Objective Questions

Q1. In analyzing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.

Ans: **Data Cleaning Steps in Power BI**

1. **Loaded the Data**
   - Imported two datasets **Doctors_Patient_Data** and **Hospitals Data** into Power BI Desktop.
2. **Opened Power Query Editor**
   - Clicked on **Transform Data** to open the Power Query Editor for cleaning and transforming the data.
3. **Checked for Missing Values**
   - Carefully inspected all columns for missing or null values.
   - Found missing values in the **pat_score** column.
4. **Handled Missing Data**
   - Replaced **null values** in the **pat_score** column with **0** to avoid skewness and maintain consistency.
5. **Verified Data Formatting**
   - Ensured that **names and text fields** were properly formatted and consistent (e.g., capitalization, no extra spaces).
   - Checked that **date and time columns** were in the correct format.
6. **Validated Data Consistency**
   - Reviewed data types (number, text, date) to make sure they matched the type of information in each column.
7. **Applied Changes**
   - Clicked **Close & Apply** to save all cleaning steps and load the cleaned data back into Power BI for further analysis.

Q2. Assess the Average Waiting Time: Analyse the patient wait times to identify the average duration a patient spends before receiving care

Ans : **Approach**

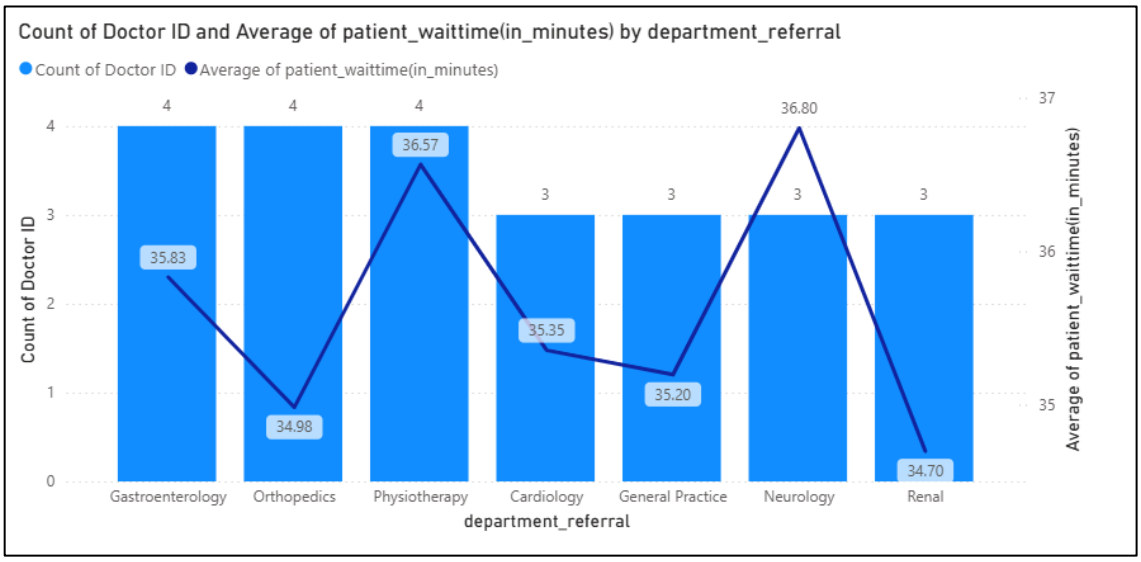
1. Calculated the **average patient waiting time** using a DAX measure:

```
1  average_wait_time = AVERAGE('Hospital ER (1)'[patient_waittime(in_minutes)])
```

2. Created a **clustered column and line chart** showing:

- **Count of Doctor ID** (by department) as columns.
- **Average patient wait time (in minutes)** as a line.

This helped compare how the number of doctors in each department relates to patient wait times.



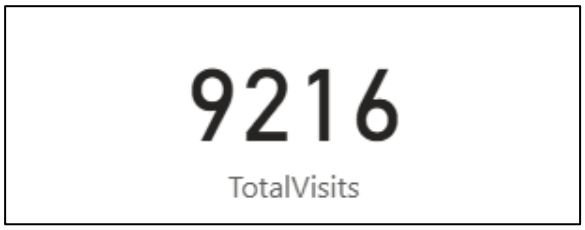
Insights:

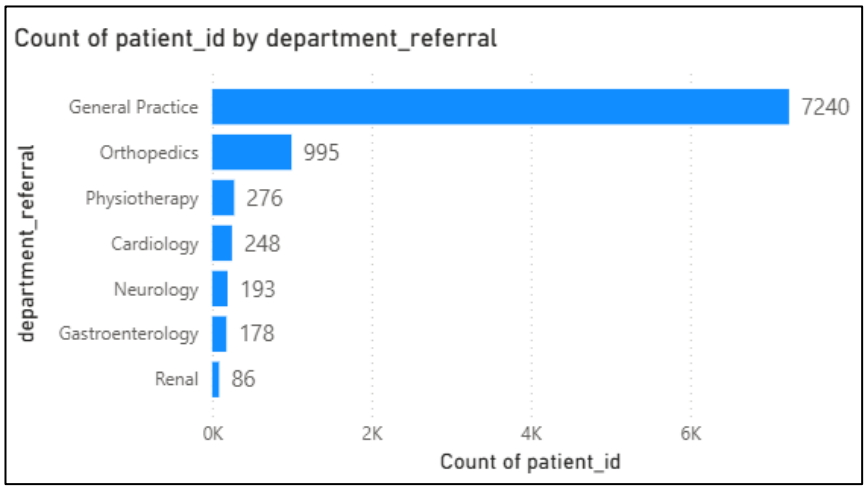
1. The **average waiting time** across departments is around **35–37 minutes**.
2. **Physiotherapy (36.57 min)** and **Neurology (36.80 min)** have slightly higher wait times.
3. Even with fewer doctors, some departments maintain similar wait times indicating **no strong correlation** between the number of doctors and waiting time.
4. The hospital may consider **reviewing appointment management** or **hiring more doctors** in busy departments to reduce waiting time further.

Q3. 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: Approach

1. Created a **DAX measure** to calculate the **total number of patient visits** across all departments.
2.
```
TotalVisits = COUNTROWS('Hospital ER (1)')
```
3. Built a **bar chart** to visualize the **count of patient visits by department referral**.
4. Added a **card visual** to display the **overall total visits (9,216)** for better context.





Insights

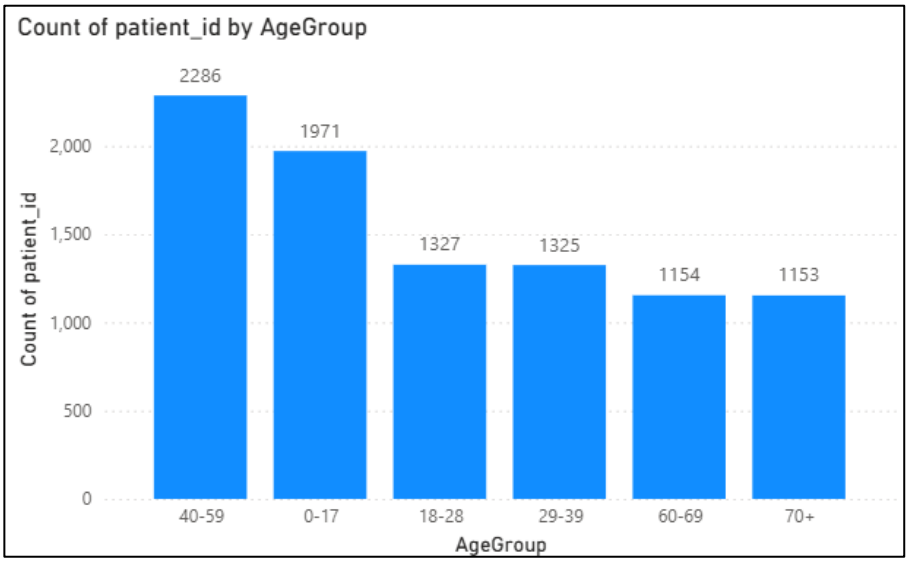
- The hospital recorded a total of **9,216 visits** across all departments.
- **General Practice** has the **highest number of visits (7,240)**, indicating it handles the majority of patient referrals.
- **Orthopedics** follows with **995 visits**, while other departments like **Physiotherapy, Cardiology, and Neurology** have much fewer.
- This shows a **high dependency on General Practice**, suggesting an **uneven distribution of patients** across departments.

Q4. Patient Visits by Age Group: Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

Ans: Approach

1. Created a **calculated column** named **AgeGroup** using a **SWITCH statement** to categorize patients into age ranges:
    - 0–17, 18–28, 29–39, 40–59, 60–69, and 70+.
2. Used a **column chart** to visualize the **count of patient visits** by each **AgeGroup**.
    - **X-axis:** AgeGroup
    - **Y-axis:** Count of Patient_ID

```
AgeGroup = SWITCH(TRUE,'Hospital ER (1)'[patient_age] < 18,"0-17",
                       'Hospital ER (1)'[patient_age] < 29,"18-28",
                       'Hospital ER (1)'[patient_age] < 40,"29-39",
                       'Hospital ER (1)'[patient_age] < 60, "40-59",
                       'Hospital ER (1)'[patient_age] < 70, "60-69",
                       'Hospital ER (1)'[patient_age] >=70, "70+",)
```



Insights:

- The **40–59 age group** recorded the **highest number of visits (2,286)**, indicating that middle-aged patients use healthcare services the most.
- The **0–17 age group** ranks second with **1,971 visits**, showing high hospital usage among children and teenagers.
- The **18–39** age groups show **moderate visit counts**, while **60+** groups have **slightly lower visits**.
- Overall, **middle-aged and young patients** are the key demographic segments driving hospital visits.

Q5. 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 Findings:

- Upon analyzing the dataset, **most columns did not contain null values**.
- However, one column **patient_sat (Customer Satisfaction)** had a **high percentage of missing values (72%)**.
- This column is important because it reflects **patients feedback** and helps identify areas for **service improvement**.

Approach Taken

- Since the **patient_sat** column had a **large number of null values**, using statistical imputation (like mean or median) could **mislead insights** and **skew the data**.
- Therefore, I chose a **simple and non-biased approach**: Replaced all **null values with 0**, to keep the dataset consistent and maintain its original structure.

