**Answers to the Objective and Subjective Questions:**

**Objective Questions:**

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

* We have cleaned the data by replacing the null value in the **Hospital ER** table with the average value of the patient\_sat\_score as volume of the null value is high and we cannot skip it from the analysis. We have changed the data format as well for the in Hospital ER, Doctor\_patient\_data Table.

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

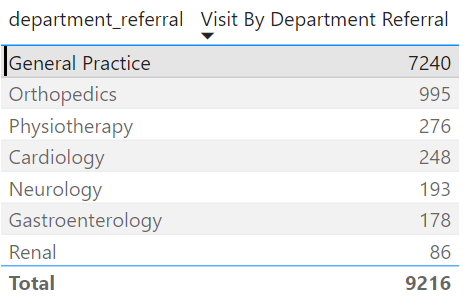
* To find the average duration of patient spends before receiving care we used the below DAX Functions:
  + **Average Wait Time = AVERAGE('Hospital ER'[patient\_waittime])**

Average Wait Time is **35.26 min**

****

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.

* In order to calculate total number of visits to each department based on referrals to understand which department are most frequently visited we have used the DAX function and below process:
  + First of all we have calculated the count of the visit as per department referrals by using below DAX Function:
    - Visit By Department Referral **= CALCULATE(COUNTROWS('Hospital ER'), GROUPBY('Hospital ER','Hospital ER'[department\_referral]))**
  + After calculating the count of the visit as per the department referral we have created a table in report view and put the field **Department Referral** and calculated Visit By Department Referral in column section and sort in from highest to lowest as shown below:



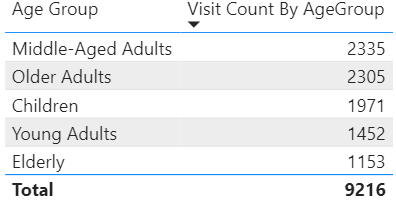
You can then see the most frequent visit is in General Practice Department which is **7240.**

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

* For calculating the patient visits according to different age groups to see which demographics utilize health services the most we have firstly created age groups for the different ages which are as follows:
  + Age less than 18 as Children
  + Age between 18 to 30 as Young Adults
  + Age between 30 to 50 as Middle-Aged Adults
  + Age between 50 to 70 as Older Adults
  + Age greater than 70 as Elderly
* After defining age group we have used the below formula in order to count the patients visit as per the age group and also to count the total number of patient count visit:
  + **Visit Count By AgeGroup** = **VAR SelectedAgeGroup = SELECTEDVALUE('Hospital ER'[Age Group])**

**RETURN**

**IF(ISINSCOPE('Hospital ER'[Age Group]),CALCULATE(COUNTROWS('Hospital ER'),'Hospital ER'[Age Group] = SelectedAgeGroup),CALCULATE(COUNTROWS('Hospital ER')))**



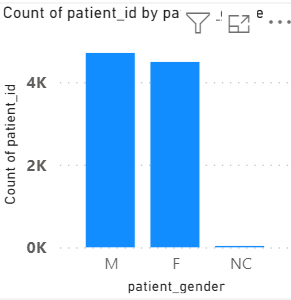
We can see that the age group of 30 to 50 which is middle-aged adults utilize the healthcare service most from the above.

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?**

* Yes there were Null values in the **patient\_sat\_score** of the **Hospital ER table.** We have replaced the same with the value as the average value of the satisfaction score as the volume of the null value was as high as 72%. We have created a new column for the patient satisfaction column by replacing the null value with the average value of the satisfaction score. As due to high volume it was not possible to eliminate that data as it impacted the analysis on higher margin.

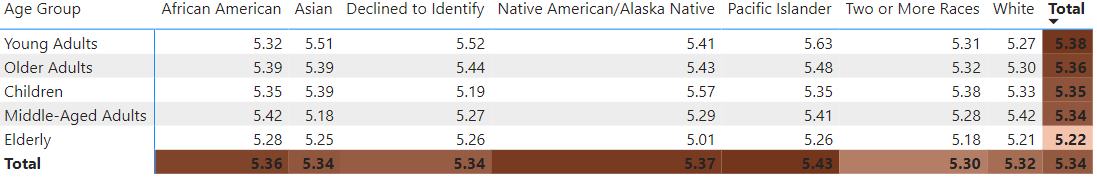
1. **Is there any relation between the number of visits and the Gender of the patients?**

* Number of **male patients** is the highest with the count of **4705**, followed by the number of **female patients** with the count of **4487** and **24 patients** are in NC(not categorized).



