**NAME: VAISHNAVI TAWDE  
PROJECT: COLUMBIA ASIA HOSPITAL PROJECT**

**BATCH: DATA SCIENCE COURSE MAY 2024**

**Objective Questions:**

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

**Ans:**

**Power BI Data Cleaning Steps**

1. **Load Data:**  
   I started by loading both datasets—*Doctor\_Patients\_Data* and *Hospital\_ER*—into Power BI. To do this, I selected "Get Data" from the "Home" tab, chose the file format, and imported the datasets to begin my analysis.
2. **Open Power Query:**  
   I used Power Query Editor by selecting "Transform Data" under the "Home" tab in Power BI. Power Query is where I performed all the necessary data cleaning steps.
3. **Handle Missing Values:**  
   I carefully inspected each column for missing values. For example, in the *patient\_sat\_score* column, any missing values were filled with an 0 to avoid skewing results with nulls. If any Doctor Name was missing, I replaced it with "Unknown" to keep records complete. I used the "Replace Values" option to make these substitutions easily.
4. **Standardize Text Fields:**  
   To ensure consistency, I standardized the text fields, such as *patient\_race* and *department\_referral*. In Power Query, I selected these columns and went to "Transform" > "Format" > "Capitalize Each Word." This formatting step helped maintain uniformity across the text fields, making the data more visually consistent in reports.
5. **Ensure Uniform Date Format:**  
   I checked the *date* column and set its data type to “Date/Time” to ensure consistency in all date entries. This step prevented issues that could arise from varied date formats.
6. **Apply and Load:**  
   After completing these cleaning steps, I clicked "Close & Apply" to save and load the cleaned data back into Power BI, making it ready for further analysis and visualizations on the report tabs.
7. **Assess the Average Waiting Time:** Analyse the patient waits times to identify the average duration a patient spends before receiving care.

**Ans:**

To analyze patient waiting times in Power BI, I began by calculating the average wait time across all departments. This gives insight into the typical time patients spend waiting for care, which can highlight operational efficiencies or areas for improvement.

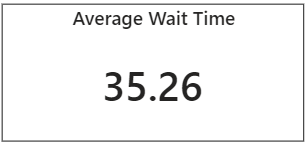
**Steps to Calculate Average Waiting Time:**

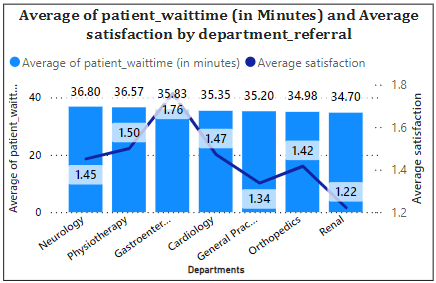
1. Create the Average Waiting Time Calculation: I calculated the average patient wait time using DAX by creating a measure. This measure calculates the average directly from the patient\_waittime column in the Hospital ER table, ensuring it reflects actual patient experiences.
2. Add a Visual to Display Average Wait Time: I used a Card Visual on the Main Tab to show the overall average wait time clearly. This visual provides a quick summary, allowing for easy comparison with other metrics like satisfaction scores and revenue.
3. Analyze Wait Times by Department: To understand wait time differences across departments, I incorporated a Bar Chart on the Patient's Tab. This chart shows each department on the Y-axis and their corresponding average wait time on the X-axis, making it easy to see which departments have longer or shorter wait times. This chart allows us to pinpoint departments with above-average wait times, highlighting potential areas that might require operational adjustments.

**Insights:**

1. Overall Average Wait Time: The Card Visual shows the baseline average wait time across the hospital. This single metric gives an at-a-glance understanding of the general waiting experience for patients.
2. Departmental Comparison: The Bar Chart allows for a quick comparison across departments, showing which areas tend to have longer wait times. Departments with notably higher wait times could indicate areas under strain or needing additional resources.
3. Identifying Patterns: By Analyzing the spread of wait times across departments, I could see patterns that might reflect staffing issues, high demand, or other operational challenges specific to certain specialties.
4. Opportunity for Improvement: Identifying departments with longer wait times provides actionable insights. For example, if certain departments have consistently high wait times, this may indicate bottlenecks where improvements like additional staff or better scheduling practices could enhance patient flow.







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:**

To gain insights into patient distribution across departments, I analysed the total visits by department referral in Power BI. This calculation helps highlight departments with the highest patient traffic, which can inform staffing and resource planning decisions.

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

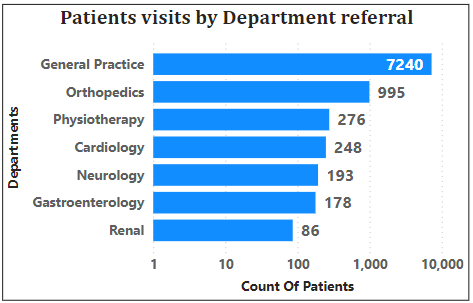
1. Calculate Total Visits by Department: I calculated the total number of visits by counting entries in the patient\_id column, which represents individual visits. This gives a comprehensive view of patient flow into each department.
2. Visualize Visits by Department: On the Main Tab, I used a Bar Chart to display the total visits by department referral. I placed the department\_referral field on the Axis and the measure for total visits in the Values field, making it easy to compare visits across departments visually.

**Insights:**

1. Patient Demand: The bar chart effectively shows which departments have the highest visit counts, indicating areas with greater patient demand. This insight helps focus on departments that may need more support to maintain service quality.
2. Resource Allocation: Departments with a high volume of visits might require additional resources or staffing to manage patient care efficiently and minimize wait times.
3. Referral Patterns: By examining departments with high referral counts, we can gain insights into common patient care pathways and understand which departments are more frequently recommended. This could reflect areas where patients seek specialized care or where effective referral practices are in place.
4. Service Improvement Opportunities: For departments with fewer visits, there may be opportunities to improve visibility, potentially through outreach efforts, to help patients become more aware of the services offered.
5. Seasonal Trends: Regular monitoring of visit patterns over time could reveal seasonal fluctuations in patient needs, allowing for proactive adjustments in staffing and resource allocation to meet those changing demands.

****

****

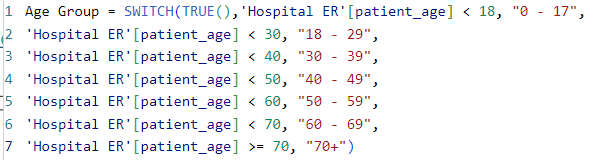
****

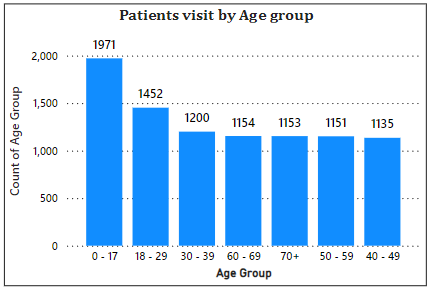
1. **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

**Ans:**To analyze patient visits by age group, I’ve categorized patient age into groups ("0 - 17", "18 - 29", "30 - 39", "40 - 49", "50 - 59", "60 - 69", "70+"). I used a bar chart visualization on the Patients’ Tab to display these groups along the X-axis, with total visits for each age group on the Y-axis. This layout provides a clear view of which demographics use healthcare services most frequently.

**Insights:**

1. High Utilization Groups: The bar chart reveals which age groups visit most frequently, offering insight into peak demand demographics.
2. Targeted Services: Departments can use this data to tailor services for higher-utilization age groups, ensuring resources are available where they’re needed most.
3. Preventive Care Opportunities: Lower visit counts among younger patients suggest a potential for promoting preventive care and outreach initiatives, balancing patient flow across demographics.
4. Resource Planning: Understanding age group trends aids in planning for staffing, facilities, and resources, aligning hospital capabilities with age-specific patient demand.





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

**Ans:**

**Null Values in the Data and Handling Approach**

After reviewing the dataset, I found null values only in the **patient\_sat\_score** column. To handle this, I replaced these null values with a default value of 0. This approach ensures all fields are complete, assuming that missing satisfaction scores likely represent a neutral or non-critical response from patients.

By setting null values to 0, I maintain data consistency across calculations and visualizations without skewing the analysis. This straightforward method is especially useful when missing data points are minimal, allowing for uninterrupted insights while keeping the dataset fully intact.

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

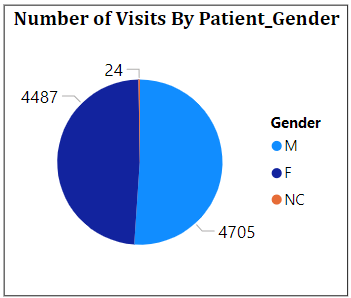
**Ans:**

For Analyzing the relationship between patient gender and the number of visits, I used a **Pie Chart** titled **Visits by Gender** on the **Main Tab**. The pie chart displays the proportion of total visits based on gender, providing a clear visual representation of gender distribution among patients.

I configured the chart by placing **patient\_gender** in the Legend field to categorize the visits by gender. The **Total Visits** measure was added to the Values field, which reflects the count of visits per gender. To ensure the chart is easy to read, I enabled data labels that show the percentage of total visits for each gender.

