

ISBLANK('Hospital Info'[patient\_sat\_score]),

AVERAGE('Hospital Info'[patient\_sat\_score]),

'Hospital Info'[patient\_sat\_score]

)

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

To determine the average time patients, wait for service.

1. Methodology:

A DAX measure was created to calculate the average of the patient\_waittime column.

1. Finding:

The average patient waiting time is 35.26 minutes.

A number on a white background

AI-generated content may be incorrect.

1. DAX Formula Used:

Average\_waiting\_time= AVERAGE('HospitalInfo'[patient\_waittime])

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

To visualize the total number of patients referred to each hospital department.

1. Methodology:

A bar chart was created. The department\_referrals column was used for the X-axis (categories). A DAX measure was created to count the number of unique patient IDs for the Y-axis (values).

1. Finding:

The chart displays the total patient volume for each department, showing which departments are busiest.

A graph of a patient referral

AI-generated content may be incorrect.

1. DAX Formula Used:

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

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

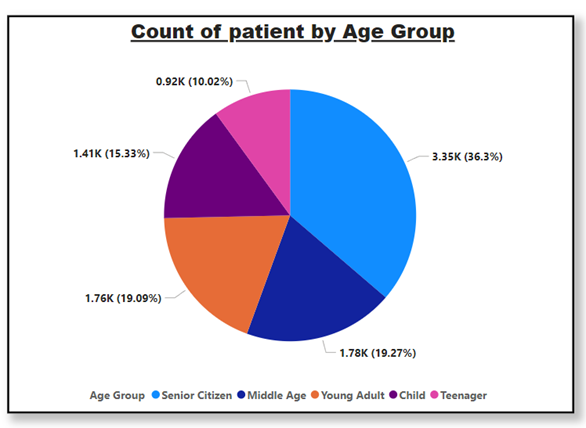
To categorize patient visits into distinct age groups to identify which demographics utilize healthcare services most frequently.

1. Methodology:

A new calculated column Age Group was created to segment patients based on their age into five categories: Child (<=12), Teenager (13-20), Young Adult (21-35), Middle Age (36-50), and Senior Citizen (>50).

1. Finding:

Senior Citizens are the largest demographic, accounting for 36.3% (3.35K) of all patient visits. The other groups are Middle Age (19.27%), Young Adult (19.09%), Child (15.33%), and Teenager (10.02%).



1. 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"

)

)

)

)

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?**
2. Objective:

To identify and remediate missing (null) values in the dataset to ensure data accuracy and prepare it for analysis.

1. Methodology:

The dataset was first inspected using the Power Query Editor to find any null or inconsistent values. Missing data was identified in the patient\_sat\_score column. The chosen approach was mean imputation, where each null value was replaced by the average (mean) score of the entire column. The final result was then rounded to one decimal place.

1. Finding:

This cleaning process resulted in a new, complete column (patient\_new\_sat\_score) with no missing data. This ensures the dataset is accurate and reliable for any analysis involving patient satisfaction.

1. DAX Formula Used:

patient\_new\_sat\_score =

ROUND(

IF(

ISBLANK('Hospital Info'[patient\_sat\_score]),

AVERAGE('Hospital Info'[patient\_sat\_score]),

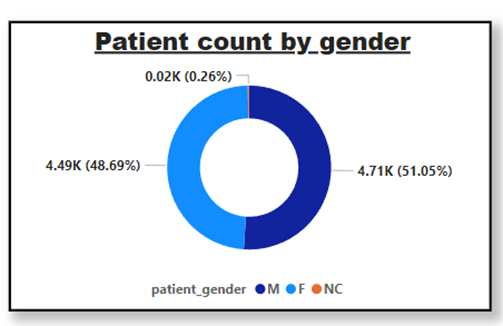
'Hospital Info'[patient\_sat\_score]

),

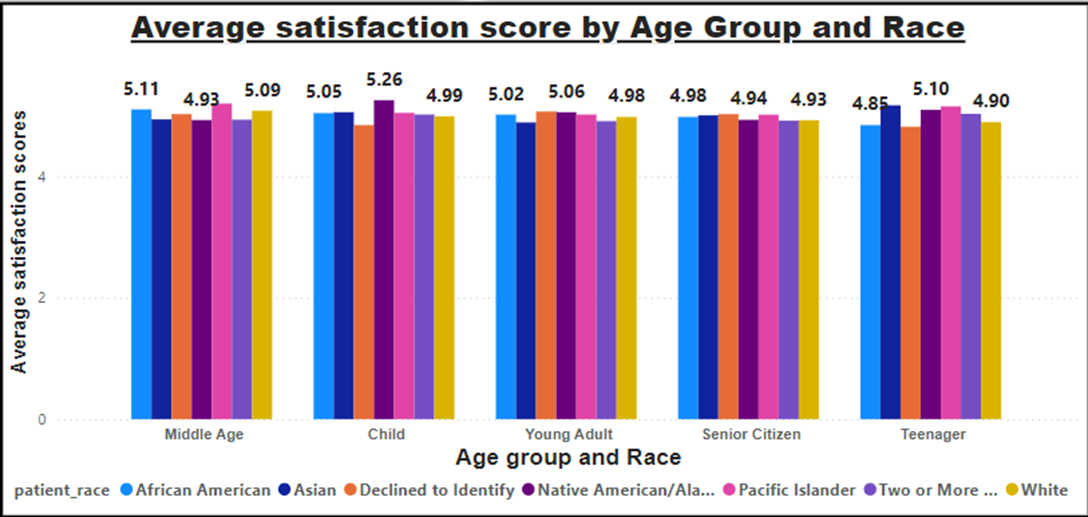
1

)

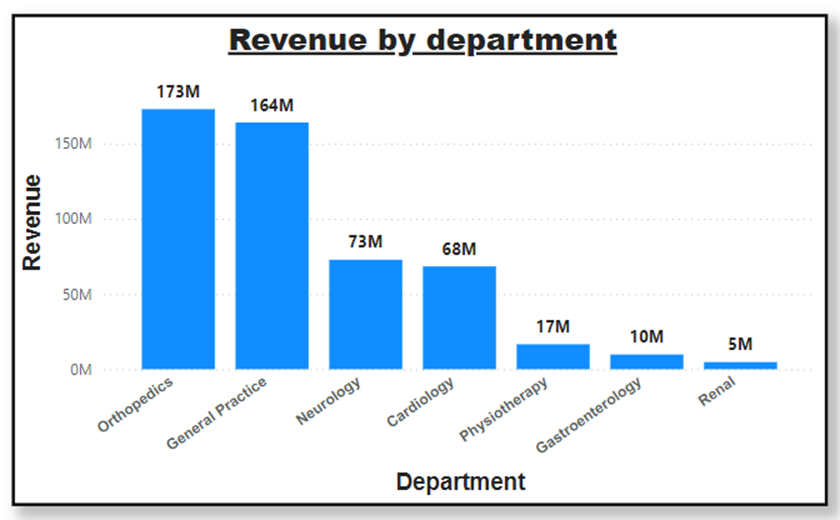
1. **Is there any relation between the number of visits and Gender of the patients.**
2. Objective: To analyse the relationship between patient gender and the total number of hospital visits.
3. Methodology: A donut chart was created to visualize the distribution of patient visits. The patient\_gender column was used to create the segments, and a count of patients was used to determine the value for each segment.
4. Finding: The analysis indicates that the number of visits is almost evenly distributed between males and females, with only a slightly higher proportion of male patients.
   1. Male (M): 4.71K visits (51.05%)
   2. Female (F): 4.49K visits (48.69%)
   3. NC: 0.02K visits (0.26%)



