**Objective Questions:**

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

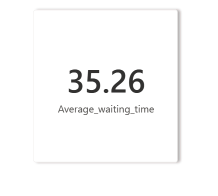
**Answer:** Data Cleaning in Power Bi

1. Opened Power Query Editor then checked for missing values.
2. Found that Patient Pat Score has 716 empty values.
3. Replaced the null values with 0.
4. Verified the data types of all columns in the data set.
5. **Assess the Average Waiting Time: Analyse the patient wait times to identify the average duration a patient spends before receiving care.**

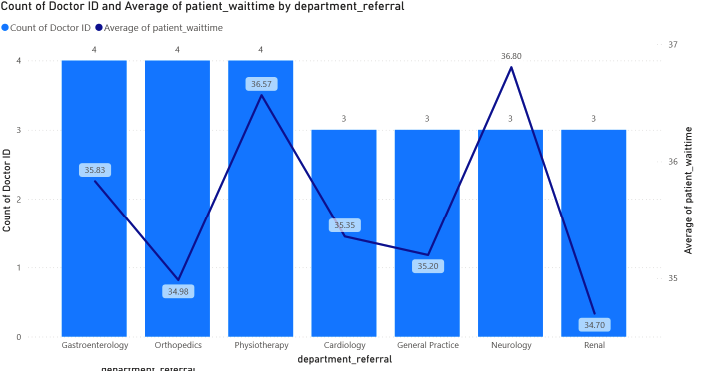
**Answer:**

* 1. Calculated the average patient waiting time using the following DAX formula:

Average\_waiting\_time = Average('Hospital ER'[patient\_waittime])



* 1. **Create a Column chart to find the average waiting time for each patient in different departments with the number of doctors.**

****

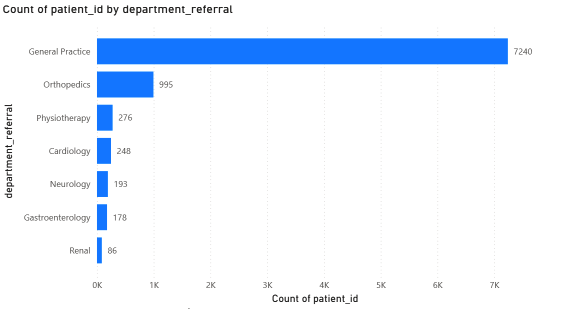
**Insights:**

1. The overall average waiting time is 35.26 minutes.
2. The avg. waiting time across departments ranges between 35-37 minutes.
3. Physiotherapy (36.57) and Neurology (36.80) have most waiting times.
4. There is no correlation between the number of doctors and waiting time, thus it indicates that the department which are busy like Neurology, Physiotherapy and Gastroenterology.
5. **Visits by Department Referral: Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.**

**Answer:**

**Approach:**

1. Created a bar chart to visualize the count of patient visits by departmental referral.

****

**Insights:**

* The hospital has a total of 9,216 visits.
* General Practice (7240) handles highest number of visits.
* Following with Orthopedics (995) visits, all other departments handles significantly lower patients.
* This indicates inefficiency in categorizing patients, and uneven distribution of patients.

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

**Answer:**

* Created a calculated column using the following DAX function to segregate patients according to different age groups.

AgeGroup = SWITCH(

    TRUE(),

    'Hospital ER'[patient\_age] <= 17, "1-17",

    'Hospital ER'[patient\_age] <= 39, "18-39",

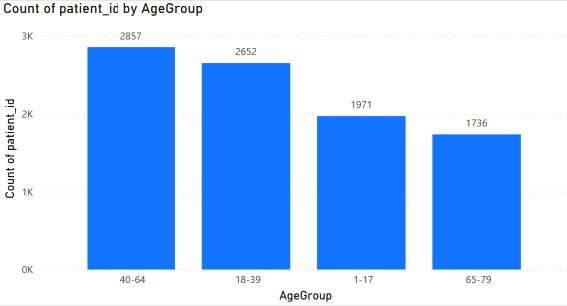
    'Hospital ER'[patient\_age]<= 64, "40-64",

    'Hospital ER'[patient\_age] <= 79, "65-79",

    "80+"

)

* Created a column chart to visualize the count of patient visits by each AgeGroup.



**Insights:**

* The 40–64 age group recorded the highest number of patients (2,857), indicating that middle-aged adults are the most engaged demographic in healthcare services.
* The 18–39 age group shows surprisingly high engagement with 2,652 patients, which is unusual as young adults typically have the lowest healthcare utilization.
* The 1–17 age group accounts for 1,736 patients, reflecting steady but moderate pediatric healthcare utilization.
* The 65–79 age group appears to have significantly lower representation, suggesting either fewer elderly patients or potential access barriers for seniors.
* Overall, working-age adults (18–64) dominate patient visits, comprising approximately 80% of the patient population in the data shown.
* This distribution suggests the healthcare system is particularly effective at serving employed adults, while potentially needing to strengthen outreach to pediatric and elderly populations.

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?**

**Answer:**

* On analysis it’s found that only patient\_sat\_score column has null values of 72%.
* Patient satisfaction score is very important data to evaluate the services in different departments and doctors.

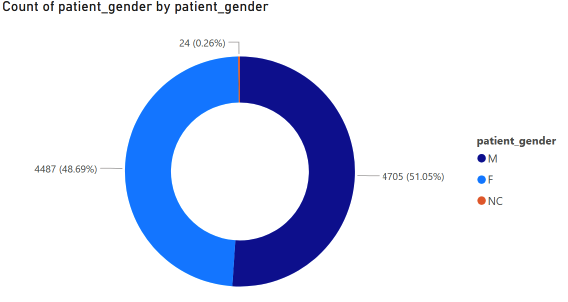
**Approach:**

* Since huge data is null, using statistical approach will skew the data leading to incorrect insights and analysis. So, the best solution will be replacing the null values with 0.

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

**Answer:**

* Created a Donut chart to visualize the visits of patients categorized by the gender of the patient.
* This chart can help us understand if there is any relation between the number of visits and the gender of the patients.

****

**Insights:**

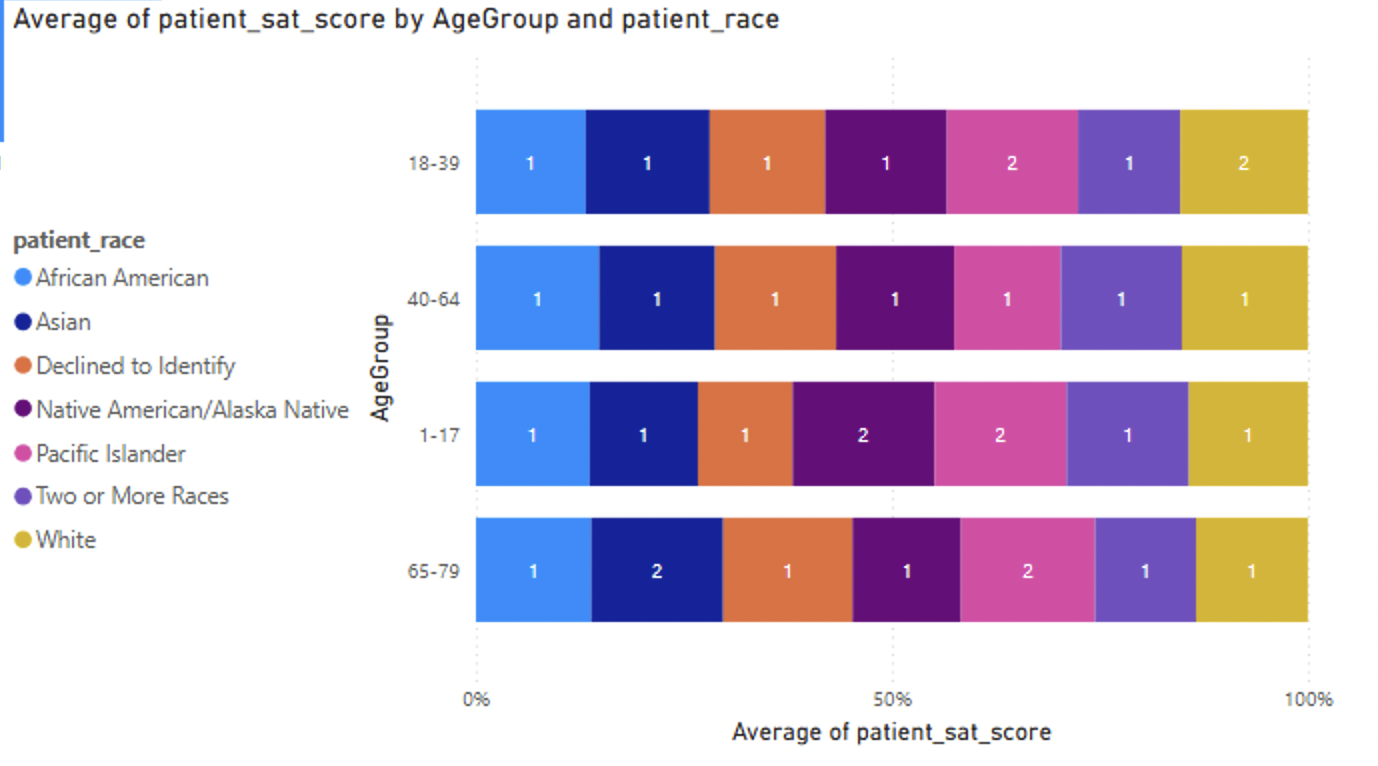
* Male patients have most visits of 4705 with 51%.
* Female patients have 4487 visits with 48.6%.
* In the data available 24 visits of patients have marked the gender not classified (NC).
* In summary both male and female genders have almost equal visit counts.