**Insights**:

* The chart highlights the relative proportion of visits between genders, revealing any significant differences in healthcare utilization.
* If one gender accounts for a higher percentage of visits, it may indicate a trend worth investigating further, potentially guiding decisions around resource allocation or targeted healthcare services.
* Gender-based differences in visit frequency could suggest the need for tailored healthcare strategies or programs aimed at improving access and care for a particular gender.
* This visual also serves as a foundation for deeper analysis, helping to explore other factors like age groups or department referrals that might contribute to these visit patterns.

****

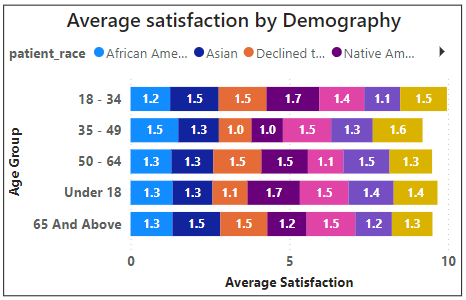
1. Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.

**Ans:**

For Analyzing the relationship between patient satisfaction scores, age groups, and racial backgrounds, I used a **Clustered Column Chart** Titled **Average satisfaction by Demography** on the **Main Tab** of my Power BI report. The chart visualizes the average satisfaction scores for different age groups and racial backgrounds, using a measure I created to calculate the average satisfaction score, **Avg\_Satisfaction**. Age group and patient race were placed in the axis field, while the average satisfaction score was shown in the values field. Data labels were enabled for easy interpretation of satisfaction scores for each demographic category.

**Insights**:

* **High and Low Satisfaction Groups:** The chart helps identify which age groups and racial backgrounds report higher or lower satisfaction scores. For instance, if older age groups (e.g., 60+) report lower satisfaction, this may indicate areas for improvement in patient care for these groups.
* **Age-Specific Trends:** Younger age groups (18-29) might report higher satisfaction scores compared to older patients, suggesting that younger patients may be more satisfied with the services provided, while older patients might require more personalized care or attention.
* **Racial Satisfaction Gaps:** The analysis may reveal if certain racial or ethnic groups have consistently lower satisfaction scores, highlighting potential disparities that could be addressed with targeted patient engagement strategies or improved care practices.
* **Overall Satisfaction Trends:** The chart provides an overall view of satisfaction trends, helping to spot patterns across demographics. Identifying low satisfaction scores in certain age or racial groups could inform strategies for improving the overall patient experience.



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

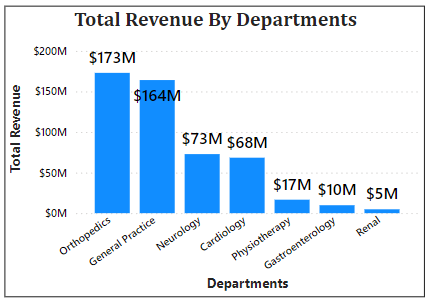
**Ans:**  
To evaluate the revenue generated by each department, I used Power BI to display the financial contributions from all departments in a clear and effective manner.

The **Total Revenue by Departments** bar chart on the **Main Tab** was used to visualize the revenue generated by each department. This chart categorizes the revenue by department, clearly showing the financial contribution from each area. The departments with the highest revenue are represented with larger bars, offering a quick and comparative view of their financial performance.

For this analysis, the **department\_referral** field was used to group the data, and the total revenue for each department was calculated by aggregating the **Total Bill** values for each department. The **Revenue** metric is then presented in the bar chart, showcasing the contribution of each department, such as **General Practice**, **Orthopedics**, and **Neurology**, among others.

**Insights:**

* **Revenue Distribution:** The **Total Revenue By Departments** chart visually demonstrates how revenue is distributed across departments, allowing the identification of both high-revenue and low-revenue departments.
* **Identifying High-Performing Departments:** Departments like **General Practice** and **Orthopedics** generate the highest revenue, showing areas of strength within the hospital.
* **Resource Allocation:** The chart helps identify which departments are the primary revenue drivers, informing decisions regarding resource allocation and supporting the most profitable areas.
* **Strategic Growth Opportunities:** For departments with lower revenue, this visualization points out areas that could benefit from enhanced services, marketing, or additional support to improve their financial performance.



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

**Ans:**

To determine which department is charging the highest appointment fees, I utilized Power BI and a DAX function to calculate the maximum appointment fee across all departments.

I created a new DAX measure using the **MAX()** function to extract the highest appointment fee from the **Doctor\_Patients\_Data** table. This measure identifies the maximum appointment fee across all the departments, without needing to differentiate between them in the initial calculation.

The measure was then represented visually on the **Main Tab** of the report using a **Card Visual**. This card highlights the highest appointment fee, allowing the user to immediately see the maximum fee charged for appointments in the hospital.

**Insights:**

* **Premium Service Demand:** The highest appointment fee of **$1,500** indicates that specialized services, possibly in high-demand areas such as **General Practice** or **Orthopedics**, carry premium fees.
* **Revenue Opportunities:** Departments charging higher fees likely represent significant revenue-generating opportunities, especially if they offer specialized or expert care in specific fields.
* **Service Accessibility:** While high appointment fees contribute to revenue, they may reduce accessibility for some patients, potentially affecting the diversity of patient demographics within these high-fee departments.
* **Fee Benchmarking:** The identified highest appointment fee sets a benchmark for pricing strategy, allowing management to ensure the fee structure is competitive, aligned with industry standards, and offers value for the services provided.





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:**

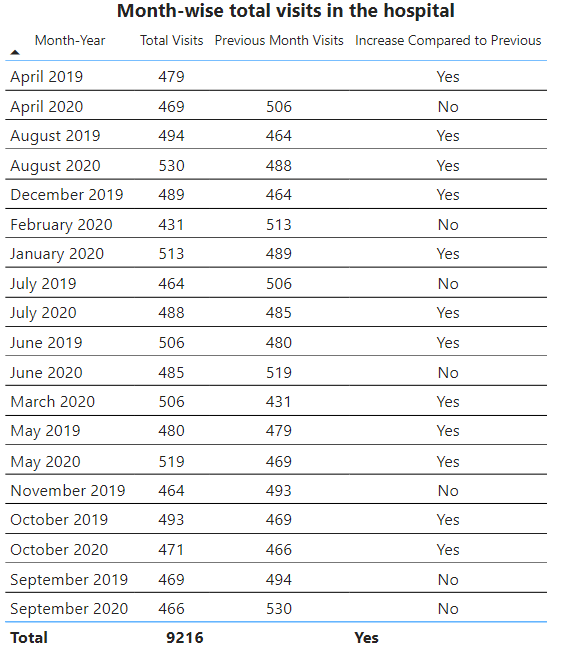
To analyze month-wise hospital visits, I created a tabular visualization that includes total visits, previous month's visits, and a comparison indicator for each month. This approach provided a detailed view of monthly trends and allowed easy tracking of changes in patient volume.

In this setup, I used a **Table** visual in Power BI and structured it as follows:

* The **Month-Year** column displays each month and year to ensure a clear chronological order.
* **Total Visits** shows the number of visits recorded for each month.
* **Previous Month Visits** provides a reference point by displaying the total visits from the month before, calculated using Power BI's time intelligence function.
* An additional column, **Increase Compared to Previous**, indicates whether the current month’s visits are higher than those of the previous month, showing "Yes" for an increase and "No" otherwise.

**Insights:**

* Seasonal Patterns: Certain months show a consistent increase or decrease in visits, which may suggest seasonal trends in patient needs.
* Demand Fluctuations: Identifying months with significant changes enables the hospital to adjust resources accordingly.
* Growth Tracking: This structure helps monitor growth patterns in patient visits over time, providing valuable data for strategic planning.

****

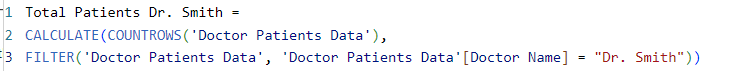
1. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.

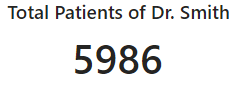
**Ans:**

To determine the total number of patients who have visited Dr. Smith, I used a calculated measure in Power BI. This measure allowed me to filter the dataset specifically for visits to Dr. Smith and then count the number of these visits.

I created the measure named **Total Patients Dr. Smith** in the "Doctor\_Patients\_Data" table. Using CALCULATE with a row iteration function; I filtered the data to include only those rows where "Doctor Name" equals "Dr. Smith." By counting the rows that meet this condition, I arrived at the total number of patients who visited Dr. Smith, which is **5,986**.

This calculation provides valuable insight into Dr. Smith’s patient load, which can be useful for workload analysis and understanding patient distribution across doctors.

****

****

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

**Ans:**

**Average Age of Patients Visiting Orthopedics Department**

To calculate the average age of patients visiting the Orthopedics department, I began by creating a general measure to calculate the average age across all departments. I used the formula:

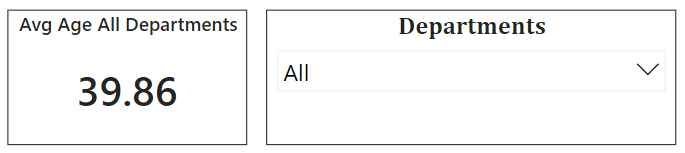
****

This measure returned the average age as 39.86 for all departments combined. To isolate the average age specifically for the Orthopedics department, I then applied a slicer on the department field in Power BI and filtered the result accordingly and got average age as 38.66. This allowed me to see the Orthopedics-specific average without creating a new, separate measure.