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.**
2. Objective: To determine the relationship between patient satisfaction scores, their age groups, and their racial backgrounds to pinpoint any areas for improvement in patient experience.
3. Methodology: A clustered bar chart titled "Average Satisfaction Score by Age Group and Race" was used for this analysis.
   1. X-Axis: Age Group (Middle Age, Child, Young Adult, Senior Citizen, Teenager)
   2. Y-Axis: Average Satisfaction Score (on a scale of 0 to 6)
   3. Legend (Categories): Race (African American, Asian, Declined to Identify, etc.) This visualization allows for a direct comparison of average satisfaction scores across different racial backgrounds within each age group.
4. Finding: The analysis shows that patient satisfaction scores are broadly consistent across all age and racial demographics, with minimal variation. All groups report an average satisfaction score clustered tightly around 4.9 to 5.1.
   1. Middle Age: Scores range from 4.93 to 5.11.
   2. Child: Scores range from 4.99 to 5.05.
   3. Young Adult: Scores are consistent, from 4.98 to 5.06.
   4. Senior Citizen: Scores show the least variation, centered at 4.94.
   5. Teenager: Scores range from 4.85 to 5.10. The data does not indicate a significant relationship between these specific demographics and patient satisfaction, suggesting a uniform patient experience.



1. **The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each**.
2. Objective: To evaluate and compare the total revenue generated by each hospital department to understand their individual financial contributions.
3. Methodology: A bar chart was created to visualize the financial performance of each department.
   1. X-Axis (Category): department\_referrals
   2. Y-Axis (Value): Sum of Total Bill
4. Finding: The analysis shows a significant variation in revenue generation across departments.
   1. Orthopaedics is the highest revenue-generating department, contributing $173M.
   2. General Practice is the second highest, with $164M in revenue.
   3. Renal recorded the lowest revenue at $5M, suggesting a potential area for further investigation.
   4. Other departmental revenues are as follows: Neurology ($73M), Cardiology ($68M), Physiotherapy ($17M), and Gastroenterology ($10M).



1. **Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.**
2. Objective: To identify the maximum appointment fee charged by each department and determine which department charges the most.
3. Methodology: A DAX measure was created to find the maximum appointment fee for each department. This measure was then used in a bar chart to compare departments and in a card visual to display the single highest fee. The CALCULATE function, combined with ALLEXCEPT, ensures that the MAX function finds the highest fee within the context of each specific department.
4. Finding: The analysis shows that the Neurology department charges the highest appointment fee at \*\*$1,500\*\*. This is followed by Cardiology ($1,200) and Physiotherapy ($1,000).
5. DAX Formula Used:

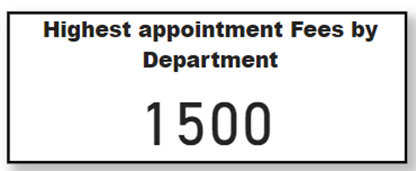
Highest Appointment Fees by Department =

CALCULATE(

MAX('Doctor Info'[Appointment Fees]),

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

)



1. Objective: To identify the maximum appointment fee charged by each department and determine which department charges the most.
2. Methodology: A DAX measure was created to find the maximum appointment fee for each department. This measure was then used in a bar chart to compare departments and in a card visual to display the single highest fee. The CALCULATE function, combined with ALLEXCEPT, ensures that the MAX function finds the highest fee within the context of each specific department.
3. Finding: The analysis shows that the Neurology department charges the highest appointment fee at \*\*$1,500\*\*. This is followed by Cardiology ($1,200) and Physiotherapy ($1,000).
4. DAX Formula Used:

Highest Appointment Fees by Department =

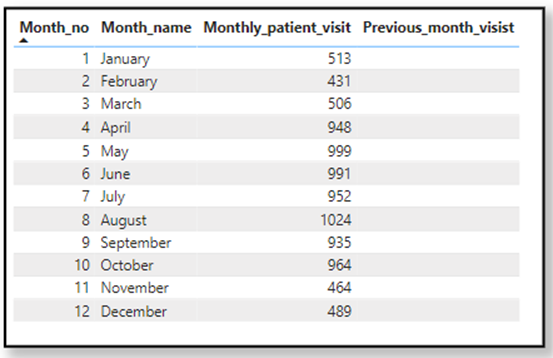
CALCULATE(

MAX('Doctor Info'[Appointment Fees]),

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

)

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.**
2. Objective: To analyze the total number of patient visits on a monthly basis and determine if the visit count increased compared to the prior month.
3. Methodology: A table visualization was created with five columns.
4. Month\_no: A calculated column on the Hospital Info table to get the month number.
5. Month\_name: A calculated column on the Hospital Info table to get the full month name.
6. Monthly\_patient\_visit: A measure to count the total patient IDs for the current month.
7. Previous\_month\_visit: A measure intended to count patient IDs from the previous month using time-intelligence functions.
8. Visit\_increased: A measure to compare the current and previous month's visits and return "Yes" or "No".
9. Finding: The Monthly\_patient\_visit column calculated correctly. However, the Previous\_month\_visit column returned blank values. Because this measure did not return any data, the Visit\_increased column was unable to perform its comparison, and the objective could not be completed.
10. DAX Formulas Used:
11. Column 1: Month\_no = MONTH('Hospital Info'[date])
12. Column 2: Month\_name = FORMAT('Hospital Info'[date],"MMMM")
13. Column 3: Monthly\_patient\_visit = COUNT('Hospital Info'[patient\_id])
14. Column 4: Previous\_month\_visist = CALCULATE(COUNT('Doctor Info'[patient\_id]), PREVIOUSMONTH(Calender[Date]))
15. Column 5: Visit\_increased = If([Monthly\_patient\_visit] > [Previous\_month\_visist], “Yes” , “No” )



1. **Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.**
2. Objective: To calculate the total number of patients who have visited Dr. Smith.
3. Methodology: Use the CALCULATE function to modify the filter context. The expression counts all patient IDs (COUNTA) while applying a filter to the 'Doctor Info' table, isolating only rows where the 'Doctor Name' is "Dr. Smith."
4. Finding: This measure will return the total number of patients associated with Dr. Smith.
5. DAX Formula Used:

Dr\_Smith\_patients =

CALCULATE(

COUNTA('Hospital Info'[patient\_id]),

'Doctor Info'[Doctor Name] = "Dr. Smith"