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.**

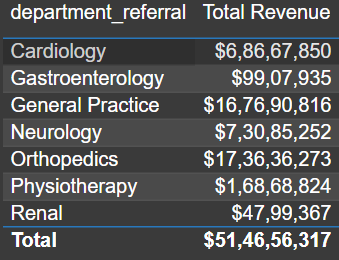
* In order to determine the relationship between patient satisfaction scores, their age groups and racial backgrounds we have used the Matrix Visualization and conditional formatting to visualize satisfaction scores by age group and their racial backgrounds.
* The darker the background, the greater the average satisfaction score and vice versa.



From the above matrix table we can see that the average satisfaction score for Elderly people is the lowest as well as forTwo or More Races too.

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

* In order calculate the department wise revenue we have to create table in Report View and drag the total bill as well as department\_referral from Doctor\_Patient\_data table into column which is as follows:

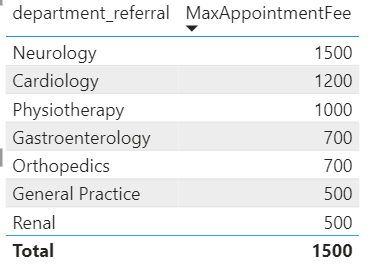


1. **Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.**

* To calculate the highest appointment fees in general as per the department we have created new measure which is as follows:

**MaxAppointmentFee = MAXX(SUMMARIZE(Doctor\_patient\_data, Doctor\_patient\_data[department\_referral], "MaxFee", MAX('Doctor\_patient\_data'[Appointment Fees])), [MaxFee])**

After creating measure, create table in visualization and put MaxAppointmentFee as well as department\_referral in the column and you will get the result as follow:



From the above table you can say that **Neurology** is charging the highest appointment fee in general which is **1500**.

1. **Create a tabular visualization in the Report view which consists of Month-wise total visits in the hospital. Add a third column in the table that consists of the previous month’s total visits for each month’s row. Also, include a column that states whether the visits in a month are greater than that of the previous month's visits.**
2. **Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.**

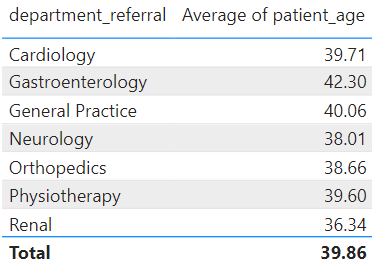
* To calculate the total number of patients who have visited Dr. Smith we have used the below DAX function:

**TotalPatientsVisitedDrSmith = CALCULATE(COUNTROWS(Doctor\_patient\_data),Doctor\_patient\_data[Doctor Name] = "Dr. Smith")**

Total number of patients who have visited Dr. Smith is **5986.**

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?**

* We can use table visualization in the report view to calculate the average age of the patients who visit the Orthopedics department. Also, we can use the same for all the department as the table will reflect the average age of all the department as shown below:



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

* Yes. In the Doctor\_patient\_data table all the data were in text as well as in number format i.e. in general format. Likewise, Appointment Fees and Total Bill should be in the whole number format but the same is in general format. In that case we can not get the numeric value in the table for the total revenue while calculating department wise total revenue. Also we have changed the Doctor Name and department referral to text data type.
* Also, Date was in date/time format in the Hospital ER table and we have changed the same to the date format.

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?**

* When we add a column in Power Query, the M language code that appears in the formula bar depends on the type of transformation we are applying to our data.

