**Objective Questions**

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

Answer:

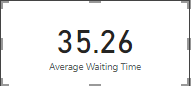
* On the initial analysis the data types of the columns were checked to ensure the integrity of the data and in the dataset all the columns were in correct data types as per the schema.
* In the column, “patient sat score” there were null values and they could show error during further analysis so the null values in the columns were replaced with 0 to ensure the integrity of the data and avoid errors.
* Date column contains both date and time, so for the date and time were extracted from the columns for conducting the analysis.
* Dataset contains initial of the first name and last name, so for the efficiency both the columns were merged and new column “Patient Name” was created, which would help in analysis.
* New column “Gender” was created so that the gender name can specified to the symbols, i.e., M – Male and F – Female.

Question 2: **Assess the Average Waiting Time:** Analyse the patient wait times to identify the average duration a patient spends before receiving care.

Answer:

Average duration a patient spends before receiving care = 35.26

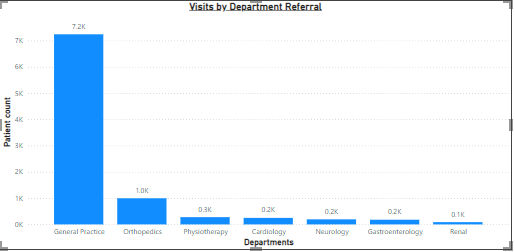
Formula used: Average Waiting Time = AVERAGE ('Hospital ER (1)'[patient\_waittime])

****

Question 3: **Visits by Department Referral:** Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.

Answer:

Visualization:



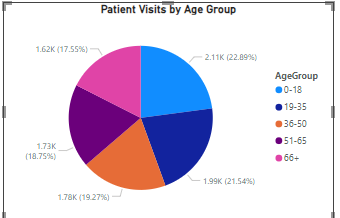
Insights:

* From the visual we can see the distribution of patients on the basis of visits to each department based on referrals
* General Practice department receives most of the patients (7.2K).
* Renal department receives least patients (0.1K).

Question 4: **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

Answer:

Visualization:



Insights:

* Patient count of different age group can be observed from the visual.
* Most number of patients are from the 0-18 age group (2110).
* 66+ age group has the least number of patients (1617).

Approach used:

* Age group column was created using the custom column feature of power bi and the patients were grouped on the basis of their age and custom groups were created.
* Formula used for custom column:

if [patient\_age] <= 18 then "0-18"

else if [patient\_age] <= 35 then "19-35"

else if [patient\_age] <= 50 then "36-50"

else if [patient\_age] <= 65 then "51-65"

else "66+"

* Column chart was created using the age group column and patient id column for the visualization.

Question 5: Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for?

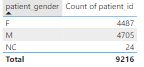
Answer:

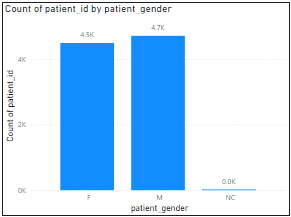
* There were null values in the “patient sat score” column.
* There are several ways to handle the null values but each is dependent on the use case, if there are null values in numerical column like age then the average of the age values can be used to fill the missing values.
* The most general way to handle null values is to remove them but the problem would be that if the values are removed then whole row would be removed which might remove the important data which would not be the best approach.
* But in this case, removing null values in “patient sat score would also remove the important data, so I replaced the null values with 0 as most of the data has null values in the column because of which mean or average of the column would not be accurate and replacing with 0 would avoid errors in the analysis and can be used for the calculations.

Question 6: Is there any relation between the number of visits and the Gender of the patients?

Answer:

Visualization:

****



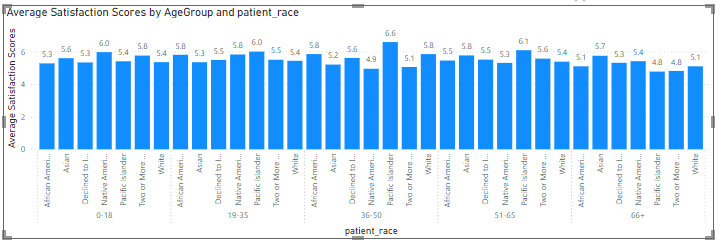
Insights:

* From the visual it can be observed that there is no relation between gender and number of visits.
* Both Male (4.5K) and Female (4.7K) have no significant difference between the number of visits, as for NC there is no mention of NC in the schema, so as per my judgement NC (24) would be the patients who might have forgot to mention their gender, even then there would no significant change as the count is very low.
* So, there is no relation between number of visits and gender.