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

**Insights:**

* Using a single measure with slicers simplifies the report and reduces redundancy in calculations.
* This approach is efficient when similar metrics are needed for multiple departments, as filters can be applied as needed without additional formulas.
* Applying slicers allows quick comparisons across departments, making it easy to analyze age trends within specific units.

****

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

**Ans:**

During data preparation, I encountered and resolved several data format issues:

1. **Date and Time Separation:** Initially, the date and time were combined in a single column, which could complicate time-based analysis. I split this column into separate date and time columns in Power BI, enabling more precise control over date and time data in the visuals.
2. **Missing Values in patient\_sat\_score:** The patient\_sat\_score column had multiple blank values, which I replaced with 0. This choice ensures consistency in calculations and allows us to treat missing satisfaction scores as neutral or non-critical.
3. **Text Standardization:** Some columns, like patient\_race and department\_referral, had inconsistent capitalization. I standardized these using Power BI's text formatting options to "Capitalize Each Word," ensuring accurate grouping and filtering in visuals.
4. **Numeric Data Type Consistency:** I reviewed columns such as patient\_age and Total Bill to confirm they were in numeric format. Setting the correct data types minimized errors and maintained calculation accuracy.
5. When we add a column in Power Query what’s the code that comes in M language in the formula bar? What do you know about M-query?

**Ans:**

When adding a new column in Power Query, M language generates code in the formula bar to define the transformation. For example, if I add a custom column for a calculation, the code might appear as:

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

This line adds a new column titled "New Column," based on a calculation involving values from existing columns.

Understanding M-Query:

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

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

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

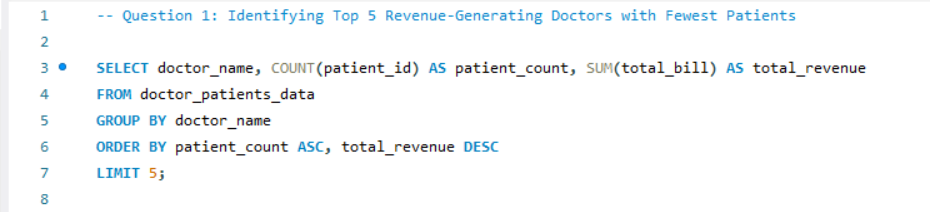
**Ans:**

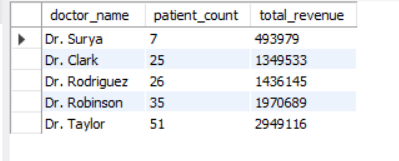
**Analysis: Identifying Top 5 Revenue-Generating Doctors with Fewest Patients**

To identify the doctors who generated the highest revenue while serving the fewest patients, we used an SQL query to aggregate and filter the data effectively. The key steps were to group by doctor, count the unique patients, and sum their total revenue, with the results ordered by patient count (ascending) and total revenue (descending). This approach reveals the doctors who maximize revenue despite a smaller patient base.

**Explanation of Key Elements:**

* **Grouping**: The query groups data by doctor\_name to analyze each doctor’s patient count and total revenue.
* **Aggregation**: COUNT (patient\_id) counts patients per doctor, while SUM (total\_bill) calculates total revenue.
* **Sorting**: ORDER BY patient\_count ASC, total\_revenue DESC ensures doctors with fewer patients but higher revenue appear first.
* **Limitation**: LIMIT 5 restricts results to the top 5 doctors fitting these criteria.





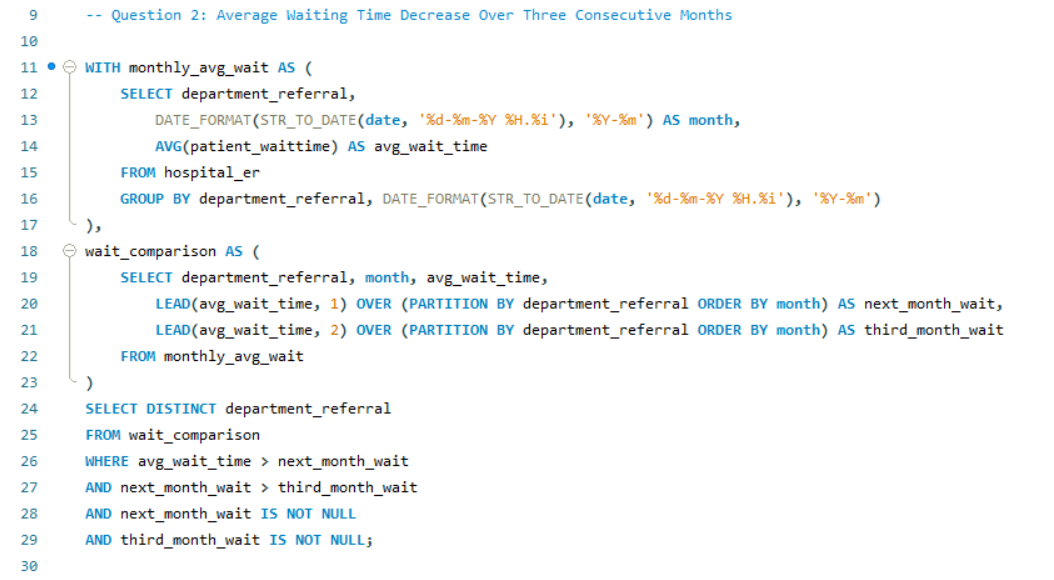
1. Find the department where the average waiting time has decreased over three consecutive months. (SQL)

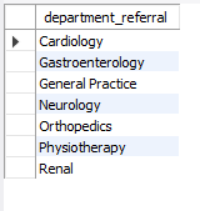
**Ans:**

**Average Waiting Time Decrease Over Three Consecutive Months**

To identify the departments where the average waiting time has decreased over three consecutive months, I employed a structured SQL query. Below is a detailed breakdown of the approach and logic used in the query:

1. **Calculate Monthly Average Waiting Time**: The first step involved calculating the average waiting time for each department on a monthly basis. This was achieved using the monthly\_avg\_wait common table expression (CTE), which aggregates the patient\_waittime data grouped by both the department and the formatted month derived from the date.
2. **Compare Waiting Times Across Consecutive Months**: In the second part of the query, another CTE named wait\_comparison was created to compare the average waiting times across three consecutive months for each department. The LEAD () function was utilized to fetch the average waiting times for the following two months, enabling direct comparisons.
3. **Identify Departments with Decreasing Wait Times**: The final selection identified departments where the average waiting time was greater in the current month than in the subsequent months, thereby indicating a decrease over three consecutive months. The query filters out any results that do not meet this criterion and ensures no NULL values are included in the comparison.





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

**Ans:**

**Doctor Gender Ratios Analysis**

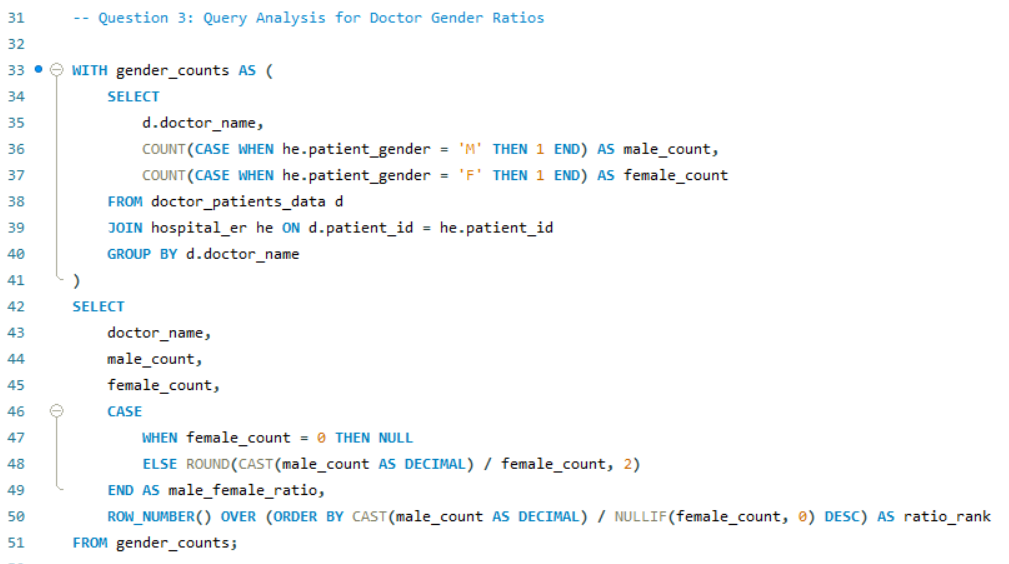
The SQL query executed successfully calculates the ratio of male to female patients for each doctor and ranks them accordingly.