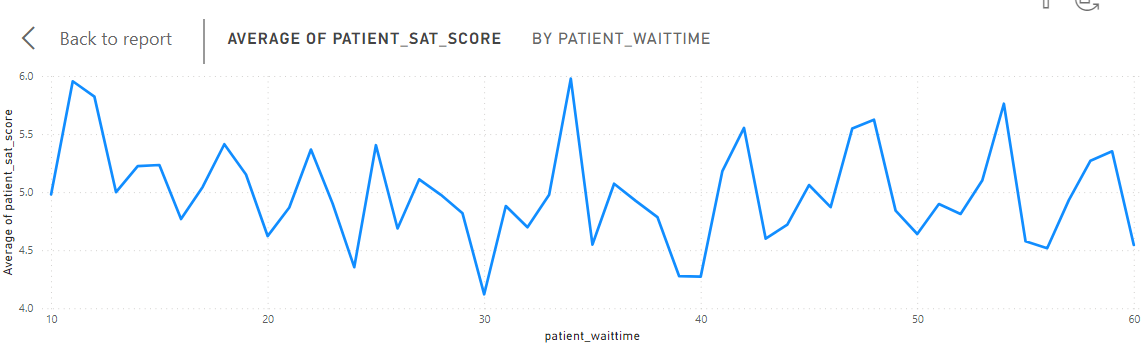
M-query is the language used by Power Query to perform data transformation tasks. It's a functional language that's designed to be expressive and easy to read, allowing us to define step-by-step transformations on our data.

Overall, M-query is a powerful tool for data preparation and transformation in Power Query, enabling us to clean, reshape, and enrich our data before loading it into our Power BI data model or other analysis tools.

**Subjective Questions**

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

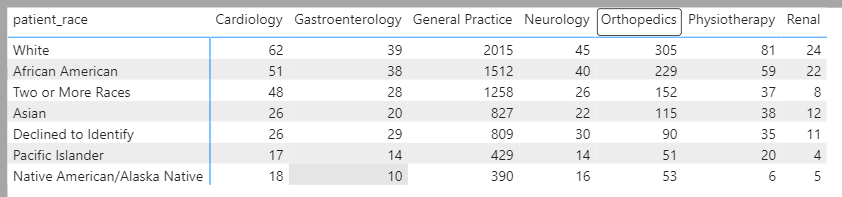
* There is no particular relation between the patient wait time and the satisfaction score where the satisfaction score is independent of the factor wait time. We can see the same from the below chart and table as well. There is no effect of the wait time on the satisfaction score.



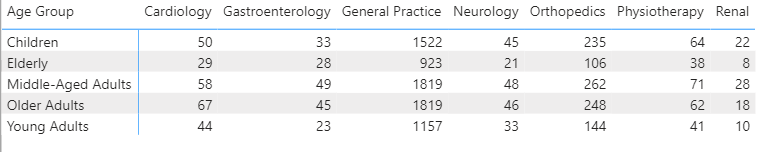
You can see in the above graph that there is no relation between patient wait time and satisfaction score.

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

* If we check the demographics effect we can consider the several demographics factors that are gender, age group , location, etc. We have checked and found the below observations considering the different demographic factors while analyzing.
  + Gender affects frequency of visits to different departments.
    - M had the highest total Visit By Department Referral at 4705, followed by F at 4487 and NC at 24.
    - General Practice in patient\_gender M made up 39.98% of Visit By Department Referral.
    - M had the highest average Visit By Department Referral at 672.14, followed by F at 641 and NC at 8.
  + Patients Race frequency of visits to different departments.
    - ﻿White in department\_referral General Practice made up 21.86% of Visit By Department Referral.﻿
    - Native American/Alaska Native have the lost number of visits among all the races.
    - Below is the table showing the visit frequency as per race to particular department:

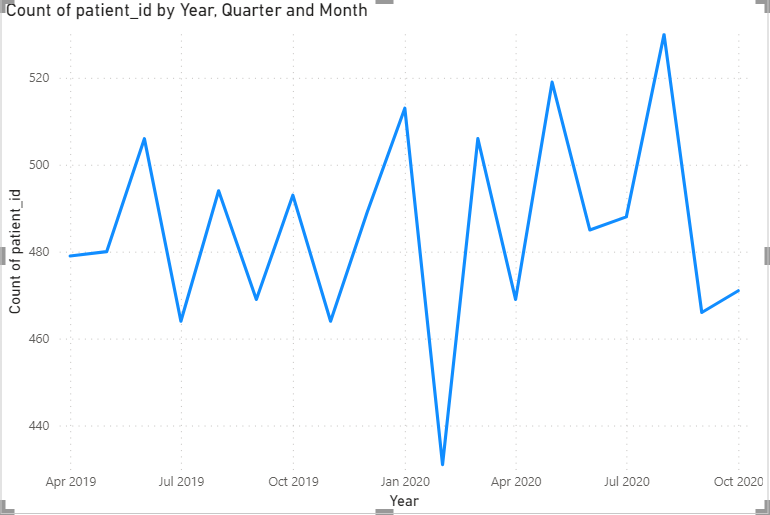


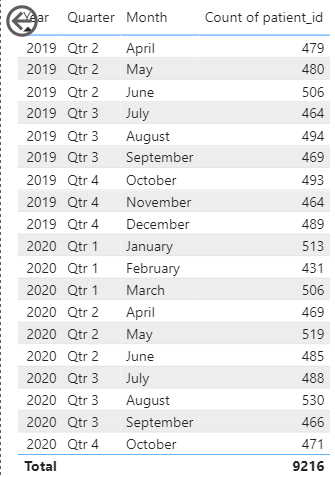
* + Age Group Patients Race frequency of visits to different departments.
    - ﻿Middle-aged Adults and Older Adults have the highest visit almost in all departments.
    - Elderly people have the least visit among all departments.



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

* Yes. There is an ups and down trend in the volume of patients visits throughout the year. You can see the same in the below chart as well as table except the first two months.

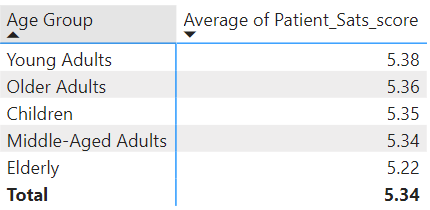




Except April and May 2019, if the patient visit goes in one month then it declines in the subsequent month and after that it rises in the other month.

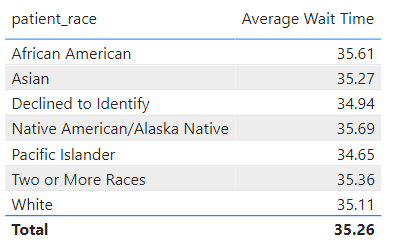
1. **Which age groups report the highest and lowest satisfaction scores?**

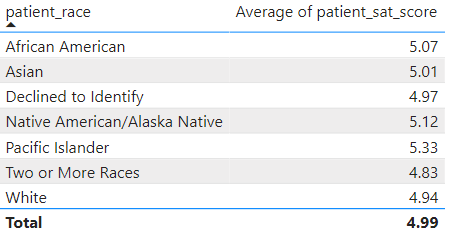
* Age Group of Young Adults i.e. age between 18 to 30 has the highest satisfaction scores while the Elderly Age Group i.e age greater than 70 has the lowest satisfaction scores.

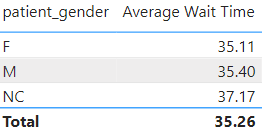
****

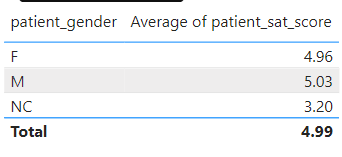
1. **Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?**

* **For identifying the claim we have considered the below factors:**
  + Patient Race vs Average Waiting Time
  + Patient Race vs Average Satisfaction Score
  + Patient Gender vs Average Waiting Time
  + Patient Gender vs Average Satisfaction Score
* As per the analysis we have found that there is no discrimination based on the racial as the average wait time as well as the average sat score for all the racial are found nearby.
* While based on gender we can say that there is discrimination with the gender NC as the average waiting time is comparatively higher than M and F. Also, the average satisfaction score for the gender NC is comparatively less than M and F.
* We can check same from the below table:



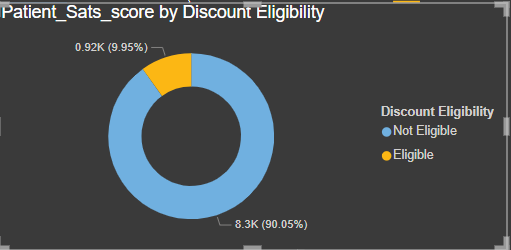






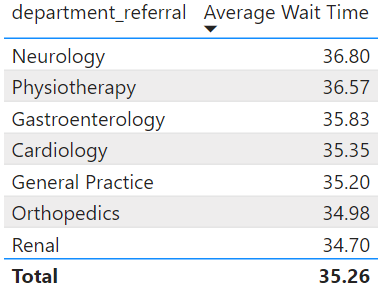
1. **The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?**

* According to me, discounts to patients can be offered considering the multiple factors such as patient visit frequency, patient satisfaction score, Appointment fees of Doctor, department with lowest patient visit in order to promote more visits of patients to such departments.
* In the given case we are considering a patient satisfaction score for offering a discount to patients as we are not aware of the cost for each patient in a particular department. Hence we can not consider the appointment or revenue of a particular department in order to consider it as a deciding factor.
* Firstly we will set the criteria for the eligibility of discount like discount will be availed to the patients who has satisfaction score more than 6.



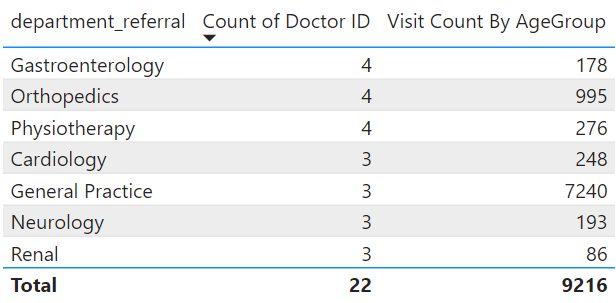
1. **The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.**

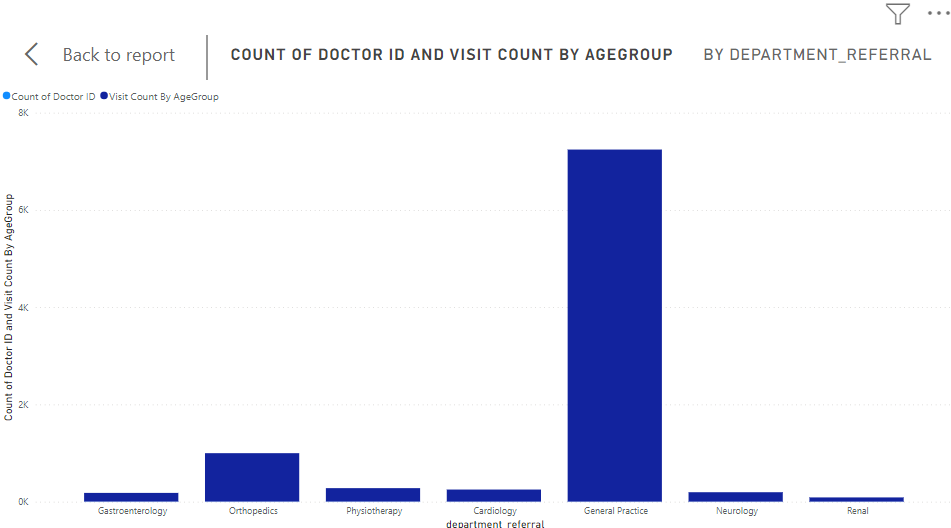
* I would personally suggest hiring the new doctors in the Neurology and Physiotherapy department as the average waiting time for both departments is higher comparatively than the other departments. In such cases, patients need to wait more and the higher average waiting time indicates that the workload is more as compared to resources.
* In case if we consider the patient visit then I would suggest hiring at least one new doctor to the General Practice department as the patient visit in the General Practice department is higher comparatively than the other departments.
* Following tables are the analysis and observations considered for this suggestion.
  + **Department vs Average Wait Time**



You can see that the average wait time in the Neurology and Physiotherapy Department is comparatively higher than the other department.

