# Objective Questions:

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

ANS:

Before loading the dataset into Power BI, we can utilize the **Power Query Editor** to perform essential data transformations and ensure data quality.

## Handling Missing Values in the Hospital ER Dataset:

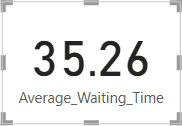
* + In the **“patient\_sat\_score”** column, nearly **72% of the values are missing (null)**, which can impact analysis and insights.
  + The best practice for handling missing values is to first investigate the **data source** to determine whether the missing data can be recovered or if a more reliable dataset is available.
  + For numerical columns like **“patient\_sat\_score”**, the most effective method is to replace null values with a statistically meaningful substitute, such as the **mean (average)** value.
  + In this case, all missing values in the **“patient\_sat\_score”** column have been replaced with an **average score of 5**, ensuring consistency in the dataset and maintaining data integrity.

By applying these transformations, we improve the dataset’s completeness, enabling more accurate visualizations and analysis in Power BI.

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

**ANS:**

**Output:**



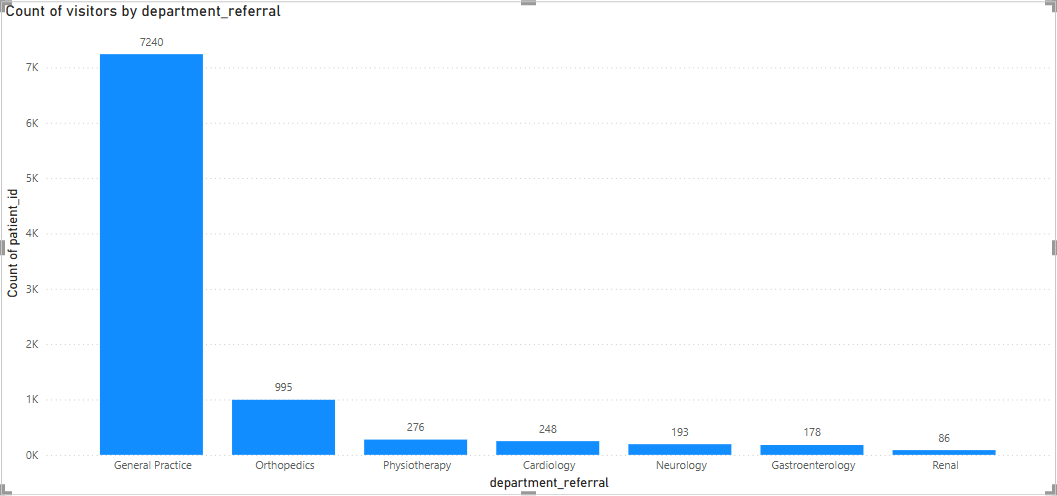
**Dax\_Fun:**

**Average\_Waiting\_Time = AVERAGE(Hospital\_ER[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.

## ANS:

**Visualization:**



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

**ANS:**

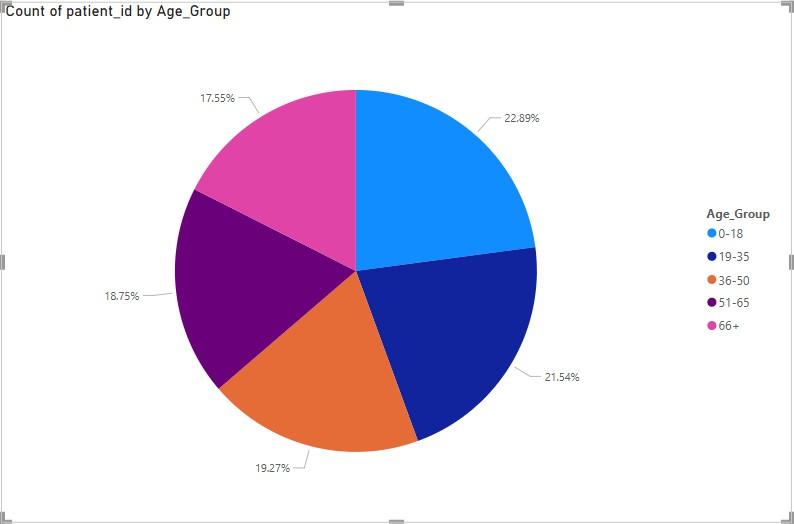
**Created Age Groups – Used Power Query Conditional Column or DAX Calculated Column to categorize patients into age groups (0-18, 19-35, 36-50, 51-65, 66+).**

**Created a Visualization – Added a Pie Chart, set "Age Group" on the Axis, and used Count of patient\_id in Values to count patient visits per age group.**

**DAX Query:**

**Age\_Group = SWITCH(TRUE(),Hospital\_ER[patient\_age]<=18,"0- 18",Hospital\_ER[patient\_age]>18 && Hospital\_ER[patient\_age]<=35,"19- 35",Hospital\_ER[patient\_age]>35 && Hospital\_ER[patient\_age]<=50,"36- 50",Hospital\_ER[patient\_age]>50 && Hospital\_ER[patient\_age]<=65,"51- 65",Hospital\_ER[patient\_age]>50,"66+")**

**Visualization:**



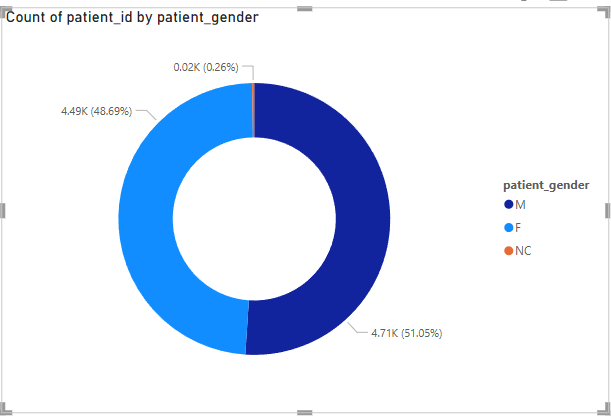
1. Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for?

ANS:

* **The “patient\_sat\_score” column contained 72% blank values, which could have a considerable impact on the accuracy of our data analysis.**
* **Ideally, addressing this issue involves revisiting the data source to retrieve the missing information.**
* **Additionally, improving the data collection process can help reduce or eliminate the occurrence of null values in the future.**
* **In this case, I first converted the blank entries in the column to “null” for consistency.**
* **Once all blank values were standardized as null, I replaced them with the column’s average value.**
* **As a result, the missing “patient\_sat\_score” values were filled with the average score of the column.**

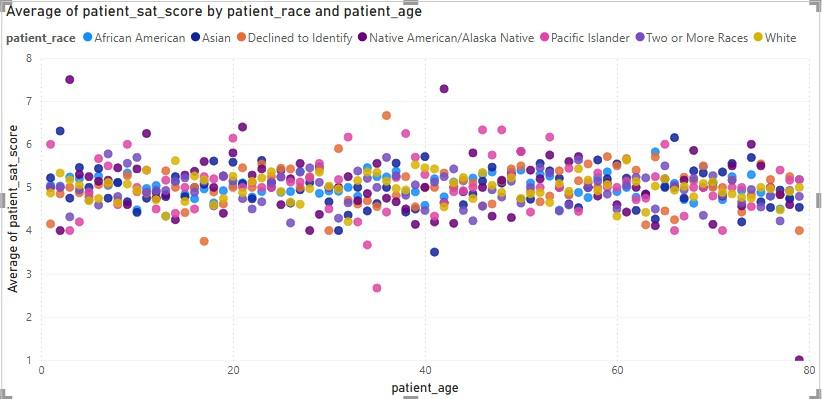
1. Is there any relation between the number of visits and the Gender of the patients?