Question 7: Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.

Answer:

Visualization:



Insights:

* As most of the data provided had null values, so the average of satisfactory scores might not provide accurate ratings.
* So, the measure was created to calculate the average satisfaction score excluding the null values i.e., 0, because if all the ratings were included the average would be so low (around 1-2) which would have drastic negative impact on the reputation of the hospital.
* From the provided dataset, we can observe that the ratings scale is from (1-10) and the average is ratings are in the range of (5.47) which is very average rating and from the visual we can understand the the relationship between patient satisfaction scores, their age groups, and racial backgrounds and can make improvements in the patient experience.

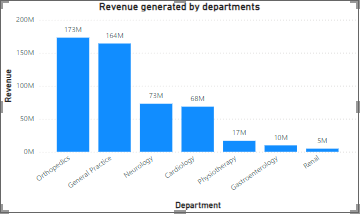
Approach used:

* Previously created age group column and race column was used to create cluster column chart along with the average satisfaction score measure, using which we can observe the relationship between patient satisfaction scores, their age groups, and racial backgrounds.

Question 8: The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.

Answer:

Visualization:



Insights:

* Revenue generated by each department can be observed from the visual.
* Orthopaedics department generates most revenue (173M).
* Renal department generates lowest revenue (5M).

Question 9: Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.

Answer:

Visualization:



Insights:

* Neurology department has the highest appointment fees (i.e., 1500).

Formula used: Department with Max Fee =

VAR MaxFee = MAXX(ALL(Sheet), Sheet[Appointment Fees])

RETURN

    CALCULATE(

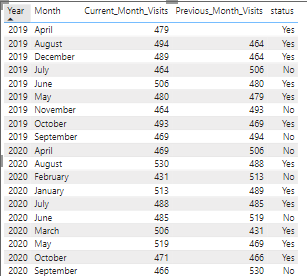
        FIRSTNONBLANK(Sheet[department\_referral], 1),

        Sheet[Appointment Fees] = MaxFee)

Question 10: Create a tabular visualization in the Report view which consists of Month-wise total visits in the hospital. Add a third column in the table that consists of the previous month’s total visits for each month’s row. Also, include a column that states whether the visits in a month are greater than that of the previous month's visits.

Answer:

Visualization:



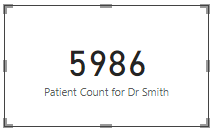
Insights:

* From the visual we can observe the visits of patients in the current month and also the number of visits in the previous month.
* On the basis of the visits, Status column indicates whether the number of visits has increased or not.

Question 11: Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.

Answer: Patient count for Dr Smith = 5986

Visualization:

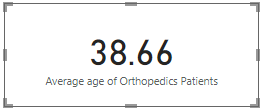


Formula used: Patient Count for Dr Smith = CALCULATE(COUNT(Sheet[patient\_id]),Sheet[Doctor Name]= "Dr. Smith

Question 12: Calculate the average age of the patients who visit the Orthopaedics department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?

Answer:

Visualization:



Formula used= Average age of Orthopedics Patients = CALCULATE(AVERAGE('Hospital ER'[patient\_age]),'Hospital ER'[department\_referral] = "Orthopedics")

Insights:

* Average age of Orthopedics patient = 38.66
* If the requirement was to find the average age of all departments, then there would be no need to use calculate function as the average age could be found directly using the average function as there were no conditions to be filtered and the formula would be something like:
* Average age of departments = AVERAGE ('Hospital ER'[patient\_age])

Question 13: Were there any data format issues in the data, and if there were/are how you handle them?

Answer:

* After analysing the dataset, no data format issues were found.
* The patient wait time was in number format but according to the schema numbers represented the wait time in minutes. So, no issues were found.
* Also new columns were created for the analysis and appropriate datatypes were provided to them.
* If there were any issues in the data format like the date and time columns were not having appropriate data types then I would have simply changed the data types to the correct format for ensuring the accurate analysis and avoid errors.

Question 14: When we add a column in Power Query what’s the code that comes in M language in the formula bar? What do you know about M-query?