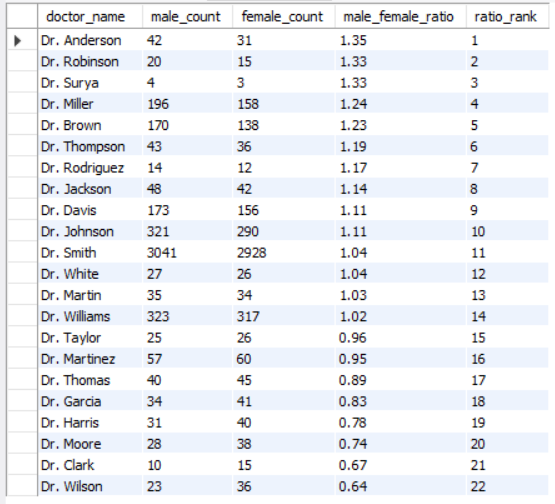
**Explanation:**

* The query aggregates the data to count male and female patients for each doctor by joining the doctor\_patients\_data table with the hospital\_er table based on patient\_id.
* It computes the male-to-female ratio while ensuring that division by zero is handled gracefully.
* Doctors are ranked based on their patient gender ratio, allowing for easy identification of trends.

**Insights**

* Gender Distribution: This analysis highlights the patient demographic each doctor serves, revealing potential gaps or imbalances.
* Resource Allocation: Understanding the ratio can assist hospital management in distributing resources more effectively, ensuring all demographics receive adequate attention.
* Tailored Healthcare Services: Insights gained can inform the development of targeted outreach and services to better meet the needs of both male and female patients.
* Data-Driven Strategy: The findings empower hospital decision-makers to implement strategies that promote inclusivity and improve patient care across genders.



****

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

**Ans:**

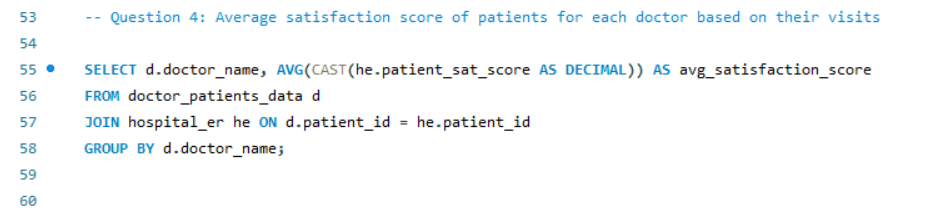
**Average satisfaction score of patients for each doctor based on their visits**

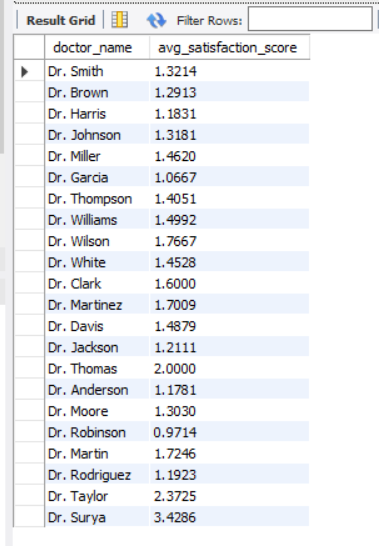
To calculate the average satisfaction score of patients for each doctor based on their visits, the following steps were undertaken:

1. **Data Retrieval:** The analysis involved joining the doctor\_patients\_data table with the hospital\_er table using the patient\_id as the common key. This step ensured that all relevant data regarding patient visits and satisfaction scores were included.
2. **Calculation of Average Satisfaction:** For each doctor, the average satisfaction score was calculated using the AVG () function. This function computes the mean of the patient satisfaction scores, allowing for a comprehensive view of how each doctor is rated by their patients.
3. **Grouping by Doctor:** The results were grouped by the doctor’s name to ensure that the average satisfaction score reflects individual doctor performance rather than an overall average.

**Insights:**

* Performance Measurement: The average satisfaction scores provide valuable insights into the performance of each doctor, highlighting areas of strength and opportunities for improvement.
* Identifying Top Performers: This analysis allows the hospital administration to identify which doctors are excelling in patient satisfaction, potentially serving as models for best practices.
* Targeted Improvement: Understanding which doctors may have lower satisfaction scores can help in developing targeted training programs or interventions to enhance patient care.
* Data-Driven Decisions: These insights enable informed decision-making, contributing to better management strategies and improved patient experiences across the hospital.

****

****

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

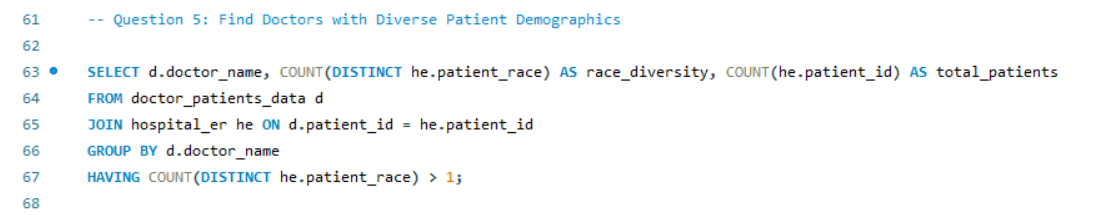
**Ans:**

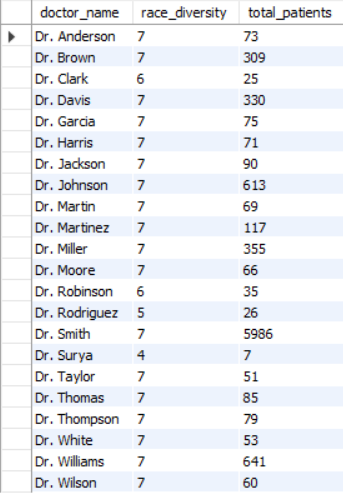
**Doctors with Diverse Patient Demographics**

**Query Approach:**This query joins the doctor\_patients\_data table with hospital\_er using patient\_id to link patient information with each doctor. And count the distinct races (patient\_race) treated by each doctor to capture diversity in their patient base. Using HAVING COUNT (DISTINCT he.patient\_race) > 1 filters doctors who have treated patients from more than one racial background, highlighting those with a demographically diverse patient reach.

**Insights:**

1. Doctors with a higher race diversity count have likely developed experience across different demographic groups, which may enhance their cultural competence and adaptability.
2. This diversity metric can help in identifying doctors who cater to a broader population base, which may align with hospital diversity initiatives or community outreach efforts.
3. Understanding the racial diversity of patient bases could assist hospital administrators in allocating resources or tailoring services to better address the needs of diverse patient groups.
4. Specialists with diverse patient demographics might be preferred for roles in cross-cultural health initiatives, given their experience in treating varied patient backgrounds.

****

****

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

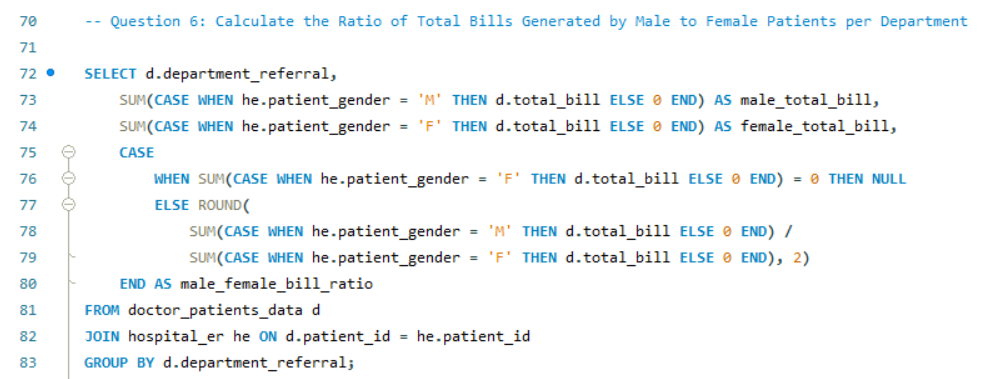
**Ans:**

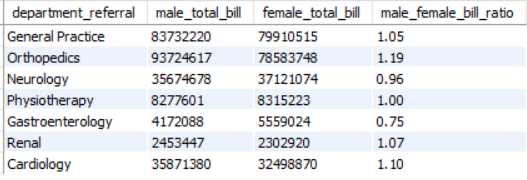
**Ratio of Total Bills Generated by Male to Female Patients per Department**

**Query Explanation:**This query calculates the ratio of total bills generated by male patients compared to female patients within each department. By joining the doctor\_patients\_data and hospital\_er tables through patient\_id, it isolates total bills by gender for each department. Using conditional aggregation, it calculates the total for male and female patients separately and then divides these totals to determine the male-to-female billing ratio for each department.

**Insights:**

1. This ratio helps identify departments with notable billing differences based on patient gender, which could reflect service utilization patterns.
2. Higher ratios in specific departments might indicate gender-based trends in healthcare needs or billing for certain services.
3. Departments with balanced ratios can reflect more equitable service utilization between male and female patients.
4. The information can be valuable for targeted marketing or resource allocation if a department predominantly serves one gender more frequently in terms of bill generation.