General Handling Strategy for Null Values

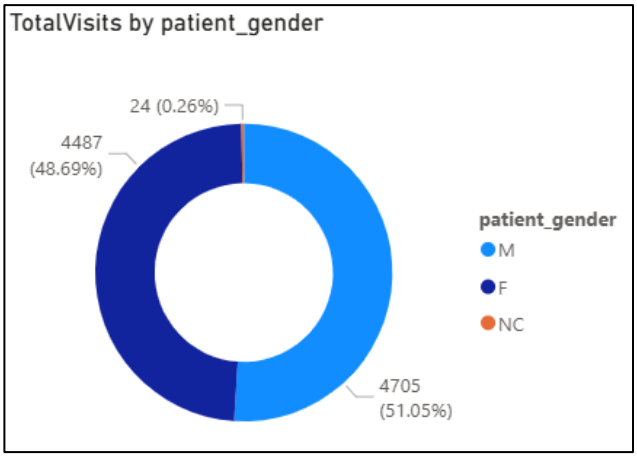
If other columns had missing data, I would:

1. **Assess the importance** of the column in the analysis.
2. Based on its relevance, apply suitable techniques such as:
    - **Replacing with 0** or **"Not Available"** for less critical columns.
    - **Using statistical imputation** (mean, median, or mode) for numeric fields when appropriate.
3. Ensure the chosen method **does not distort the dataset** or **bias the results**.

Q6. Is there any relation between the number of visits and the Gender of the patients?

Ans: **Approach**

1. Created a **Doughnut Chart** to visualize the **total number of visits by patient gender**.
2. Used **Patient_Gender** as the legend and **Count of Patient_ID (Total Visits)** as the value.
3. This visualization helps understand which gender group contributes more to hospital visits, even if it doesn't show a direct correlation.



Insights

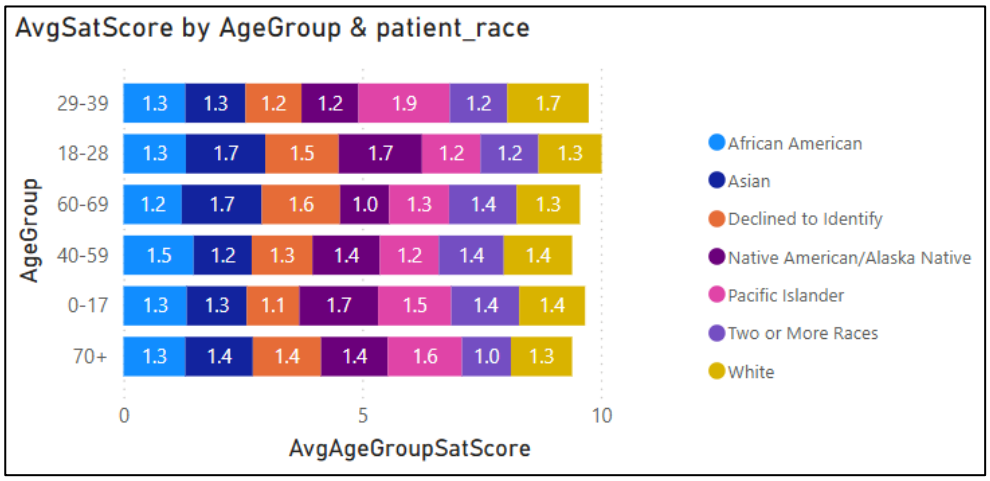
- **Male patients** account for the **highest number of visits (4,705 – 51.05%)**.
- **Female patients** closely follow with **4,487 visits (48.69%)**, showing a nearly equal distribution.
- A very small portion (**24 visits – 0.26%**) belongs to patients with **gender not classified (NC)**.
- Overall, **both genders visit the hospital in almost equal proportions**, indicating **balanced healthcare utilization** between male and female patients.

Q7. 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 **Approach:**

1. Created a **bar chart** to analyze how **patient satisfaction scores** vary across **age groups** and **racial backgrounds**.
2. Plotted:
    - **X-axis:** Average Patient Satisfaction Score (PatientSatScore)
    - **Y-axis:** AgeGroup
    - **Legend:** Patient_Race
3. This visualization helps understand how satisfaction differs among demographic segments.

Note: that the **PatientSatScore column contains around 72% null values**, which can limit the reliability of insights.



Insights

- Overall **average satisfaction scores** are **close across all age groups and races**, showing minimal variation.
- Slightly higher satisfaction is observed among some groups in the **18–28** and **40–59** age ranges.
- Since the **PatientSatScore data is mostly incomplete (72% null values)**, these findings may not fully represent actual patient experience.
- To gain meaningful insights, the hospital should **improve data collection for satisfaction feedback** and **encourage more patients to provide ratings**.

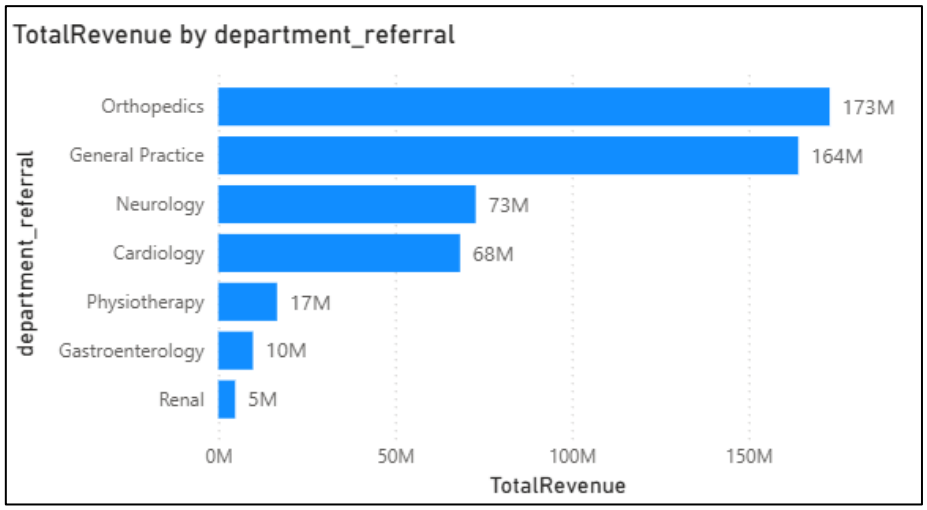
Q8. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.

Ans Approach

1. Created a **DAX measure** to calculate the total revenue generated by each department:

```
TotalRevenue = SUM(Sheet1[Total Bill])
```

2. Used a **bar chart** to visualize revenue distribution across departments.
   - **X-axis:** TotalRevenue
   - **Y-axis:** Department_Referral
3. This helped identify which departments contribute most to the hospital's overall revenue.



Insights

- **Orthopedics** department generated the **highest revenue (₹173M)**, followed closely by **General Practice (₹164M)**.
- **Neurology (₹73M)** and **Cardiology (₹68M)** contribute moderately to the total revenue.
- **Physiotherapy, Gastroenterology, and Renal** departments contribute **comparatively lower revenue**.
- The results show that **Orthopedics and General Practice are the hospital's top-performing departments**, indicating high patient demand and service utilization.

Q9. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.

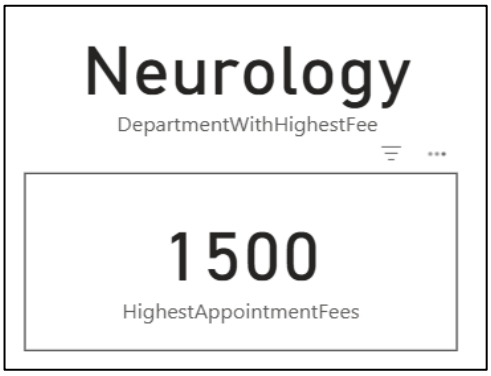
Ans: Approach:
First, I created a measure to calculate the highest appointment fee using an aggregation DAX function:

```
HighestAppointmentFees = MAX(Sheet1[Appointment Fees])
```

Next, to identify which department this highest fee belongs to, I created another DAX measure:

```
DepartmentWithHighestFee =
VAR MaxFee = [HighestAppointmentFees]
RETURN
CALCULATE(
    SELECTEDVALUE(Sheet1[Department_Referral]),
    Sheet1[Appointment Fees] = MaxFee
)
```

This combination helps in determining both the maximum fee and the corresponding department name.



Insights:

- The analysis reveals that the **Neurology department** charges the **highest appointment fee**, which is **$1,500**.
- This could indicate that the department deals with more specialized or complex medical consultations, justifying higher charges compared to other departments.

Q10. 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 Approach: To analyze month-over-month trends in patient visits, I have created a following DAX measures:

1. **TotalVisits** : calculates the total number of patient visits each month:

```
TotalVisits = COUNTROWS('Hospital ER (1)')
```

2. **LastMonthVisits**: retrieves the total visits of the **previous month**, functioning similar to the **LAG()** concept:

```
LastMonthVisits =
VAR CurrentMonth =
    MAX ('Hospital ER (1)'[MonthNumber] )
RETURN
CALCULATE (
    [TotalVisits],
    FILTER (
        ALL ( 'Hospital ER (1)' ),
        'Hospital ER (1)'[MonthNumber] = CurrentMonth - 1
    )
)
```

3. **IncreaseComparedToPrevious** : compares the current month's visits with the previous month and flags if visits have increased:

```
IncreaseComparedToPrevious = IF([TotalVisits] > [LastMonthVisits],"Yes","No")
```

Then I created a **tabular visualization** showing:

- **Month**
- **TotalVisits**
- **LastMonthVisits**
- **IncreaseComparedToPrevious**

Month	TotalVisits	LastMonthVisits	IncreaseComparedToPrevious
January	513		Yes
February	431	513	No
March	506	431	Yes
April	948	506	Yes
May	999	948	Yes
June	991	999	No
July	952	991	No
August	1024	952	Yes
September	935	1024	No
October	964	935	Yes
November	464	964	No
December	489	464	Yes
Total	**9216**	**464**	**Yes**

Insights:

- The hospital recorded **fluctuating visit patterns** throughout the year.
- **January → February** saw a slight **drop**, indicating possibly seasonal or post-holiday effects.
- A **steady rise** was observed from **March to May**, with the highest month being **May (999 visits)** showing strong patient inflow during mid-year.
- After May, a **small decline** appeared in **June and July**, before rebounding in **August**, which again shows **increased visits (1,024)** possibly due to seasonal ailments or specific health campaigns.
- The **lowest patient turnout** occurred in **November (464)**, suggesting reduced hospital activity towards the year-end.
- Overall, the pattern alternates between increases and drops, suggesting a **cyclic or seasonal visit behavior** that could help in **resource and staffing optimization**.