**ANS:**



* **The donut chart above shows that Male and Female patients have a nearly equal share of visits.**
* **Male patients account for 4,705 visits (51.05%), while female patients contribute 4,487 visits (48.69%) of the total.**
* **A small proportion of patients, categorized as “Non-Confirmed (NC)” because they either chose not to disclose their gender or do not align with the two primary gender categories, represent 0.26% of the visits, totaling 26 visits.**

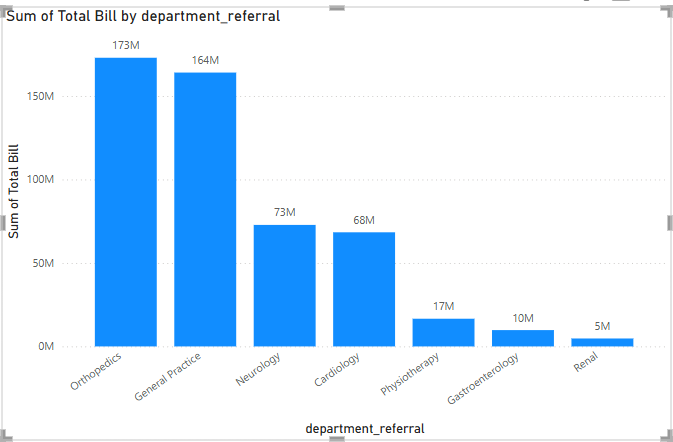
1. Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.

ANS:

* The patient **satisfaction scores** consistently range between **4.0** and **6.0** across all age groups.
* Similarly, patients from various racial backgrounds also maintain satisfaction scores within this range, regardless of age.
* This indicates that satisfaction scores remain unaffected by variations in age or racial background.

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

ANS:



* + The **Orthopedics** department generates the highest revenue (**$173M**), followed closely by **General Practice** (**$164M**).
  + The **Renal** department generates the lowest revenue (**$5M**).

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

ANS:



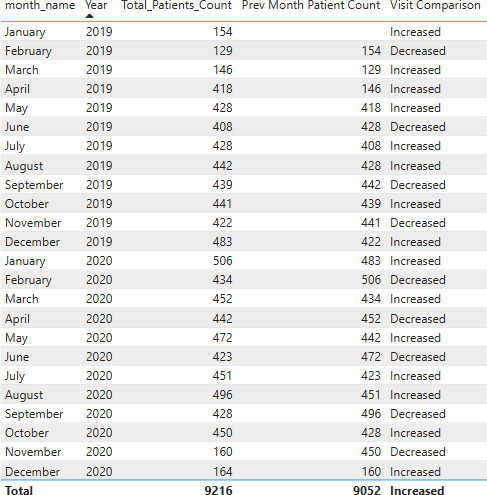
DAX\_Query:

## Dept\_High\_Appointment\_Fee = LOOKUPVALUE(Doctor\_Patients\_Data[department\_referral],Doctor\_Patients\_Data[ Appointment Fees],MAX(Doctor\_Patients\_Data[Appointment Fees]))

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.

ANS:

Tabular Visualization



First Created Date Table which starts from minimum date and ends at maximum date. then Joined the Date table

For Prev Month Patient Count I used this DAX function:

Prev Month Patient Count = CALCULATE([Total\_Patients\_Count],DATEADD('Date Table'[Date],-1,MONTH))

For comparison I used this DAX function: Visit Comparison =

IF(

[Total\_Patients\_Count] > [Prev Month Patient Count], "Increased",

"Decreased"

)

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

ANS:



Formula:

## patient visited Dr Smith = CALCULATE([Total\_Patients\_Count],Doctor\_Patients\_Data[Doctor Name]="Dr. Smith")

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?

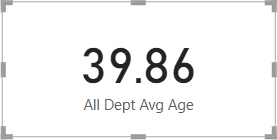
ANS:



DAX Query:

## Avg\_Age\_Patients\_Who\_Visits\_Orthopedic = CALCULATE(AVERAGE(Hospital\_ER[patient\_age]),Hospital\_ER[department\_referr al]="Orthopedics")

However, if we calculate the overall average age including all the departments then we need not use the CALCULATE() DAX function and use the AVERAGE() on age directly



DAX Query:

## All Dept Avg Age = AVERAGE(Hospital\_ER[patient\_age])

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

ANS:

* + The **Date** column in the *Hospital ER* table, along with the *Appointment Fees* and

*Total Bill* columns in the *Doctor Patient Data* table, had data format issues.

* + To address this, we selected the affected columns and updated their data types using the **Structures** tab in Table View.
  + Specifically, we changed the *Appointment Fees* and *Total Bill* columns from **Whole Number** to **Currency**.
  + Since Columbia is in the US, we assumed the monetary values were in **USD** and formatted both columns accordingly.

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?

ANS:

* + **M-query language** is used in Power BI to manipulate tables and transform datasets.
  + Whenever columns are added, removed, or modified in a dataset, Power BI generates the corresponding **M-query** and displays it in the formula bar.
  + The **Power Query UI** automatically creates M-query code as you perform transformations on your data.
  + For instance, when adding a new column in Power Query, the following M-query might appear in the formula bar:

## DAX

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

* + In this example, the M-query combines the **patient\_first\_initial** and **patient\_last\_name** columns into a new column called **Full Name**, using the **Column from Examples** feature in Power Query.

**Q15. Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)**

## Ans:



* **Dr. Harris** earned **26,682,124** from just **71 patients**, showing the highest revenue per patient among the group.
* **Dr. Brown** generated **54,075,456** revenue from **309 patients**, placing them among the top for efficiency.
* **Dr. Davis** earned **57,413,306** from **330 patients**, maintaining a high revenue-to- patient ratio.
* **Dr. Miller** generated revenue of **61,322,460** from just **355 patients**, suggesting a high revenue per patient and significant efficiency.
* **Dr. Smith** has the highest revenue of **135,679,687** and managed **5,986 patients**, indicating efficient revenue generation with a relatively high patient count.



**Q16. Find the department where the average waiting time has decreased over three consecutive months. (SQL).**

**Ans:** I have used the below query to identify the same. The departments with consistently decreasing average waiting times over three consecutive months were identified. This indicates improved efficiency and better patient flow management, contributing to enhanced service quality.

