**Columbia Asia Hospital analysis**

**Problem Statement**

* **Assess the hospital's revenue generation**
* **Insights about suitable departments for new hires**
* **Strategies and suggestions for patient discounts**

**Objective Questions:**

**Q1)** In analysing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.

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

**Replaced null values with average of Patient sat score column to clean the data and avoid the inconsistencies.**

**Q2) Assess the Average Waiting Time:** Analyse the patient waits times to identify the average duration a patient spends before receiving care.

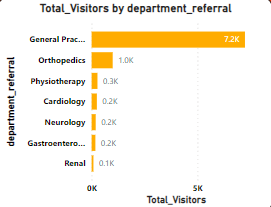
**Average waiting Time of the patients is 35.26** 

formula to calculate Average wait time

Average\_Waiting\_Time = AVERAGE('Hospital ER'[patient\_waittime])

**Q3) Visits by Department Referral:** Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.

To address the question of calculating the total number of visits by department, I created a Stacked Bar Chart. In this chart, the department-referral combinations are represented on the Y-axis, while the total daily visits are shown on the X-axis.

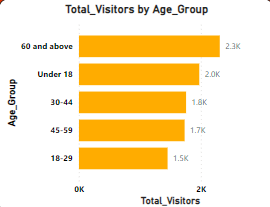
****Formula to calculate total daily visits.

Thus, we can see that the General Practice department has the highest number of visits compared to all other departments.

Total\_Visitors = COUNT('Hospital ER'[patient\_id])

**Q4) Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

To address the question of segregating patient visits by different age groups, I utilized a Stack Bar Chart. In this chart, the patient age groups are displayed on the Y-axis, and the total visits are depicted on the X-axis.

**** Formula to calculate Patient age group.

Age\_Group = SWITCH(TRUE(),

                'Hospital ER'[patient\_age] < 18,"Under 18",

                'Hospital ER'[patient\_age] < 30, "18-29",

                'Hospital ER'[patient\_age] < 45, "30-44",

                'Hospital ER'[patient\_age] < 60, "45-59",

                'Hospital ER'[patient\_age] >= 60, "60 and above",

                BLANK())

As a result, we find that the patient age group of 60 and above has the highest number of visits among all the age groups.

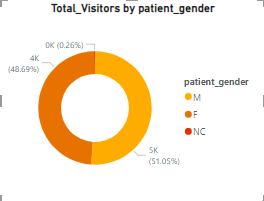
**Q5)** Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for?

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

Formula to remove null values in patient sat score column.

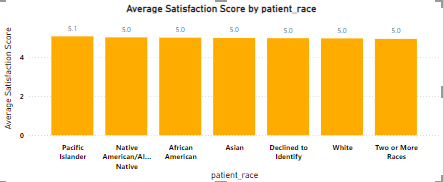
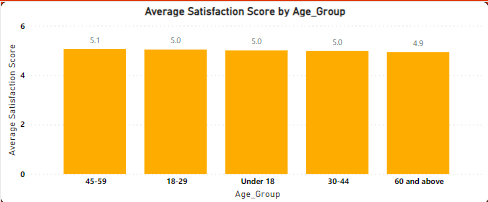
PsatisfactionScore = ROUND(IF(ISBLANK('Hospital ER'[patient\_sat\_score]),AVERAGE('Hospital ER'[patient\_sat\_score]),'Hospital ER'[patient\_sat\_score]),0)

**Q6)** Is there any relation between the number of visits and the Gender of the patients?

To address the question of the total number of visits by gender, I employed a Donut Chart. In this chart, gender is represented in the legend, and the total daily visits are shown in the values. ****

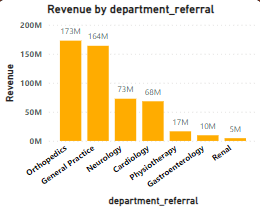
This Donut Chart illustrates that both male and female patients visit the hospital. Therefore, hospital visits are independent of gender.

**Q7)** 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.



**As we can observe in the graph, the average satisfaction score of patients above 60 is lower than that of other age groups. To address this issue, we can conduct a survey to understand the specific problems faced by these patients and improve their overall experience.**

**Q8)** The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.

To evaluate the total revenue of each department, I used a Stacked Column Chart. In this chart, the departments are displayed on the X-axis, while the total revenue is represented on the Y-axis.

Formula to calculate total revenue.