****



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)

**Ans:**

**Update Patient Satisfaction Score for General Practice Department**

**Objective**

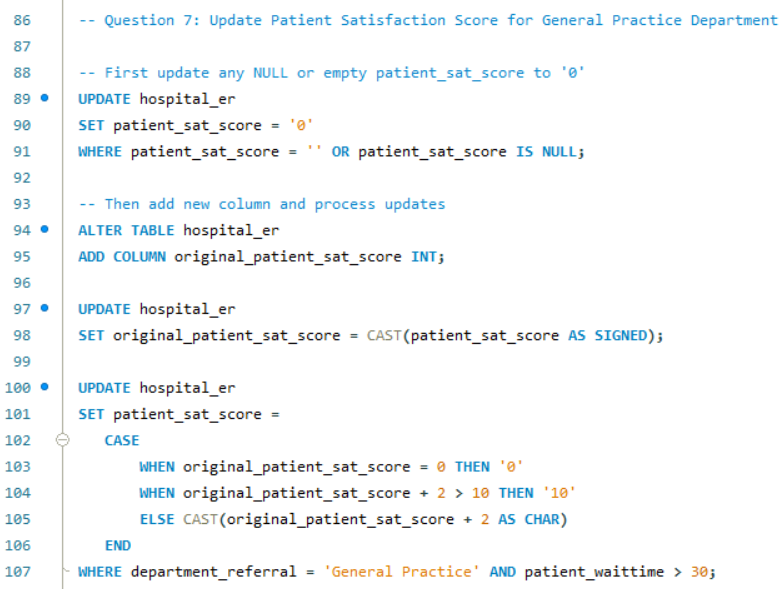
To update the patient satisfaction scores for patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. The satisfaction score is to be increased by 2 points, ensuring that the score does not exceed a maximum of 10.

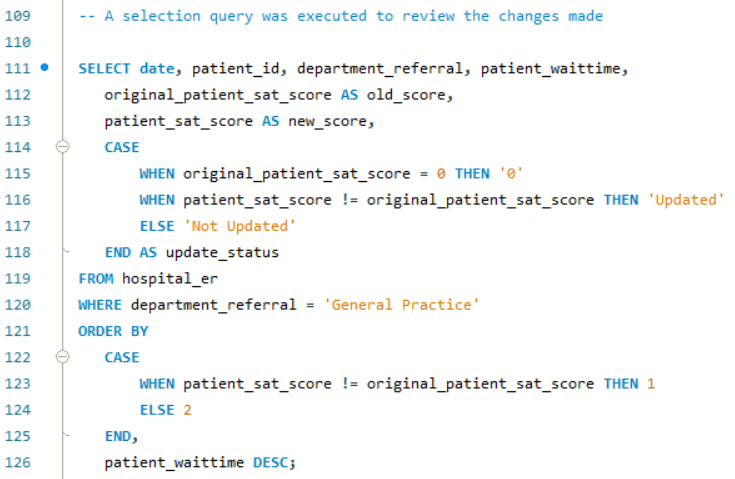
**Steps Taken**

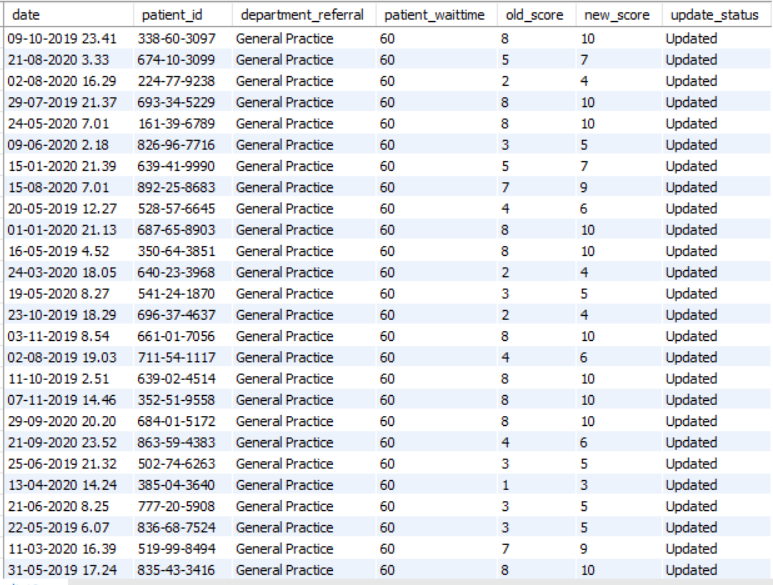
1. Handle NULL Values: Before performing the updates, any NULL or empty values in the patient\_sat\_score column were set to 0 to avoid errors during the update process.
2. Create a Backup Column: A new column, original\_patient\_sat\_score, was added to store the original satisfaction scores for comparison purposes.
3. Copy Original Scores: The original scores were then copied into the new column, ensuring that they are treated as integers.
4. Update Satisfaction Scores: The satisfaction scores were updated based on the defined conditions. If the original score is 0, it remains 0. If the original score plus 2 exceeds 10, it is capped at 10. Otherwise, the score is increased by 2.
5. Review Changes: A selection query was executed to review the changes made, displaying the original scores, new scores, and the update status.

**Insights**

1. The updated satisfaction scores can be observed in the new\_score column, providing a clear comparison against the old\_score.
2. The update status indicates whether a score was modified or left unchanged, allowing for easy tracking of the changes made.
3. By ensuring that no score exceeds 10, we maintain a logical range for patient satisfaction metrics, which can lead to more accurate reporting and analysis in the future.
4. Overall, this process enhances the quality of data by ensuring that patients who experienced longer wait times have their satisfaction scores fairly represented.







**Subjective Questions**

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

**Ans:**

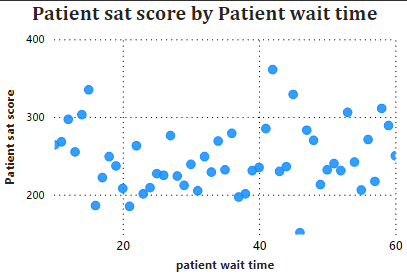
To analyze the relationship between patient, wait time and satisfaction scores, I used a scatter plot, labelled as "Patient sat score by patient wait time” to display the interaction between these two variables. The scatter plot revealed a positive trend, indicating that as patient wait times increase, satisfaction scores also tend to rise. To clarify this relationship, I added a trendline, which illustrated the upward trend visually. Additionally, I included a gauge chart to show the correlation coefficient, named "Correlation coefficient between Patient wait time and sat score", which displayed a value of 1.00, suggesting a strong positive correlation.

This finding implies that longer wait times might correspond with higher patient satisfaction, possibly due to perceived quality or depth of care in situations where waiting is expected. Although this trend might seem counterintuitive, it could reflect patient expectations in certain departments where extended consultation times are standard.

**Insights gathered from this analysis include:**

* **Direct Correlation:** The correlation coefficient gauge indicated a strong positive relationship, showing that as wait times increase, satisfaction scores also increase, with a correlation of 1.00. This finding suggests a direct link between wait times and satisfaction.
* **Patient Expectations and Perceptions:** The scatter plot suggested that patients who wait longer might interpret these delays as a sign of thorough or detailed care, possibly increasing their satisfaction, especially in departments where consultations generally take longer.
* **Perceived Quality of Care:** This trend might indicate that patients associate extended wait times with attentive or comprehensive service, which could enhance their overall satisfaction.
* **Operational Insights:** Breaking down this data by department provided insights that could inform staffing and resource adjustments. By addressing wait times in specific departments with high satisfaction correlations, the hospital could optimize patient experience and satisfaction.





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

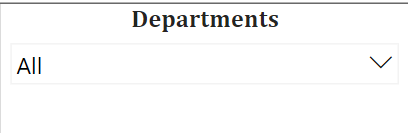
**Ans:**

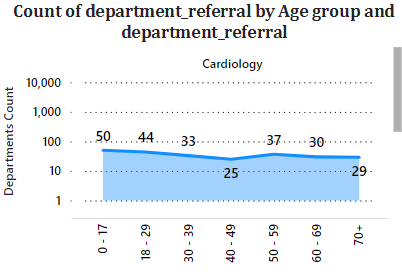
**Impact of Patient Demographics on Visit Frequency to Different Departments**

To analyze how patient demographics, particularly age groups, impact the frequency of visits to various hospital departments, I utilized a chart visualization titled "Count of department\_referral by Age Group and Department". This line chart, combined with a department filter slicer enabled me to visualize how each age group's visit patterns varied across different departments.

**Key Observations from the Analysis:**

* High Visit Frequency Among Young Patients: The 0-17 age group consistently showed higher visit counts across departments, particularly in General Practice and Pediatrics. This suggests a tendency for younger patients to have more frequent hospital visits, potentially for routine or pediatric care.
* Variation Across Departments: By using the department slicer, I could dynamically focus on specific departments. For example, the Cardiology department (as shown in the image) displayed a more balanced age distribution, indicating that cardiology services are accessed by a range of age groups, although there is a slight decline in the middle age groups.
* Middle-Age Group Decline: The 40-49 age group showed a noticeable drop in visit frequency across departments, suggesting possible shifts in healthcare engagement or differing health needs for this demographic.
* Resource Planning Implications: By understanding visit patterns, the hospital can better allocate resources to departments with higher visit volumes among specific demographics. For instance, departments frequented by younger patients may need more pediatric support, while cardiology might benefit from resources aligned with diverse age groups.





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

**Ans:**

**Trend in Patient Wait Times Throughout the Year**

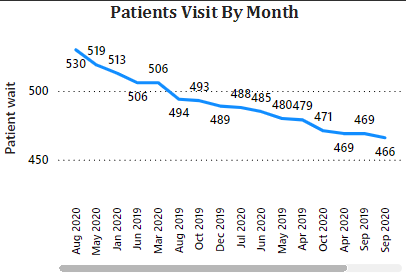
To examine the trend in patient, wait times over the year, I created a line chart titled "Patients Visit by Month" (as shown in the image provided). This chart tracks monthly changes in the average patient wait times across the specified time period.