Q11. Using 'Calculate' and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.

Ans Approach:

1. To determine how many patients visited **Dr. Smith**, a **CALCULATE** function was used along with a row-level filter.
2. The following DAX measure was created:

```
DrSmithsTotalVisitPatients = CALCULATE(COUNTA(Sheet1[patient_id]),
                                       Sheet1[Doctor Name] = "Dr. Smith")
```

3. The **CALCULATE** function changes the filter context to include only those rows where the doctor's name is *"Dr. Smith"*, and the **COUNTA** function counts all unique patient visit entries for that doctor.
4. Finally, the measure was displayed using a **card visualization** to show the total number of visits made to Dr. Smith.

Insights:

- **Dr. Smith** recorded a total of **5,986 patient visits**, indicating a **high patient engagement** and trust level compared to other doctors.
- This may reflect Dr. Smith's **specialization**, **experience**, or **popularity** within the hospital.
- The insight can help hospital management in **workload distribution** or **resource planning**, ensuring that doctors with higher visit counts are provided additional support or scheduling adjustments.

Q12. 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?

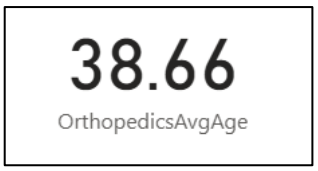
Ans: Approach:

1. To calculate the **average age of patients in the Orthopedics department**, the following DAX measure was used:

```
OrthopedicsAvgAge = CALCULATE(AVERAGE('Hospital ER (1)'[patient_age]),'Hospital ER (1)'
[department_referral] = "Orthopedics")
```

2. Yes The approach will be different if we want to calculate the metric if we required all department here we use **ALL()** function which removes all the filters from the table and caculates the overall departments age
3. To calculate the **average age across all departments**, the measure was slightly modified to ignore any department filters:

```
avgAge = CALCULATE(AVERAGE('Hospital ER (1)'[patient_age]),ALL('Hospital ER (1)'))
```



Insights:

- The **average patient age in the Orthopedics department** is **38.66 years**, while the **overall hospital average** is **39.86 years**.
- Orthopedic patients are slightly **younger** than the overall patient population, suggesting that **middle-aged adults** are more prone to bone, joint, or injury-related issues.
- This insight helps management in **resource planning** and **specialized care programs** for this specific age group.

Q13 Were there any data format issues in the data, and if there were/are how you handle them?

Ans: **Approach**

If there were any **data format issues** in the dataset such as incorrect **date/time formats**, **currency mismatches**, or **inconsistent text capitalization** I applied **data transformation techniques** to ensure consistency and accuracy before performing any analysis.

For example:

- If the **date** column had inconsistent formats, I converted them to a **standard Date/Time data type** using the **Power Query "Change Type"** option or DAX conversion functions like DATEVALUE() or FORMAT().
- If there were **currency or numeric format** issues, I standardized them by changing the **data type to Decimal/Whole Number** and aligning the format settings.
- For **text-related issues**, I used **text functions** like UPPER(), LOWER(), or PROPER() to make the data **consistent and structured**.
- If null or blank values appeared due to format mismatches, I replaced or filtered them using **Power Query "Replace Values"** or IF(ISBLANK(), ...) logic in DAX.

This process helped ensure that the dataset was **clean, uniform, and ready for reliable analysis** without any data type conflicts or inconsistencies.

Q14. 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 we add a new column **in Power Query, the** M language code **that appears in the formula bar generally starts with** Table.AddColumn**.**
For example:

= Table.AddColumn(#"Previous Step", "New Column Name", each [ExistingColumn] * 2)

This line means a **new column** is being added to the table based on a logic or expression provided after each.

M language, also known as **Power Query Formula Language**, is the **scripting language behind Power Query**.
It is used to perform **data transformation, cleaning, and preparation** before loading the data into Power BI.
M is a **case-sensitive, functional language** designed to work with data tables — it executes step by step (each step feeding the next).

Some key points about **M-query**:

- It is used in the **Power Query Editor**.
- Each transformation in Power Query creates a new **step** written in M.
- You can view or edit these steps in the **Advanced Editor**.
- Common functions include Table.AddColumn, Table.TransformColumnTypes, Table.RenameColumns, and Table.SelectRows.

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

Ans: Approach: To perform this analysis, I followed these steps:

1. **Joined the tables:** Combined the doctor_patients_data and hospital_er tables using an **INNER JOIN** on the common key patient_id to bring doctor, patient, and billing details together.
2. **Aggregated the data** : Applied the **GROUP BY** clause on doctor_name to calculate:
    - **Total Revenue:** using SUM(total_bill)
    - **Patient Count:** using COUNT(DISTINCT patient_id)
3. **Ordered the results**: Used the **ORDER BY** clause to sort doctors by:
    - **Highest total revenue (DESC)**, and
    - **Fewest patients (ASC)** to meet the question's condition.
4. **Limited the output**: Used the **LIMIT 5** command to extract only the **top 5 doctors** who fit the criteria.

SQL QUERY:

```
SELECT dp.doctor_name,SUM(total_bill)/1000000 as total_revenue_in_millions,
COUNT(distinct he.patient_id) as total_patients
FROM doctor_patients_data dp
INNER JOIN hospital_er he
ON dp.patient_id = he.patient_id
GROUP BY dp.doctor_name
ORDER BY SUM(total_bill) desc,COUNT(distinct he.patient_id)
limit 5;
```

OUTPUT:

doctor_name	total_revenue_in_millions	total_patients
Dr. Smith	135.6797	5986
Dr. Miller	61.3225	355
Dr. Davis	57.4133	330
Dr. Brown	54.0755	309
Dr. Harris	26.6821	71

Insights:

- **Dr. Smith** generated the **highest total revenue (₹135.67M)** with **5,986 patients**, indicating strong patient inflow and consistent revenue generation.
- **Dr. Miller** and **Dr. Davis** followed with **₹61.32M** and **₹57.41M** revenue respectively, but with significantly **fewer patients (355 and 330)** suggesting **higher consultation or treatment charges**.
- **Dr. Brown** also showed similar trends, maintaining **₹54.07M** revenue with **309 patients**.
- **Dr. Harris**, though having the **fewest patients (71)**, still contributed **₹26.68M**, showing **high revenue per patient** possibly due to **specialized or high-cost treatments**.

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

Ans: **Approach**

1. Calculated the **average waiting time** for each department by month using the AVG() and GROUP BY functions.
2. Used **window function LAG()** to fetch the waiting time of the **previous two months** for every department.
3. Applied a **filter condition** where the current month's average waiting time was **less than both previous months**, indicating a consistent decrease.
4. Displayed only the **distinct department names** that met this trend of decreasing wait times.

SQL QUERY:

```
With mycte as (
SELECT department_referral,EXTRACT(MONTH FROM date) as month_num,AVG(patient_waittime) as
current_month_avg_time
FROM hospital_er
GROUP BY department_referral,EXTRACT(MONTH FROM date)
ORDER BY department_referral,EXTRACT(MONTH FROM date)
),
mycte2 as (
SELECT department_referral,month_num,current_month_avg_time,LAG(current_month_avg_time,1) OVER(PARTITION BY
Department_referral ORDER BY month_num) as prev_month_avg_time,
LAG(current_month_avg_time,2) OVER(PARTITION BY Department_referral ORDER BY month_num)
as prev_2month_avg_time
FROM mycte
)

SELECT distinct department_referral
FROM mycte2
WHERE current_month_avg_time > prev_month_avg_time AND prev_month_avg_time > prev_2month_avg_time
```

OUTPUT:

department_referral
Cardiology
Gastroenterology
General Practice
Neurology
Orthopedics
Physiotherapy
Renal

Insights:

- Departments such as **Cardiology**, **Orthopedics**, and **Renal** consistently reduced their **average patient waiting time** over three consecutive months.
- This indicates **improved operational efficiency**, better **patient flow management**, and possibly **optimized scheduling or staffing** strategies during those months.

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

Ans Approach:

1. **Segregate male and female counts:**
    - Two CTEs (mycte and mycte2) were created to calculate the total number of ***male*** and ***female*** patients per doctor using **COUNT()** grouped by doctor_name.
    - This separation ensures gender-based aggregation before combining results.
2. **Join both CTEs:**
    - Joined the two CTEs on doctor_name to bring male and female counts together for comparison.
3. **Compute ratio and rank:**
    - Calculated the **male-to-female ratio** as (male_patient_count / female_patient_count) and rounded to 2 decimals.
    - Used the RANK() window function to rank doctors in descending order based on this ratio (higher male proportion = higher rank).

SQL QUERY:

```
With mycte as (
SELECT dp.doctor_name,COUNT(hr.patient_gender) as male_patient_count
FROM hospital_er hr
INNER JOIN doctor_patients_data dp
ON hr.patient_id = dp.patient_id
WHERE patient_gender = 'M'
GROUP BY dp.doctor_name
),
mycte2 as (
SELECT dp.doctor_name,COUNT(hr.patient_gender) as female_patient_count
FROM hospital_er hr
INNER JOIN doctor_patients_data dp
ON hr.patient_id = dp.patient_id
WHERE patient_gender = 'F'
GROUP BY dp.doctor_name
)

SELECT ct.doctor_name,ct.male_patient_count,ct2.female_patient_count,
ROUND(ct.male_patient_count*1.0 / ct2.female_patient_count,2) as male_female_ratio,
RANK() OVER (ORDER BY (ct.male_patient_count * 1.0 / ct2.female_patient_count) desc) as ratio_rank
FROM mycte ct
INNER JOIN mycte2 ct2
ON ct.doctor_name = ct2.doctor_name
```

OUTPUT:

doctor_name	male_patient_count	female_patient_count	male_female_ratio	ratio_rank
Dr. Anderson	42	31	1.35	1
Dr. Robinson	20	15	1.33	2
Dr. Surya	4	3	1.33	2
Dr. Miller	196	158	1.24	4
Dr. Brown	170	138	1.23	5
Dr. Thompson	43	36	1.19	6
Dr. Rodriguez	14	12	1.17	7
Dr. Jackson	48	42	1.14	8
Dr. Davis	173	156	1.11	9
Dr. Johnson	321	290	1.11	10
Dr. Smith	3041	2928	1.04	11
Dr. White	27	26	1.04	12
Dr. Martin	35	34	1.03	13
Dr. Williams	323	317	1.02	14
Dr. Taylor	25	26	0.96	15
Dr. Martinez	57	60	0.95	16

Insights:

- **Dr. Anderson** has the highest male-to-female ratio (**1.35**), indicating he sees significantly more male patients.
- **Dr. Robinson** and **Dr. Surya** follow closely with ratios of **1.33**, suggesting similar trends.
- Doctors like **Dr. Taylor (0.96)** and **Dr. Martinez (0.95)** show near-equal gender representation.
- At the lower end, **Dr. Clark (0.67)** and **Dr. Wilson (0.64)** treat more **female** patients than male.
- Overall, most doctors maintain a fairly balanced patient ratio, with only a few showing noticeable gender skew.