)



1. **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?**

Analysis 1: Average Age for Orthopedic Department:

1. Objective: To calculate the average age of patients who visit the Orthopedic department.
2. Methodology: A specific DAX measure was created to be displayed in a card visual. The CALCULATE function was used to modify the filter context of a simple AVERAGE function. This filter restricted the calculation to include only rows where the department\_referral column was "Orthopedics."
3. Finding: The average age of patients visiting the Orthopedic department is 39.
4. DAX Formula Used:

Orthopedic\_avg\_age =

CALCULATE(

AVERAGE('Hospital Info'[patient\_age]),

'Hospital Info'[department\_referral] = "Orthopedics"

)

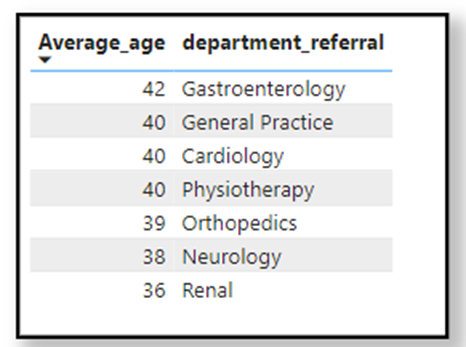
A white rectangular sign with black text and black letters

AI-generated content may be incorrect.

Analysis 2: Average Age for All Departments:

1. Objective: To calculate the average patient age for all departments and compare them.
2. Methodology: Yes, the approach is different and more flexible. Instead of hard-coding a filter (like "Orthopedics"), a general measure was created to calculate the average age of all patients. This single measure was then placed in a table visual. The department\_referral column was added to the table, which automatically "slices" the general average by each department.
3. Finding: The table visual successfully displays the average age for every department, showing Gastroenterology (42) has the highest average age and Renal (36) has the lowest.
4. DAX Formula Used:

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



1. **Were there any data format issues in the data, and if there were/are how you handle them?**
2. Objective: To identify and resolve any data format and type issues within the Hospital Info table to ensure data integrity and usability for analysis.
3. Methodology: The data was inspected, and two key issues were identified and corrected:
   1. Date Column: The date column was incorrectly formatted as text. Its data type was changed to "Date" to enable time-based analysis.
   2. Patient Satisfaction Score Column: The patient\_sat\_scores column was also in text format and contained blank spaces (nulls), which would disrupt calculations.
      1. First, the column's data type was converted from text to a number.
      2. Second, the resulting blank/null values were filled using mean imputation (replacing them with the average value of the entire column).
4. Finding: These transformation steps successfully cleaned and standardized the dataset. All columns now have the correct data types, and missing values in the patient\_sat\_scores column have been handled, making the data reliable for accurate analysis and insight generation.
5. **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?**
6. Objective: To explain the role of the M-query language in Power Query and to identify the specific code that appears in the formula bar when adding a new column.
7. Methodology: The answer is broken into two parts:
   1. A definition of the M-query language as the backend language for data transformation in Power Query, which is automatically generated by UI actions.
   2. A specific M-query code example is provided to show how a new column ("Full Name") is created by merging two existing columns ("patient\_first\_inital" and "patient\_last\_name").
8. Finding:
   1. The M-query language is used to manipulate tables and perform transformations on a dataset.
   2. Every action taken in the Power Query UI (adding, removing, or manipulating columns) generates a corresponding M-query step, which is visible in the formula bar.
   3. The specific function used for adding a column is Table.AddColumn. This function was used (for example, via "Column from Examples") to merge two patient name columns into a new "Full Name" column.
9. M-Query Formula Used:

= Table.AddColumn(

#"Replaced Value",

"Full Name",

each Text.Combine({[patient\_first\_inital], " ", [patient\_last\_name]}),

type text

)

1. **Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)**
2. Objective: To identify the top 5 doctors who generated the most revenue while having the fewest patients, using a SQL query.
3. Methodology: The SQL query aggregates data from the doctor table, grouping by Doctor Name. For each doctor, it calculates two metrics:
4. Revenue: The sum of Total Bill.
5. no\_of\_patients: The count of DISTINCT patient\_id. The results are then sorted, first by Revenue in descending order (highest revenue first) and then by no\_of\_patients in ascending order (fewest patients first). Finally, the query limits the output to the top 5 results.
6. Finding: The query successfully identifies the top 5 doctors based on the criteria. The output shows Dr. Smith has the highest revenue ($1,356,796,87) from 5,986 patients. Dr. Miller, Dr. Davis, Dr. Brown, and Dr. Harris follow, showing high revenue relative to a much smaller patient volume (e.g., Dr. Harris generated $2,668,212 from only 71 patients).
7. SQL Query Used:

SELECT

`Doctor Name` AS Doctor,

SUM(`Total Bill`) AS Revenue,

COUNT(DISTINCT patient\_id) AS no\_of\_patients

FROM doctor

GROUP BY Doctor

ORDER BY Revenue desc, no\_of\_patients asc

limit 5;

A screenshot of a computer

AI-generated content may be incorrect.

1. **Find the department where the average waiting time has decreased over three consecutive months. (SQL)**
2. Objective: To identify any hospital department where the average patient waiting time has decreased for three consecutive months, using SQL.
3. Methodology: A multi-step query using Common Table Expressions (CTEs) was executed:
   * 1. AvgWaitTimeByMonth (CTE 1): First, the average patient\_waittime was calculated for each department\_referral and date (month).
     2. WaitTimeWithLag (CTE 2): The LAG window function was used to create new columns (PrevMonthAvg, TwoMonthsAgoAvg, ThreeMonthsAgoAvg) that pull the average wait time from the previous three months for each department.
     3. Final SELECT: The main query filters the results from the second CTE. It uses a WHERE clause to find rows where AvgWaitTime < PrevMonthAvg AND PrevMonthAvg < TwoMonthsAgoAvg AND TwoMonthsAgoAvg < ThreeMonthsAgoAvg.
     4. Final GROUP BY: The results are grouped by department\_referral to ensure each department is listed only once.
4. Finding: The query successfully identified the departments that achieved a consistent decrease in average patient wait times over a three-month period. As shown in the result grid, these departments are Cardiology, Gastroenterology, General Practice, Neurology, Orthopedics, Physiotherapy, and Renal.
5. SQL Query Used:

WITH AvgWaitTimeByMonth AS (

SELECT

department\_referral,

`date`,

ROUND(AVG(patient\_waittime), 2) AS AvgWaitTime

FROM

hospital

GROUP BY

department\_referral, `date`

),

WaitTimeWithLag AS (

SELECT

department\_referral,

`date`,

AvgWaitTime,

LAG(AvgWaitTime, 1) OVER (PARTITION BY department\_referral ORDER BY `date`) AS PrevMonthAvg,

LAG(AvgWaitTime, 2) OVER (PARTITION BY department\_referral ORDER BY `date`) AS TwoMonthsAgoAvg,

LAG(AvgWaitTime, 3) OVER (PARTITION BY department\_referral ORDER BY `date`) AS ThreeMonthsAgoAvg

FROM

AvgWaitTimeByMonth

)

SELECT

department\_referral

FROM

WaitTimeWithLag

WHERE

AvgWaitTime < PrevMonthAvg

AND PrevMonthAvg < TwoMonthsAgoAvg

AND TwoMonthsAgoAvg < ThreeMonthsAgoAvg

GROUP BY

department\_referral;

A screenshot of a medical list

AI-generated content may be incorrect.