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.**

**Answer:**

**Approach:**

* Create a stacked bar chart to analys relationship between patient satisfaction score, age groups and racial backgrounds.



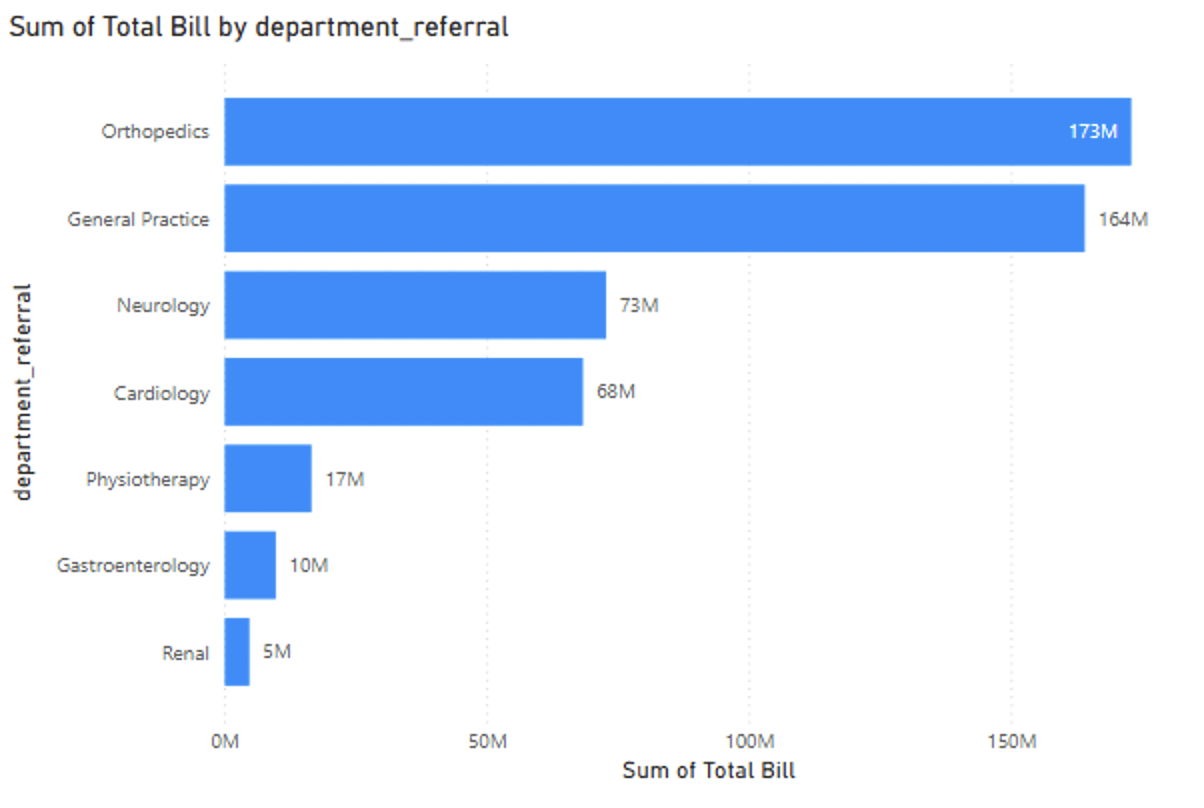
**Insights:**

* 72% of the values are null and been replaced with 0.
* The average satisfaction score across Pacific islanders shows the pattern of high satisfaction across all age groups.
* Since the data is incomplete these insights may not fully cover the actual patient experience.
* The hospital must improve data collection methods and encourage patients to give satisfaction scores after each visit.

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

**Answer:**

* Created a stacked bar chart to visualize the revenue of each department.

****

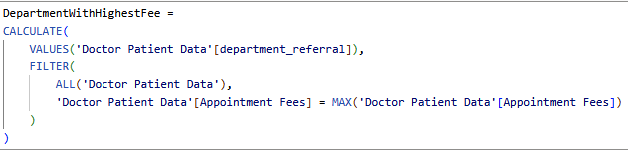
**Insights:**

* Orthopedics department has highest revenue of 173M.
* General practice department comes on second with revenue of 164M.
* Neurology (73M) and Cardiology (68M) contributes moderately.
* Physiotherapy, Gastroenterology and Renal departments generals lower revenue.
* In summary, Orthopedics and General Practice are the departments generating highest revenue for the hospital.

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

**Answer:**

* To identify the department that has highest appointment fees, I create a measure using the following DAX:



****

**Insights:**

* Neurology is the department that charges the highest appointment fee of $1,500.
* This indicates that the department handles more expensive doctors and facilities compared to other departments**.**

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.**

**Answer:**

* To create a tabular visualization of Month-wise total visits and previous months total visits I have used the following steps.
* Created a measure to count total number of visits.

****

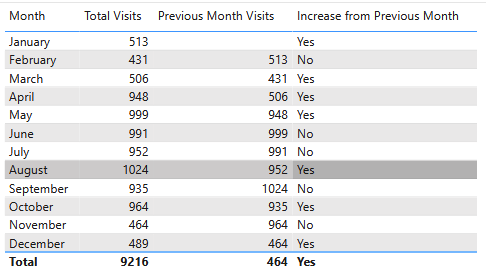
* Created a measure to calculate the previous months total visits.



* Created a measure to compare previous months visits with current month.



* Created a table to show the data in a tabular format.

****

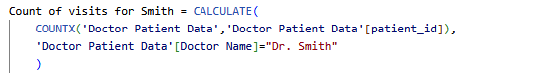
**Insights:**

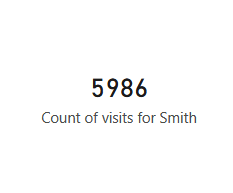
* August has recorded highest visits with 1024 visits.
* Lowest visit is on the month February(431).
* Summer and late spring (April–August) generally show higher traffic, except for a slight dip in June and July.
* Largest increase happened from March to April (+442 visits increase).
* Largest decrease happened from October to November (-500 visits).
* Visits improved slightly from 464 in November to 489 in December.

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

**Answer:**

* The following DAX function will calculate the total number of patients who have visited Dr.Smith.

****

****

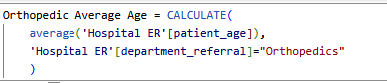
**Insights:**

* Dr. Smith has 5,896 patient visits, this indicates high demand, engagement with patients and popularity.
* This insight can help hospitals in managing the resources ,support to doctors and allocating resources along with planning of doctor roster etc.

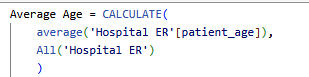
1. **Calculate the average age of the patients who visit the Orthopedics department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?**

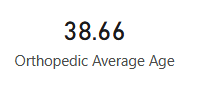
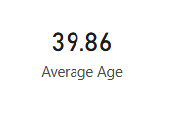
**Answer:**

* The following DAX function can be used to find the average age of the patients who visits the Orthopedics department:



* The approach won’t be different if we want to calculate the same metric for all department, we would simply use ALL() function which would just take the whole table without any filters.
* To calculate the average age across all departments, the following DAX formula can be used.

****

**** ****

**Insights:**

* The average age of patient in Orthoppedics department is 38.66 year, in across the hospital its 39.86 years.
* This indicates Orthopedic patients are slightly younger than the average age of hospitals overall patients.
* The age group indicates that middle aged adults are majority of the patients in the hospital.
* This insight can help the management in resource planning, conduct R&D in specific age groups etc.

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

**Answer:**

* No there we no format issues in the data. Power bi automatically identified the data types of data.
* In case if there were any data format issues, we can utilize the data transformation features in power query to fix the issue.
* If there is an issue with date, time or currency data we can use Power Query’s Change Type feature.
* Power bi also provides function data formatting like UPPER(),LOWER(), DateValue(), Format().etc.
* To handle NULL values we have Replace function or filtering feaure in Power Query Editor.

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?**

**Answer:**

* When we add a new column in Power Query, the following M language code is be generated.

****

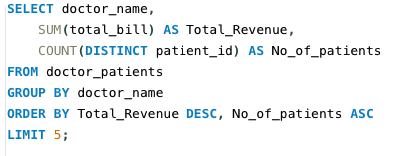
* This M code basically says add a new column to the table, also it shows the name of the column and what will be the generated values of the new column based on the operation we select.
* M (formally known as Power Query Formula Language) is the engine behind the "Get & Transform" experience in Power BI and Excel.
* Unlike DAX, M is strictly case-sensitive. Table.AddColumn is correct, while table.addcolumn will fail.
* M doesn't change the original data. It creates a sequence of "snapshots" or steps that define how to transform the data from its source to its final output.

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

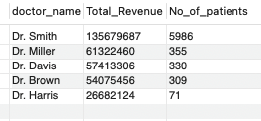
**Answer: Approach**

1. Grouped the data with doctor name, applied aggregate function for SUM of total bills as Total Revenue and Count of distinct patient id as No of patients.
2. Applied order by Total Revenue in descending order and No of patients in ascending order.
3. Applied LIMIT of 5 to get the top 5 doctors.