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

Ans Approach:

1. **Joined the tables:**
    - Combined hospital_er and doctor_patients_data using patient_id to align each patient's satisfaction score with their respective doctor.
2. **Count and aggregate:**
    - Used COUNT(DISTINCT hr.patient_id) to get **unique patient visits** per doctor.
    - Calculated the **average satisfaction score** with **AVG(COALESCE(hr.patient_sat_score, 0))** ensuring null scores are treated as zero to prevent skewed averages.
3. **Group by doctor:**
    - Applied GROUP BY dp.doctor_name to compute these metrics at the doctor level.

SQL QUERY:

```
SELECT dp.doctor_name,COUNT(distinct hr.patient_id) as patient_visits,
AVG(coalesce(hr.patient_sat_score,0)) as avg_sat_score
FROM hospital_er hr
JOIN doctor_patients_data dp
ON hr.patient_id = dp.patient_id
GROUP BY dp.doctor_name
```

OUTPUT:

doctor_name	patient_visits	avg_sat_score
Dr. Anderson	73	1.17808
Dr. Brown	309	1.29126
Dr. Clark	25	1.60000
Dr. Davis	330	1.48788
Dr. Garcia	75	1.06667
Dr. Harris	71	1.18310
Dr. Jackson	90	1.21111
Dr. Johnson	613	1.31811
Dr. Martin	69	1.72464
Dr. Martinez	117	1.70085
Dr. Miller	355	1.46197
Dr. Moore	66	1.30303
Dr. Robinson	35	0.97143
Dr. Rodriguez	26	1.19231

Insights:

- **Dr. Surya (3.43)** and **Dr. Taylor (2.37)** achieved the **highest satisfaction averages**, indicating excellent patient experience.
- **Dr. Martin (1.72)**, **Dr. Martinez (1.70)**, and **Dr. Wilson (1.77)** also show strong satisfaction levels.
- Doctors with **high visit volumes** like **Dr. Smith (5986 visits)** and **Dr. Johnson (613)** maintain **consistent satisfaction (~1.3)**, reflecting stable performance across a large patient base.
- **Dr. Robinson (0.97)** and **Dr. Garcia (1.07)** have relatively **lower satisfaction**, which could point to service quality issues or patient expectation gaps.
- Overall, satisfaction varies widely suggesting opportunities to identify **best practices from top-rated doctors** and **improve consistency** across departments.

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

Ans: **Approach:**

1. **Join the datasets:**
    - Used an INNER JOIN between doctor_patients_data and hospital_er on patient_id to connect each doctor with their patients' demographic details.
2. **Calculate unique races per doctor:**
    - COUNT(DISTINCT hr.patient_race) finds how many **different races** each doctor has treated.
3. **Compare against total diversity:**
    - Divided by the total number of distinct races in the hospital (from a subquery: SELECT COUNT(DISTINCT patient_race) FROM hospital_er).
    - Multiplied by 100 to express it as a **percentage diversity score**.
4. **Group by doctor:**
    - Aggregated results at the doctor level using GROUP BY dp.doctor_name.

SQL QUERY:

```
SELECT dp.doctor_name,
ROUND(COUNT(distinct hr.patient_race) * 100.0 /  (SELECT COUNT(distinct patient_race) FROM hospital_er)) as
diversity_of_doctors_patient_race
FROM doctor_patients_data dp
INNER JOIN hospital_er hr
ON dp.patient_id = hr.patient_id
GROUP BY dp.doctor_name;
```

OUTPUT:

doctor_name	diversity_of_doctors_patient_race
Dr. Anderson	100
Dr. Brown	100
Dr. Clark	86
Dr. Davis	100
Dr. Garcia	100
Dr. Harris	100
Dr. Jackson	100
Dr. Johnson	100
Dr. Martin	100
Dr. Martinez	100
Dr. Miller	100
Dr. Moore	100
Dr. Robinson	86
Dr. Rodriguez	71

Insights:

1. Most doctors (like **Dr. Smith, Dr. Johnson, Dr. Miller, Dr. Brown**) achieved a **100% diversity score**, meaning they have treated patients from **all available racial groups** showing broad and inclusive patient engagement.
2. A few doctors such as **Dr. Robinson (86%)**, **Dr. Rodriguez (71%)**, and **Dr. Surya (57%)** show **lower diversity**, possibly due to **department specialization, patient location bias, or demographic concentration**.
3. The results highlight that **most doctors handle diverse populations**, which is a positive indicator for inclusivity and equal access to medical care.
4. **Further analysis** could explore whether higher diversity correlates with **higher satisfaction or revenue metrics** for deeper insights.

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

Ans Approach:

1. **Created two CTEs for gender-based aggregation:**
    - **mycte** → Calculates the **total bill amount from male patients**, grouped by department.
    - **mycte2** → Calculates the **total bill amount from female patients**, grouped by department.
2. **Used INNER JOIN** to combine the two datasets on department_referral, allowing direct comparison of male and female billing totals.
3. **Calculated the ratio:**
    - Used: **ROUND(ct.total_bill_by_male_patient_in_million * 1.0 / ct2.total_bill_by_female_patient_in_million, 2)**

      This gives a **clean decimal ratio** of male-to-female billing, rounded to two decimal places.

4. **Normalized values to millions** (/1000000) to keep results easier to interpret.

SQL QUERY:

```
With mycte as (
SELECT hr.department_referral,SUM(dp.total_bill)/1000000 as total_bill_by_male_patient_in_million
FROM hospital_er hr
INNER JOIN doctor_patients_data dp
ON hr.patient_id = dp.patient_id
WHERE hr.patient_gender = 'M'
GROUP BY hr.department_referral
),
mycte2 as (
SELECT hr.department_referral,SUM(dp.total_bill)/1000000 as total_bill_by_female_patient_in_million
FROM hospital_er hr
INNER JOIN doctor_patients_data dp
ON hr.patient_id = dp.patient_id
WHERE hr.patient_gender = 'F'
GROUP BY hr.department_referral
)

SELECT ct.department_referral,ROUND(ct.total_bill_by_male_patient_in_million*
1.0/ct2.total_bill_by_female_patient_in_million,2) as ratio_total_bill_male_to_gemale
FROM mycte ct
INNER JOIN mycte2 ct2
ON ct.department_referral = ct2.department_referral
```

OUTPUT:

department_referral	ratio_total_bill_male_to_gemale
Cardiology	1.10
Gastroenterology	0.75
General Practice	1.05
Neurology	0.96
Orthopedics	1.19
Physiotherapy	1.00
Renal	1.07

Insights:

1. **Orthopedics (1.19)** and **Cardiology (1.10)** show **higher billing ratios for male patients**, suggesting that **male patients might be generating more revenue** in these departments possibly due to more frequent or costlier treatments.
2. **General Practice (1.05)** and **Renal (1.07)** maintain near-equal spending patterns between genders.
3. **Gastroenterology (0.75)** indicates **higher billing from female patients**, meaning women could be more likely to seek or require care in that department.
4. **Physiotherapy (1.00)** shows a **balanced gender ratio**, suggesting equitable patient spending patterns.
5. Overall, the ratios highlight how **gender distribution impacts departmental revenue**, useful for **targeted health outreach and service planning**.

Q21. Update the patient satisfaction score for all patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. Increase their satisfaction score by 2 points, but ensure that the satisfaction score does not exceed 10. (SQL)

Ans: **Approach**

- **Created a new column** instead of overwriting the original data for maintaining data integrity and easy comparison.
- Then applied a **conditional update** that targets only patients under *General Practice* with **waiting time > 30 mins**, ensuring focused modification.
- The use of **COALESCE** handled null satisfaction scores gracefully, while **LEAST** prevented scores from exceeding 10.

SQL QUERY:

```
ALTER TABLE hospital_er
ADD COLUMN updated_sat_score DECIMAL(5,2);

UPDATE hospital_er
SET updated_sat_score =
    CASE
        WHEN LOWER(department_referral) = 'general practice' AND patient_waittime > 30
            THEN LEAST(COALESCE(patient_sat_score, 0) + 2, 10)
        ELSE COALESCE(patient_sat_score, 0)
    END;

SELECT department_referral,patient_waittime,coalesce(patient_sat_score,0) as old_sat_score,updated_sat_score
FROM hospital_er
WHERE department_referral = "General Practice";
```

OUTPUT:

department_referral	patient_waittime	old_sat_score	updated_sat_score
General Practice	39	10.0	10.00
General Practice	27	0.0	0.00
General Practice	55	9.0	10.00
General Practice	31	8.0	10.00
General Practice	59	0.0	2.00
General Practice	43	0.0	2.00
General Practice	23	0.0	0.00
General Practice	42	1.0	3.00
General Practice	51	0.0	2.00
General Practice	53	0.0	2.00
General Practice	45	0.0	2.00
General Practice	49	0.0	2.00
General Practice	57	0.0	2.00
General Practice	35	0.0	2.00
General Practice	55	0.0	2.00
General Practice	40	2.0	4.00
General Practice	25	0.0	0.00

Subjective Questions

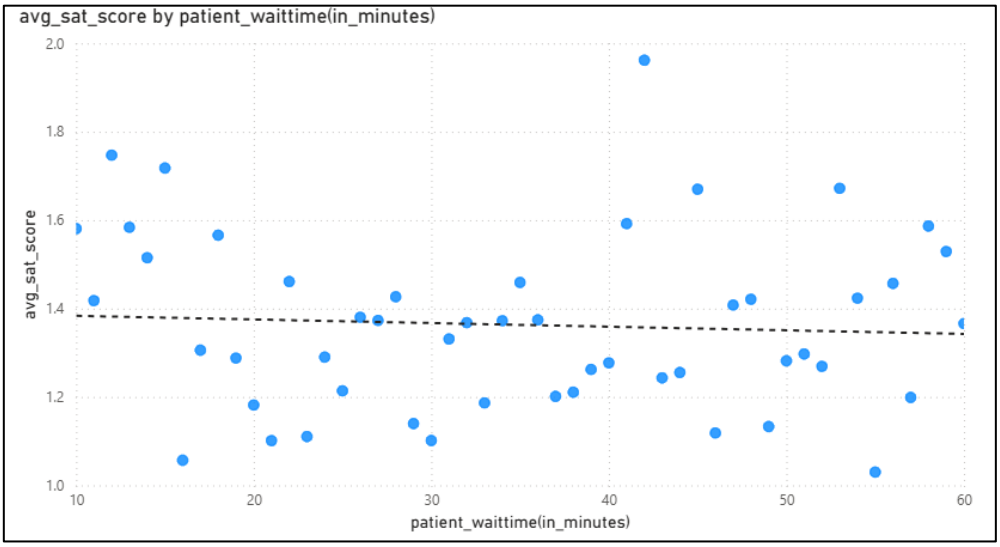
Q1. What is the relation between patient wait time and satisfaction scores?