1. **Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)**
2. Objective: To determine the ratio of male-to-female patients for each doctor and then rank the doctors based on this ratio (from highest to lowest).
3. Methodology: A multi-step query using Common Table Expressions (CTEs) was executed:
   * 1. patientCount (CTE 1): This CTE joins the hospital and doctor tables. It then groups by Doctor\_Name and uses a SUM(CASE ...) expression to pivot the data, creating separate counts for male (male\_count) and female (female\_count) patients.
     2. DoctorRatios (CTE 2): This CTE takes the counts from patientCount and calculates the Male\_To\_Female\_ratio by dividing male\_count by female\_count, rounding the result to two decimal places.
     3. Final SELECT: The main query selects all data from the DoctorRatios CTE and uses the DENSE\_RANK() window function to assign a Ranking based on the Male\_To\_Female\_ratio in descending order.
4. Finding: The query successfully ranked all doctors by their male-to-female patient ratio. Dr. Anderson has the highest ratio at 1.35 (42 males to 31 females), ranking #1. The ranking continues down to Dr. Wilson, who has the lowest ratio at 0.64.
5. SQL Query Used:

with patientCount as (

select

d.`Doctor Name` as Doctor\_Name,

sum(case when h.patient\_gender = "M" then 1 else 0 end) male\_count,

sum(case when h.patient\_gender = "F" then 1 else 0 end) female\_count

from hospital as h

join doctor as d on h.patient\_id = d.patient\_id

group by Doctor\_Name

),

DoctorRatios as (

select

Doctor\_Name,

male\_count,

female\_count,

round((male\_count/female\_count),2) as Male\_To\_Female\_ratio

from patientCount

)

select

Doctor\_Name,

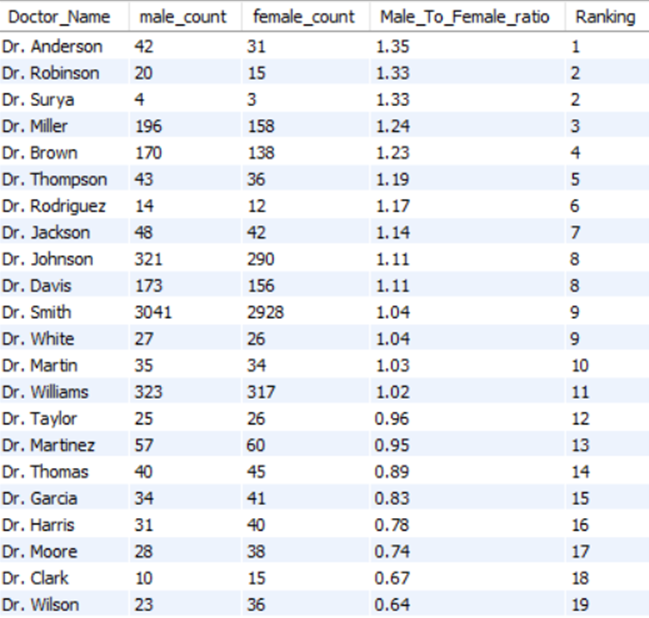
male\_count,

female\_count,

Male\_To\_Female\_ratio,

Dense\_rank() over( order by Male\_To\_Female\_ratio desc) as Ranking

from DoctorRatios;



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

1. Objective: To calculate the average patient satisfaction score for each doctor based on their associated visits, using SQL.
2. Methodology: A SQL query was executed to join the hospital and doctor tables on their common patient\_id. The data was then grouped by Doctor Name. A CASE statement was used during the AVG aggregation to handle data quality issues: if a patient\_sat\_score was blank (""), it was replaced with a default value of 5. The final average score was rounded to two decimal places and sorted in descending order.
3. Finding: The query successfully ranked all doctors by their average patient satisfaction score. Dr. Taylor has the highest average score at 5.71, followed by Dr. Surya at 5.57. Dr. Robinson has the lowest average satisfaction score at 4.69.
4. SQL Query Used:

select

d.`Doctor Name` as Doctor\_Name,

round(avg(case when h.patient\_sat\_score = "" then 5 else h.patient\_sat\_score end),2) as patient\_sat\_score

from hospital as h

join doctor as d

on h.patient\_id = d.patient\_id

group by Doctor\_Name

order by patient\_sat\_score desc;



1. **Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)**
2. Objective: To find doctors who have treated patients from different races and calculate the diversity (number of unique races) of their patient base.
3. Methodology: A SQL query was executed to join the hospital and doctor tables on their common patient\_id. The data was then grouped by Doctor\_Name, and COUNT(DISTINCT h.patient\_race) was used to calculate the number of unique patient races treated by each doctor. The HAVING clause was applied to filter this list, including only those doctors who have treated more than one race. Finally, the results were sorted in descending order.
4. Finding: The query successfully identified all doctors who treat a diverse patient base. A large number of doctors (including Dr. Anderson, Dr. Brown, and Dr. Davis) have the highest diversity, having treated patients from all 7 racial categories. The list descends to Dr. Surya, who has treated 4 different races.
5. SQL Query Used:

select

d.`Doctor Name` as Doctor\_Name,

count(distinct h.patient\_race) as differnet\_race\_count

from hospital as h

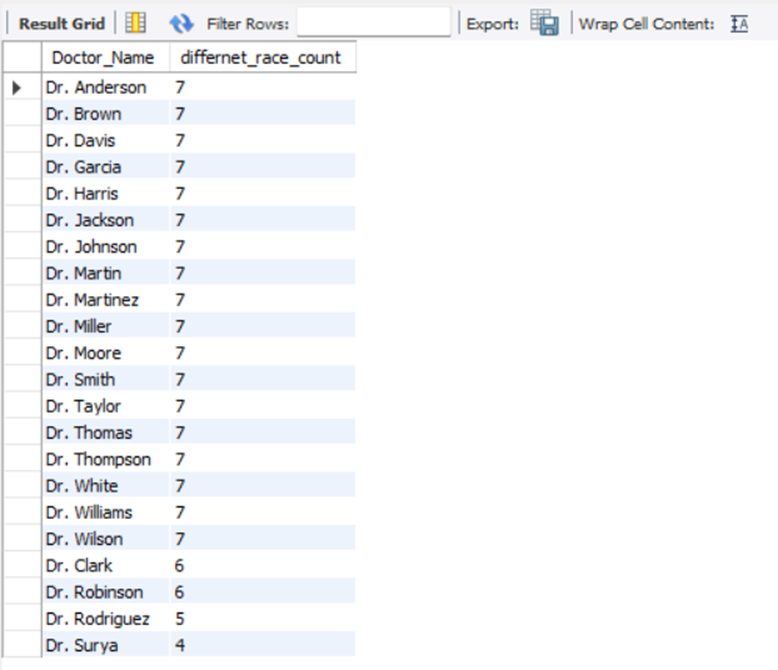
join doctor as d

on h.patient\_id = d.patient\_id

group by Doctor\_Name

having count(distinct h.patient\_race) > 1

order by differnet\_race\_count desc;