**Output:**



**Q17.Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)**

**Ans.** I have used the below query to identify the same. The ratio of male to female patients for each doctor is calculated and used to rank doctors in descending order based on this ratio. This analysis provides insights into patient demographics for each doctor, helping to identify trends or biases in patient distribution.

## Output:

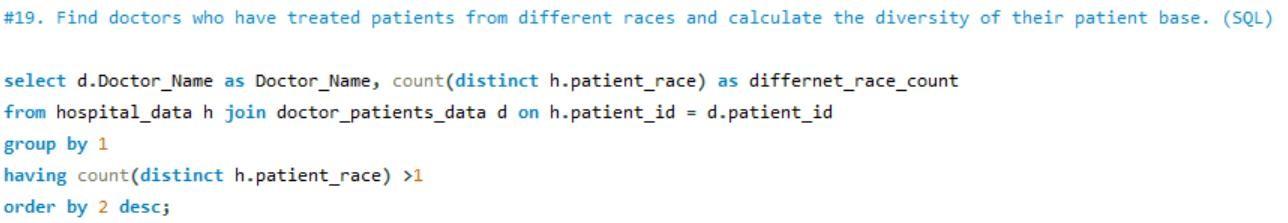
**Q18. Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)**

**Ans.** I have used the following query for the above need. The average patient satisfaction score for each doctor is calculated, considering visits where missing scores are replaced with a default value of 5. This provides an overall measure of how patients rate their experience with each doctor, ranked by the highest satisfaction scores.

**Output:**

**Q19. Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)**

**Ans.** The ratio of total bills generated by male patients to those generated by female patients is calculated for each department. This analysis offers insights into the financial contributions of each gender, enabling the identification of spending trends across various departments.





**Q20. Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)**

**Ans.** The ratio of total bills generated by male patients to female patients is calculated for each department. This provides insights into the financial contribution from each gender, helping to identify trends in patient spending across different departments.

**Output:**

****

**Q21. Update the patient satisfaction score for all patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. Increase their satisfaction score by 2 points, but ensure that the satisfaction score does not exceed**

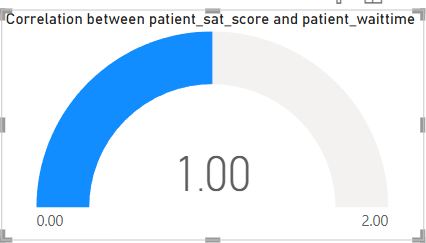
**10. (SQL)**

**Ans:** In the "General Practice" department, patients with waiting times longer than 30 minutes have their satisfaction scores increased by 2 points. If the adjusted score surpasses 10, it is limited to a maximum of 10. This approach ensures fairness while keeping scores within the allowed range.

# Subjective Questions

1. What is the relation between patient wait time and satisfaction scores? ANS:
   * A **quick measure** was created in Power BI, using **department\_referral** as the category, **patient\_sat\_score** as Measure1, and **patient\_waittime** as Measure2.
   * Power BI automatically generated a **DAX formula** to compute the correlation between these two measures.
   * The correlation was visualized using a **Gauge Chart**, with a range set from **0 to 2**, and the observed correlation value was **1**.
   * The analysis revealed a **moderate positive correlation (value = 1)**, indicating that patient wait times may influence satisfaction scores, though not significantly.
   * Departments that effectively reduce wait times or manage patient expectations could see **higher satisfaction levels**, but factors like **service quality and communication** likely play a more significant role in shaping patient perceptions.

## Visualization:



**DAX Query:**

Correlation between patient\_sat\_score and patient\_waittime =

VAR CORRELATION\_TABLE = VALUES('Hospital\_ER'[department\_referral]) VAR COUNT =

COUNTX(

KEEPFILTERS( CORRELATION\_TABLE), CALCULATE(

COUNTA('Hospital\_ER'[patient\_sat\_score])

\* SUM('Hospital\_ER'[patient\_waittime])

)

)

VAR SUM\_X = SUMX(

KEEPFILTERS( CORRELATION\_TABLE),

CALCULATE(COUNTA('Hospital\_ER'[patient\_sat\_score]))

)

VAR SUM\_Y =

SUMX(

KEEPFILTERS( CORRELATION\_TABLE),

CALCULATE(SUM('Hospital\_ER'[patient\_waittime]))

)

VAR SUM\_XY = SUMX(

KEEPFILTERS( CORRELATION\_TABLE), CALCULATE(

COUNTA('Hospital\_ER'[patient\_sat\_score])

\* SUM('Hospital\_ER'[patient\_waittime]) \* 1.

)

)

VAR SUM\_X2 = SUMX(

KEEPFILTERS( CORRELATION\_TABLE),

CALCULATE(COUNTA('Hospital\_ER'[patient\_sat\_score]) ^ 2)

)

VAR SUM\_Y2 = SUMX(

KEEPFILTERS( CORRELATION\_TABLE),

CALCULATE(SUM('Hospital\_ER'[patient\_waittime]) ^ 2)

) RETURN

DIVIDE(

COUNT \* SUM\_XY - SUM\_X \* SUM\_Y \* 1., SQRT(

( COUNT \* SUM\_X2 - SUM\_X ^ 2)

\* ( COUNT \* SUM\_Y2 - SUM\_Y ^ 2)

)

)

**Insights:**

**Moderate Positive Correlation (~1):**

* **A correlation value of 1 suggests a *perfect positive linear relationship*—but you noted it as *moderate*, which is contradictory. It’s worth rechecking the data or clarifying the strength (perfect = 1.0, moderate = ~0.5–0.7).**
* **If the correlation is *exactly 1*, this could point to a data anomaly or a non-random pattern (e.g., derived or dependent values).**

**Wait Time Affects Satisfaction:**

* **Longer wait times may be linked to lower patient satisfaction, or vice versa.**
* **However, correlation does not imply causation; satisfaction could be influenced by multiple factors (e.g., service quality, doctor interaction, communication).**

**Departmental Impact:**

* **Some departments may consistently outperform others in managing wait times and expectations.**
* **Identifying these departments can help replicate best practices across others.**

**Recommendations:**

**Focus on Departments with High Wait and Low Satisfaction:**

* **Prioritize operational improvements in departments with both high average wait times and low patient satisfaction.**

**Patient Communication and Expectation Setting:**

* **While reducing wait time helps, clear communication about delays and proactive engagement may boost satisfaction more significantly.**

**Introduce Smart Scheduling Tools:**

* **Optimize doctor-patient schedules or waiting room load using AI-powered time slotting.**

**Further Multivariate Analysis:**

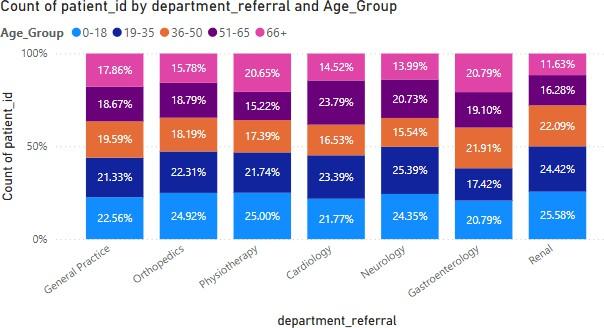
* **Consider building a multiple regression model in Power BI or Excel that includes other predictors:**
  + **Service quality**
  + **Doctor ratings**
  + **Appointment duration**
  + **Demographics (age, gender, admin\_flag)**