Revenue = SUM(Sheet1[Total Bill])

**In the graph, Orthopaedics and General Practice departments show the highest revenue generation. This emphasizes their significant financial contribution to the healthcare facility.**

**Q9)** Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.

Formula Used

Highest\_Appointment\_Fees = CALCULATE(MAX('Hospital ER'[department\_referral]),Sheet1[Appointment Fees] = MAX(Sheet1[Appointment Fees]

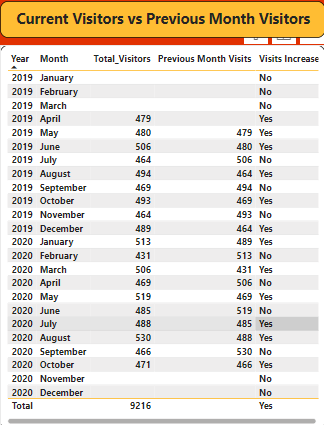
))

**Neurology is the department which is charging highest appointment fees.**

**I have used Card to visualize it**



**Q10)** 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.



**Q11)** Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.

**The total number of patients visited by Dr smith is 5986**

Formula used to calculate

Total Patients Visited Dr. Smith =

CALCULATE(

    DISTINCTCOUNT('Hospital ER'[patient\_id]),

    FILTER(

        'Sheet1',

        'Sheet1'[Doctor Name] = "Dr. Smith"

    )

)



**Q12)** Calculate the average age of the patients who visit the Orthopaedics department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?

**The average age of the patients who visited the Orthopaedics department is 38.66**

Average Age Orthopedics =

CALCULATE(

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

    FILTER(

        'Hospital ER',

        RELATED('Sheet1'[department\_referral]) = "Orthopedics"

    )

)

****

**Yes, the approach uses to calculate the average age of all the department will be different.To, calculate all department’s age, we will not give any filter context in.**

**Q13)** Were there any data format issues in the data, and if there were/are how you handle them?

**No, there were no data format issue in the data.**

**Q14)** When we add a column in Power Query what’s the code that comes in M language in formula bar? What do you know about M-query?

**= Table.AddColumn(#"PreviousStepName", "NewColumnName", each [Column1] + [Column2])**

* **When we add a column in Power Query the above code comes in M language in formula bar.**
* **When you add a column in Power Query, the code in the formula bar is written in the M language. It specifies the previous step, the new column's name, and the transformation logic. M-query, also known as the Power Query Formula Language, is used for data transformation tasks. It's flexible, integrates with various sources, optimized for performance, and readable. It's a powerful tool in Power Query for shaping and preparing data for analysis in Power BI**

**Subjective Questions**

**Q1)** What is the relation between patient wait time and satisfaction scores?

Formula Used to calculate Corelation

patient\_waittime and patient\_sat\_score correlation for department\_referral =

VAR \_\_CORRELATION\_TABLE = VALUES('Hospital ER'[department\_referral])

VAR \_\_COUNT =COUNTX(KEEPFILTERS(\_\_CORRELATION\_TABLE),CALCULATE(

SUM('Hospital ER'[patient\_waittime])\* SUM('Hospital ER'[PsatisfactionScore])))

VAR \_\_SUM\_X =SUMX(KEEPFILTERS(\_\_CORRELATION\_TABLE),CALCULATE(SUM('Hospital ER'[PsatisfactionScore])))

VAR \_\_SUM\_Y =SUMX(KEEPFILTERS(\_\_CORRELATION\_TABLE),CALCULATE(SUM('Hospital ER'[PsatisfactionScore])))

VAR \_\_SUM\_XY =SUMX(KEEPFILTERS(\_\_CORRELATION\_TABLE),CALCULATE(SUM('Hospital ER'[patient\_waittime])

\* SUM('Hospital ER'[PsatisfactionScore]) \* 1.))

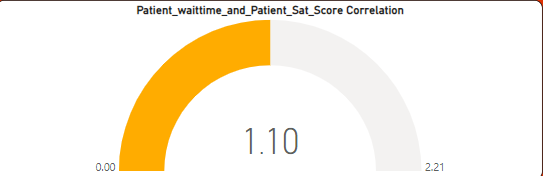
VAR \_\_SUM\_X2 =SUMX(KEEPFILTERS(\_\_CORRELATION\_TABLE),CALCULATE(SUM('Hospital ER'[patient\_waittime]) ^ 2))