1. **Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)**
2. Objective: To calculate the ratio of total bills generated by male patients to female patients for each department.
3. Methodology: A SQL query was executed to join the hospital and doctor tables on patient\_id. The data was then grouped by department\_referral. Conditional aggregation (SUM(CASE WHEN ... END)) was used to create separate sums for male (Male\_total\_bill) and female (Female\_total\_bill) patient bills. Finally, the male total was divided by the female total and rounded to two decimal places to determine the Male\_To\_Female\_Ratio.
4. Finding: The query successfully calculated the billing ratio for all departments. Orthopedics has the highest ratio at 1.19 (meaning revenue from males is 19% higher than from females), while Gastroenterology has the lowest at 0.75 (meaning revenue from females is higher). Physiotherapy is perfectly balanced with a ratio of 1.00.
5. SQL Query Used:

select

h.department\_referral,

sum(case when patient\_gender = "M" then d.`Total Bill` end) as Male\_total\_bill,

sum(case when patient\_gender = "F" then d.`Total Bill` end) as Female\_total\_bill,

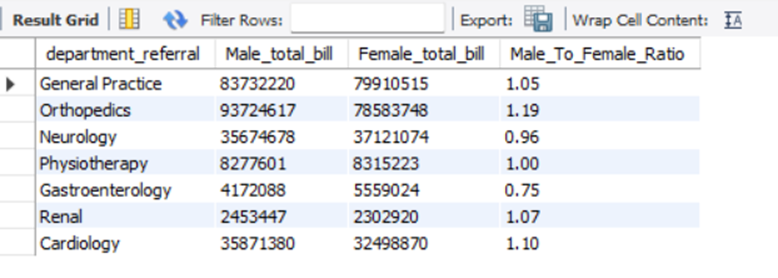
round(sum(case when patient\_gender = "M" then d.`Total Bill` end) / sum(case when patient\_gender = "F" then d.`Total Bill` end),2) as Male\_To\_Female\_Ratio

from hospital as h

join doctor as d

on h.patient\_id = d.patient\_id

group by h.department\_referral;



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)**
2. Objective: To update the patient\_sat\_score for all patients who visited the "General Practice" department and waited more than 30 minutes. The score should be increased by 2 points, but not exceed a maximum score of 10.
3. Methodology:
   1. Disable Safe Updates: The SET SQL\_SAFE\_UPDATES = 0 command was used to allow the UPDATE query to run without a WHERE clause on a key column.
   2. Run Update Query: An UPDATE query was executed on the hospital table.
   3. Filter Rows: The WHERE clause filtered the data to find rows where LOWER(department\_referral) = "general practice" AND patient\_waittime > 30.
   4. Set New Score: The SET command used the LEAST function to assign the new score. This function takes the lower value between (patient\_sat\_score + 2) and 10, effectively capping the score at 10.
   5. Re-enable Safe Updates: SET SQL\_SAFE\_UPDATES = 1 was run to restore safe update mode.
   6. Finding: The query successfully updated the target rows. Patients meeting the criteria had their satisfaction scores increased, ensuring no score went above the 10-point maximum.
4. SQL Query Used:

-- Disable Safe Updates

SET SQL\_SAFE\_UPDATES = 0;

UPDATE hospital

SET patient\_sat\_score = LEAST(patient\_sat\_score + 2,10)

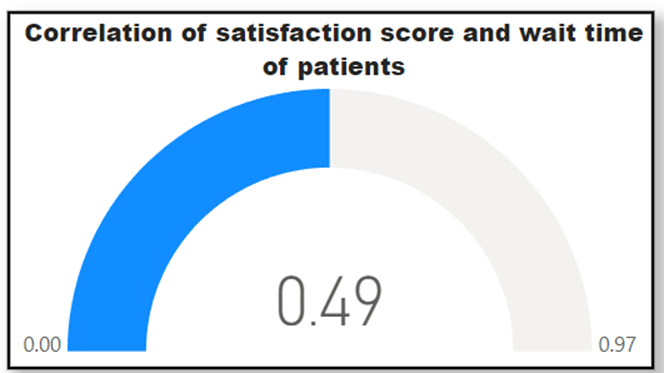
WHERE LOWER(department\_referral) = "general practice" AND patient\_waittime > 30;

-- Re-enable Safe Updates

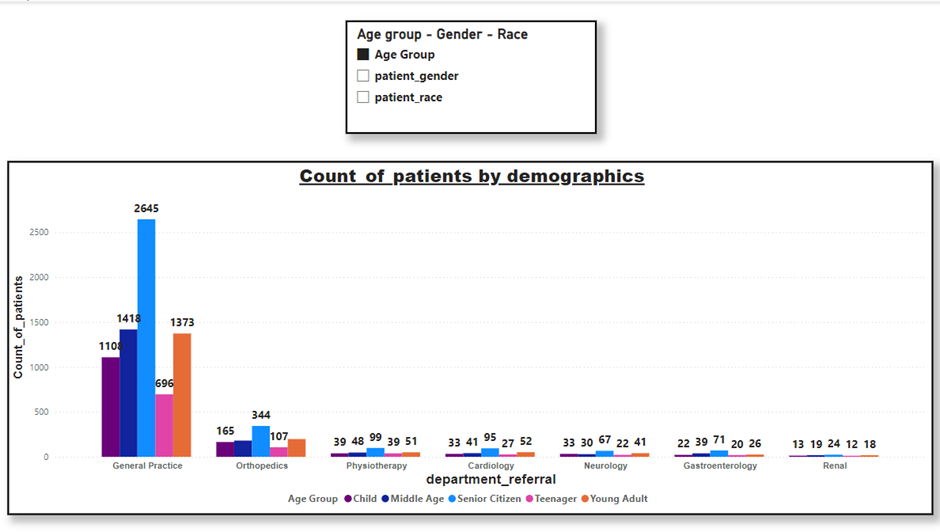
SET SQL\_SAFE\_UPDATES = 1;

**subjective questions**

1. **What is the relation between patient wait time and satisfaction scores?**
2. Objective: To measure the statistical relationship between patient wait times and their corresponding satisfaction scores, in order to understand if wait time is a key driver of satisfaction.
3. Methodology: Power BI's Quick Measure feature was used to avoid writing a complex DAX formula.
   1. The "Correlation Coefficient" calculation was selected.
   2. patient\_sat\_score was set as Measure X.
   3. patient\_waittime was set as Measure Y.
   4. The analysis was grouped by department\_referral (Category).
4. Finding: The analysis found a correlation coefficient of +0.49, which was displayed on a Gauge Chart.
5. Interpretation: This value indicates a moderate positive correlation.
6. Implication: This is a counterintuitive insight. It suggests that as patient wait times increase, satisfaction scores also have a slight tendency to increase. This strongly implies that wait time is not the primary factor driving patient dissatisfaction. It is possible that longer wait times are associated with more thorough care, or that other factors like staff communication and quality of service have a far greater impact on the final satisfaction score.