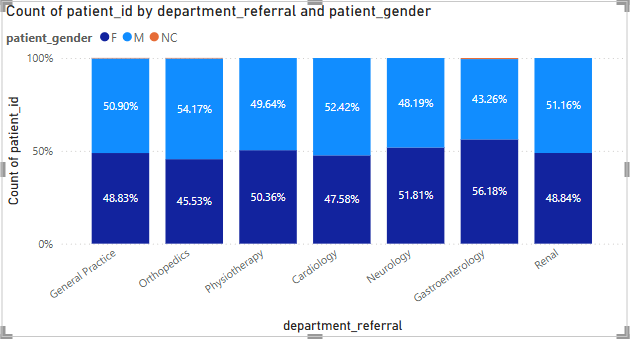
1. How do patient demographics affect the frequency of visits to different departments?

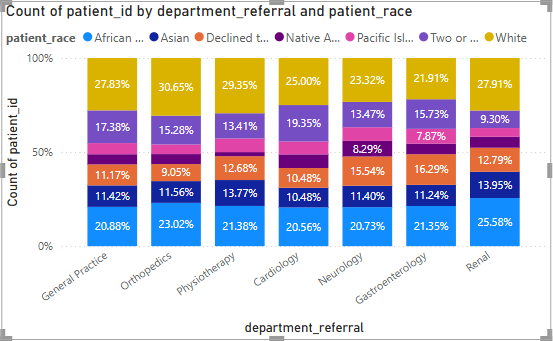
ANS:

* + The dataset was structured to include **Department, Patient Count, and demographic attributes** such as **Age Group, Race, and Gender**.
  + A **Stacked Column Chart** was created with **Department** on the **x-axis** and **Patient Count** on the **y-axis** to analyze demographic distribution across departments.
  + This visualization helps assess the **impact of patient demographics** on visit frequencies, illustrating how different demographic groups contribute to the overall patient count in each department.
  + The approach provides insights into whether **specific departments attract certain demographic groups** and helps identify potential **patterns or disparities** in departmental usage based on patient characteristics.
  + **Insight: Demographics have minimal influence** on the number of visits to different departments. Instead, **visit frequency is primarily driven by patients' health conditions and the hospital's location**, rather than demographic factors.

Visualization:







**Insights:**

* **Demographics (age, race, gender) have minimal impact on departmental visit frequency.**
* **Patient health condition and hospital location are the main drivers of department visits.**
* **All departments receive a diverse mix of demographic groups, indicating equitable access.**

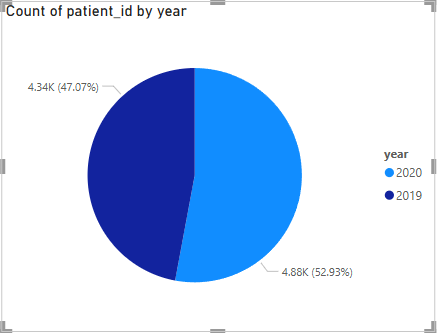
**Recommendations:**

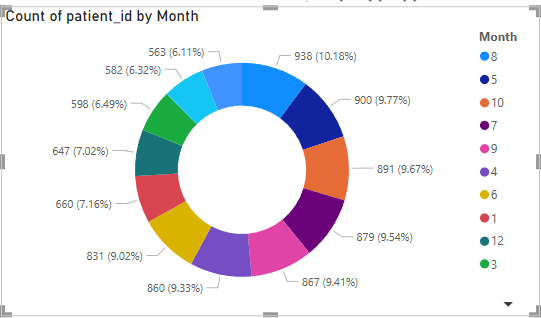
* **Focus outreach and resources based on health condition trends, not just demographics.**
* **Use location-based data (e.g., patient residence) to improve access planning and facility placement.**
* **Ensure all departments are equipped to handle diverse patient needs, regardless of demographic group.**
* **Continue monitoring for any emerging disparities, especially if patterns shift over time.**
* **Use the demographic breakdown to guide staff training (e.g., language support, cultural sensitivity).**

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

ANS:

Visualization:





* + A **Pie Chart** was utilized to examine the **distribution of patient visits over time**, providing a **clear breakdown of how visit volumes were spread across different periods**.
  + By implementing a **Year and Month Name hierarchy**, the visualization enables **drill-up and drill-down functionality**, allowing a **high-level (yearly) view** while also offering **more granular (monthly) insights** into patient visit patterns.
  + The **Count\_of\_Patients** measure was plotted to represent the proportion of visits per time segment, making it easier to **detect shifts in patient volumes** across different timeframes.
  + This approach helps identify **seasonal variations, emerging trends, and potential anomalies** in patient visits, offering valuable insights for **healthcare planning and hospital operations**.
  + The **visual representation of patient distribution** over time can reveal whether certain months experience a **higher influx of patients**, potentially due to **seasonal outbreaks, policy shifts, or external factors**.

## Insights:

* + **2019 recorded 4,338 patient visits**, which rose to **4,878 in 2020**, reflecting a **7.5% increase** in patient volume.
  + This growth may indicate a **higher demand for healthcare services**, possibly influenced by **greater health awareness, expanded medical offerings, or external events like the COVID-19 pandemic**.
  + The distribution of visits across different months suggests **potential seasonal trends**, which could help **optimize hospital resource management, staffing decisions, and service readiness**.

**Recommendations:**

* + **Use line or area charts to clearly track patient visit trends over months.**
  + **Identify peak months (e.g., flu season) to prepare hospital resources in advance.**
  + **Use trend lines to highlight rising or falling patient visit patterns.**
  + **Monitor seasonal spikes to adjust OPD schedules and doctor availability.**
  + **Segment data by department to see which units face higher patient loads.**

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

ANS:

**Insights from the Visualization: Average Patient Satisfaction Score by Age Group**

This **horizontal bar chart** presents the **average patient satisfaction score (patient\_sat\_score) across different age groups**. The key observations and insights are as follows:

## Highest Satisfaction in the 36-50 Age Group (5.05)

* + The **36-50 age group** has the **highest average satisfaction score (5.05)** among all age groups.
  + This suggests that **patients in this age range may be the most content with healthcare services**, possibly due to a balance of expectations and healthcare experiences.

## Slightly Lower Satisfaction in the 66+ Age Group (4.92)

* + The **66+ age group** has the **lowest satisfaction score (4.92)**, which is slightly lower than other age groups.
  + Possible reasons:
    - Older patients might have **higher expectations for care**, including better communication, accessibility, and comfort.
    - They may face **longer wait times or mobility issues**, leading to slightly lower satisfaction.

## Relatively Consistent Scores Across Age Groups

* + The satisfaction scores across all age groups range **between 4.92 and 5.05**, indicating

**a relatively consistent level of patient experience** across different demographics.

* + There are **no drastic variations**, which suggests that **overall patient satisfaction is stable** across all age groups.