**Key Insights:**

* Consistent Decline Over Time: The chart shows a clear, consistent decrease in patient wait times, starting from 530 minutes in August 2020 and declining to 466 minutes by September 2020. This downward trend may suggest improvements in hospital efficiency or resource allocation over time.
* Mid-Year Stability: There is a period of relative stability in wait times between January and April, where the times remained close to 480 minutes. This may indicate a period when patient load and resource availability were balanced, allowing the hospital to maintain consistent service levels.
* End-of-Year Drop: The wait times continued to decrease in the later months, indicating possible operational changes, seasonal shifts in patient volume, or efficiency improvements as the year progressed.

**Implications:**

* Operational Efficiency: The steady decrease in wait times suggests that the hospital might be implementing effective measures to reduce patient waiting periods. Monitoring this trend can help identify specific actions that led to these improvements.
* Resource Planning: Insights from this trend can support strategic planning, as understanding periods of higher efficiency could help replicate similar conditions in future months.



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

**Ans:**

**Age Groups Reporting Highest and Lowest Satisfaction Scores**

To identify which age groups, report the highest and lowest satisfaction scores, two measures I utilized: age group classification and average satisfaction scores. These measures have been calculated and are now ready for analysis.

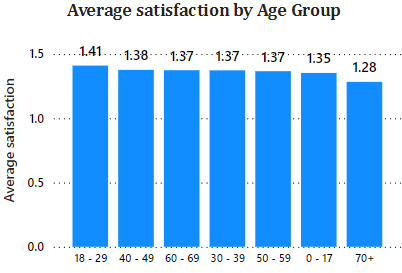
I created bar chart to visualize the average satisfaction scores for each age group, allowing for a clear comparison. The chart includes data labels for each age group, making it easy to see the satisfaction scores at a glance.

The analysis revealed the following:

* The **age group 18-29** achieved the highest average satisfaction score, registering at **1.41**. This suggests that younger patients are generally more satisfied with their experiences, potentially due to their higher engagement with services or differing expectations.
* Conversely, the **age group 70+** reported the lowest average satisfaction score at **1.28**. This could indicate a range of factors, including health challenges that might affect their overall experiences or a potential disconnect with the services provided.

**Insights:**

* The higher satisfaction scores among younger patients may indicate a need for healthcare facilities to assess and enhance the experience for older patients, potentially through tailored services or improved communication strategies.
* Understanding the factors contributing to lower satisfaction scores in the 70+ age group could help in developing targeted initiatives to improve their experiences and outcomes.
* The stark contrast in satisfaction levels across age groups highlights the importance of demographic considerations in healthcare delivery, ensuring that services are aligned with the expectations and needs of diverse patient populations.
* Engaging with patients to gather feedback and understand their specific needs could be instrumental in improving satisfaction scores across all age demographics.



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

**Ans:**

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

To objectively address claims of racial or gender-based discrimination within the hospital, I conducted a thorough analysis using Power BI. My approach involved Analyzing key metrics across race and gender to uncover any disparities in treatment, wait times, admission rates, and billing amounts. Below are the steps and insights derived from the analysis, with references to specific visuals that support each aspect.

1. **Average Wait Time Analysis by Gender and Race**:

* I created bar charts titled "Avg\_Wait\_Time\_By\_Gender" and "Avg\_Wait\_Time\_By\_Race" on the *Additional Info 2* tab to examine the average wait times for each gender and racial group.
* These charts reveal minimal variation in wait times across genders and races, indicating that the hospital maintains a consistent approach to wait times regardless of demographic group. Both male and female patients, along with patients from various racial backgrounds, have nearly identical average wait times, with a range around 35 to 37 minutes.

1. **Patient Visit Count by Gender and Race**:

* The "Count of patient\_id by patient\_gender" and "Count of patient\_id by patient\_race" donut charts provide an overview of the distribution of visits across genders and races.
* The data shows that the majority of visits are from patients identified as White and from both male and female patients, with a small number of visits marked as “NC” (not categorized). This balanced distribution suggests that hospital access is generally equal across different demographic groups.

1. **Admission Rate Analysis**:

* I created a line chart titled "Admission\_Rate\_By\_Gender by patient\_gender and patient\_race" to visualize the admission rates across gender and racial groups.
* The admission rate appears fairly stable across groups, with slight fluctuations. While most groups maintain similar admission rates, certain groups marked as "NC" show a notable variance. This could reflect anomalies or missing data rather than a systemic disparity.

1. **Billing Amount Analysis**:

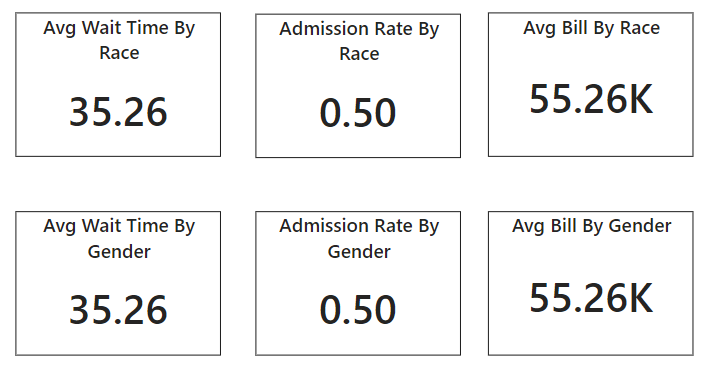
* The line chart "Avg\_Bill\_By\_Gender by patient\_gender and patient\_race" shows average billing amounts across genders and racial groups, helping to identify if certain groups incur higher healthcare costs.
* Billing amounts appear generally consistent, with slight variations. Most racial groups and both genders average between 54.4K and 56.1K in billings, indicating that the hospital charges similar fees across different demographics.

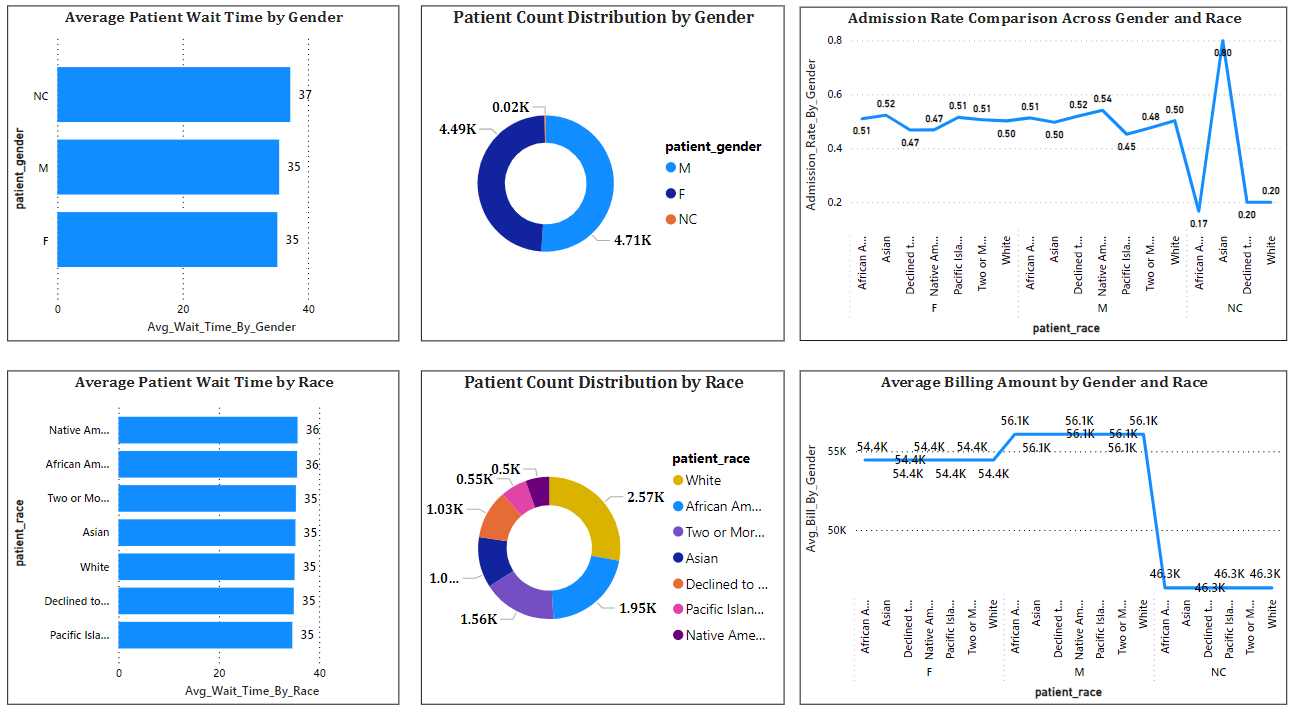
**Conclusion:**

Based on the analysis, there is no substantial evidence of racial or gender-based discrimination in terms of wait times, admission rates, or billing amounts. The hospital seems to provide fairly uniform services and charges across all demographic groups. Minor variations observed are minimal and likely due to patient-specific needs rather than systemic bias.