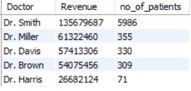
Answer:

* **M-query** is the language used in Power Query within Power BI for data transformation.
* It’s a functional language, meaning that it operates on functions and transformations to modify the data in queries.
* When you add a new column in **Power Query**, the code in **M Language** gets automatically written in the formula bar. This code is what Power Query uses to generate the transformation steps and it helps in monitoring the transformations.

Question 15: Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)

Answer:

Visualization:



SQL code:

select `Doctor Name` as Doctor, sum(`Total Bill`) as Revenue,

count(distinct patient\_id) as no\_of\_patients

from doctor

group by Doctor

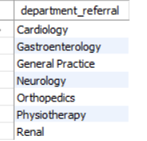
order by Revenue desc , no\_of\_patients asc

limit 5;

Question 16: Find the department where the average waiting time has decreased over three consecutive months. (SQL)

Answer:

Visualization:



SQL code:

with cte as

(select department\_referral, `date`,

round(avg(patient\_waittime), 2) as average\_wait\_time

from hospital

group by department\_referral, `date`),

cte2 as

(select department\_referral, `date`, average\_wait\_time,

lag(average\_wait\_time, 1) over (partition by department\_referral order by `date`) as previous\_month\_average,

lag(average\_wait\_time, 2) over (partition by department\_referral order by `date`) as previous\_2\_month\_average,

lag(average\_wait\_time, 3) over (partition by department\_referral order by `date`) as previous\_3\_month\_average

from cte)

select department\_referral

from cte2

where average\_wait\_time < previous\_month\_average

and previous\_month\_average < previous\_2\_month\_average

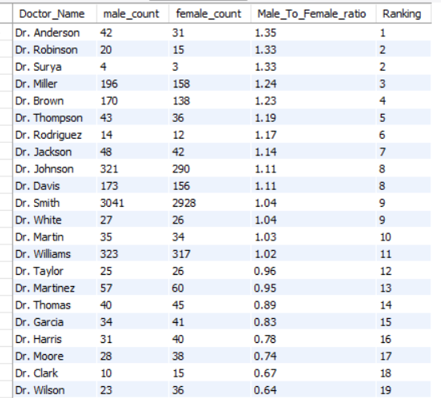
and previous\_2\_month\_average < previous\_3\_month\_average

group by department\_referral;

Question 17: Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)

Answer:

Visualization:



SQL code:

with cte as

(select d.`Doctor Name` as Doctor\_Name,

sum(case when h.patient\_gender = "M" then 1 else 0 end) as male\_count,

sum(case when h.patient\_gender = "F" then 1 else 0 end) as female\_count

from hospital h

inner join doctor as d on h.patient\_id = d.patient\_id

group by Doctor\_Name),

cte2 as

(select Doctor\_Name, male\_count, female\_count,

round((male\_count/female\_count),2) as Male\_To\_Female\_ratio

from cte)

select Doctor\_Name, male\_count, female\_count, Male\_To\_Female\_ratio,

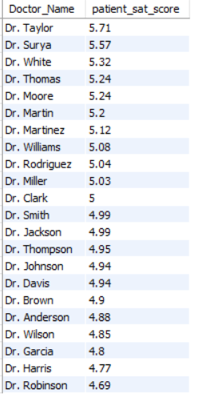
dense\_rank() over(order by Male\_To\_Female\_ratio desc) as Ranking

from cte2;

Question 18: Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)

Answer:

Visualization:



SQL code:

select d.`Doctor Name` as Doctor\_Name,

round(avg(case when h.patient\_sat\_score = "" then 5 else h.patient\_sat\_score end),2) as patient\_sat\_score

from hospital h

inner join doctor d

on h.patient\_id = d.patient\_id

group by Doctor\_Name

order by patient\_sat\_score desc;

Question 19: Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)

Answer:

Visualization:



SQL code:

select d.`Doctor Name` as Doctor\_Name,

count(distinct h.patient\_race) as differnet\_race\_count

from hospital h

inner join doctor d

on h.patient\_id = d.patient\_id

group by Doctor\_Name

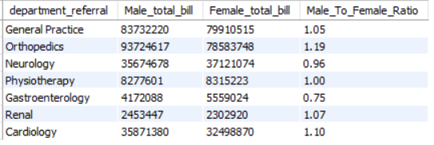
having count(distinct h.patient\_race) > 1

order by differnet\_race\_count desc;