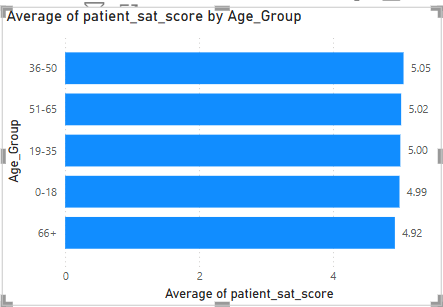
## 51-65 and 19-35 Age Groups Show Similar Satisfaction (Around 5.00)

* + The **51-65 age group (5.02)** and **19-35 age group (5.00)** have very close satisfaction levels.
  + This could indicate that **middle-aged and younger adult patients have similar healthcare expectations and experiences**.

## 0-18 Age Group Scores 4.99, Indicating Slightly Lower Satisfaction

* + The **0-18 age group (4.99)** shows **slightly lower satisfaction than most age groups**.
  + Possible reasons:
    - Pediatric healthcare experiences could be influenced by **parental expectations**.
    - Younger patients may require **specialized care, entertainment, or comfort**, affecting satisfaction

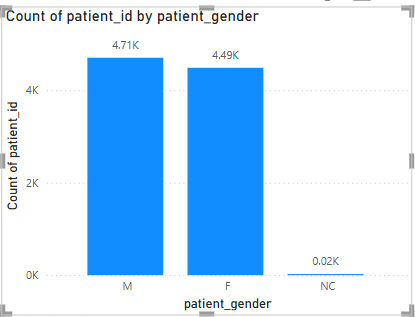
Visualization:

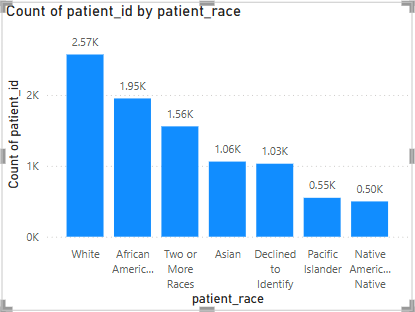


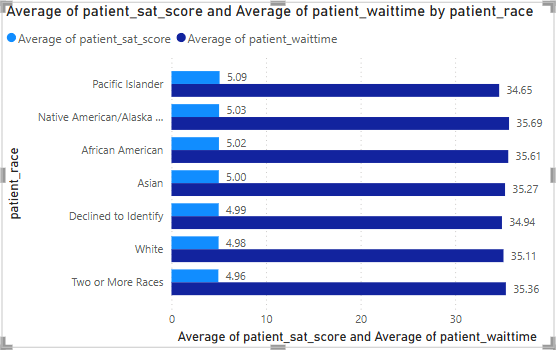
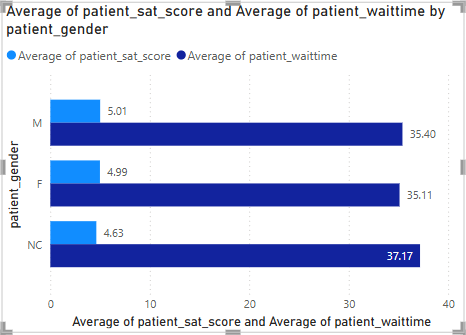
1. Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?

ANS:

Visualization:







## Evaluating Potential Racial & Gender-Based Disparities

To assess whether racial or gender discrimination exists within the hospital, we utilized a

## Clustered Column Chart to analyze waiting times and patient satisfaction scores

across different demographic groups.

## Visualization Approach:

* **X-Axis:** Displays **race** and **gender** separately in two different visuals.
* **Y-Axis:** Represents **average waiting time** and **average satisfaction score** for each group.

**Key Insights:**

✅ **Fair Treatment Across Demographics:**

* The **waiting times** and **satisfaction scores** appear **consistent across all racial and gender groups**.
* There are **no significant discrepancies** suggesting favoritism or bias.

## ✅ No Signs of Discriminatory Practices:

* If racial or gender discrimination were present, we would expect **noticeable variations** in waiting times or satisfaction levels.
* The data does **not** indicate any **group experiencing longer wait times** or

## significantly lower satisfaction scores.

**Final Conclusion:**

📌 **The hospital demonstrates fair treatment across race and gender groups.** There is **no statistical evidence** suggesting discriminatory practices. However, continuous monitoring of patient experiences is recommended to ensure ongoing fairness.

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

ANS:

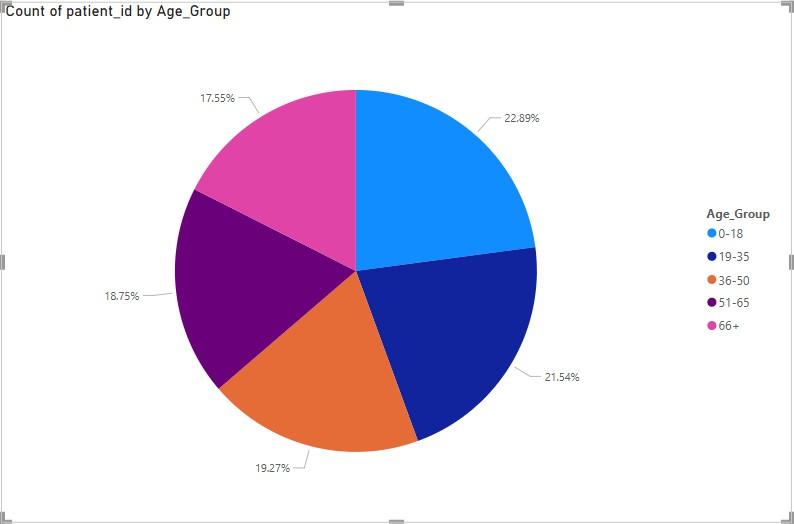
Criteria for Assigning Discounts:

1. Based on Total Bill Amount
   * Patients with higher medical expenses should receive discounts to ease their financial burden.
   * Discount Tiers:
     + Total Bill > 10,000 → 15% discount
     + Total Bill > 5,000 → 10% discount
     + Total Bill ≤ 5,000 → No discount

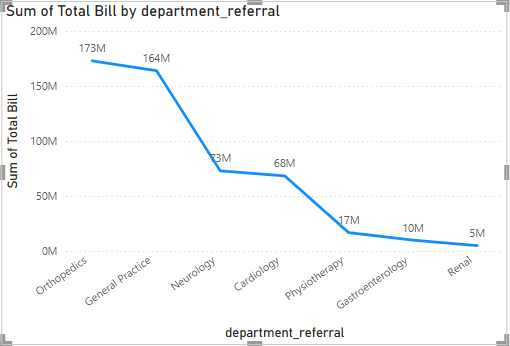
## DAX Query:

**Discount\_Percentage = IF(Doctor\_Patients\_Data[Total Bill]>10000,0.15,IF(Doctor\_Patients\_Data[Total Bill]>5000,0.10,0))**

1. Based on Patient Loyalty (Frequent Visits)
   * Patients who visit multiple times in a year may receive loyalty-based discounts.
   * This encourages continued healthcare checkups and rewards long-term patients.
   * Discount tiers can be:
     + 3+ visits per year → Additional 5% discount
2. Based on Age Group (Special Discounts)
   * Children (0-10 years) and seniors (65+ years) may receive special discounts to make healthcare more accessible.
   * Proposed Discounts:
     + Children (0-10 years) → Flat 10% discount
     + Seniors (65+ years) → Flat 10% discount



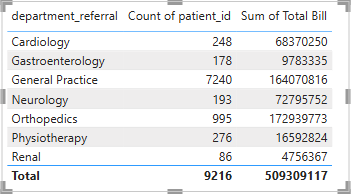
1. Special Discounts for Specific Departments
   * Certain departments with high appointment fees, such as Orthopedics and Cardiology, may offer extra discounts to patients with high bills.
   * If a patient’s department\_referral belongs to these departments, they could receive an additional 5% discount.