Ans Approach:

1. Calculated a **measure** for average satisfaction score:

```
avg_sat_score = AVERAGE('Hospital ER (1)'[patient_sat_score])
```

2. Used a **Scatter Plot** to study the relationship between **patient wait time** (X-axis, set to *Don't Summarize*) and **average satisfaction score** (Y-axis, measure).
3. Added a **trendline** in the Analytics pane to identify correlation patterns.
4. **Note** that around **72% of satisfaction score values were null**, which were replaced with **0** before visualization this data treatment may have influenced the trend slightly.



Insights

1. The **trendline shows a mild negative correlation**, suggesting that **higher wait times lead to lower satisfaction**.
2. A few **outliers** show good satisfaction even with long wait times possibly due to better service experience or specific cases.
3. The **large proportion of null-to-zero replacements** may slightly bias the overall downward trend, as these zeros reduce the average satisfaction level.

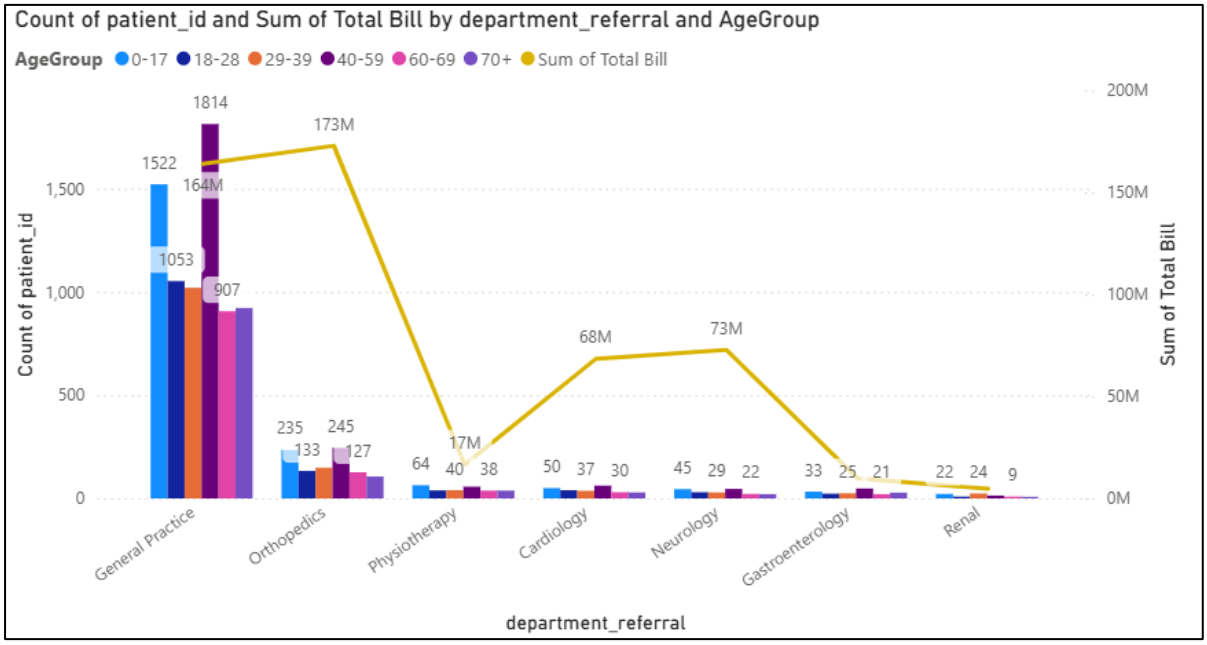
Recommendations

1. **Reduce patient wait times** through better scheduling and queue management.
2. **Improve communication** to set realistic wait-time expectations.
3. **Reassess missing data:** If feasible, collect or estimate satisfaction feedback rather than defaulting to zero to improve accuracy.
4. **Drill down by department** to see where this relationship is strongest.

Q2. How do patient demographics affect the frequency of visits to different departments?

Ans: **Approach**

- Created a **column and line combo chart** in Power BI.
- **X-axis:** department_referral (department names).
- **Y-axis (Column):** Count of patient_id to represent visit frequency.
- **Legend:** AgeGroup to show demographic distribution across departments.
- **Y-axis (Line):** Sum of Total Bill to visualize spending trends alongside visit frequency.
- Chose to exclude gender from this view since it had minimal impact on visit patterns compared to age segmentation.



Insights:

1. **General Practice** shows the **highest patient count across all age groups**, especially among the **29–59** and **60–69** brackets suggesting it serves as the primary care entry point.
2. **Orthopedics** also has substantial visits, particularly from older age groups (40–69), aligning with age-related musculoskeletal conditions.
3. Departments like **Cardiology** and **Neurology** have **lower visit counts** but **higher total bills**, indicating more specialized, high-cost treatments.
4. Younger age groups (0–28) are primarily concentrated in **General Practice**, while elderly patients are more frequent in **Orthopedics** and **Neurology**.

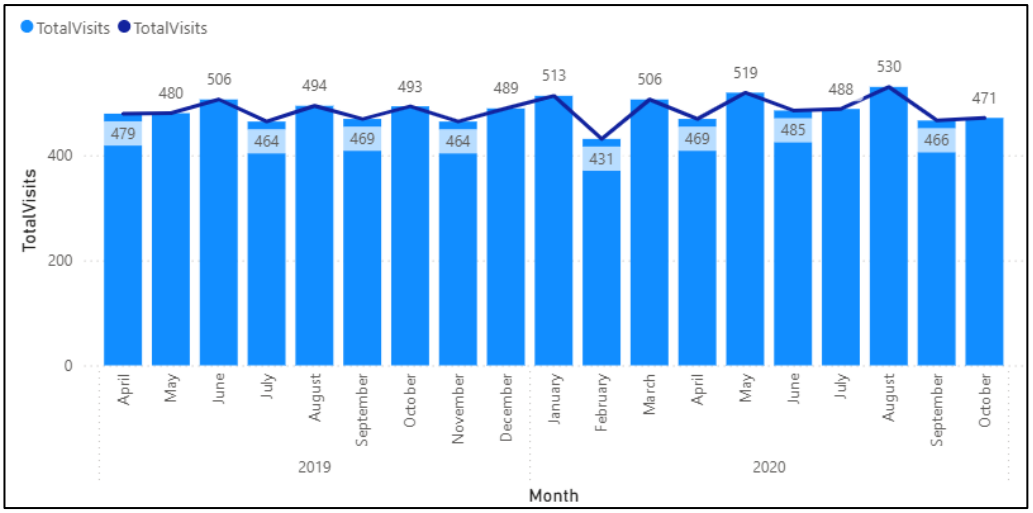
Recommendations:

1. **Resource Allocation:** Strengthen staff and capacity in **General Practice and Orthopedics**, where patient traffic is highest.
2. **Preventive Care Programs:** Introduce **age-specific health initiatives** e.g., wellness checkups for adults (40–59) and preventive physiotherapy for seniors.
3. **Departmental Cost Analysis:** Since specialized departments (like Neurology) have **fewer patients but higher bills**, review **cost structures** and **treatment efficiency**.
4. **Expand Data Depth:** Incorporate **gender and race dimensions** in future analysis for more granular insights, even if initial results show weaker influence.

Q3. Is there a noticeable trend in the volume of patient visits throughout the year?

Ans: Approach

- Created a **column and line combo chart** in Power BI to observe patient visit trends over time.
- **X-axis:** Month (with Year included for a continuous time series).
- **Y-axis (Column):** Measure TotalVisits representing the total number of patient visits each month.
- **Y-axis (Line):** Same measure TotalVisits to show the overall trend across months and years.
- Time-based data was sorted chronologically to ensure trend clarity.



Insights

- Patient visits **fluctuate moderately** throughout the year with **no extreme seasonal spikes**.
- A **slight dip in February** is observed in both years possibly due to fewer working days or seasonal patterns.
- The **highest visits occurred in August 2020 (530 visits)**, indicating a mid-year peak.
- Overall, **average monthly visits remain consistent** around 470–510, reflecting steady hospital engagement and balanced patient inflow.
- The slight rise in 2020 compared to 2019 may indicate improved outreach, awareness, or patient retention.

Recommendations

- **Capacity Planning:** Maintain a **consistent staffing level** throughout the year since visit volumes remain stable, with minor adjustments during peak months like August.
- **Investigate February Dips:** Review operational or external factors (e.g., holidays, weather, scheduling practices) contributing to lower visits in February.
- **Enhance Seasonal Preparedness:** Even with steady visits, ensure **readiness for minor fluctuations** in high-demand months.
- **Monitor Year-on-Year Growth:** Continue tracking this trend to measure **growth in patient base** and evaluate the impact of new hospital initiatives or campaigns.