Question 20: Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)

Answer:

Visualization:



SQL code:

select h.department\_referral,

sum(case when patient\_gender = "M" then d.`Total Bill` end) as Male\_total\_bill,

sum(case when patient\_gender = "F" then d.`Total Bill` end) as Female\_total\_bill,

round(sum(case when patient\_gender = "M" then d.`Total Bill` end) / sum(case when patient\_gender = "F" then d.`Total Bill` end),2) as Male\_To\_Female\_Ratio

from hospital h

inner join doctor d

on h.patient\_id = d.patient\_id

group by h.department\_referral;

Question 21: Update the patient satisfaction score for all patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. Increase their satisfaction score by 2 points, but ensure that the satisfaction score does not exceed 10. (SQL)

Answer:

SQL code:

update hospital

set patient\_sat\_score = least(patient\_sat\_score + 2,10)

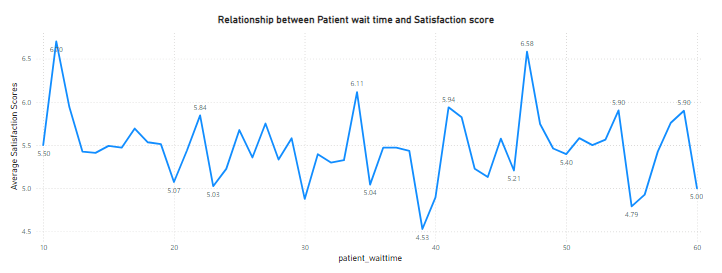
where lower(department\_referral) = "general practice" and patient\_waittime > 30;

**Subjective Questions**

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

Answer:

Visualization:



Insights:

* From the above visual we can understand the relationship between patient wait time and satisfaction score.
* We can observe that the satisfaction score is mostly uniform during the period of wait time but highest satisfactory score is obtained during the 10 mins wait time (6.70) and lowest around 40 mins (4.53).
* The ratings might not be accurate as 75% of the ratings were missing and analysis has been made on the basis of remaining 25%, so the hospital should also encourage patients to provide the ratings so that the accurate analysis can be made.

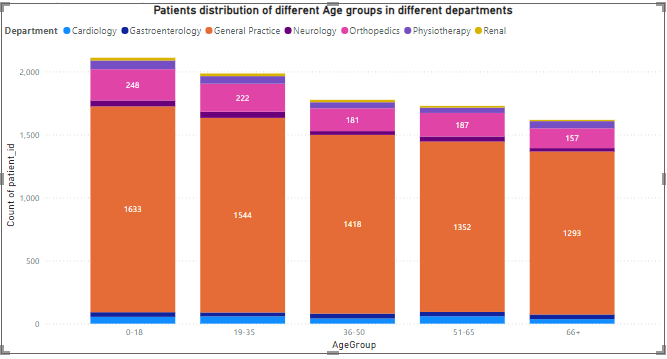
Approach used:

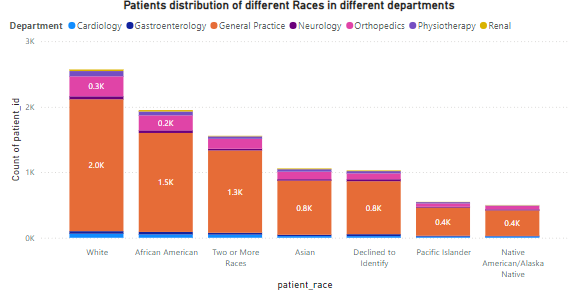
* Scatter chart was created using the wait time column and average satisfaction score measure.
* Average satisfaction score measure was used because the 75% of the ratings data is null so to make the analysis remaining 25% data was used to make somewhat accurate analysis.

Question 2: How do patient demographics affect the frequency of visits to different departments?

Answer:

Visualization:





Insights:

* From the visuals we can observe the distribution of patients in the departments on the basis of their age and race.
* This will help in understanding the demographics of the patients.
* This will also help in improving the facilities on the basis of the observations.