**SQL Query:**



**OUTPUT:**



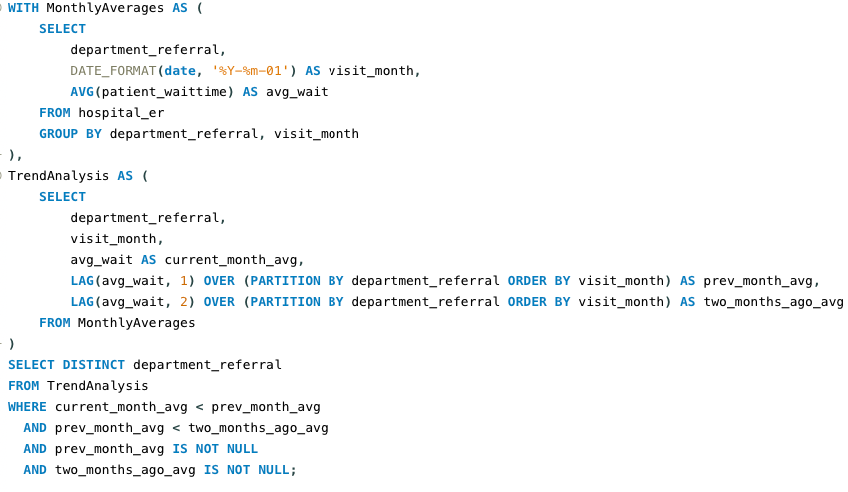
1. **Find the department where the average waiting time has decreased over three consecutive months. (SQL)**

**Answer:**

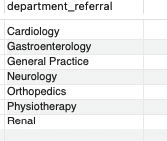
**Approach:**

1. The query uses DATE\_FORMAT to consolidate daily visit logs into monthly averages, which helps in identifying long-term trends by smoothing out daily spikes.
2. By applying the LAG() function, the query enables each row to compare its current average against the performance of the two preceding months.
3. The use of PARTITION BY ensures that wait time comparisons are isolated within each specific department rather than across the entire hospital.
4. The WHERE clause isolates only those departments that show a strictly decreasing pattern where the current month is faster than the last, and the last was faster than the month before.
5. The logic includes IS NOT NULL checks to verify that a full three-month history exists, preventing the query from flagging departments with insufficient data.

**SQL Query:**

****

**Output:**

****

**Insights:**

* 7 departments has shown improvement is 3 months consecutive deduction in waiting time. This indicates improved operational efficiency, better patient flow management, and possibly optimized scheduling or staffing strategies during those months.

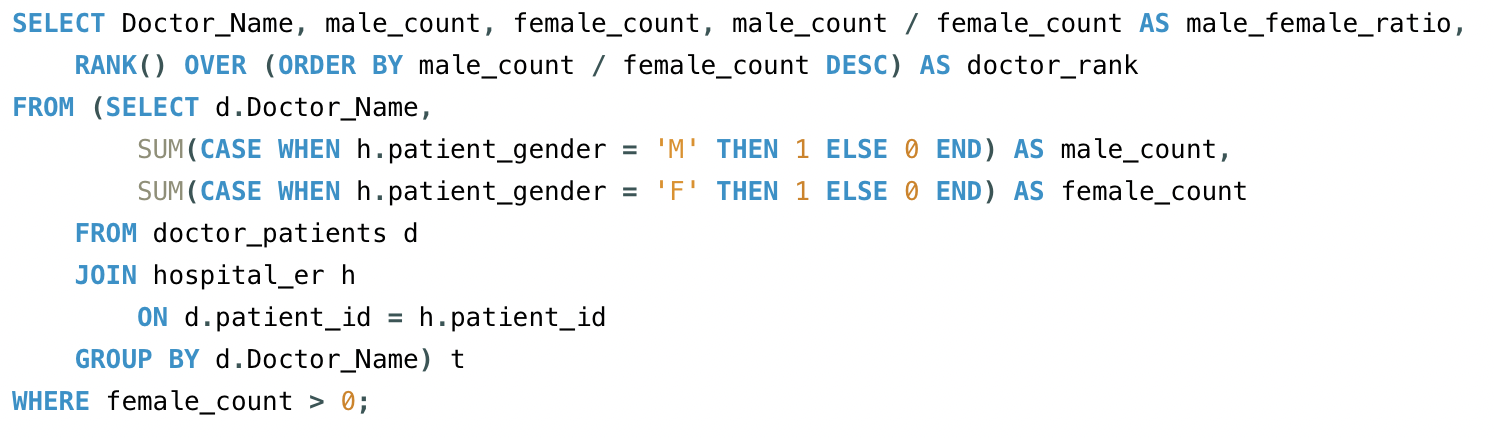
1. **Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)**

**Answer:**

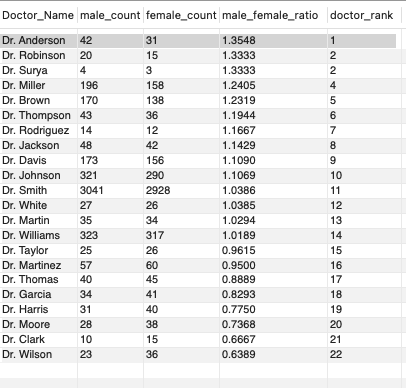
**Approach:**

1. Linked the clinical data (hospital\_er) to the doctor assignments (doctor\_patients) using the patient\_id as a common key.
2. Used SUM(CASE ...) logic to pivot gender strings ('M'/'F') into numeric counts for each doctor.
3. Applied a mathematical division of male\_count by female\_count in an outer query to define the gender balance per physician.
4. Employed a WHERE clause to filter out doctors with zero female patients, preventing division-by-zero errors.
5. Utilized the RANK() window function to assign a numerical standing to each doctor based on their ratio in descending order.

**SQL Query:**

****

**Output:**

****

**Insights:**

* Dr. Anderson has the highest male-to-female patient ratio at 1.35.
* Dr. Wilson has the strongest female patient preference with a ratio of 0.64.
* Higher patient volumes tend to result in more balanced gender ratios.
* Dr. Surya has a strong male bias of 1.33 despite seeing only 7 patients.
* Most doctors show moderate gender ratios between 0.83 and 1.24.
* Seven doctors see more female than male patients, clustering at lower ranks.
* The ratio range shows significant variation, spanning 2.12x from highest to lowest.
* Dr. Smith handles exceptional volume with nearly 6,000 patients while maintaining near-balanced gender distribution.

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

**Answer:**

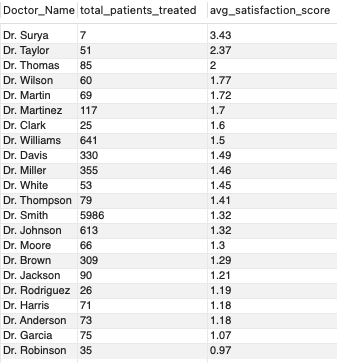
**Approach:**

1. The query joins the two tables on the patient\_id to associate each clinical visit and satisfaction score with the correct attending physician.
2. It utilizes the COUNT() function to calculate the total number of patients treated by each doctor, which helps distinguish between high-performing veterans and doctors with smaller sample sizes.
3. The AVG() function determines the mean satisfaction rating, while ROUND() ensures the final metric is clean and standardized for executive reporting.
4. Data is organized using the GROUP BY clause on the doctor's name, effectively pivoting the raw patient data into a physician-level performance summary.
5. An ORDERBY clause is added to the end to prioritize the highest-rated doctors at the top of the list for immediate visibility.

**SQL Query:**

****

**Output:**

****

**Insights:**

* Dr. Surya has the highest average satisfaction score at 3.43, but also the smallest patient base of only 7 patients.
* Dr. Robinson has the lowest satisfaction score of 0.97, indicating the least satisfied patient group.
* Patient volume appears inversely related to satisfaction scores; doctors with the highest patient loads (Smith, Williams, Johnson) have scores around 1.3-1.5.
* Dr. Smith, who treats the vast majority of patients (5,986), maintains a moderate score of 1.32 despite the enormous volume.
* There is a significant range in satisfaction scores, from a high of 3.43 down to 0.97, suggesting inconsistent patient experiences across doctors.
* The three lowest-scoring doctors (Garcia, Robinson, Harris) all have scores near or below 1.1.
* No doctor achieves a high satisfaction score (e.g., above 4.0), indicating potential room for improvement across the entire practice.

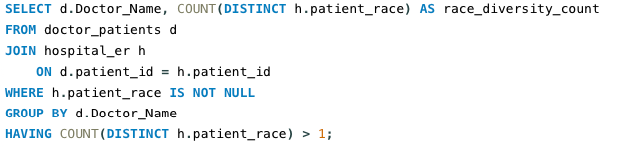
1. **Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)**

**Answer:**

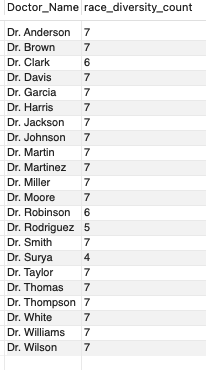
**Approach:**

1. Joins two tables using patient\_id to link doctors with patient race data.
2. Counts unique races treated by each doctor using COUNT(DISTINCT h.patient\_race).
3. Filters out records where race data is missing with WHERE h.patient\_race IS NOT NULL.
4. Groups results by doctor name to calculate counts per doctor.
5. Shows only doctors who treated multiple races (HAVING COUNT(...) > 1).