* Secondly if we consider the patient visit as a deciding factor then we will suggest at least one new doctor in General Practice. You can see the same in the below table and chart.





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

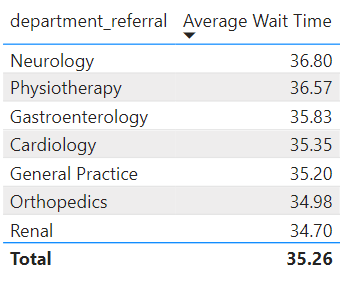
* In order to determine the profitability of any business we require at least two elements which are cost and revenue.
* In the given case, we have the revenue generated from the visit of the patient but there is no cost element. Hence, we cannot determine the profitability

We can just calculate the total revenue generated from the visit of patient by the below formula:

**Total Revenue = SUM(Doctor\_patient\_data[Revenue])**

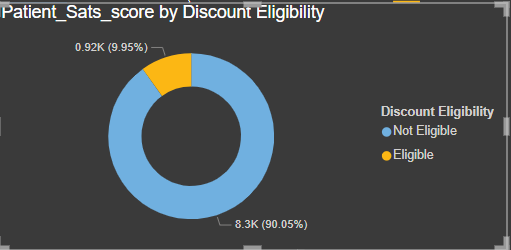
1. **Any Department for which the waiting time is oddly large?**

* We can not say that there is an oddly large waiting time in any department as compared to the average timing but the average waiting time for Neurology and Physiotherapy as well as Gastroenterology is little high compared to the average waiting timing for all the departments. We can see the same from below table:



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

* According to me, discounts to patients can be offered considering the multiple factors such as patient visit frequency, patient satisfaction score, Appointment fees of Doctor, department with lowest patient visit in order to promote more visits of patients to such departments.
* In the given case we are considering a patient satisfaction score for offering a discount to patients as we are not aware of the cost for each patient in a particular department. Hence we can not consider the appointment or revenue of a particular department in order to consider it as a deciding factor.
* Firstly we will set the criteria for the eligibility of discount like discount will be availed to the patients who has satisfaction score more than 6.



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?**

* To establish two shifts for the doctors in the General Practice department and effectively allocate staff, follow these steps:
* **Analyze Patient Visits:** Review historical data on patient flow throughout the day to identify peak times when the most doctors are needed. This could involve analyzing appointment schedules, emergency room visits, and walk-in patient numbers.
* **Consult Staff Preferences:** Survey the doctors in the General Practice department to understand their preferences regarding shift timing. Some doctors may prefer earlier shifts while others may prefer later shifts based on personal obligations or work-life balance considerations.
* **Consider Peak Hours:** Determine the hospital's peak hours based on patient traffic and adjust shift timings accordingly. For example, if the hospital experiences high patient volumes in the morning and afternoon, consider establishing shifts that cover these periods.
* **Create Balanced Shifts:** Divide the doctors into two shifts that provide adequate coverage throughout the day. Aim to distribute doctors evenly between the shifts to ensure sufficient staffing during peak hours and maintain continuity of care.
* **Rotate Shifts:** Implement a rotation system to ensure fairness and prevent burnout among doctors. Rotate doctors between the two shifts periodically to give them exposure to different patient populations and work environments.
* **Account for Breaks and Rest:** Schedule breaks and rest periods within each shift to prevent fatigue and maintain productivity. Ensure that doctors have sufficient time for meals, rest, and downtime between patient appointments.
* **Monitor and Adjust:** Continuously monitor the effectiveness of the two-shift policy and adjust as needed based on patient demand, staff feedback, and operational considerations. Solicit feedback from doctors and hospital administrators to identify areas for improvement and make necessary adjustments.
* Implementing a two-shift policy for doctors in the General Practice department can be beneficial for the hospital in several ways:
* **Improved Patient Access:** By aligning staffing with peak hours, the hospital can ensure that patients have timely access to medical care and reduce waiting times for appointments.
* **Enhanced Efficiency:** Distributing doctors across two shifts can optimize resource utilization and improve operational efficiency, leading to better patient throughput and satisfaction.
* **Flexibility and Adaptability:** A two-shift policy provides flexibility to accommodate changes in patient demand and adapt to fluctuations in workload throughout the day.
* **Better Work-Life Balance:** By offering a variety of shift options, the hospital can support doctors in achieving a better work-life balance, which can contribute to higher job satisfaction and retention rates.