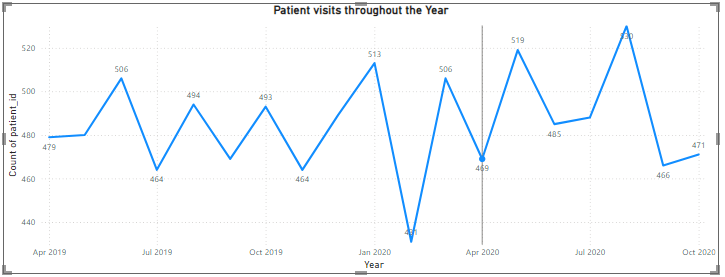
Approach used:

* Stacked column chart was created for observing the distribution of patients.
* Columns department referral, age group, race and count of patient id were used for the charts.

Question 3: Is there a noticeable trend in the volume of patient visits throughout the year?

Answer:

Visualization:



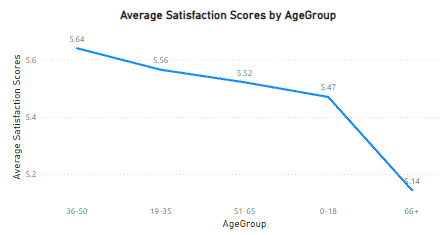
Insights:

* From the visual we can observe the visits of patients throughout the year.
* Patients visits peaked on August 2020 (530).
* Least number of patients visited on February 2020 (430).
* The trend cannot be analysed as the data is inconsistent which affects the analysis.

Question 4: Which age groups report the highest and lowest satisfaction scores?

Answer:

Visualization:



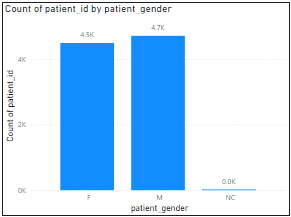
Insights:

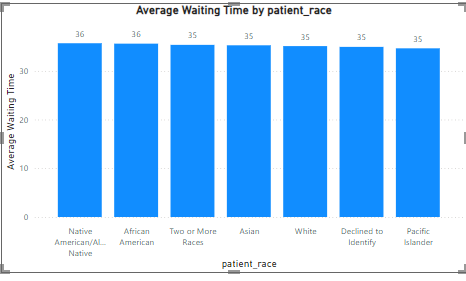
* From the visual we can observe the satisfaction score for different age groups.
* Line chart was created using the average satisfaction score measure and age group column.
* Age group 36-50 reports the highest satisfaction score (5.64).
* Age group 66+ reports the lowest ratings (5.14).

Question 5: Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?

Answer:

Visualization:





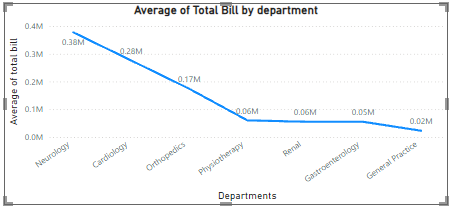
Insights:

* From the above visual of patient distribution patients, we can observe that the number of male and female patients does not have significant difference between them.
* From the visual average waiting time by patient race we can clearly observe that all the races have similar waiting time indicating that no race is being treated differently.
* This clearly indicates that the medical care is provided irrespective of the gender and race and there are no signs of racism and gender inequality in the hospital.

Question 6: The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?

Answer:

Visualization:



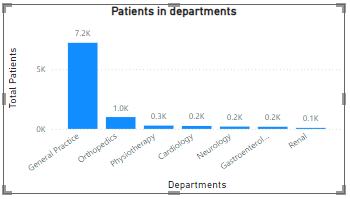
Insights:

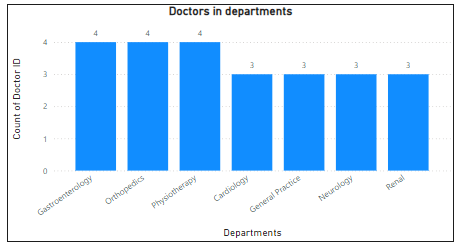
* From the visual we can observe the revenue generated from each department.
* Neurology (0.38M) and Cardiology (0.28M) are the highest revenue generating departments indicating higher number of patient visits.
* From the previous analysis we had observed that neurology has the highest appointment fee i.e., (1500)
* Considering all the factors, we can provide discounts to the patients from neurology and cardiology departments.

Question 7: The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.