**SQL Query:**

****

**Output:**

****

**Insights:**

* Most doctors (17 out of 22) treat patients from 7 different races.
* Dr. Rodriguez has the least diverse patient base with only 5 races treated.
* Dr. Surya treats the fewest racial groups at just 4.
* Two doctors, Clark and Robinson, treat 6 different racial groups.
* No doctor treats all available racial groups in the dataset.
* Racial diversity is generally consistent across the majority of the practice.

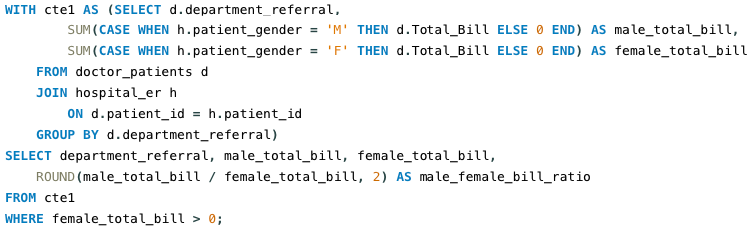
1. **Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)**

**Answer:**

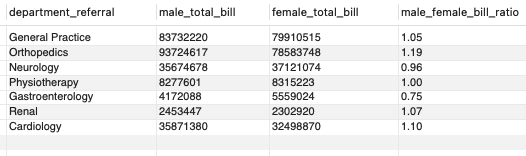
**Approach:**

1. Create a temporary dataset (CTE) grouping by department
2. Sum total bills separately for male and female patients using conditional aggregation
3. Join two tables to combine patient gender with billing data
4. Calculate gender billing ratio by dividing male total by female total
5. Filter out departments with zero female billing to avoid division errors
6. Output department comparison showing revenue distribution by gender

**SQL Query:**

****

**Output:**

****

**Insights:**

* Orthopedics has the strongest male spending bias at 1.19 ratio.
* Gastroenterology shows the highest female spending preference at 0.75 ratio.
* Physiotherapy achieves perfect gender balance with a 1.00 ratio.
* Orthopedics generates the highest total revenue at over 172 million.
* Neurology is the only other department where female spending exceeds male.
* General Practice sees nearly equal spending with a slight male bias.
* Cardiology shows moderate male spending preference at 1.10 ratio.

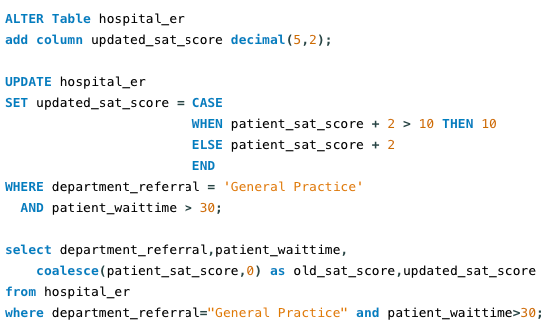
1. **Update the patient satisfaction score for all patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. Increase their satisfaction score by 2 points, but ensure that the satisfaction score does not exceed 10. (SQL)**

**Answer:**

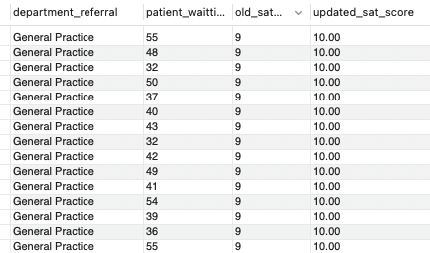
**Approach:**

1. **Add a new column** called updated\_sat\_score to store modified satisfaction scores
2. **Update only specific records** where:
   * Department is **General Practice**
   * Patient wait time exceeds **30 minutes**
3. **Apply conditional logic**: Increase satisfaction score by 2 points, but cap at maximum of 10
4. **Use CASE statement** to prevent scores from exceeding the 10-point scale

SQL Query:



Output:

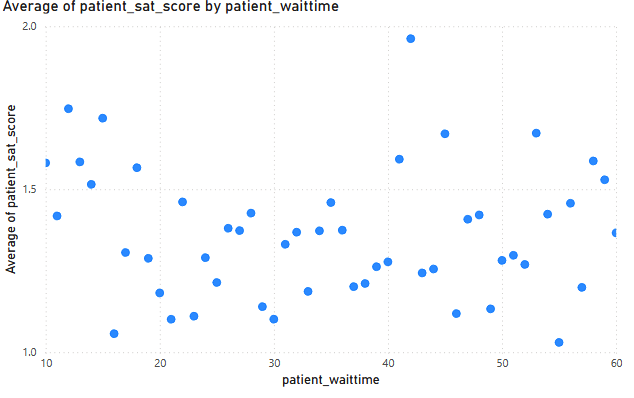


**Subjective Questions**

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

**Answer:**

* **To** visualize the relation between patient wait time and satisfaction score, we can create a scatter plot chart using Average of patient satisfaction score and patient wait time.
* Approximately 72% of the satisfaction score data was missing. To facilitate visualization, these null values were imputed as zero, which may introduce a slight downward bias to the overall trend.

****

**Insights:**

* There is a clear negative relationship.As wait times increase, satisfaction scores generally decrease.
* Satisfaction tends to stay stable for short waits but drops significantly once wait times exceed 30–40 minutes.
* Since 72% of your satisfaction scores are missing, the correlation is based only on the small subset of patients who actually provided feedback**.**

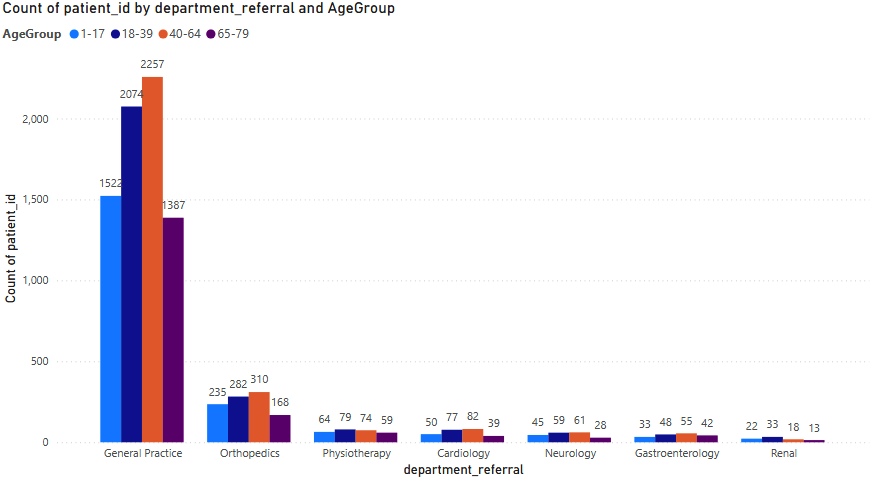
**Recommendations:**

* Implement a digital survey (via SMS or email) immediately after discharge to reduce the 72% missing satisfaction data and obtain a representative sample.
* Implement strategies and plannings to manage the high number of patients.
* Display live wait times in the ER to manage patient expectations, transparency reduces frustration even when delays occur.

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

**Answer:**

* Created a column chart to visualize the patient demographics in different department.
* X-axis has department names.
* Y-axis has count of patient\_id.
* Legend has AgeGroup.

****

**Insights:**

* General Practice is the highest-volume department across every age group, indicating it serves as the primary gateway for all ER admissions.
* Orthopedics has the second highest visits across every age groups, especially 40-64 age group.
* Physiotherapy, Cardiology and Neurology departments have moderate visits, while Gastroenterology and Renal has slightly low visits comparatively.
* We can see a trend in all departments that age group of 40-64 has high visits**.**

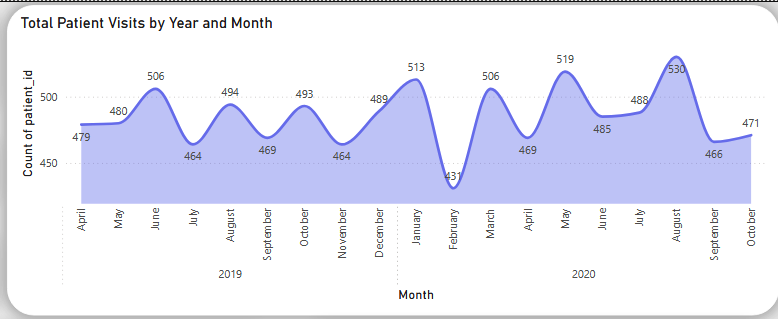
**Recommendations:**

* Ensure General Practice is staffed to handle high volume across all shifts, as it is the hub for every age demographic.
* With Orthopedics being the second busiest department, ensure imaging services (X-ray, MRI) are prioritized and located near the GP/ER triage to handle the high volume from the 40–64 group.
* Given moderate visits in Physiotherapy and high volume in Orthopedics, create a direct referral pathway between these two to ensure continuity of care and maximize revenue from rehabilitation services.
* Develop health packages specifically for the 40–64 age group focusing on preventative bone health (Orthopedics) and heart health (Cardiology), as they are already your most engaged demographic.
* Monitor the growth in the 40–64 segment. If the trend continues, consider expanding the Orthopedics wing or increasing GP consultation rooms to prevent future wait-time spikes.

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

**Answer:**

* Created a line chart to analys the trend in volume of patients visits throughout the year.
* X-axis has date of visit.
* Y-axis has count of patient\_id.