1. **How do patient demographics affect the frequency of visits to different departments?**
2. Objective: To explore how patient demographics (such as Age Group, Race, and Gender) influence the frequency of visits to various hospital departments.
3. Methodology: A Stacked Column Chart was created to visualize the demographic breakdown of patients within each department.
   1. X-Axis: department\_referral
   2. Y-Axis: Count\_of\_patients
   3. Legend: A Field Parameter was used to allow the report user to dynamically switch the legend's category between Age Group, patient\_gender, and patient\_race.
4. Finding: The analysis suggests that demographics do not directly affect the frequency of visits to a specific department. The number of visits appears to be driven by the patient’s individual health needs and the services offered by the department. The demographic distribution seen within each department (e.g., in General Practice, which has the highest volume) largely reflects the overall demographic mix of the hospital's total patient population, rather than indicating that a specific department is preferred by a specific demographic group.

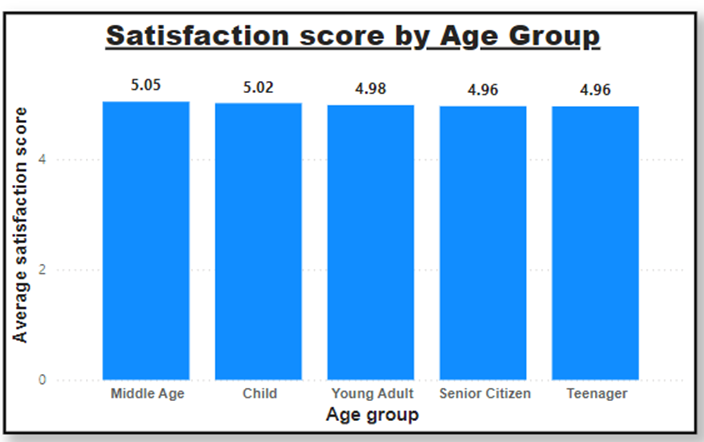


1. **Is there a noticeable trend in the volume of patient visits throughout the year?**
2. Objective: To analyze the volume of patient visits on a monthly basis to identify any seasonal trends or noticeable patterns throughout the year.
3. Methodology: A line chart was created to visualize the total count of patients for each month.
   1. X-Axis: Month\_name
   2. Y-Axis: Count\_of\_patients
4. Finding: Yes, there is a distinct seasonal trend in patient visits.
   1. Low Season: The year begins and ends with a low volume of patients (January-March and November-December). The quietest month of the year is February (431 visits).
   2. High Season: There is a sharp increase in patients in April, and the hospital experiences a sustained high volume from April (948 visits) through October (964 visits).
   3. Peak: The absolute peak in patient visits occurs in August (1024 visits).
   4. Trend Change: A sharp decline in visits occurs after October, moving into the low season.

A graph showing the number of patients

AI-generated content may be incorrect.

1. **Which age groups report the highest and lowest satisfaction scores?**
2. Objective: To identify which age groups report the highest and lowest average patient satisfaction scores.
3. Methodology: The average satisfaction score was calculated for each of the five defined age groups (Child, Teenager, Young Adult, Middle Age, Senior Citizen).
4. Finding:
   1. Highest Score: The Middle Age group (ages 36-50) reports the highest average satisfaction score at 5.05.
   2. Lowest Score: The Senior Citizen (age >50) and Teenager (ages 13-20) groups share the lowest average score at 4.96.
   3. Observation: The difference between the highest and lowest scores is minimal (0.09), indicating that satisfaction levels are generally consistent across all age groups.
5. Key Insight: While middle-aged individuals report slightly higher satisfaction, the overall stability of scores suggests that age is not a significant driver of patient dissatisfaction.



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?**
2. Objective: To analyze patient data to determine if there is any evidence of racial discrimination in either the patient population served or the patient satisfaction reported.
3. Methodology: A two-part analysis was conducted using patient demographic and satisfaction data:
   1. Patient Demographics: A pie chart ("Count of patients by patient\_race") was created to visualize the percentage distribution of the total patient count, segmented by racial background.
   2. Patient Experience: A table was created to compare the Average\_satisfaction\_score for each patient\_race category.
4. Finding: The data does not support the claim of racial discrimination. The analysis shows the hospital serves a highly diverse population and provides a consistent quality of experience for all racial groups.
5. Analysis of Patient Demographics (Diversity)

The pie chart of patient counts shows that the hospital's patient base is highly diverse and not dominated by any single group.

* White patients account for 27.9% of the total patient base.
* The other 72.1% of patients come from various backgrounds, including African American (21.17%), Two or More Races (16.89%), Declined to Identify (11.18%), Asian (11.5%), Pacific Islander (5.96%), and Native American/Alaska Native (5.4%).

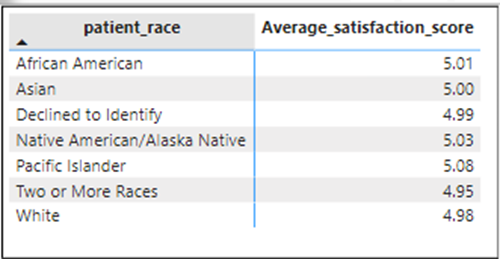
A pie chart with numbers and a few different colored circles

AI-generated content may be incorrect.

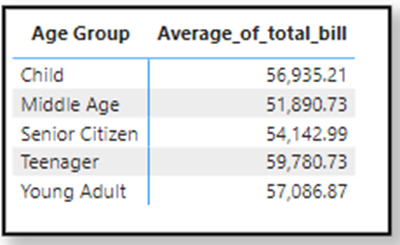
2. Analysis of Patient Satisfaction (Experience)

The table of satisfaction scores shows that there is no significant variation in patient experience based on race. The scores are remarkably consistent across all groups.