Answer:

Visualization:





Insights:

* From the visual we can see the number of patients in different departments.
* General Practice (7.2K) has the highest count of patients followed by Orthopedics (1.0K).
* Renal (0.1K) department has the lowest number of patients.
* Considering all the factors new doctors should be hired in the departments with higher number of patients.
* So, doctors should be hired in General practice, Orthopedics and Physiotherapy department.
* General Practice only has 3 doctors tending around (7.2K) patients, so the doctors should be hired in this department.

Question 8: Is the hospital profitable? How will you determine the profitability?

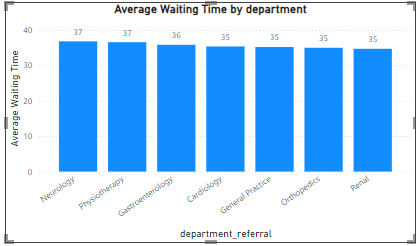
Answer:

* Appropriate data like doctor salary, maintenance costs and faculty salaries are not available to conduct the analysis.
* Only the revenue generated can be calculated but not the operational costs.
* So, the analysis cannot be conducted.

Question 9: Any Department for which the waiting time is oddly large?

Answer:

Visualization:



Insights:

* From the visual we can observe average wait time in different departments.
* Almost all the departments have similar wait time.
* So, no department has oddly large waiting time.

Question 10: Come up with strategies to provide discounts to the patients.

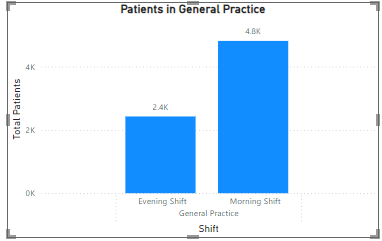
Answer:

* As previously mentioned in question 6, neurology and cardiology department are the highest revenue generating departments and they also have high appointment fees.
* So, they are appropriate departments to provide discounts.
* Medical campaigns can be started in which during certain days discounts can be provided which would also help in getting more patients.
* This would help in maximizing the profits and would be beneficial to hospitals and also people could access healthcare facilities in affordable prices.

Question 11: Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also, will this 2-shift policy be helpful for the hospital?

Answer:

Visualization:



Insights:

* From the visual we can observe the distribution of patients on the basis of timings that is morning or evening.
* Morning shifts (4.8K) have more patients than evening shifts (2.4K).
* From the previous analysis we had observed that general practices department only have 3 doctors which would be hectic for doctors as the workload would be much higher for 3 doctors and they could not work efficiently.
* So, new doctors should be hired and they should be split in different shifts so that the doctors can have enough rest and could work efficiently.
* 2-shift policy would definitely be helpful in the medical field a bit of negligence can cost the life of an individual so by introducing this policy doctors can have enough rest and it would help them in working efficiently.

Question 12: What do you understand by Power BI gateway? What are its use cases?

Answer:

* A **Power BI Gateway** is a bridge that facilitates secure data transfer between on-premises data sources and the Power BI service (cloud).
* It allows users to access on-premises data without moving it to the cloud.
* The gateway provides secure and seamless integration for data refreshes, ensuring that cloud-based Power BI reports and dashboards are up-to-date with the data residing on local servers.

Use cases:

* Scheduled Data Refresh
* Real-Time Data Access (Direct Query)
* Access to On-Premises Data
* Sharing and Collaboration with On-Premises Data

Question 13: How would you approach this problem, if the objective and subjective questions weren't given?

Answer:

If the guidelines were not provided following would be my approach:

* Firstly, I would have cleaned the data, i.e., I would have checked whether all the columns had appropriate data types.
* I would have addressed the missing values and would have used appropriate methods to deal with them and also would have created a separate date and time columns for efficiency.
* Then I would have created a separate table to store all the measures created for better efficiency.
* Then I would have created basic metrics, analysing them and then use them for further analysis.
* I would have created charts accordingly to the business requirements such as doctors in different departments, revenue generated by each department and many more.
* Then I would have compiled all the necessary visuals to create an informative report.

Question 14: Can you analyse and write the type of relationship between the doctor id and department, is it one-to-one?

Answer:

* No, doctor id and department does not have one to one relationship.
* Relationship between doctor id and department is one to many.
* Because multiple doctors may work in the "General Practice" department, but each doctor is typically assigned to just one department at a time.