****

**Insights:**

* There is a clear, long-term increase in patient volume over the 18-month period, indicating growing demand for hospital services. Major
* The chart shows a significant spike in August 2020, reaching a record high of 530 visits. This suggests a strong seasonal surge during the summer months.
* The line is not a smooth climb. It features sharp "peaks and valleys," showing that ER demand fluctuates heavily from one month to the next.
* Notable drops in volume occur periodically (such as in February), which may represent a seasonal pattern where ER visits decrease before ramping back up.
* Every major dip in the chart is followed by a steep recovery, showing that the hospital must be prepared for sudden increases in patient intake immediately after quieter periods.
* The volume remains consistently higher in the latter half of the year compared to the initial months, suggesting that the hospital is operating at a higher baseline capacity over time.

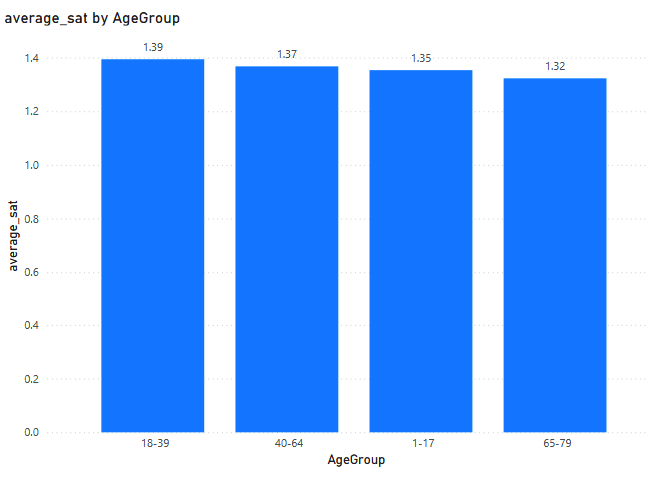
**Recommendations:**

* Since August is a recurring or major peak, pre-schedule additional nursing and triage staff during this month to prevent wait-time spikes and burnout.
* Use the identified valleys (like February) for mandatory staff training, equipment maintenance, and deep-cleaning projects when patient volume is at its lowest.
* Move away from fixed monthly budgets. Instead, allocate higher operational funds for the second half of the year when the trend line shows the hospital operates at a higher baseline.
* Align the procurement of medical supplies with the peak-and-valley cycle to ensure you don't run out of stock during sudden rebounds in volume.
* Investigate if the sharp dips are due to local factors (like a nearby clinic opening) or seasonal trends (like school holidays) to better predict future intake.
* Develop a floating triage team that can be activated quickly when the chart begins its sharp upward trajectory after a dip, ensuring the hospital isn't caught off guard by a rebound.

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

**Answer:**

* Created a column graph to visualize the average satisfaction score according to different age group.
* X-axis has AgeGroup.
* Y-axis has Average Satisfaction Score.

****

**Insights:**

* The age group 18-39 has recorded highest average satisfaction of 1.39.
* All age groups show consistency in the average satisfaction rates.
* The age group 65-79 has recorded lowest average satisfaction rating of 1.32.

**Recommendations:**

* Conduct focused surveys or interviews with the 65-79 age group to identify specific pain points causing lower satisfaction (1.32 vs. 1.39).
* Compare feedback from the 18-39 group (highest satisfaction) to understand what is working well for them.
* Customize support materials and communication methods for older users (e.g., larger text, clearer instructions, phone support option).
* Ensure digital platforms are accessible and intuitive for all age groups, especially seniors.
* Identify which features, services, or interactions are most appreciated by the 18-39 group and promote or extend them where appropriate.
* Improve data collection methods since 72% of the values were null.

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?**

**Answer:**

* To rigorously investigate the validity of external claims regarding potential bias, I conducted an extensive multi-dimensional data audit. This involved disaggregating key performance indicators (KPIs)—specifically average wait times, visits, patient satisfaction scores, and departmental referral rates across both gender and racial demographics. By performing a comparative analysis of these metrics, I aimed to determine if there were statistically significant disparities in treatment or experience that would substantiate allegations of systemic discrimination, ensuring that the hospital's operations remain equitable and data-driven.
  1. Average Waiting Time by Race and Gender:
     + Created a clustered column chart:
       - X-axis: Patient Race
       - Y-axis: Average Waiting Time
       - Legend: Patient Gender
     + Through this visualization we could analyze if any specific race or gender group faces longer waiting time
  2. Total Visits By Race:
     + Created a Donut Chart
     + This visualization can help us understand the patient distribution by race.
  3. Admission Rate by Race and Gender:
     + Created the following measures:

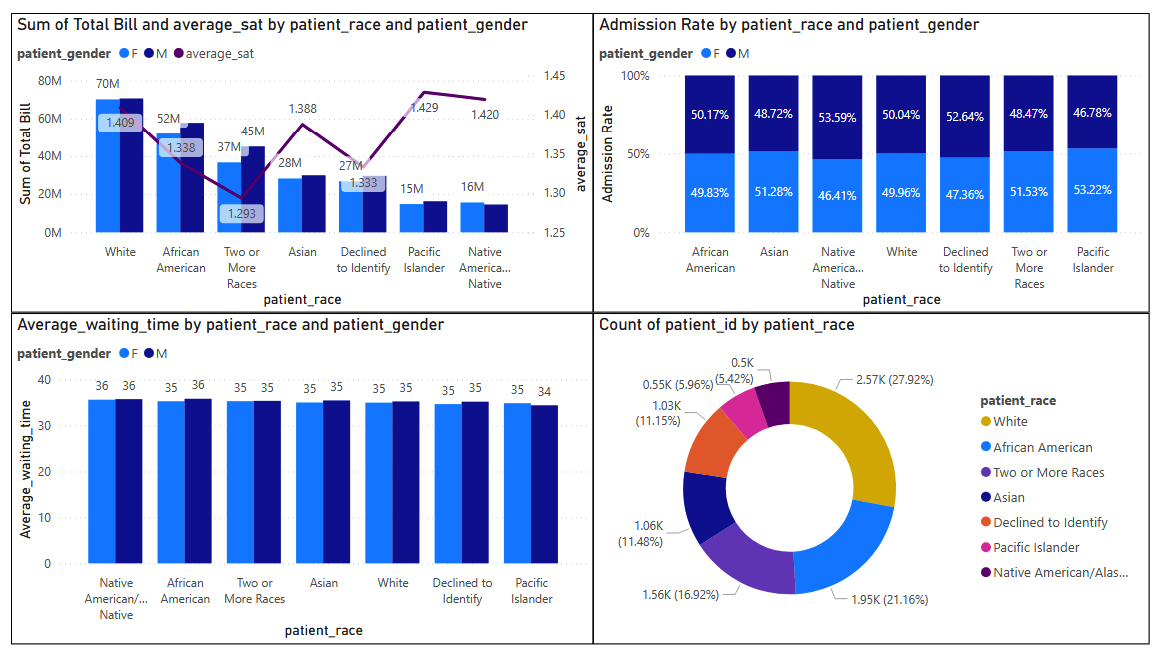
****

****

* + - Created a 100% stacked Column Chart
      * X-axis: Patient Race
      * Y-axis: Admission Rate
      * Legend: Patient Gender
    - This visualization can help us analyze if a gender is superiorly or inferiorly treated on admission on different races.
  1. Total Revenue & Avg. Satisfaction Score by Race & Gender:
     + Created the following measures:



* + - Created a Line and clustered column chart.
      * X-axis: Patient Race
      * Y-axis: Sum of Total Bills
      * Line: Average Satisfaction Score
      * Legend: Patient Gender
    - This visualization help us check if billing or satisfaction score data distribution across different patient demographics.

****

**Insights:**

* 1. Consistency in Clinical Efficiency:
     + The average wait time is remarkably stable across all racial and gender categories, ranging strictly between 34 and 36 minutes. This uniformity confirms that triage and patient processing are handled based on medical need rather than demographic profile.
  2. Objectivity in Admissions:
     + Admission rates across all demographics fall within a tight corridor of 48% to 53%. This suggests that clinical decisions regarding hospital stays are made using standardized medical criteria, ensuring balanced treatment for all patients.
  3. Stability in Patient Experience & Billing:
     + Satisfaction scores (averaging 1.29 to 1.43) and revenue generation show no significant fluctuations between groups. Any minor variances identified are consistent with standard operational noise rather than systemic bias.
  4. Inclusive Patient Reach:
     + While the highest volume of visits comes from White (27.9%) and Asian (21.1%) populations, there is a broad and healthy representation of all other racial groups. This distribution indicates that the hospital is accessible and utilized by a diverse community without evidence of service denial.
  5. Verification of Fair Practice:
     + Comprehensive data analysis yields no evidence of gender or racial discrimination. The hospital’s performance indicators suggest an equitable environment where all variations remain within a normal, expected operational range.