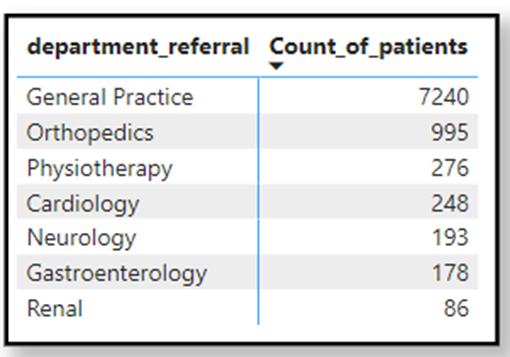
* Pacific Islander: 5.08 (Highest)
* Native American/Alaska Native: 5.03
* African American: 5.01
* Asian: 5.00
* Declined to Identify: 4.99
* White: 4.98
* Two or More Races: 4.95 (Lowest)

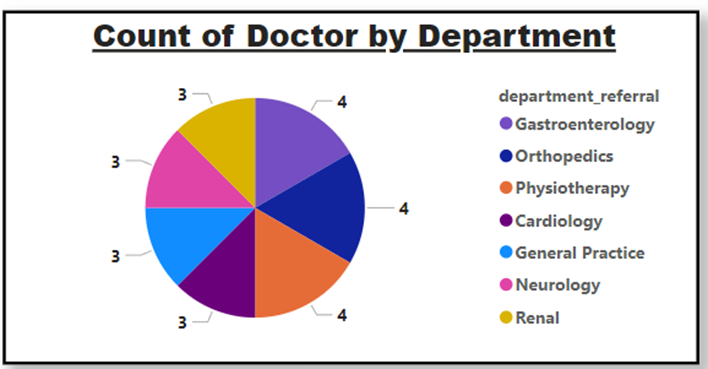


1. Conclusion: The minimal difference in scores (a range of only 0.13) and the diverse patient population strongly refute any claims of racial discrimination.
2. **The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?**
3. Objective: To determine a data-driven basis for assigning patient discounts, in order to support the hospital's management in this initiative.
4. Methodology: This approach analyses the average hospital bill for each defined Age Group. The goal is to identify which patient segment has the highest average spending, as this group represents a high-value customer base.
5. Finding: The analysis of average spending by age group reveals:
   1. Teenager: $59,780.73 (Highest)
   2. Young Adult: $57,086.87
   3. Child: $56,935.21
   4. Senior Citizen: $54,142.99
   5. Middle Age: $51,890.73 (Lowest)
   6. The Teenager age group (ages 12-20) is the highest-spending segment on average.

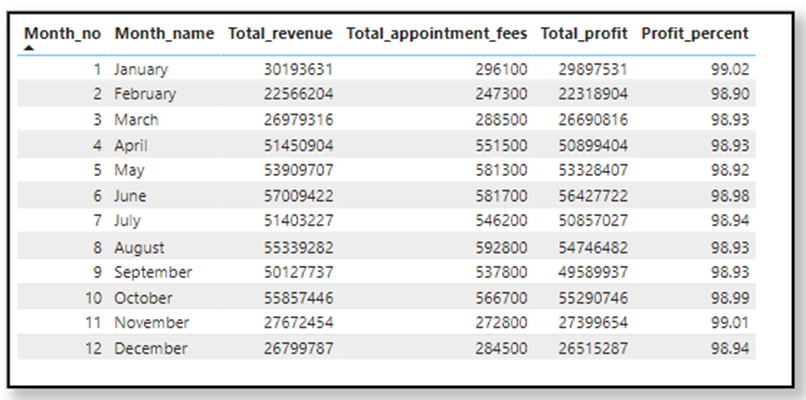


1. Strategic Rationale & Recommendation: It is recommended to offer targeted discounts to the Teenager age group.
   1. High-Value Customers: This group already represents the hospital's "biggest customer" in terms of average bill.
   2. Long-Term Loyalty: By offering discounts and attracting patients from this younger demographic, the hospital has an opportunity to build positive relationships and create lifelong customers.
2. **The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.**
3. Objective: To provide a data-driven recommendation on which departments should receive the 2-3 new doctors the hospital plans to hire.
4. Methodology: The analysis was conducted by comparing two key metrics:
   1. The current number of doctors staffed in each department
   2. The total volume of patient visits for each department. The goal was to identify departments with a high patient load that may be understaffed.
5. Finding: A significant imbalance was found in the General Practice department.
   1. Patient Load: The General Practice department has the highest patient volume in the hospital by an extremely large margin, with 7,240 total visits. The second-busiest department, Orthopedics, has only 995 visits.
   2. Staffing: Despite handling the vast majority of patients, the General Practice department currently has only 3 doctors, which is comparable to departments with a fraction of its patient load.

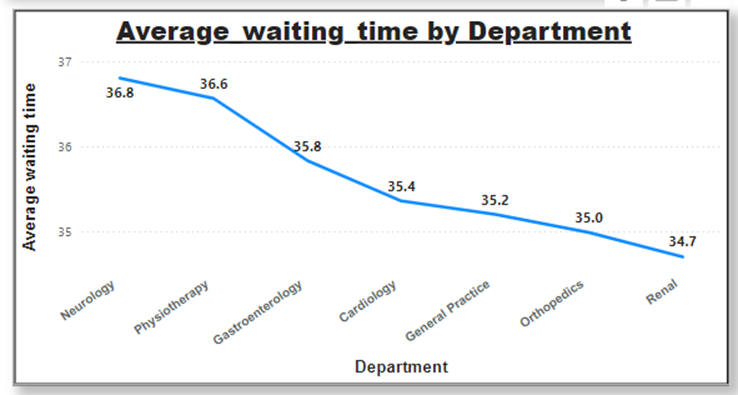




1. Recommendation: It is strongly recommended to allocate all 3 new doctor positions to the General Practice department. The data shows this department is handling a disproportionately high patient-to-doctor ratio, and additional staff is necessary to manage the existing workload and maintain the quality of care.
2. **Is the hospital profitable? How will you determine the profitability?**
3. Objective: To determine the hospital's profitability by analyzing its total revenue against its total costs (represented by appointment fees).
4. Methodology: To assess profitability, four key DAX measures were created and analyzed on a monthly basis:
   1. Total Revenue: Calculated by summing the Total Bill amounts for all patients.
   2. Total Appointment Fees: Calculated by summing the appointment fees for all patients (treated as the primary cost for this analysis).
   3. Total Profit: Derived by subtracting the Total Appointment Fees from the Total Revenue.
   4. Profit Percentage: Calculated by dividing the Total Profit by the Total Revenue to find the profit margin.



1. Finding: Yes, the hospital is highly profitable. The data table shows that the profit margin is consistently stable and high throughout the entire year, remaining above 98% in every single month. This indicates a very strong and efficient financial performance.
2. **Any Department for which the waiting time is oddly large?**
3. Objective: To identify any departments with unusually or oddly large average patient waiting times compared to the others.
4. Methodology: A line chart was created to visualize and compare the average waiting time for each department.
   1. X-Axis: Department
   2. Y-Axis: Average Waiting Time (in minutes)
5. Finding: The analysis shows that while Neurology has the highest average waiting time (36.80 minutes), the variation across all departments is minimal. The difference between the highest and lowest average wait time is only 2.1 minutes (ranging from 34.70 to 36.80 minutes).