VAR \_\_SUM\_Y2 =SUMX(KEEPFILTERS(\_\_CORRELATION\_TABLE),CALCULATE(SUM('Hospital ER'[PsatisfactionScore]) ^ 2))

RETURN

DIVIDE(\_\_COUNT \* \_\_SUM\_XY - \_\_SUM\_X \* \_\_SUM\_Y \* 1.,

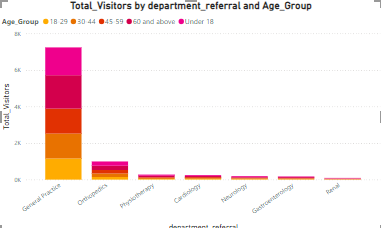
SQRT((\_\_COUNT \* \_\_SUM\_X2 - \_\_SUM\_X ^ 2)\* (\_\_COUNT \* \_\_SUM\_Y2 - \_\_SUM\_Y ^ 2)))



This value suggests a neutral or average correlation between patient satisfaction scores and the waiting time they experienced.

Q2) How do patient demographics affect the frequency of visits to different departments?

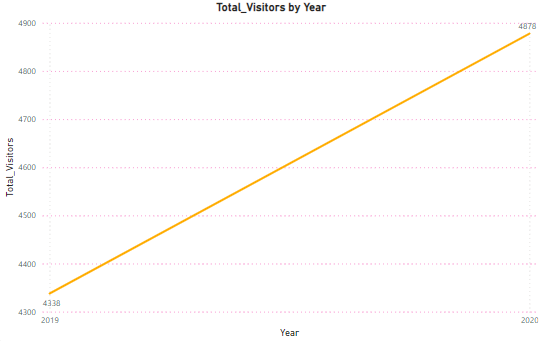
As, the question ask about affecting the frequency of patient demographics. Therefore I have used **Stacked Column Chart** and take total daily visits in Y-axis and department referral in X-axis.



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

Q3) Is there a noticeable trend in the volume of patient visits throughout the year?

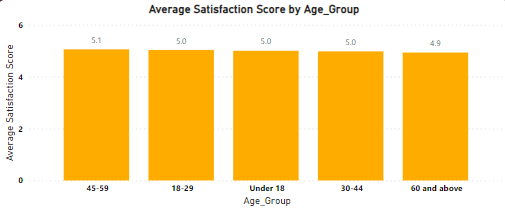
As, the question ask is there any noticeable trend of patient visit in year wise. Therefore I have used line chart and take year in X-axis and total visitors in Y-axis.



**The graph illustrates a consistent rise in patient volume throughout the year, indicating increasing demand for healthcare services. This trend underscores the importance of resource management and service delivery to meet the growing needs of patients effectively.**

Q4) Which age groups report the highest and lowest satisfaction scores?

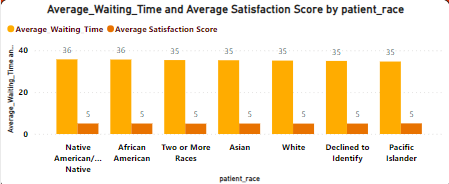
As, the question ask the highest and lowest satisfaction score by patient age group. Therefore I have used Stacked **Column Chart** and take Patient age group in X-axis and customer satisfaction score in Y-axis.



**The graph shows that patients aged 45-59 have the highest satisfaction scores, while those aged 60 and above have the lowest. This highlights the importance of addressing age-specific needs to improve overall patient satisfaction.**

Q5) 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?

As, the question ask outside the hospital someone claims that is discrimination in hospital. Therefore I have used Cluster Column **Chart** and take Race in X-axis and average waiting time and average satisfaction score in Y-Axis



**By examining the patient satisfaction scores and waiting times displayed in the chart, we can conclude that there is no evidence of racial discrimination occurring in the hospital.**

Q6) The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?

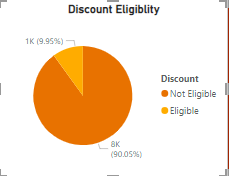
Discounts to patients can be offered based on several factors, such as patient visit frequency, patient satisfaction score, doctor's appointment fees, and promoting departments with the lowest patient visits to encourage more traffic to those areas.

In this case, we're focusing on patient satisfaction scores to determine discount eligibility because we lack information on the costs associated with each patient in a specific department. Therefore, appointment fees or department revenue cannot be used as deciding factors.

We will set the eligibility criteria for the discount, where patients with a satisfaction score greater than 6 will qualify for the discount. To calculate discount eligibility, the following DAX function is used.