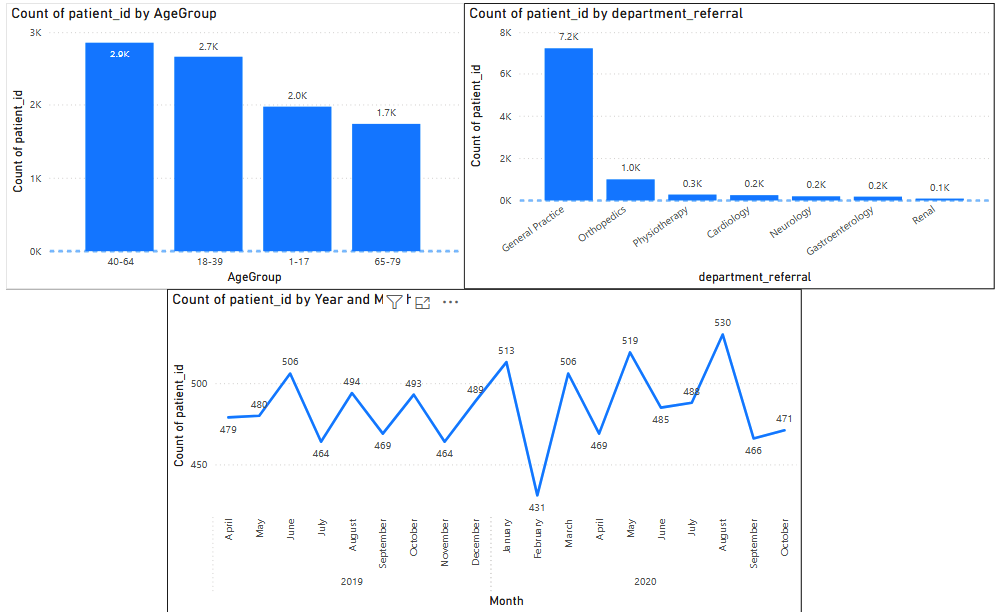
**Recommendations:**

* 1. Institutionalize a Demographic Equality Dashboard:
     + Establish a permanent monitoring system to track key performance indicators including wait times, admission rates, revenue generation, and satisfaction levels disaggregated by race and gender. Monthly reviews of this data will ensure any emerging disparities are identified and corrected in real-time.
  2. Conduct Targeted Departmental Audits:
     + Perform deep-dive reviews in specific departments where minor statistical variations occur. This ensures that triage protocols, admission criteria, and billing practices remain standardized and free from localized operational bias.
  3. Optimize Patient Feedback Channels:
     + Revitalize the survey process to improve response rates across all demographics. By capturing a more representative sample of patient sentiment, the hospital can better understand and address subtle differences in the "lived experience" of diverse patient groups.
  4. Implement Proactive Bias Mitigation Training:
     + Integrate unconscious bias and sensitivity workshops into regular staff professional development. These programs should focus on treatment prioritization and patient communication to ensure every interaction is rooted in clinical objectivity and empathy.
  5. Commit to Transparent Public Reporting:
     + Launch a Quarterly Fairness and Inclusion Report. Sharing high-level equity metrics with the community not only builds public trust but also reinforces the hospital’s commitment to accountability and inclusive excellence**.**

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

**Answer:**

* The discounts can be offered to increase the overall revenue based on Seasonality, Department and age group basis.
  1. Seasonality basis:
     + Created a line chart.
       - X-axis: Year, Month
       - Y-axis: Count of patient id
     + This visualization can help us understand the trends over the past, to identify the peak and dip period.
  2. Age group Basis:
     + Created a Column Chart
       - X-axis: Age Group
       - Y-axis: Count of Patient id
     + Help us identify the age groups with high volume patients and low volume of patients, the plan the discount offer to patient based on age groups.
  3. Department Basis:
     + Created a column Chart:
       - X-axis: Department Referral
       - Y-axis: Count of patient id
     + This visualization can help us identify the department that needs a boost in number of patients.

****

**Solution:**

* + There is a notable decline in visits during February. This dip represents underutilized staff and equipment, making it the ideal time to offer seasonal health check-up discounts to stimulate demand.
  + Age group 40-64 and 18-39 represent your highest traffic. Discounts for these groups should focus on "loyalty" or "membership" models to ensure they choose your facility every time they need care.
  + Since age group of 65-79 has lowest visit volume consider targeted discounts Senior Citizen Discount can lower the barrier to entry for these specific groups.
  + Departments like Physiotherapy, Gastroenterology, and Renal show significantly lower patient counts. Offering referral-based discounts can help boost these specific departments and maximize the use of specialized medical staff.

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

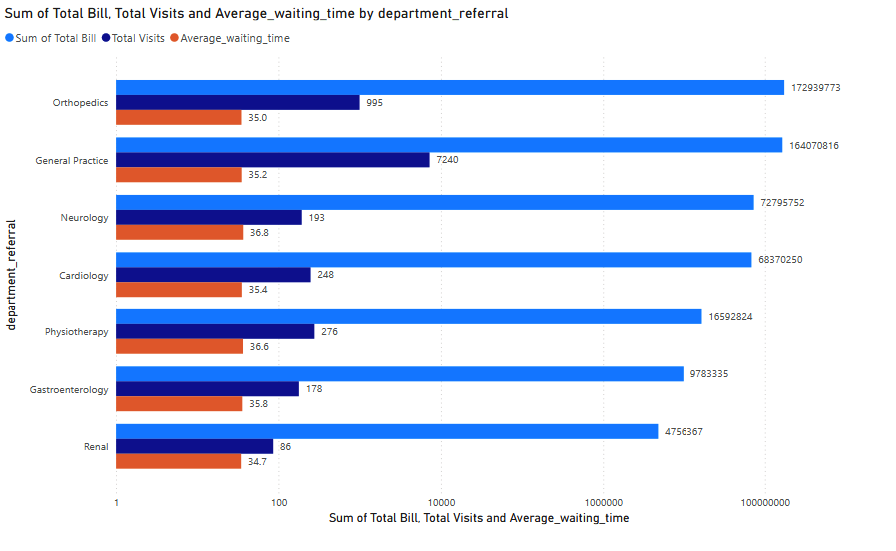
**Answer:**

To determine which departments require additional medical staffing, I conducted a multidimensional analysis focusing on three critical performance pillars:

* + Total Revenue: To assess the financial contribution and economic impact of each department.
  + Total Visits: To quantify actual patient demand and the resulting clinicalworkload.
  + Average Wait Time: To measure service efficiency and identify where staff capacity is failing to meet patient volume.

**I integrated these metrics into a comparative Bar Chart, plotting the three variables against the departmental referrals. This visual correlation allowed me to pinpoint pressure points specifically departments exhibiting high patient volume or excessive wait times alongside stagnant revenue or efficiency indicating a clear necessity for strategic hiring to alleviate doctor burnout and improve patient throughput.**

* X-axis: TotalRevenue, TotalVisits, Average Wait Time
* Y-axis: Department Referral

****

**Insights:**

* General Practice (High Capacity Pressure): With a dominant volume of 7,240 visits and substantial revenue, this department acts as the hospital’s primary engine. However, the 35.19-minute wait time highlights a heavy patient load that risks staff burnout and diminished service quality if not closely monitored.
* Orthopedics (High-Value Specialization): This department generates significant revenue despite a relatively low visit count (995). This suggests a model focused on complex, high-billing procedures or longer, specialized appointments rather than high-volume throughput.
* Neurology & Cardiology (Staffing Bottlenecks): These departments exhibit a concerning trend of low volume (193 & 248 visits) paired with peak wait times (up to 36.8 mins). This disparity strongly suggests either a severe shortage of specialists or inherently lengthy consultation requirements that current staffing levels cannot efficiently meet.
* Physiotherapy & Gastroenterology (Stable Operations): These units maintain a healthy equilibrium, with moderate wait times and consistent patient traffic. Current staffing appears well-aligned with departmental demand, representing a sustainable workload.
* Renal (Low Demand): Recording the lowest metrics in both visits (86) and revenue, the Renal department currently operates well within its capacity. Additional recruitment for this department is not a strategic priority at this time.

**Recommendations:**

* General Practice:
  + Reasoning: This is the highest-volume department (7,240 visits). A single hire here has a "force multiplier" effect. By clearing the "front door" of the hospital faster, you improve the flow for every other department.
  + Impact: Directly reduces the 35.19-minute wait time for the largest portion of your patient base.
* Neurology :
  + Reasoning: This department has the worst wait time (36.8 mins) in the entire hospital despite having very few patients. This is a clear indicator of a critical staff shortage.
  + Impact: Prevents specialized cases from "bottlenecking" and reduces the risk of patients leaving without being seen due to extreme delays.
* Cardiology :
  + Reasoning: Similar to Neurology, Cardiology has high wait times (35.3 mins) and specialized needs. Heart health is time-sensitive, a delay here is a significant clinical risk.
  + Impact: Improves service speed for high-risk patients and balances the workload for specialized care**.**

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

**Answer:**

* To determine the hospital's financial health in the absence of a direct Profit column, I utilized a Proximal Profitability Analysis. This approach evaluates financial viability by correlating revenue generation with operational efficiency.

1. Average Revenue Per Doctor by department:
   * + Created a new measure



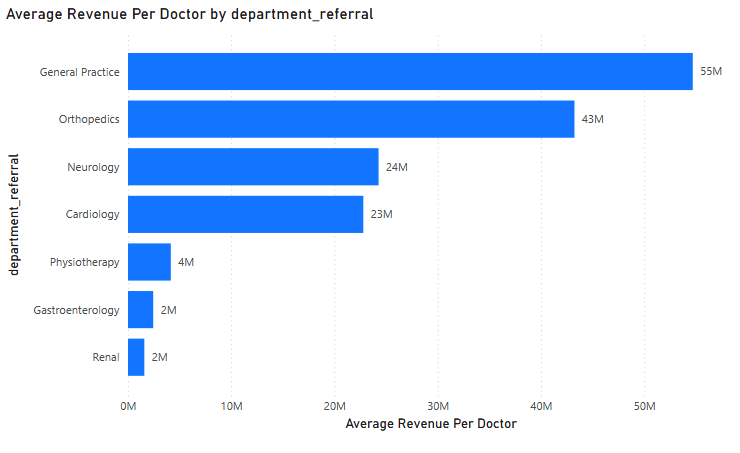
* + - Created a bar chart to visualize the revenue per doctor in departments.