Overall, a well-designed two-shift policy for doctors in the General Practice department can help the hospital meet patient needs effectively while also promoting staff satisfaction and productivity.

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

* Power BI Gateway is a tool provided by Microsoft as part of the Power BI suite. It serves as a bridge between on-premises data sources and Power BI services in the cloud. Its primary function is to facilitate secure data transfer between on-premises data sources and Power BI cloud services, ensuring that Power BI reports and dashboards can access and refresh data from on-premises sources in real-time or on a scheduled basis.

Here's a breakdown of its key components and use cases:

* **Data Connectivity:** Power BI Gateway allows users to connect to various on-premises data sources such as SQL Server databases, SharePoint lists, Excel files, Oracle databases, and more. It establishes a secure connection between these data sources and the Power BI service in the cloud.
* **Data Refresh:** One of the primary use cases of Power BI Gateway is to enable automatic data refresh for Power BI reports and datasets that are based on on-premises data sources. Users can schedule data refreshes to ensure that their reports and dashboards reflect the most current data.
* **DirectQuery**: Power BI Gateway supports DirectQuery mode, which allows Power BI reports to query on-premises data sources in real-time. This is particularly useful for scenarios where users need access to up-to-the-minute data without having to import it into Power BI datasets.
* **Live Connection:** Another use case is enabling live connections to on-premises Analysis Services models or SQL Server databases. This allows users to interact with the data in real-time directly from the on-premises source.
* Data Security and Compliance: Power BI Gateway ensures that data transferred between on-premises data sources and Power BI services is encrypted and secure. It also helps organizations maintain compliance with data governance policies by allowing them to control access to on-premises data sources.

Overall, Power BI Gateway plays a crucial role in integrating on-premises data sources with Power BI cloud services, enabling organizations to leverage the benefits of both on-premises and cloud-based analytics and reporting capabilities.

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

* If the objective and subjective questions weren't given we would follow the below approach to gain the insights:
* **Understand the Problem:** Begin by thoroughly understanding the requirements and objectives of the analysis. Clarify what insights or outcomes are desired from the data.
* **Data Exploration:** Explore the available data to identify relevant metrics and dimensions that could provide insights into the performance of doctors in the General Practice department. This may include data on patient visits, appointment scheduling, doctor availability, and patient satisfaction.
* **Define Key Performance Indicators (KPIs):** Based on the objectives of the analysis, define KPIs that can measure the performance of doctors in the General Practice department. These KPIs could include metrics such as patient wait times, appointment adherence, patient outcomes, and doctor productivity.
* **Data Preparation:** Prepare the data for analysis by cleaning, transforming, and aggregating it as needed. This may involve joining multiple datasets, handling missing or erroneous data, and creating calculated fields.
* **Analysis and Visualization:** Use data visualization tools or software (Power BI) to analyze the data and create visualizations that help identify trends, patterns, and insights related to doctor performance. This could include charts, graphs, dashboards, and reports.
* **Identify Patterns and Trends:** Analyze the data to identify any patterns or trends related to doctor performance in the General Practice department. Look for correlations between different variables and investigate any outliers or anomalies.
* **Generate Insights and Recommendations:** Based on the analysis, generate insights into the performance of doctors in the General Practice department and identify areas for improvement. Develop recommendations or action plans to address any issues or challenges identified.
* **Iterative Analysis and Refinement:** Iterate on the analysis process as needed, refining the approach and exploring additional data sources or variables to gain deeper insights into doctor performance. Continuously monitor and evaluate the effectiveness of any interventions or changes implemented.

By following this approach, we can still effectively analyze the performance of doctors in the General Practice department and derive actionable insights to improve their performance, even without using Power BI's objective and subjective questions.

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

* Yes, we can analyze the type of relationship between the doctor id and department in the model view by dragging the doctor id and dropping it on the department it will ask to create the relationship and we can check the type of relationship over there.
* There are many to many relationships between doctor id and department.