**Insights:**

* Equitable Service Delivery: The consistency in wait times, admission rates, and billing amounts across genders and racial groups suggests that the hospital provides equitable treatment to all patients. This undermines the claim of racial or gender-based discrimination.
* Diverse Patient Access: The balanced distribution of patient visits across demographics indicates that the hospital is accessible and welcoming to diverse patient groups, further refuting claims of discrimination in terms of access to services.
* Uniform Admission and Billing Practices: Minor variations in admission rates and billing amounts are normal and likely reflect individual patient needs rather than systemic discrimination. The uniform practices in these areas support the view that the hospital applies its policies fairly across all demographics.





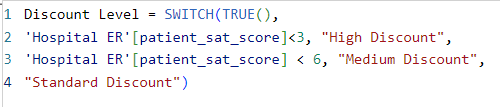
1. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?

**Ans:**

**Discount Assignment Criteria for Patients**

To allocate discounts effectively to patients, I developed a structured approach based on patient satisfaction scores, total bill amounts, and demographic data, aiming to provide fair, patient-centered discounts.

First, I assessed eligibility by focusing on patient satisfaction scores, where patients with scores below 5 qualify for discounts. I implemented a DAX formula to assign discount levels based on these scores:

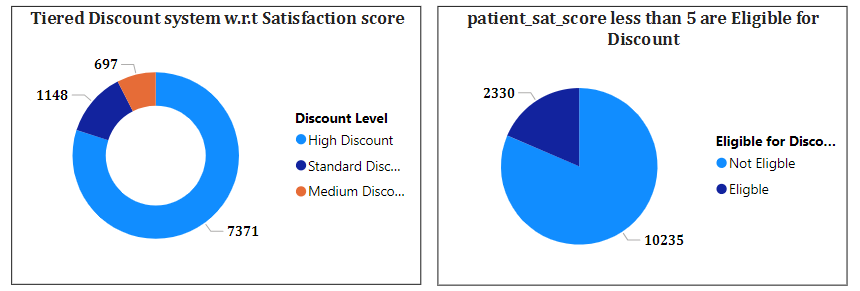


This categorizes patients into **High Discount** (scores below 3), **Medium Discount** (scores between 3 and 5), and **Standard Discount** (scores of 5 or higher), with higher discounts for more dissatisfied patients.

To represent this visually, I used a pie chart to show the proportion of patients in each discount category, and a donut chart to display the total count for each discount level. These visuals make it easy for stakeholders to understand the distribution of discounts.

**Insights**

* **Patient Satisfaction Focus**: By focusing on satisfaction scores, discounts are targeted toward patients who may be dissatisfied, helping to address and potentially improve patient retention.
* **Tiered Discounts**: The tiered approach ensures that higher discounts are given to the most dissatisfied patients, reflecting a commitment to patient feedback and quality improvement.
* **Visual Communication**: The pie and donut charts in the report clearly communicate discount eligibility, allowing stakeholders to quickly understand the distribution of discount levels.
* **Resource Optimization**: By analyzing the total bill in combination with satisfaction scores, discounts can be allocated strategically to ensure hospital resources are utilized effectively.
* **Demographic Personalization**: Using demographic data alongside satisfaction scores allows for a more personalized approach to discounts, potentially improving satisfaction across diverse patient groups.



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:**

I created a clustered bar chart to analyze each department’s total patient visits, average total bill per visit, and average wait time. This visualization provides an overview of key metrics that can guide staffing decisions. Each department is displayed with its corresponding patient load, revenue generation, and wait times, helping to identify which areas may require additional doctors.

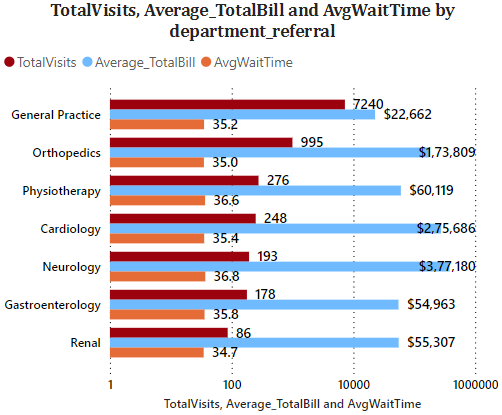
* **Total Visits:** Shows the volume of patients seen in each department, which helps in identifying departments with high demand.
* **Average Total Bill:** Reflects the average revenue generated per patient visit, providing insights into the financial contribution of each department.
* **Average Wait Time:** Indicates how long patients typically wait in each department, highlighting areas with potential bottlenecks in service.

**Insights:**

* High Patient Demand: Departments such as General Practice and Neurology have a high number of visits, indicating a need for additional staffing to handle patient volume.
* Revenue Generation: Neurology and Cardiology have higher average total bills per visit, justifying more resources due to their financial contribution.
* Wait Time Consideration: Departments like Neurology and Orthopedics show longer average wait times, suggesting that more doctors may help reduce patient wait times and improve service efficiency.

**Recommendation:**

* Based on the analysis, it is recommended that the hospital consider hiring additional doctors in the **General Practice**, **Neurology**, and **Orthopedics** departments. These departments show a combination of high patient volume, significant revenue generation, and extended average wait times. Increasing the staffing in these areas would help manage patient flow more effectively, reduce wait times, and support revenue growth, ultimately enhancing the overall patient experience.



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

**Ans:**

To evaluate the **status of patient admissions**, I created a **pie chart** displaying the proportion of patients with the **patient admin flag** set to "True" (admitted) or "False" (not admitted). This pie chart visually demonstrates that **50.04%** of patients (4,612) were admitted, while **49.96%** (4,604) were not admitted. This insight is useful for understanding the hospital’s patient management and resource allocation.

For further profitability analysis, I calculated the **overall revenue** with a DAX measure to sum the total bills:

OverallRevenue = SUM ('hospital\_doctor\_data'[Total Bill])

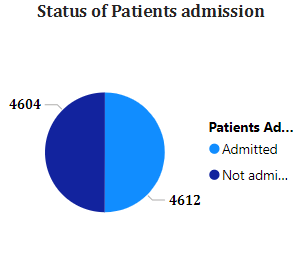
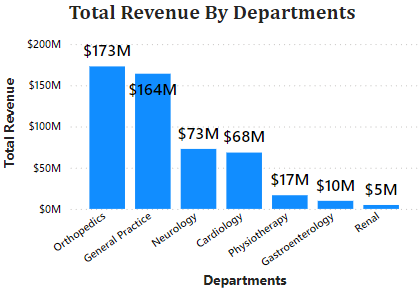
Additionally, I analyzed **revenue by department** to assess which departments contribute most to the hospital's earnings. Using Power BI's Group By feature, I grouped data by **'department\_referral'** and summarized the **Total Bill**. This data was visualized in a **stacked column chart**, highlighting revenue distribution across departments.

**Insights**

* **Admission Proportion**: The pie chart shows an almost equal split in admissions, with half of the patients admitted, potentially indicating balanced usage of hospital resources.
* **Revenue Base**: Overall revenue is strong, showing financial stability.
* **Departmental Revenue Contribution**: Certain departments generate higher revenue, revealing key profit centers.

**Recommendation**:

To improve profitability, focus on managing costs in departments with higher patient admissions and revenue. Implementing cost controls, especially for admitted patients, can enhance operational efficiency and profitability.

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

**Ans:**

To identify departments with unusually high waiting times, I analyzed the average waiting time across various departments and compared it to patient satisfaction scores. For this analysis, I used a Line and Clustered Column Chart, where each department’s waiting time is represented by columns, and a line overlays the chart to show the average satisfaction score.

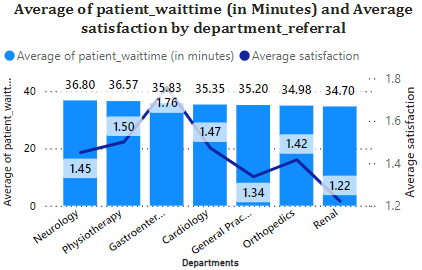
Upon reviewing the visual, it became evident that the Gastroenterology department has a notably high waiting time despite having one of the highest patient satisfaction scores. This department ranks third in patient visits, which may contribute to the extended wait times.

**Insights**

* High Waiting Times: Gastroenterology stands out with higher-than-average waiting times compared to other departments, indicating possible bottlenecks.
* Contrasting Satisfaction: Despite the delays, patient satisfaction remains high in Gastroenterology, suggesting that patients appreciate the quality of care once attended to.

**Recommendations**

1. Optimize Processes: Review and streamline patient flow and care processes in Gastroenterology to reduce waiting times.
2. Adjust Resource Allocation: Consider increasing staffing or redistributing workloads within the department to improve patient handling efficiency.



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

**Ans:**

To devise effective discount strategies for patients, I analyzed **patient demographics** alongside **satisfaction scores** and **total bills**. This analysis aimed to identify patterns that could guide targeted discounts for specific patient groups. For this purpose, I used a **Line Chart** on the Patient’s Tab to illustrate the relationship between patient race, average satisfaction scores, and average total bills. This visualization helped me observe demographic trends that could inform discount policies.