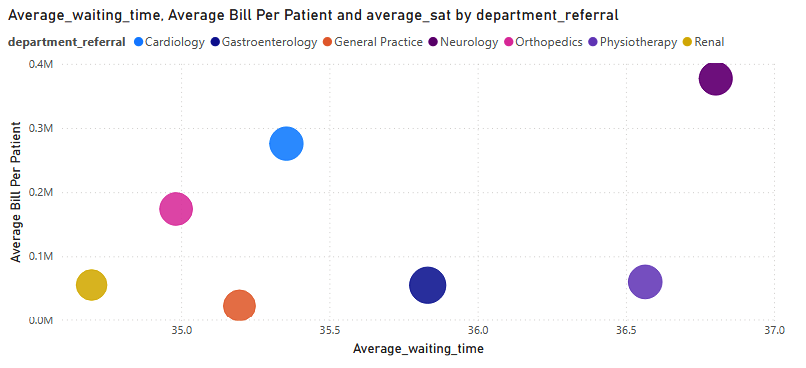
1. Avg. Revenue per patient Vs. Wait time by department:
   * Created the following measures:





* Created a Scatter chart:
  + X-axis: Average waiting time
  + Y-axis: Average Bill per Patient
  + Legend: Department Referral
  + Size: Average Satisfaction



****

**Insights:**

* General Practice and Orthopedics emerge as the hospital's most financially efficient units. They demonstrate the highest revenue-per-physician ratios, signaling optimal productivity and strong alignment between staffing levels and patient demand.
* Cardiology and Neurology generate moderate revenue, but their financial performance appears hindered by significant patient wait times. While these departments are high-value, operational bottlenecks may be preventing them from reaching their full revenue potential.
* Physiotherapy, Gastroenterology, and Renal departments exhibit lower revenue metrics both per doctor and per patient. This suggests these areas may be currently underutilized or operating as lower-margin support services, requiring a review of their cost-to-revenue structures.
* A clear trend exists where departments with extended wait times also report higher average billings. This suggests that long wait times in these areas are likely driven by the specialized nature of the care and the increased duration of complex clinical consultations.

**Conclusion**:

* By merging these two analyses, we can conclude that the hospital’s profitability is likely driven by high-margin specialties (like Orthopedics) and high-volume efficiency (in General Practice). However, the departments with peak wait times (Neurology and Cardiology) represent "hidden losses" where the cost of specialized staff is not being fully optimized due to throughput delays.

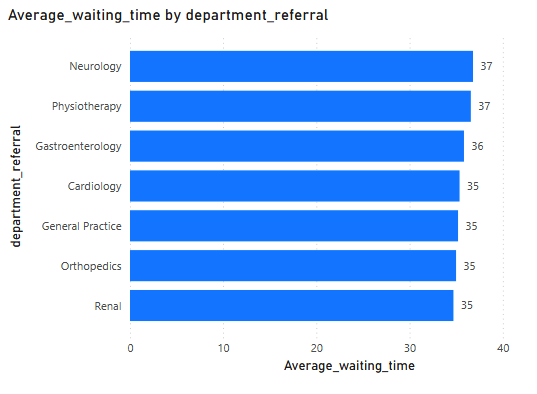
**Recommendations:**

* Direct capital and resource investments toward General Practice and Orthopedics. These departments demonstrate the highest productivity and revenue-per-physician, making them the primary engines of the hospital’s financial stability.
* Implement targeted operational improvements in Neurology and Cardiology. By streamlining triage and consultation workflows, the hospital can reduce excessive wait times, thereby increasing patient throughput and capturing "at-risk" revenue currently lost to delays.
* Conduct a structural review of the Physiotherapy, Gastroenterology, and Renal departments. Strategies should focus on either revitalizing service offerings to drive volume or reallocating staff to higher-demand areas to minimize idle labor costs.
* Ensure that the pursuit of high-margin care does not compromise service quality. The hospital must synchronize billing objectives with experience metrics, ensuring that premium specialized treatments are delivered with efficiency to maintain high satisfaction and long-term patient loyalty.

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

**Answer:**

* Created a Bar Char
  + X-axis: Department Referral
  + Y-axis: Average Waiting Time

****

**Insights:**

* While wait times across the hospital remain relatively uniform, a granular analysis reveals that Neurology and Physiotherapy exhibit the highest delays, averaging approximately 37 minutes.

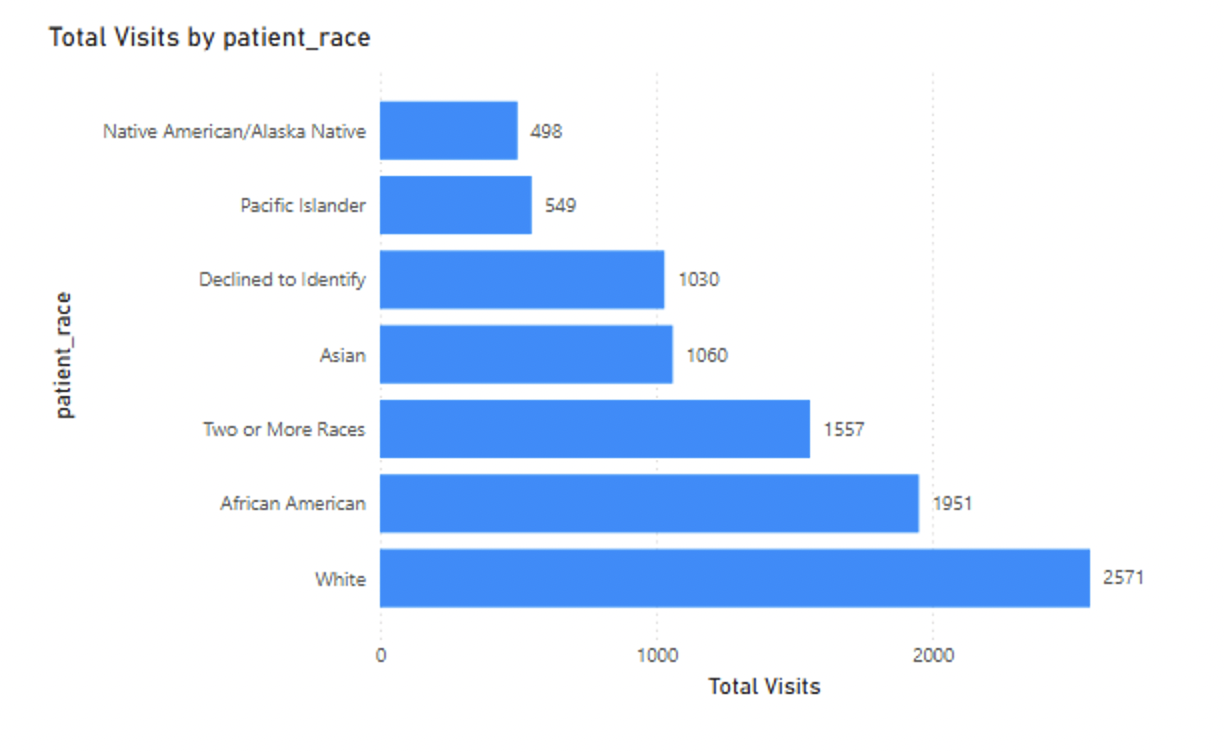
**Recommendations:**

* Realign shifts in high-traffic departments like General Practice and Orthopedics to match peak patient volume, ensuring staff capacity meets revenue demand.
* Establish a prioritization system to separate urgent clinical cases from routine visits, preventing specialized departments like Neurology from bottlenecking.
* Deploy automated booking and SMS reminders to minimize "no-shows" and reduce physical overcrowding in waiting areas.
* Use targeted surveys to pinpoint if delays are caused by paperwork or backlogs, allowing for specific rather than general fixes.

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

**Answer:**

* Building upon the existing discount strategy which optimizes for age demographics, seasonal trends, and departmental demand. I have integrated Patient Race as a new analytical dimension. This expansion aims to identify high-volume racial demographics and deploy targeted discounts that drive both patient engagement and health equity across diverse populations.
* Created a Bar Chart:
  + X-axis: Total Visits
  + Y-axis: Patient Race
* We aimed to identify racial groups with low visit volumes so that targeted incentives could be used to encourage participation and bridge gaps in healthcare access.

****

**Insights:**

* White (2,571 visits) and African American (1,951 visits) demographics represent the hospital’s most active patient segments, demonstrating established trust and high utilization of services.
* The Two or More Races (1,557) and Asian (1,060) groups exhibit stable, moderate engagement levels, forming a significant portion of the patient base.
* Pacific Islander (549) and Native American/Alaska Native (498) groups report the lowest visit frequencies. These figures may point to specific accessibility gaps or a need for targeted community outreach to improve participation.
* Patients who Declined to Identify (1,030 visits) show a mid-range engagement level, highlighting a subset of the population where demographic-specific insights remain unavailable.