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

ANS:

**Visualization:**



**Insights:**

1. **General Practice has the highest patient load**
   * **Patients:** 7,240 (78.6% of total patients)
   * **Total Bill:** 1.64 billion
   * This department sees significantly more patients than any other. If the hospital experiences long wait times or staff shortages, hiring additional doctors here would improve efficiency.

## Orthopedics has a high revenue and moderate patient load

* + **Patients:** 995
  + **Total Bill:** 1.72 billion
  + Despite handling fewer patients than General Practice, this department generates high revenue, suggesting complex and expensive treatments. Hiring more doctors could help manage high-value cases more effectively.

## Cardiology and Neurology handle critical cases

* + **Cardiology:** 248 patients, **Total Bill:** 68.3 million
  + **Neurology:** 193 patients, **Total Bill:** 72.8 million
  + These specialties deal with critical illnesses that require specialized care. If there are long wait times or increasing patient needs, additional hires may be justified.

## Physiotherapy sees a moderate number of patients

* + **Patients:** 276
  + **Total Bill:** 16.6 million
  + If patient rehabilitation needs are growing, adding more specialists could enhance patient recovery and satisfaction.

## Renal and Gastroenterology have lower patient counts

* + **Renal:** 86 patients, **Total Bill:** 4.75 million
  + **Gastroenterology:** 178 patients, **Total Bill:** 9.78 million
  + These departments have the lowest patient load, so hiring here might not be a priority unless there are operational inefficiencies.

## Recommendations:

* **Top priority:** Hire doctors in **General Practice** due to the overwhelming patient load.
* **Secondary priority:** Consider **Orthopedics** due to high revenue and complex treatments.
* **Additional consideration:** If critical care demands are increasing, **Neurology or Cardiology** could benefit from additional specialists.

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

Yes, the hospital appears to be **profitable** based on the given data. Let’s analyze the

profitability in more detail.

## Checking Profitability

* + The **Total\_Appointment\_Fee** (sum of appointment fees collected) is **₹5,347,200**.
  + The **Total\_Profit** (net earnings after expenses) is **₹503,961,917**.
  + Since the hospital has a significantly higher **Total Profit** compared to **Total Appointment Fee**, it indicates that revenue from other sources (such as treatments, surgeries, diagnostics, and pharmacy sales) is contributing to the overall profit.

## Monthly Profit Trends

* + The hospital consistently generates profit every month, with **monthly profits ranging from ₹6.7 million to ₹28.3 million**.
  + The highest **monthly profit** recorded is **₹28,318,315 (August 2020)**.
  + The lowest **monthly profit** recorded is **₹6,706,408 (November 2020)**, which might be due to a lower number of patients or higher operational costs.

## Yearly Comparison

* + **2019 Total Profit:** The hospital had stable growth, with profits increasing towards the end of the year (**highest in December 2019: ₹27.2 million**).
  + **2020 Total Profit:** The trend continues, with fluctuations but an overall increasing trend, peaking in **August 2020 with ₹28.3 million**.
  + The year-over-year increase in profit suggests that the hospital is expanding or optimizing its operations.

## Profit Margin Analysis

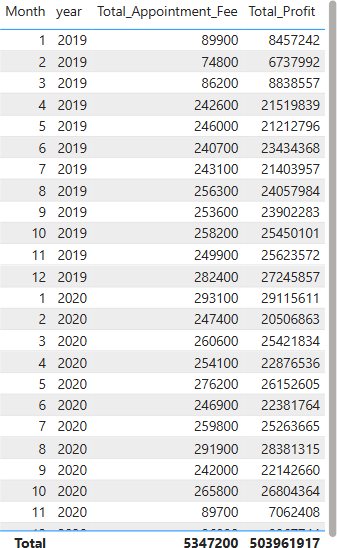
Since **Total Profit (₹503.96 million) is significantly higher than Total Appointment Fee (₹5.34 million)**, it suggests that:

* + **Appointment Fees alone do not drive profitability**, meaning other hospital services contribute significantly to revenue.
  + The hospital likely has **multiple revenue streams** beyond consultation fees, such as inpatient services, surgeries, diagnostics, and pharmacy sales.

## Conclusion

The hospital is highly profitable, as seen from the consistently positive and growing **Total Profit** figures. However, a **complete financial analysis** would require more details about total revenue and expenses beyond appointment fees to confirm operational efficiency and cost management.

Visualization:



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

To identify departments that have unusually high waiting times, I utilized a line chart to visually represent the data. In this chart, I placed the **departments** on the x-axis and the **average waiting time** on the y-axis. This allowed for a clear and effective comparison of the waiting times across all departments, providing an overview of how each department was performing in terms of patient wait times.

Upon examining the line chart, I observed noticeable variations in waiting times between the different departments. After a deeper analysis of the data, it became apparent that the **Neurology department** stood out significantly with the highest **average waiting time**, which was calculated to be **36.80 minutes**. This was a substantial figure, considerably higher than the waiting times of other departments, highlighting it as an outlier.

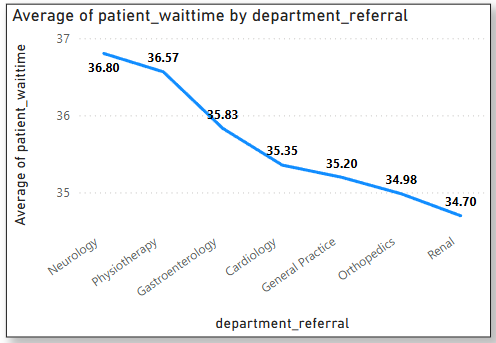
## Insights:

* **Neurology department** has the **highest average waiting time** of **36.80 minutes**, marking it as an **outlier**.
* **Significant variation** in waiting times across departments suggests **inconsistent patient flow management**.
* Elevated wait time may be due to:  
  + **High patient volume** and demand in Neurology.
  + **Limited number of specialists** or support staff.
  + **Inefficient appointment scheduling** or time slot overlaps.