1. Conclusion: This difference is very insignificant. Therefore, we can conclude that no particular department has a relatively high or oddly large waiting time, and the patient wait experience is consistent across the hospital.
2. **Come up with strategies to provide discounts to the patients.**
3. Objective: To implement a discount strategy for high-spending patients, as intended by hospital management.
4. Methodology: A DAX calculated column was created to categorize each patient. The formula calculates the total bill for a patient across all their visits (using ALLEXCEPT to group by patient\_id). If the patient's total bill exceeds $10,000, they are marked as "Eligible" for the discount; otherwise, they are "Ineligible."
5. Finding: The donut chart reveals that under this rule, the vast majority of patients would qualify for a discount.
   1. Eligible: 8.23K patients (89.29%)
   2. Ineligible: 0.99K patients (10.71%) This suggests the $10,000 threshold may be too low if the goal is to target a more exclusive group of high-spending patients.

A blue circle with black text

AI-generated content may be incorrect.

1. DAX Formula Used:

Discount\_eligibility =

IF(

CALCULATE(

SUM('Doctor Info'[Total Bill]),

ALLEXCEPT('Doctor Info', 'Doctor Info'[patient\_id])

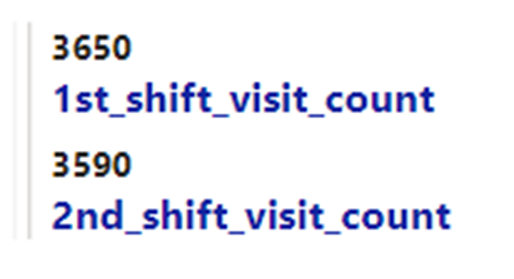
) > 10000,

"Eligible",

"Ineligible"

)

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?**
2. Objective: To create a two-shift system for doctors in the "General Practice" department, determine the shift timings, recommend how to divide the doctors, and assess if this policy would be beneficial.
3. Methodology:
   1. The day was divided into two 12-hour shifts:
      1. Shift 1: 0:00 to 12:00 (Midnight to Noon)
      2. Shift 2: 12:00 to 24:00 (Noon to Midnight)
   2. DAX measures were created to calculate the total count of patient visits specifically for the "General Practice" department, filtering for each of these two time blocks.
   3. The results were visualized on a multi-row card to compare patient volume between the shifts.
4. Finding: By comparing the two measures, the analysis shows that Shift 1 (0:00 - 12:00) has a higher patient volume than Shift 2 (12:00 - 24:00).



1. Recommendation & Rationale:
   1. Staff Allocation: The General Practice department currently has 3 doctors. Based on the findings, it is recommended to assign 2 doctors to Shift 1 (the busier shift) and 1 doctor to Shift 2.
   2. Benefit: Yes, this policy would be helpful. It distributes the workload more evenly in line with patient demand. This should lead to reduced doctor fatigue, lesser wait times for patients, and, consequently, better overall patient satisfaction.
2. DAX Formulas 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"

)

1. **What do you understand by Power BI gateway? What are its use cases?**
2. What is the Power BI Gateway?

The Power BI Gateway is a software tool that acts as a secure bridge or "gate" between your on-premises data sources (data stored locally on your company's servers) and the Power BI cloud service.

Its primary function is to enable secure data transfer, allowing you to access and refresh your local data in your cloud-based Power BI reports and dashboards.

1. Key Use Cases

* Data Refresh: Automatically refreshes your Power BI reports and dashboards with the latest data from your on-premises sources, ensuring your reports are always up-to-date.
* Secure Data Transfer: Ensures that all data moving between your local servers and the Power BI cloud is encrypted and secure.
* Access On-Premises Data: Allows you to connect to and use data from local sources like SQL Server, local Excel files, or other databases directly in your Power BI reports.
* Hybrid Data Integration: Lets you combine on-premises data with cloud-based data (e.g., from Azure) into a single, unified report.
* Direct Query: Enables Power BI to send real-time queries directly to your on-premises data source, which is crucial when you need to see live data without importing it.

1. Types of Gateways

* Personal Gateway: Designed for individual use. It allows one person to connect to sources and refresh their own data.
* Enterprise (Standard) Gateway: Designed for organizations. It allows multiple users to connect to multiple on-premises data sources and is the standard for business use.

1. Key Benefits

* Security: Provides a secure, reliable, and encrypted connection, which is essential for protecting sensitive company data.
* Cost-Effectiveness: Allows you to access and analyze your data directly without having to first move it all to an expensive cloud storage solution.
* Automation: Enables scheduled refreshes for dashboards, ensuring that decision-makers always have the most current information from on-premise sources.

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

If no specific objective or subjective questions were provided, I would approach the problem using a goal-oriented, exploratory data analysis framework.

1. General Data Analysis Workflow

First, I would follow a standard process to prepare and understand the data:

* Data Import: Begin by importing the dataset into Power BI.
* Data Transformation (Power Query): Use the Power Query Editor to transform the data. This would involve filtering irrelevant data, merging tables, and shaping the data to fit a clean, structured model.
* Data Cleaning: Address any data quality issues, such as handling missing or inconsistent values, correcting data types, and standardizing formats across the dataset.
* KPI Identification: Identify the Key Performance Indicators (KPIs) most relevant to the hospital's goals, such as revenue, patient wait times, satisfaction scores, and patient volume.
* KPI Analysis: Analyze each KPI in detail, exploring trends, relationships, and patterns to derive meaningful insights.

1. Specific Strategic Analysis

After establishing the baseline, I would move on to generating actionable insights for key business areas:

On Hiring Strategy:

* I would not limit the hiring analysis to just doctors. I would look at the "bigger picture" for each department, including nursing and support staff.
* This comprehensive view helps distribute the workload evenly and allows for better-informed hiring or firing decisions based on the hospital's true operational needs.

On Patient Discounts:

* My strategy would remain the same: give discounts to the age group that provides the most business (i.e., has the highest average bill).

The rationale for this is two-fold:

* Immediate Benefit: It encourages high-value customers, leading to increased cash flow.
* Long-Term Strategy: Catering to this specific high-revenue group could pave the way for creating a new, specialized department, which in turn would attract more of that same target demographic.

1. Final Deliverables

* Report & Dashboard Creation: I would build comprehensive reports and dashboards using a mix of visualizations (bar charts, line graphs, tables) to clearly present the findings.
* Interactivity: I would add slicers and filters to the reports, enabling users to dynamically explore the data based on different dimensions (like department, age group, or month).
* Actionable Recommendations: Finally, I would review the completed report for accuracy and clarity and present actionable recommendations to management based on the insights gathered.

1. **Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?**
2. Objective: To analyze and define the database relationship between the doctor\_id and department entities and determine if it is one-to-one.
3. Methodology / Analysis: The relationship is defined by analyzing the business rules from the perspective of both entities:
   1. From the Doctor's perspective: Each doctor\_id is associated with only one department.
   2. From the Department's perspective: Each department can have many doctors assigned to it.
4. Finding:
   1. No, the relationship is not one-to-one.
   2. This is a Many-to-One ($M:1$) relationship from the doctors' perspective (many doctors belong to one department).
   3. Looked at the other way, it is a One-to-Many ($1:M$) relationship from the department's perspective (one department has many doctors).