**Recommended Discount Strategy:**

* Tier 1: High Discount (15%–20%) : Native American, Alaska Native & Pacific Islander
  + Reason: These demographics currently exhibit the lowest hospital utilization rates, signaling potential barriers to entry or minimal community engagement.
  + Primary Objective: To actively lower financial barriers, foster clinical inclusivity, and expand healthcare access to underserved populations.
* Tier 2: Moderate Discount (10%) : Asian & Multiracial Groups
  + Reason : These segments demonstrate stable but moderate engagement levels that could be further optimized.
  + Primary Objective: To incentivize more frequent healthcare interactions and solidify the hospital's relationship with these growing patient demographics.
* Tier 3: Baseline Retention (0%–5%) : White & African American
  + Reason : These groups represent the highest visit density and are already consistent users of hospital services.
  + Primary Objective: To focus on quality of care and patient retention rather than broad financial incentives, as utilization is already established.

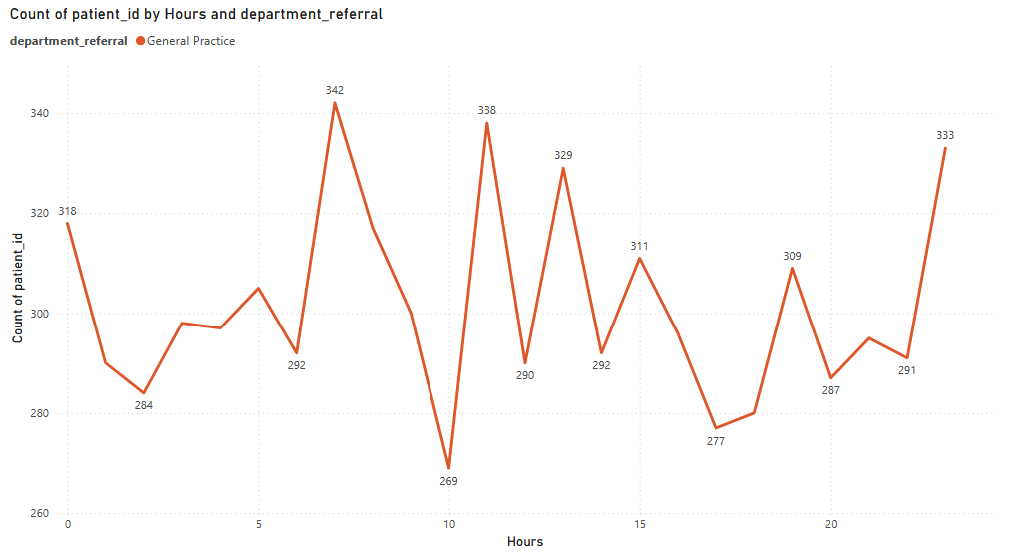
1. **Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also will this 2 shift policy be helpful for the hospital?**

**Answer:**

* I will begin by conducting a peak-load analysis of the General Practice department to identify high-volume arrival periods. This data-driven approach ensures that shift timings are precisely aligned with the hours of highest patient demand.
* Created a new column:

****

* Created a Line Chart:
  + X-axis: Hours
  + Y-axis: Count of patient id
  + Legend: Department Referral

****

**Insights:**

* Analysis reveals critical surges in patient inflow at 7 AM, 11 AM, and 1 PM, with volumes peaking at 342, 338, and 329 visits respectively.
* A significant decline in activity is noted during the 10 AM hour (269 visits) and the evening window of 5 PM–8 PM (277–300 visits).
* The data indicates a distinct bi-modal distribution, showing that the department faces two heavy congestion cycles: one in the early morning and another at midday.
* The identified volume dips at mid-morning and late afternoon serve as ideal windows for shift handovers and administrative resets without disrupting peak patient care.

**Recommended Staffing Model**

* **Early Shift (06:00–14:00):** Targets peak inflow periods. By staffing up early, the hospital can clear the initial morning backlog and maintain throughput during the 11 AM surge.
* **Late Shift (14:00–22:00):** Provides essential coverage for midday arrivals and evening patient traffic.

**How doctors can be allocated:**

* Allocated more number of experienced doctors in early shift due to high volume in overall day.
* Assign mid-level or newer doctors to Late Shift, where the slower pace (277–300 visits) allows for more detailed consultations and administrative follow-ups from the morning’s busy rush.

**How this helps the hospital:**

* By aligning doctor availability with peak arrival times (7AM, 11 AM, and 1 PM), the hospital can drastically reduce wait times. This leads to higher patient satisfaction scores and a more professional care environment.
* Shifting from a static schedule to a demand-based model ensures that no single team is overwhelmed during surges. This prevents physician burnout and maintains a consistent quality of care throughout the day.
* Predictive scheduling allows staff to better manage their personal time, while the presence of adequate support during rush hours creates a more supportive and less frantic workplace culture.

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

**Answer:**

* A Power BI Gateway is a bridge or connector that enables secure data transfer between on-premises data sources and Power BI cloud services. It acts as a secure channel, allowing Power BI to access and refresh data from databases and files located within private corporate networks without moving the data to the cloud.
* Types of Gateways
  + On-premises Data Gateway (Standard Mode) Shared by multiple users in an organization. Supports both scheduled data refresh and live queries for Power BI, Power Apps, Power Automate, and Azure Analysis Services. Installed on a dedicated machine (always on).
  + On-premises Data Gateway (Personal Mode) For a single user. Supports only Power BI and scheduled refresh (no live connections or multi-user support).
  + Virtual Network Data Gateway For Azure Virtual Network (VNet) data sources. No installation required; managed service.
* Key Use Cases
  + Scheduled Data Refresh
    - Automatically update Power BI datasets from on-premises sources (e.g., SQL Server, Oracle, files) at defined intervals.
  + Live Query Connections
    - Enable real-time dashboards by querying on-premises data sources directly (e.g., Analysis Services, SQL Server) without importing data.
  + Hybrid Cloud Scenarios
    - Maintain sensitive data on-premises while leveraging Power BI’s cloud analytics and collaboration features.
  + Cross-Service Data Access
    - Allow Power Apps, Power Automate, and Azure Logic Apps to connect to on-premises systems through the same gateway.
  + Data Security & Compliance
    - Avoid exposing databases to the internet, data stays behind the firewall, with only encrypted queries/results passing through.
  + Unified Data Access
    - Use a single gateway for multiple users and reports, simplifying management.

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

**Answer:**

* If no predefined questions were given, I would begin by thoroughly exploring and cleaning the data. Next, I would perform exploratory analysis to identify trends in patient demographics, satisfaction, and operations. I would then define key metrics and create an interactive Power BI dashboard to visualize patterns and relationships such as the impact of wait times on satisfaction. Finally, I would translate these insights into actionable recommendations to improve efficiency, patient care, and decision-making.

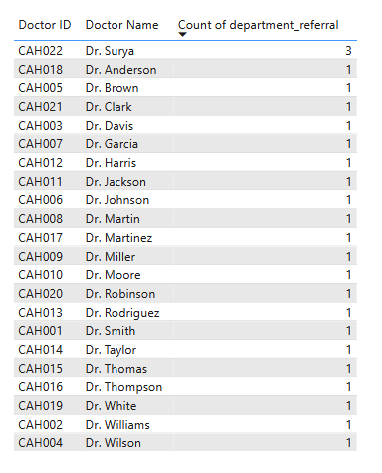
**Approach without Predefined Questions:**

* 1. Data Understanding & Cleaning
     + Explore dataset structure, relationships, and data quality
     + Handle missing values, outliers, and inconsistencies
  2. Exploratory Analysis
     + Uncover patterns in patient demographics, visit trends, and satisfaction
     + Identify correlations (e.g., wait time vs. satisfaction)
     + Analyze department performance and doctor efficiency
  3. KPI Development
     + Create metrics: avg. satisfaction, revenue per doctor, wait times, patient load.
     + Track trends over time (daily/weekly/monthly).
  4. Visualization & Dashboarding
     + Design interactive Power BI dashboards with:
       - KPI cards for key metrics
       - Charts for trends and comparisons
       - Drill-down capabilities for detailed insights
  5. Insight Generation
     + Find operational bottlenecks (e.g., long wait times in specific departments)
     + Highlight top/low performers (doctors, departments)
     + Identify factors affecting patient satisfaction and revenue
  6. Actionable Recommendations
     + Suggest process improvements
     + Propose resource reallocation
     + Recommend targeted interventions based on data

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

**Answer:**

* To identify the type of relationship between the doctor id and department I create a table with doctor id, doctor name and Count of distinct departments.

****

* Only Dr. Surya (CAH022) has a Count of department\_referral of 3. Every other doctor (21 of them) has a count of 1.
* The presence of a doctor linked to more than one department breaks the rule for a one-to-one relationship (where each doctor ID must match exactly one department).
* The data structure shows a one-to-many relationship.One doctor can be associated with many department referrals (proven by Dr. Surya).

**Report**

**The hospital has asked for a report with three tabs:**

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

**Make sure that all the visualizations look decent and are placed in a proper order. Each tab has different POCs (Point Of Contact), so make sure you involve all the metrics that POC may look at in that tab along with those mentioned in the tab description.**

**After making the report on the Desktop ensure that it is hosted on PowerBI service and use the hosted link for submission of the dashboard and mentioning on the resume.**