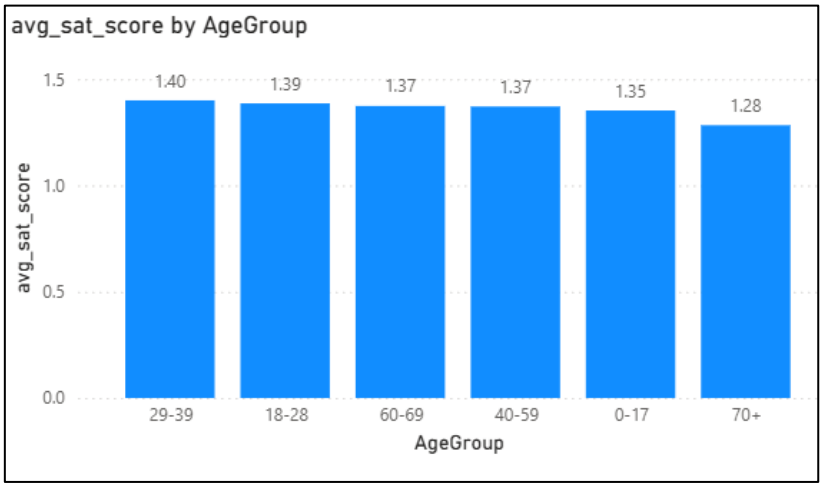
Q4. Which age groups report the highest and lowest satisfaction scores?

Ans: Approach

- Created a DAX measure:

```
avg_sat_score = AVERAGE('Hospital ER (1)'[patient_sat_score])
```

- Used a **column chart** to visualize satisfaction by age group.
- **X-axis:** AgeGroup
- **Y-axis:** avg_sat_score
- Added data labels to display exact average satisfaction values for better readability.



Insights

- The **29–39 age group** reports the **highest average satisfaction (1.40)**, closely followed by **18–28 (1.39)**.
- Satisfaction gradually **declines with increasing age**, reaching the **lowest average (1.28)** among patients **aged 70+**.
- The trend suggests that **younger and middle-aged patients are slightly more satisfied** with their hospital experience than older ones.
- Although differences are small, the **consistent downward pattern** across age groups indicates an age-related perception gap in patient experience.

Recommendations

- **Enhance elderly patient experience:** Implement age-friendly services such as clearer communication, mobility support, and faster service for seniors.
- **Conduct targeted feedback sessions:** Understand **specific pain points** for the 60+ and 70+ groups to address issues like comfort, staff interaction, or accessibility.
- **Maintain satisfaction for younger groups:** Continue the strategies that resonate well with younger demographics (e.g., digital updates, shorter waits).
- **Reassess data collection:** Since earlier you mentioned handling 72% null satisfaction values by replacing with 0, consider improving **feedback collection efficiency** to ensure these averages accurately reflect patient sentiment.

Q5. 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 Approach: To verify whether the claim of racial or gender-based discrimination in the hospital is valid, I performed **three key analyses** focusing on patient outcomes and experience metrics.

1. Average Wait Time by Race & Gender

- **Measure Used:**

```
average_wait_time = AVERAGE('Hospital ER (1)'[patient_waittime(in_minutes)])
```

- **Visualization:** Clustered column chart
  - **X-axis:** Patient Race
  - **Legend:** Patient Gender
  - **Value:** Average Wait Time
- **Purpose:** To check if any specific race or gender group faces longer wait times before treatment, indicating possible bias in service speed.

2. Admission Rate by Race & Gender

- **Measures Used:**

```
Admission_Count = CALCULATE(COUNTROWS('Hospital ER (1)'), 'Hospital ER (1)'
[patient_admin_flag] = TRUE())
```

```
Admission_Rate = DIVIDE( [Admission_Count],[TotalVisits])
```

- **Visualization:** 100% Stacked Column Chart
    - **X-axis:** Patient Race
    - **Legend:** Patient Gender
    - **Value:** Admission Rate
- **Purpose:** To analyze if certain races or genders are less likely to be admitted compared to others, which could indicate bias in decision-making.

3. Total Revenue & Avg Satisfaction Score by Race & Gender

- **Measures Used:**

```
avg_sat_score = AVERAGE('Hospital ER (1)'[patient_sat_score])
```

```
TotalRevenue = SUM(Sheet1[Total Bill])
```

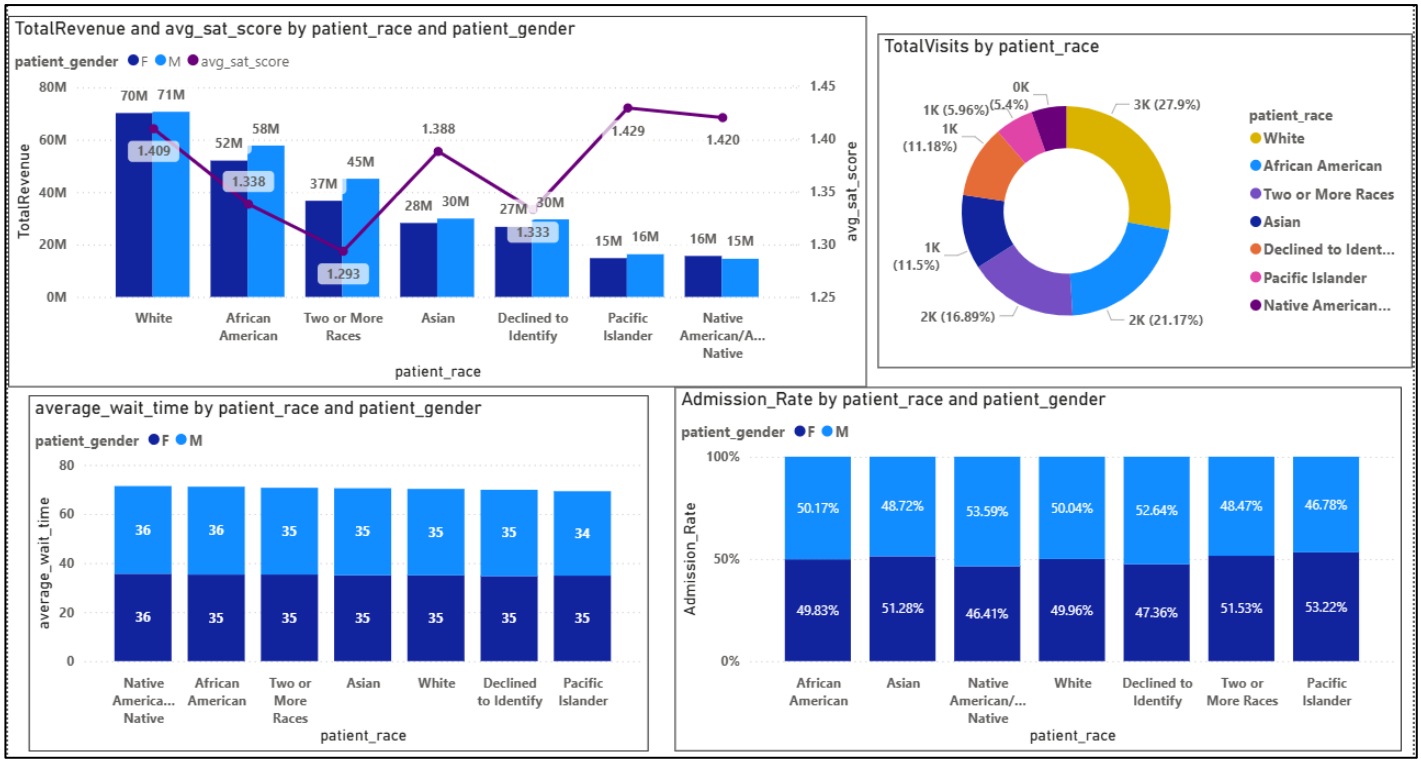
- **Visualization:** Combo (Column + Line Chart)

- **X-axis:** Patient Race
- **Column (Value):** Total Revenue
- **Legend:** Patient Gender
- **Line (Value):** Avg Satisfaction Score

Purpose: To check if billing or satisfaction levels differ across demographic groups, possibly reflecting treatment inequality.

4. Total Visits by Race

- **Visualization:** Donut Chart
- **Purpose:** To understand the patient distribution by race and ensure balanced hospital accessibility.



Insights

1. **Wait Time:**
   - Average wait time remains consistent (34–36 minutes) across all races and genders, showing no delay bias.
2. **Admission Rate:**
   - Admission rates are balanced between **48–53%**, suggesting fair admission decisions across demographics.
3. **Revenue & Satisfaction:**
   - Revenue and satisfaction are stable across all groups (avg. satisfaction 1.29–1.43).
   - Minor differences appear operational rather than discriminatory.
4. **Visit Distribution:**
   - Majority of visits come from White (27.9%) and Asian (21.1%) groups, but all other races are reasonably represented.
   - No evidence of exclusion or service denial to any race.
5. **Overall:**
   - No significant gender or racial discrimination observed.
   - Small variations are within normal operational range.

Recommendations

1. **Maintain Fairness Monitoring System:**
    - Implement a **Demographic Equality Dashboard** tracking Wait Time, Admission Rate, Revenue, and Satisfaction by Race & Gender on a monthly basis.
2. **Periodic Department-Level Audit:**
    - Review departments where slight differences persist to ensure consistent triage, admission, and billing processes.
3. **Enhance Patient Feedback Mechanism:**
    - Encourage detailed satisfaction surveys across all demographic groups to capture subtle experience differences.
4. **Bias Awareness Training:**
    - Conduct diversity and sensitivity workshops for staff to minimize unconscious bias in patient interaction and treatment prioritization.
5. **Transparent Reporting:**
    - Publish a short quarterly "Fairness and Inclusion Report" to maintain public trust and demonstrate transparency.

Q6. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?

Ans: Approach To determine how discounts should be allocated, I decided to analyze the **AgeGroup** based on three key metrics:
Total Revenue, **Total Visits**, and **Average Appointment Fees**.

Based on these metrics, I planned to identify which age groups can be allocated for discounts, their respective **discount category (High / Medium / Low)**, and **eligibility (Yes/ No)**.

Step 1: Creating DAX Measures

To begin, I created the following **DAX measures**:

```
TotalRevenue = SUM(Sheet1[Total Bill])
```

Represents the total income generated

Average Appointment Fees

```
AvgAppointmentFees = AVERAGE(Sheet1[Appointment Fees])
```

Indicates the average consultation fee paid by patients**.**

Total Visits

```
TotalVisits = COUNTROWS('Hospital ER (1)')
```

Reflects the engagement level and frequency of visits from each age group.

Step 2: Visualization

Then, I created a **Combo Chart (Column & Line)** to visualize:

- **Columns:** Total Revenue
- **Line:** Total Visits and Avg Appointment Fees

This helped to easily compare revenue and engagement across different **Age Groups**.