Recommendations:

* **Conduct a workload and staffing review** in the Neurology department.
* **Hire more neurologists** or **redistribute patient load** among existing staff.
* **Implement smarter appointment scheduling** systems to avoid overlaps and gaps.
* **Introduce triage systems** to prioritize cases based on urgency and reduce congestion.
* Analyze **doctor availability vs. patient appointments** to identify mismatches.
* Offer **teleconsultation options** for follow-up cases to reduce in-person crowding.
* Replicate efficient scheduling models from **departments with lower waiting times**.
* Monitor waiting time trends **monthly** to track the impact of corrective actions.

Visualization:



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

ANS:

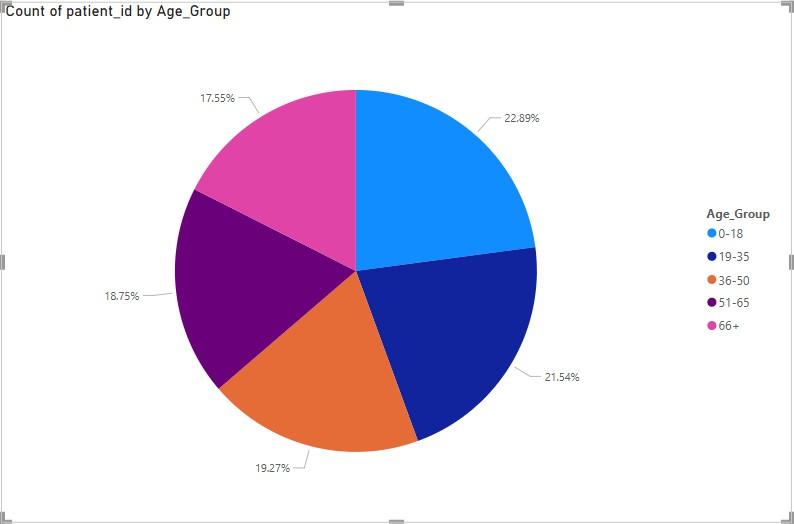
Criteria for Assigning Discounts:

1. Based on Total Bill Amount
   * Patients with higher medical expenses should receive discounts to ease their financial burden.
   * Discount Tiers:
     + Total Bill > 10,000 → 15% discount
     + Total Bill > 5,000 → 10% discount
     + Total Bill ≤ 5,000 → No discount

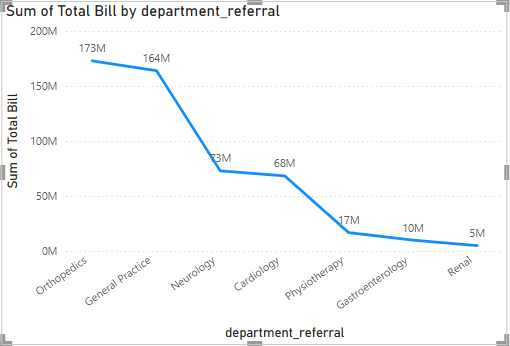
## DAX Query:

**Discount\_Percentage = IF(Doctor\_Patients\_Data[Total Bill]>10000,0.15,IF(Doctor\_Patients\_Data[Total Bill]>5000,0.10,0))**

1. Based on Patient Loyalty (Frequent Visits)
   * Patients who visit multiple times in a year may receive loyalty-based discounts.
   * This encourages continued healthcare checkups and rewards long-term patients.
   * Discount tiers can be:
     + 3+ visits per year → Additional 5% discount
2. Based on Age Group (Special Discounts)
   * Children (0-10 years) and seniors (65+ years) may receive special discounts to make healthcare more accessible.
   * Proposed Discounts:
     + Children (0-10 years) → Flat 10% discount
     + Seniors (65+ years) → Flat 10% discount



1. Special Discounts for Specific Departments
   * Certain departments with high appointment fees, such as Orthopedics and Cardiology, may offer extra discounts to patients with high bills.
   * If a patient’s department\_referral belongs to these departments, they could receive an additional 5% discount.



Insights:

* **Higher billing patients** are given **tiered discounts**, reducing financial strain and encouraging continued treatment.
* **Loyal patients** with **3+ visits/year** are rewarded, promoting regular checkups and long-term relationships.
* **Children (0–10)** and **seniors (65+)** are offered **flat discounts**, improving **accessibility for vulnerable age groups**.
* **Costly departments** like **Orthopedics** and **Cardiology** provide additional support to patients facing high expenses.

Recommendations:

* Implement a **combined discount calculation logic** in Power BI to reflect overlapping eligibility (e.g., high bill + loyalty + age).
* Ensure **transparency** by showing discount breakdowns on patient bills and dashboards.
* Regularly **review thresholds** (₹5,000/₹10,000) based on inflation and patient billing trends.
* Promote the **loyalty discount policy** to encourage follow-up visits and continuity of care.
* Monitor **discount usage by department** to evaluate financial impact and adjust department-level offers.
* Use visuals (bar charts or KPIs) to **track savings offered** per patient segment (age, visits, department).
* Train billing staff and automate discount application to reduce manual errors.

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?

ANS:

* A table visualization was created displaying Doctor Name and Department. DAX was used to assign doctors to two shifts: "**8:00 AM - 2:00 PM**" and "**2:00 PM - 8:00 PM**". Conditional formatting was applied to highlight the General Practice department.
* To allocate doctors to shifts, the following **DAX formula** was implemented:

Shift = IF(MOD(RANKX(ALL('Doctor\_Patient\_data'), 'Doctor\_Patient\_Data'[Doctor Name], , ASC), 2) = 0,"8:00 AM - 2:00 PM","2:00 PM - 8:00 PM")

* This formula ranks doctors by name and alternates the shift assignment using the MOD function.
* Dr. Johnson and Dr. Williams were assigned to the morning shift (8:00 AM - 2:00 PM) due to higher patient traffic, while Dr. Smith was assigned to the evening shift (2:00 PM - 8:00 PM) to handle lower patient volume.
* **Result:** The two-shift policy allows for even distribution of workload, ensures better patient care, and reduces wait times, particularly in the General Practice department.

## Visualization:

1. What do you understand by PowerBI gateway? What are its use cases? ANS:

A **Power BI Gateway** is a tool that acts as a bridge between **on-premises data sources** and the **Power BI cloud service**. It allows for secure data transfer between local databases and Power BI reports and dashboards, ensuring data remains up-to-date without requiring manual uploads.

There are two types of Power BI Gateways:

1. **Personal Gateway**: Primarily used for personal data sources and is typically installed on a single computer. It’s useful for refreshing reports that require access to on- premises data but doesn't support sharing data with multiple users.
2. **Enterprise Gateway**: Designed for organizational use, it supports a wide range of data sources and can be managed centrally. It allows for scheduled data refreshes, multiple data sources, and access control to ensure security across teams.