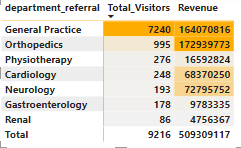
Discount = IF('Hospital ER'[patient\_sat\_score]>6, "Eligible", "Not Eligible")

To analyse that I have used **Donut Chart** and take Eligible for discount in Legend and Total visitors in values.



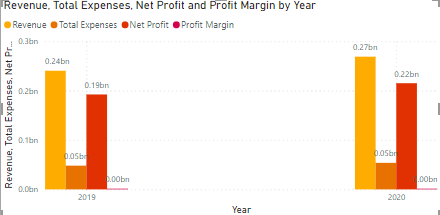
There are 9.95% patient who has patient satisfaction score is more than 6 which are eligible for discount.

Q7) The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.



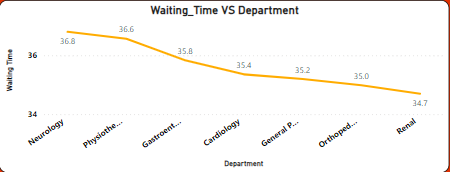
**As depicted in the graph, both the general practice and orthopaedics departments exhibit the highest revenue and visitor count among all departments. Based on this observation, I recommend hiring additional doctors for these departments to enhance their performance. By increasing the staffing levels, particularly with qualified professionals, these departments can effectively manage the influx of patients, optimize service delivery, and potentially further increase revenue. This proactive measure aligns with the goal of maintaining high-quality healthcare services and meeting the demands of patients seeking treatment in these specialized areas.**

Q8) Is the hospital profitable? How will you determine the profitability?



**The hospital is not profitable. To determine its profitability, we need to calculate the total expenses, net profit, and profit margin. This involves summing the appointment fees to get the total expenses, summing the total bills to get the total revenue, and then calculating the net profit by subtracting the total expenses from the total revenue. The profit margin is determined by dividing the net profit by the total revenue and expressing it as a percentage. By analyzing these metrics, we can identify areas where the hospital can manage expenses or increase revenue to improve its financial performance.**

Q9) Any Department for which the waiting time is oddly large?



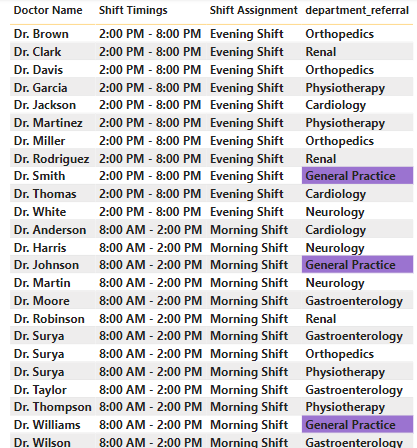
**As depicted in the graph, the waiting time across different departments shows minimal variation, suggesting that the waiting time is relatively consistent across all departments. This indicates that the waiting time is not significantly high for any particular department compared to others. Overall, the relatively uniform distribution of waiting times across departments suggests that patients are not experiencing excessively long wait times regardless of the department they visit.**

Q10) Come up with strategies to provide discounts to the patients.

Offering discounts to patients can be an effective strategy for healthcare providers to enhance patient satisfaction, foster patient loyalty, and attract new patients. Here are a few strategies to consider:

* **Discount on Total Bill**: Offer discounts for bills exceeding a certain amount, such as $3000 or $5000.
* **High Satisfaction Scores**: Provide discounts to patients who have a satisfaction score higher than 6.
* **Frequent Visitors**: Reward patients who visit frequently with discounts.

Q11) 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?

**Dr. Jackson and Dr. Williams will handle the morning shift from 8:00 AM to 2:00 PM due to higher patient traffic during these hours. Dr. Smith will be assigned to the evening shift from 2:00 PM to 8:00 PM. This allocation optimizes patient care and ensures efficient resource utilization while supporting the well-being of our medical staff.The two-shift policy in the General Practice department improves patient flow and boosts hospital revenue. With high visitor numbers, splitting shifts ensures smoother operations and shorter wait times. Aligning doctor shifts with peak demand optimizes patient care and revenue generation. Overall, the policy efficiently addresses demand peaks while enhancing financial performance.**

Q12) What do you understand by PowerBI gateway? What are its use cases?