Step 3: Key Findings

Based on the visual insights:

1. The **40–59** and **0–17** age groups generated the **highest total revenue** and had **maximum visits**, showing strong engagement these groups don't require heavy discounts.
2. The **60–69** and **70+** age groups showed **lowest revenue and visits**, indicating the need for higher discounts to encourage checkups.
3. The **29–39** and **18–28** groups showed **moderate performance**, so a medium-level discount is suitable to improve their engagement.

Step 4: Creating Discount Category

Next, I created a DAX column to categorize discounts:

```
DiscountCategory =
SWITCH(
    TRUE(),
    'Hospital ER (1)'[AgeGroup] IN {"40-59", "0-17"}, "Low",
    'Hospital ER (1)'[AgeGroup] IN {"29-39", "18-28"}, "Medium",
    'Hospital ER (1)'[AgeGroup] IN {"60-69", "70+"}, "High",
    "Low"
)
```

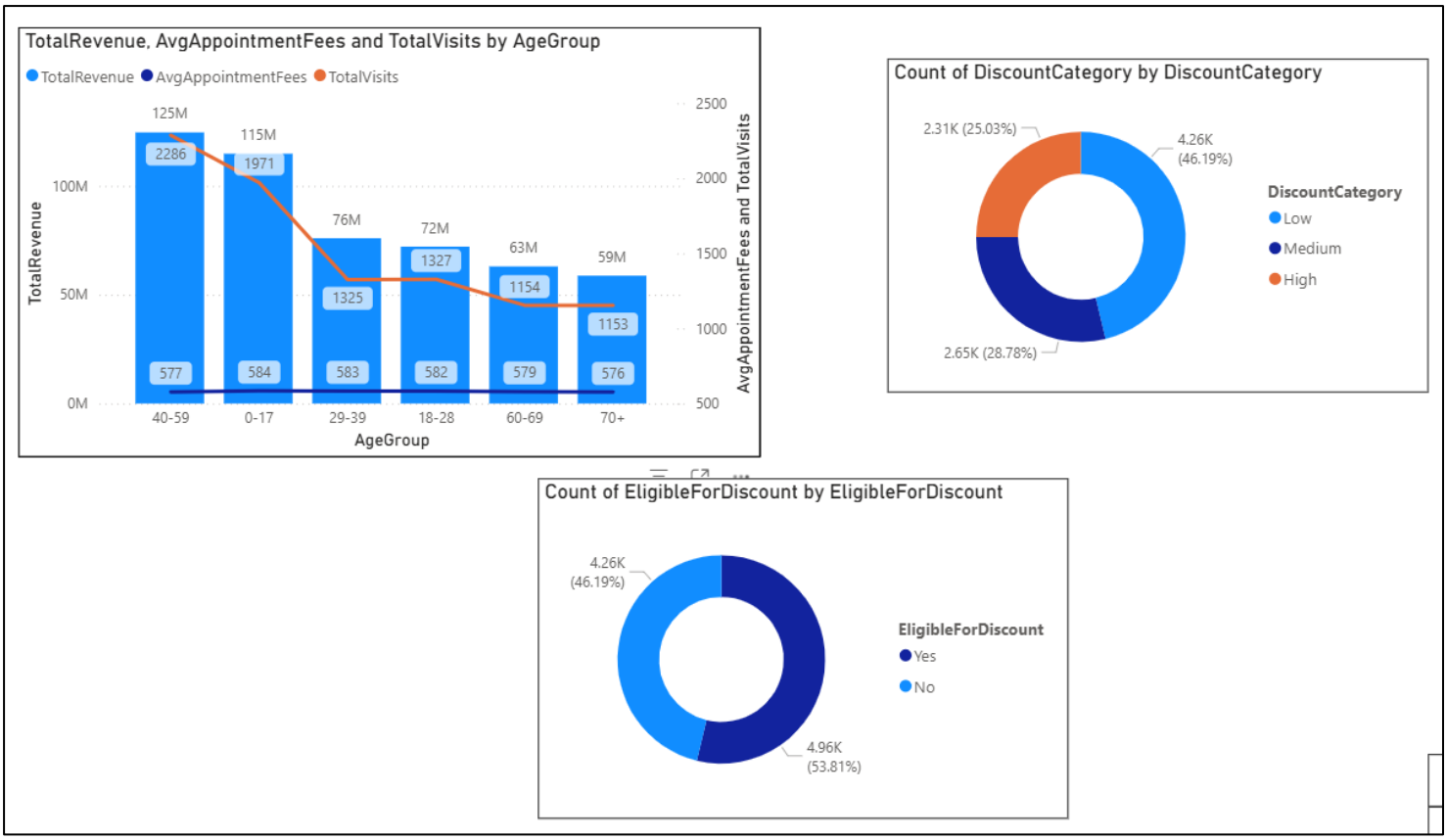
This categorization helps to clearly define which age groups deserve higher or lower discounts.

Step 5: Identifying Discount Eligibility

- To identify eligible patients for discounts, a new **DAX column (EligibleForDiscount)** was created using conditions based on **patient satisfaction score and age**.

```
EligibleForDiscount =
SWITCH(
    TRUE(),
    'Hospital ER (1)'[patient_sat_score] >= 5 || 'Hospital ER (1)'[patient_age] >=
    60, "Yes",
    "No"
)
```

- Patients with a **satisfaction score ≥ 5** or **age ≥ 60** were marked as **“Yes” (eligible)**, while others were marked as **“No”**.
- This approach helps to **target elderly and highly satisfied patients**, making the discount distribution more logical and data-driven



Conclusion: Based on the analysis, the discount allocation can be categorized into three groups

1. **High Discounts (15–20%)**: Applicable for **60–69** and **70+** age groups.

- Aim: Encourage **elderly patients** to visit more frequently for **regular health checkups**.

2. **Medium Discounts (around 10%)**: Applicable for **18–28** and **29–39** age groups.

- Aim: Motivate **working-age adults** to **increase engagement** and **improve visit frequency**.

3. **Low or No Discounts**: Applicable for **0–17** and **40–59** age groups.

- Aim: These groups already show **high revenue and visit contribution**; focus on **retention and satisfaction** rather than new incentives.

Q7. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.

Ans: Approach

To identify which departments need additional doctors, I analyzed three key metrics:

- **TotalRevenue** : Indicates how much revenue each department contributes.

```
TotalRevenue = SUM(Sheet1[Total Bill])
```

- **TotalVisits** :Reflects patient demand and department workload.

```
TotalVisits = COUNTROWS('Hospital ER (1)')
```

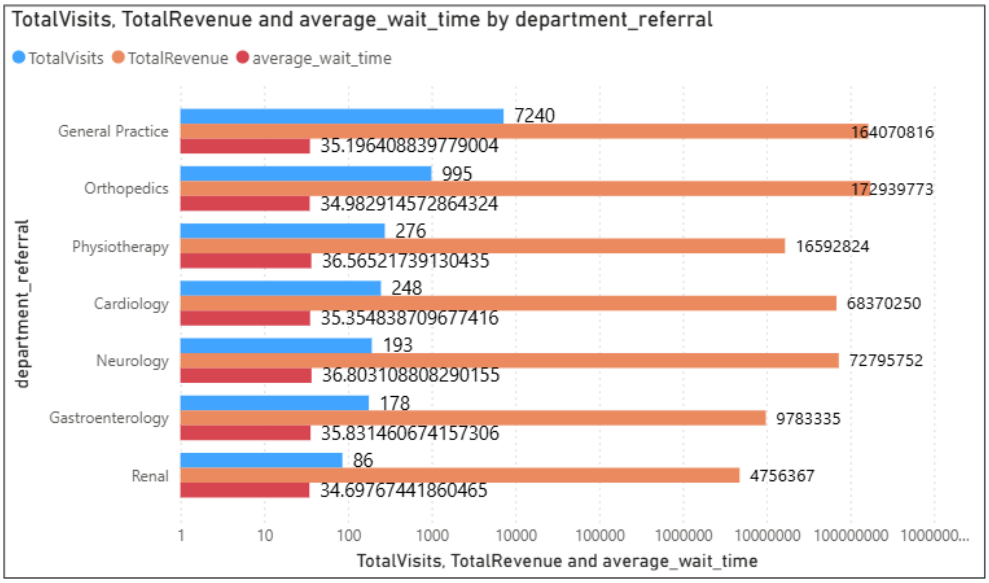
- **average_wait_time** :Represents patient waiting time and helps assess staff workload and service efficiency.

```
average_wait_time = AVERAGE('Hospital ER (1)'[patient_waittime(in_minutes)])
```

Then, I visualized these three metrics using a **Bar Chart**, with

- **X-Axis:** TotalRevenue, TotalVisits, average_wait_time
- **Y-Axis:** department_referral

This helped identify departments that have **high patient visits** but **low doctor availability** or **high wait times**, indicating a potential need for hiring.



Insights:

From the visualization

- **General Practice** has the **highest number of visits (7240)** and **high total revenue**, but also a **moderate wait time (35.19 mins)** → indicating **heavy patient load**.
- **Orthopedics** has **high revenue** but **very low visits (995)** → meaning fewer but possibly longer appointments or specialized treatments.
- **Neurology** and **Cardiology** show **lower visits (193 & 248)** but **high average wait times (36.8 & 35.3 mins)**, suggesting **doctor shortage or longer consultation durations**.
- **Physiotherapy** and **Gastroenterology** have **moderate wait times** and **medium-level visits**, showing manageable workloads.
- **Renal** department has **the lowest visits (86)** and **lowest revenue**, so it currently may not need additional doctors.

Recommendations:

Based on the analysis of workload (TotalVisits) **and** service pressure (average_wait_time)**:**

1. **General Practice**
    - **Reason:** Highest patient visits and revenue, indicating heavy demand.
    - **Recommendation:** Hire at least **one additional doctor** to reduce patient waiting times and manage the inflow effectively.
2. **Neurology**

- **Reason:** High average wait time despite lower patient volume, suggesting longer consultation durations or fewer doctors.
- **Recommendation:** Add **one doctor** to improve service speed and reduce patient delays.

3. **Cardiology**
   - **Reason:** Similar pattern high wait times and specialized care needs.
   - **Recommendation:** Hire **one more doctor** to handle time-intensive cases efficiently.

Q8. Is the hospital profitable? How will you determine the profitability?