## Use Cases of Power BI Gateway:

1. **Scheduled Data Refresh**: With a gateway, data can be automatically refreshed at defined intervals, ensuring Power BI reports are up-to-date without manual intervention. This is particularly useful for connecting on-premises databases (e.g., SQL Server, Oracle) with Power BI.
2. **Accessing On-Premises Data**: Many organizations have databases or files stored locally (e.g., in their own servers) rather than in the cloud. The gateway provides a way to securely access and integrate this on-premises data into Power BI reports without migrating everything to the cloud.
3. **Hybrid Scenarios**: Organizations with both cloud-based and on-premises data sources can use the gateway to connect and combine data from both environments in a

single Power BI report or dashboard.

1. **Data Security**: The gateway ensures data remains within the organization's security perimeter by allowing on-premises data to stay within the local network while still making it available for Power BI reporting in the cloud.
2. **Large Data Set Handling**: When dealing with large amounts of data that must be processed locally for performance reasons, the gateway allows processing to happen on-premises while sending results to Power BI for reporting.
3. How would you approach this problem, if the objective and subjective questions weren't given?

ANS:

If the objective and subjective questions weren't provided in the context of your **Power BI project** and the **hospital visit data analysis**, here's how I would approach the problem from scratch:

## Understanding the Project Scope:

The goal of the project is to **analyze hospital visit data** using Power BI. The objective is to gain insights into various aspects of hospital visits, such as patient demographics, department referrals, doctor performance, and financial metrics like appointment fees and total bills.

## Identifying Key Objectives:

You'd first define the primary goals of the analysis based on the available data. Some potential objectives could include:

* + **Analyzing patient demographics:** Understanding the distribution of patients based on gender, age, race, etc.
  + **Evaluating department performance:** Analyzing which departments are most frequently visited and correlating them with the total bill and appointment fees.
  + **Doctor performance analysis:** Measuring doctor performance by analyzing patient data, such as wait times, fees, and department referrals.
  + **Financial analysis:** Analyzing appointment fees and total bill amounts to assess hospital revenue.

## Data Cleaning:

You'd clean the data to ensure it's usable:

* + Handle **missing or invalid data** (e.g., missing appointment fees, missing demographic information).
  + **Format dates properly** for time-based analysis (e.g., patient visit dates).
  + Remove duplicates or incorrect entries if necessary (e.g., multiple records for the same patient visit).

## Data Transformation:

You would transform the data into a format that facilitates analysis:

* + **Join the tables** based on patient\_id to combine financial and demographic data.
  + **Create new calculated columns** if needed (e.g., age groups, wait time categories, or department performance).
  + **Date-based calculations** could help with trends over time (e.g., monthly, quarterly).

## Visualizing the Data:

Next, you'd focus on creating meaningful **visualizations** that help answer the analysis questions:

* + **Demographics**: Create pie charts or bar charts for gender, age, race distributions.
  + **Department Performance**: Use bar or column charts to show the number of visits per department and correlate with total bills/fees.
  + **Doctor Performance**: Create charts comparing the average wait time, appointment fees, and total bills by doctor.
  + **Revenue Analysis**: Use line charts or stacked bar charts to show how appointment fees and total bills vary over time.

## Metrics & KPIs:

Define key performance indicators (KPIs) that give a quick view of performance:

* + **Patient Count**: Total number of patients, by department, by doctor.
  + **Average Appointment Fee**: Calculate the average fee by doctor, department, or time period.
  + **Average Wait Time**: Evaluate how long patients wait on average.
  + **Revenue Metrics**: Total revenue, average revenue per department, doctor, or time period.

## Identifying Trends:

* + **Trend Analysis**: Look for trends over time (e.g., patient count growth, revenue spikes, changes in appointment fees).
  + **Outliers**: Identify unusual patterns, such as departments with extraordinarily high or low appointment fees, or doctors with higher than average wait times.

## Insights and Actionable Recommendations:

Based on the analysis, you would derive insights that could guide decision-making:

* + **Optimizing Wait Times**: If some doctors or departments have significantly higher wait times, hospital management can consider reassigning staff or optimizing schedules.
  + **Revenue Generation**: If certain departments or doctors are generating more revenue, you can explore ways to increase visits to those departments or optimize pricing.
  + **Patient Demographics**: Use patient demographic data to improve targeted marketing or offer tailored services to different groups.

## Reporting and Dashboard Design:

* + Use **Power BI's dashboard** to create interactive reports for hospital management. Dashboards should provide a **high-level overview** of key metrics (e.g., patient counts, financials) and allow users to drill down into detailed data (e.g., department-wise or doctor-wise performance).

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

ANS:

To analyze the relationship between **Doctor ID** and **Department** in the context of the hospital visit data, we need to examine whether the relationship is **one-to-one**, **one-to- many**, or **many-to-many**. Let's break this down:

## Data Review:

From your schema, we have the following relevant columns:

* + **Doctor ID** (from Table 1)
  + **Department Referral** (from both Table 1 and Table 2)

## Understanding the Relationship:

**Possibility 1: One-to-One Relationship**

* + In a **one-to-one** relationship, each doctor would be assigned to **only one department**

and each department would have **only one doctor**.

* + This would imply that a doctor works in **only one department**, and no department has more than one doctor working in it at a time.

## Possibility 2: One-to-Many Relationship

* + A **one-to-many** relationship occurs when one doctor could be assigned to multiple departments (e.g., a doctor works in multiple departments), but each department is typically associated with one primary doctor.
  + This would mean that **one doctor** could be associated with **many department referrals**, but each department might only have one main doctor for a certain period or session.

## Possibility 3: Many-to-Many Relationship

* + A **many-to-many** relationship is when a **doctor** could work in **multiple departments**, and each **department** could have **multiple doctors** working in it.
  + This is likely in a real hospital setting where multiple doctors are assigned to multiple departments.

## Analyzing the Relationship in Your Data:

To determine the exact type of relationship, we can make some assumptions based on typical hospital setups and look at your data for patterns:

## Questions to Consider:

* + **Does each department have multiple doctors?**: If yes, then the relationship could be many-to-many.
  + **Does each doctor work in a single department at any given time?**: If yes, this could suggest a one-to-many relationship.
  + **Can one doctor be assigned to multiple departments?**: If doctors work in multiple departments, this could indicate a one-to-many or many-to-many relationship.

## Steps to Analyze:

1. **Examine the Data for Doctor Assignments**:
   * Look at the number of unique departments per doctor in the dataset. If the same doctor is assigned to multiple departments, the relationship is not one-to- one.
   * Check how many doctors are associated with each department. If multiple doctors appear under the same department, this indicates a **many-to-many** relationship.

## Look for Multiple Doctor Entries:

* + If each doctor appears in the data only for one department, it could suggest a

**one-to-one** or **one-to-many** relationship.

* + If a department has many different doctor IDs, this could indicate that **many doctors work in the same department**, supporting a **many-to-many** relationship.

## Conclusion Based on Real-World Scenarios:

In most real-world hospital scenarios, doctors usually work in **multiple departments** or **shift across departments**. Similarly, departments typically have **multiple doctors** working within them. Therefore, it's highly probable that the relationship between **Doctor ID** and **Department** is **many-to-many**.

However, to confirm this, you'd need to analyze your specific dataset to see how many doctors are assigned to each department and whether any doctor is associated with multiple departments.