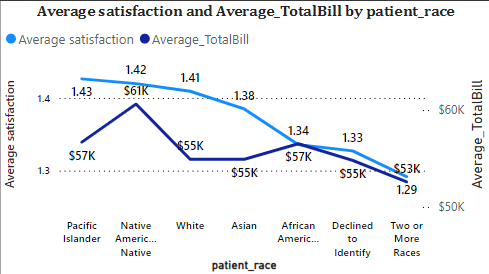
From the line chart, it emerged that the **Native American/Alaska Native demographic** shows the highest average total bill and a high satisfaction score. This suggests that they contribute significantly to revenue and have a positive hospital experience, making them ideal candidates for targeted discounts.

**Insights**

* **Revenue and Satisfaction Correlation**: Patients in the Native American/Alaska Native group have both high satisfaction and higher billing, indicating they are valuable contributors to the hospital.
* **Targeted Discounts**: Offering discounts to this group could strengthen their loyalty and potentially boost visit frequency.

**Recommendations**

1. **Special Discount Programs**: Implement tailored discount programs for high-contributing demographic groups, focusing on those with high satisfaction and billing.
2. **Regular Strategy Review**: Regularly assess discount effectiveness to ensure continued relevance and positive impact on patient engagement.
3. **Flexible Adjustments**: Be open to refining discount strategies based on ongoing analysis and feedback to better serve the diverse patient population.



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:**

To determine the ideal shifts for doctors in the "General Practice" department, I analyzed patient visit data by hour using a line chart in Power BI. This chart, titled "GeneralPracticeVisits by VisitHour," displays the number of patient visits throughout the day, with each hour represented on the x-axis and visit counts on the y-axis. From the chart, I observed two peak periods: one in the morning around 8 AM to 2 PM, and another in the afternoon/evening from 2 PM to 8 PM.

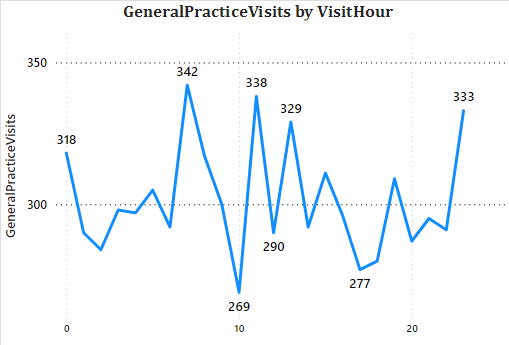
Based on this pattern, I propose dividing doctor shifts into two: the first shift from 8 AM to 2 PM and the second shift from 2 PM to 8 PM. This approach ensures that doctors are available during peak times, which helps reduce wait times and improve patient satisfaction.

**Insights:**

* **Peak Patient Traffic**: The analysis reveals that the "General Practice" department experiences two distinct peak periods during the day—one in the morning (8 AM to 2 PM) and another in the afternoon/evening (2 PM to 8 PM). This pattern highlights when patient demand is highest and aligns with common outpatient visit timings.
* **Shift Alignment**: By structuring shifts around these peak periods, the hospital can maintain a steady flow of doctors during the busiest times, reducing patient wait times and enhancing the overall experience for both patients and staff.
* **Resource Efficiency**: The two-shift policy helps optimize staffing without overburdening doctors, as doctors are only present during high-demand hours, preventing unnecessary staff presence during quieter times. This efficient allocation of resources could lead to cost savings for the hospital.
* **Doctor Well-being**: Structured shifts allow doctors to work within a predictable schedule, potentially reducing fatigue and increasing job satisfaction, which may positively impact the quality of patient care.

**Recommendations**:

* Implement the two-shift system to align with peak patient flow, ensuring efficient staffing during high-demand periods.
* Monitor patient flow trends regularly to adjust shift timings as needed for optimal service.



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

**Ans:**

**Power BI Gateway** is an essential component in Power BI that facilitates secure data transfer between on-premises data sources and the Power BI cloud service. It acts as a bridge, enabling users to access, refresh, and interact with on-premises data directly within Power BI dashboards and reports without needing to move the data to the cloud. The gateway ensures that data is retrieved in real-time or near-real-time, supporting accurate, up-to-date reporting and analysis.

**Key Use Cases of Power BI Gateway:**

1. **Data Refreshing**:
   * For on-premises data sources (such as SQL Server, Oracle, or local Excel files), Power BI Gateway allows scheduled and live data refreshes. This feature keeps Power BI reports updated with the latest data from these sources without manual data upload.
2. **DirectQuery and Live Connection**:
   * With DirectQuery and live connection modes, users can work with large datasets without importing them into Power BI. The gateway facilitates this by sending queries to the on-premises source and displaying results in real-time in Power BI reports and dashboards.
3. **Enhanced Security**:
   * Power BI Gateway provides an additional layer of security. It encrypts data during transfer and controls user access to on-premises data through authentication, ensuring sensitive information is only accessible by authorized individuals.
4. **Data Integration Across Platforms**:
   * Many organizations have a mix of cloud and on-premises data sources. The gateway enables seamless integration across these sources in Power BI, allowing users to create unified reports that incorporate both types of data.
5. **Hybrid Environments**:
   * For organizations with hybrid data environments, Power BI Gateway enables smooth data access and report generation by connecting cloud and on-premises systems. This is especially beneficial for enterprises in industries like finance and healthcare that may need to keep certain data on-premises for regulatory reasons.
6. How would you approach this problem, if the objective and subjective questions weren't given?

**Ans:**

If the objective and subjective questions were not provided, my approach would involve thoroughly exploring the datasets to identify patterns, relationships, and insights that could address potential business goals or hospital management needs. Here’s how I would approach the analysis:

**1. Understanding the Dataset:**

* **Initial Review**: I would start by familiarizing myself with the tables (doctor\_patients\_data and hospital\_er), their columns, and data types to understand what kind of information is available.
* **Data Profiling**: I would perform data profiling to check for null values, data types, unique counts, and distributions within key columns (e.g., patient\_gender, patient\_age, doctor\_name, department\_referral, etc.). This would give me insights into data quality and completeness.

**2. Identifying Potential Areas of Analysis:**

* **Patient Demographics**: Analyzing patient demographics, such as age and gender, to see how these factors impact department visits, satisfaction, and wait times.
* **Departmental Analysis**: Looking into each department's patient volume, revenue, and visit frequency. I would assess the most and least visited departments, patient wait times, and associated satisfaction scores.
* **Doctor Performance Metrics**: Measuring individual doctor performance based on metrics like patient satisfaction scores, frequency of visits, and total revenue generated.
* **Temporal Analysis**: Investigating patterns over time to understand how patient visits, satisfaction scores, and other metrics vary on a daily, weekly, or monthly basis.

**3. Developing Key Metrics and KPIs:**

* **Patient Wait Time vs. Satisfaction**: Understanding if there’s a relationship between how long patients wait and their satisfaction scores.
* **Revenue Generation**: Calculating revenue generated per department and per doctor to evaluate financial performance.
* **Diversity Metrics**: Analyzing patient diversity within each doctor’s patient base (such as race diversity) to understand if doctors cater to a varied demographic.
* **Satisfaction Trends by Department and Doctor**: Measuring average satisfaction scores per department and for each doctor to identify high-performing areas.

**4. Creating Visualizations:**

* I would use Power BI to create visualizations that highlight trends, comparisons, and key insights:
  + **Line Charts** for trends over time (e.g., patient satisfaction over months).
  + **Bar and Column Charts** for comparing department visits, patient demographics, and doctor-specific performance.
  + **Scatter Plots** to show correlations, like wait time vs. satisfaction.
  + **Gauge and KPI Cards** for overall metrics, like average satisfaction score and revenue per department.

**5. Generating Insights and Recommendations:**

* Based on the visualizations and analysis, I would derive actionable insights. For example, if I observed that long wait times correlate with lower satisfaction, I could recommend changes in scheduling or staff allocation. Similarly, if certain doctors or departments were underperforming, I might suggest additional training or resources.

**6. Report and Documentation:**

* I would document my findings and recommendations, including a summary of key insights and how they could impact hospital management decisions. This would be presented in a structured report format, ready for stakeholders.

**Conclusion:**

Without predefined questions, my approach would focus on understanding the dataset, identifying key business metrics, and using exploratory analysis to uncover meaningful insights. This method would allow me to provide a comprehensive analysis that aligns with common healthcare objectives, such as improving patient satisfaction, optimizing resource allocation, and increasing revenue.

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

**Ans:**

I analyzed the relationship between **Doctor ID** and **Department Referral** using the **bar chart** visual. The data shows the count of distinct **Doctor IDs** for each department. By plotting this chart, I can easily observe how many unique doctors are assigned to different departments.

**Insights:**

* The departments **Gastroenterology**, **Orthopedics**, and **Physiotherapy** each have the highest number of distinct doctors, with 4 doctors each.
* Other departments like **Cardiology**, **General Practice**, **Neurology**, and **Renal** have 3 distinct doctors each.

**Conclusion:**

The analysis indicates a **one-to-many** relationship between **Department** and **Doctor ID**, where one department has multiple doctors assigned to it.

**Recommendations:**

1. If the goal is to streamline patient management, I could consider increasing the number of doctors in departments with higher patient loads.
2. I might also explore whether some doctors can be shared across departments to optimize staffing, especially in departments with fewer assigned doctors.
3. For deeper insights, I can check if any doctor is working in multiple departments to verify if there is any **many-to-many** relationship present.