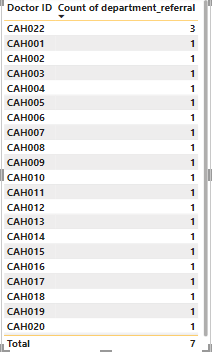
**The Power BI Gateway is a vital component facilitating seamless integration between on-premises data sources and the Power BI service. Its primary function is to enable scheduled data refreshes, ensuring that Power BI reports and dashboards reflect the latest information from on-premises systems. Moreover, it supports Direct Query connections, allowing users to access real-time data without importing it into Power BI, which is particularly beneficial for monitoring operational metrics. Additionally, the Gateway enables live connections to on-premises Analysis Services models, enabling interactive exploration of on-premises data directly within Power BI. It also supports hybrid scenarios by seamlessly integrating data from on-premises and cloud environments, providing centralized management capabilities, ensuring data security through encryption and authentication mechanisms, and enforcing data governance policies to comply with regulatory requirements and internal standards. In summary, the Power BI Gateway plays a crucial role in facilitating data connectivity, security, and governance for organizations leveraging Power BI for analytic and reporting.**

**Q13)How would you approach this problem, if the objective and subjective questions were not given?**

Without the objective and subjective questions, I would thoroughly analyze the data by examining each column and identifying patterns. I would extract as many insights as possible from the data and use these insights to build the dashboard.

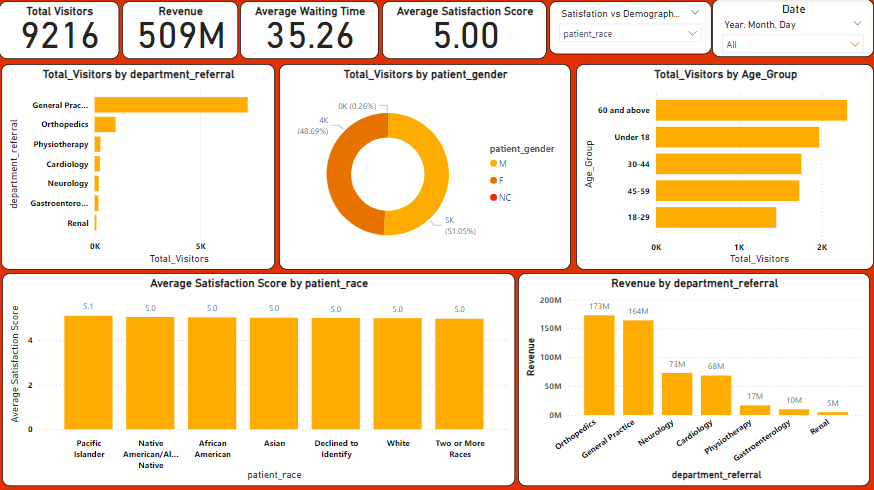
**Q14)Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?**

No, it is many to many. because one department has multiple doctors, also one doctor is in multiple departments.

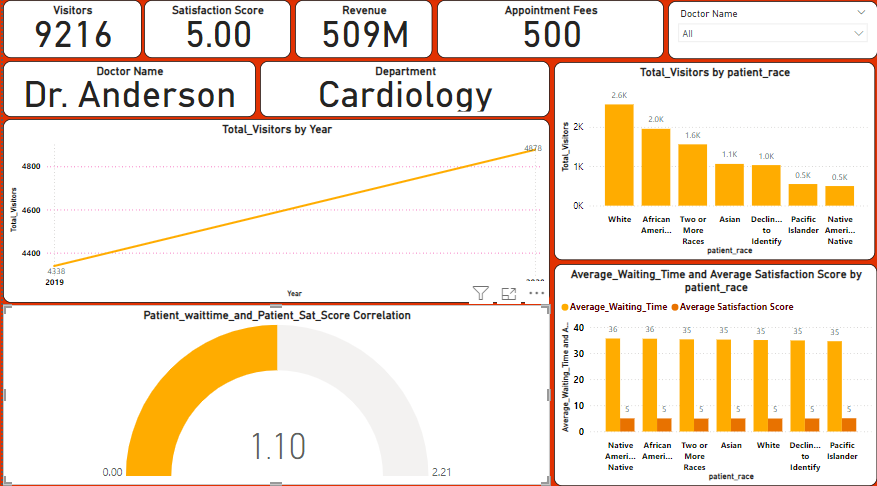


**Report:-**

The hospital has asked for a report with three tabs:

**Main Tab**

**Doctors’ Tab**



* **Patients’ Tab**