Ans: Approach As of now, there is no specific column directly indicating hospital profitability in the dataset.
So, to assess profitability indirectly, I performed **two key analyses**:

1. **Average Revenue per Doctor (by Department Referral)**
   - Created a measure:

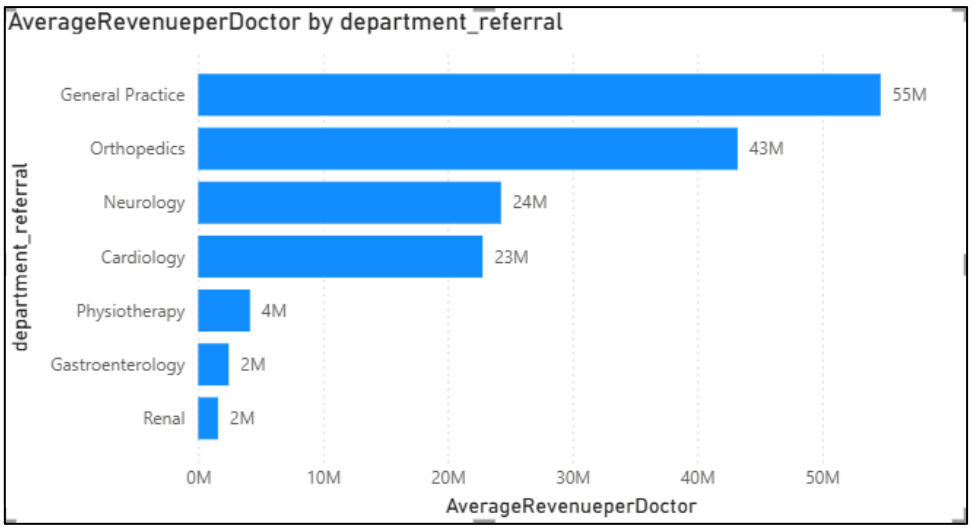
```
AverageRevenueperDoctor = SUM(Sheet1[Total Bill]) / DISTINCTCOUNT(Sheet1[Doctor Name])
```

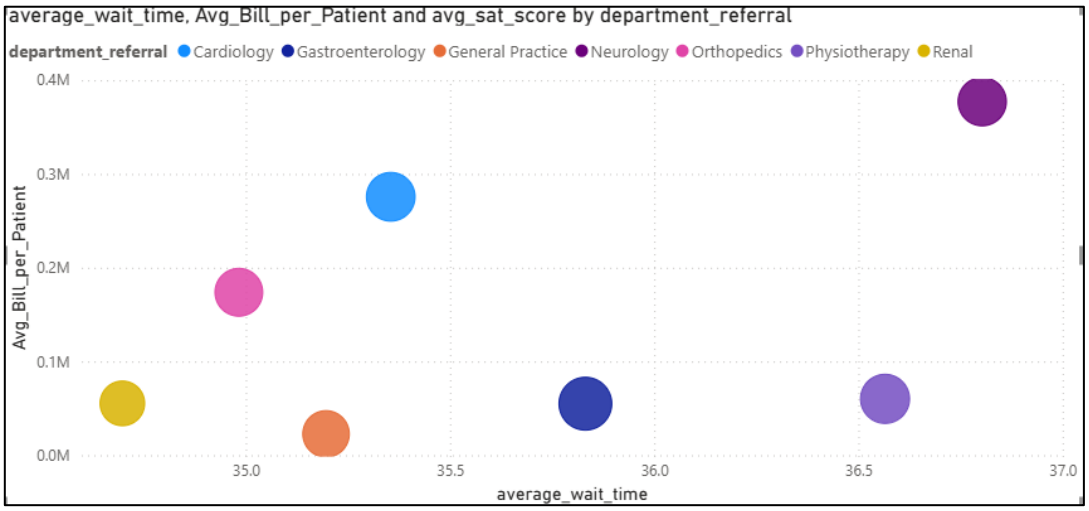
   - Used a **bar chart** to visualize which departments generate higher revenue per doctor, helping assess **productivity and efficiency**.
2. **Average Revenue per Patient vs Average Wait Time (by Department)**
   - Calculated two measures:

```
Avg_Bill_per_Patient = DIVIDE([TotalRevenue], DISTINCTCOUNT('Hospital ER (1)'
[patient_id]))
```

```
average_wait_time = AVERAGE('Hospital ER (1)'[patient_waittime(in_minutes)])
```

   - Used a **bubble chart**, where:
     - X-axis → Average Wait Time
     - Y-axis → Avg Bill per Patient
     - Bubble size → Average Satisfaction Score
     - Color → Department Referral
   - This helps visualize the balance between **revenue efficiency, patient experience, and service speed**.





Insights

- **General Practice and Orthopedics** show the **highest revenue per doctor**, indicating these departments are highly productive and financially efficient.
- **Cardiology and Neurology** generate moderate revenue but might be affected by longer patient wait times.
- **Physiotherapy, Gastroenterology, and Renal** have both **lower revenue per doctor and per patient**, suggesting lower profitability or underutilization.
- Departments with **higher average wait times** tend to have **higher bills per patient**, which might indicate specialized or longer treatments.

Recommendations

- **Focus investments** in high-performing departments like **General Practice and Orthopedics**, as they show strong profitability per doctor.

- **Improve operational efficiency** in departments with long wait times (e.g., Neurology, Cardiology) to enhance both satisfaction and throughput.
- **Review underperforming departments** (e.g., Physiotherapy, Gastroenterology, Renal) either by improving service offerings or optimizing staff allocation.
- **Balance profitability and patient satisfaction** by ensuring that higher revenue per patient doesn't come at the cost of excessive waiting times or poor experience.

Q9. Any Department for which the waiting time is oddly large?

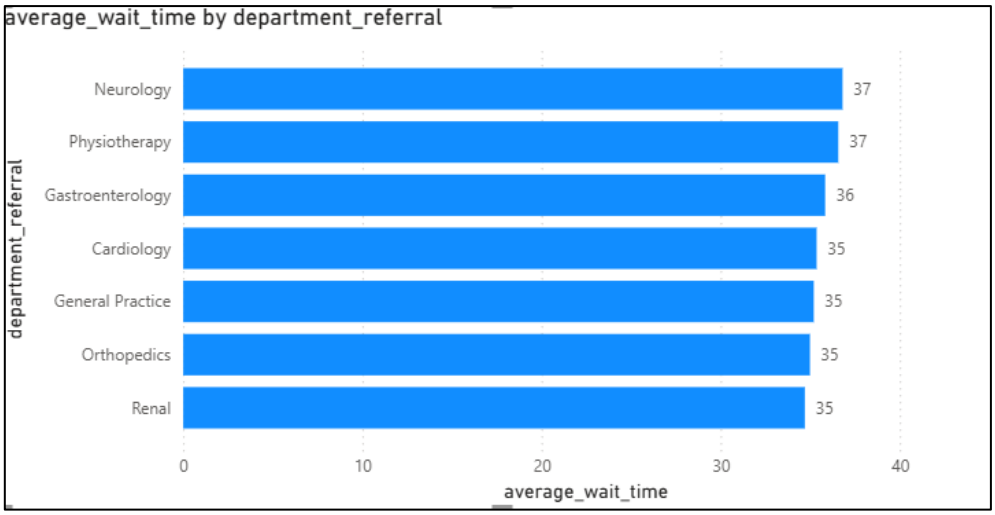
Ans: Approach: To analyze departments with unusually high waiting times,
I created a **measure** for the average patient wait time:

```
average_wait_time = AVERAGE('Hospital ER (1)'[patient_waittime(in_minutes)])
```

Then, I used a **bar chart** visualization:

- **X-axis:** Department Referral
- **Y-axis:** Average Wait Time

This helps identify departments with above-average patient waiting durations.



Insights:

- There are no such departments which indicates oddly higher wait times all of them have equivalent wait times but , the **Neurology** and **Physiotherapy** departments have **an slightly higher wait times (~37 mins)** compared to other departments, indicating possible bottlenecks in patient handling or appointment scheduling.

Recommendations:

- Review staffing and scheduling patterns in all departments or targeting those departments who has highest no.of visits and total revenue to balance patient load more effectively.
- Introduce **appointment prioritization** or **triage systems** to manage urgent vs. routine cases efficiently.
- Encourage **digital appointment booking and reminders** to reduce overcrowding and missed slots.
- Monitor **patient feedback** from these departments to identify specific causes of long waits (e.g., diagnostic delays, admin processes).

Q10.Come up with strategies to provide discounts to the patients.

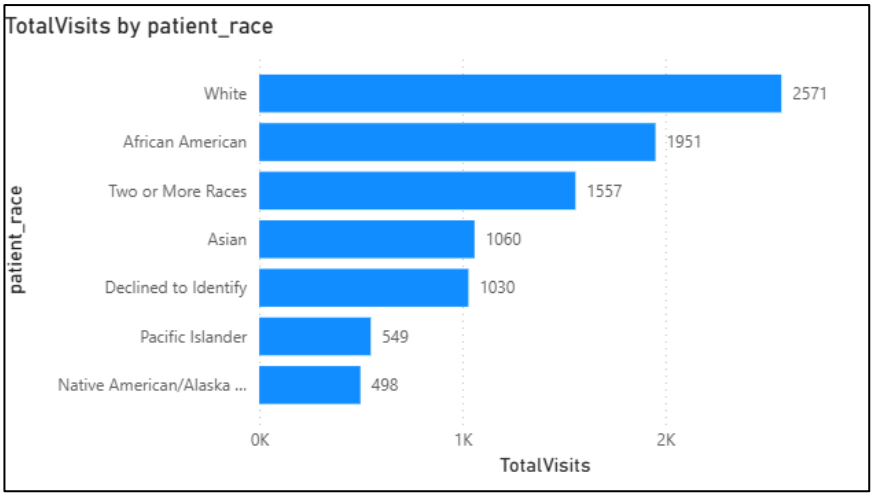
Ans Approach:

In my previous analysis, I designed the discount strategy using **Age Group** as the key parameter, where discounts were categorized based on **Total Revenue, Total Visits, and Average Appointment Fees**.
To extend this analysis further, I decided to explore the **Patient Race** dimension to understand which racial groups contribute the most visits and where targeted discounts could enhance engagement and inclusivity.

For this, I created a **bar chart visualization** in Power BI with:

- **X-Axis:** TotalVisits
- **Y-Axis:** patient_race

The goal was to identify racial groups with **lower visit frequencies**, where discounts or promotional offers could encourage higher participation and improve healthcare accessibility.



Insights

From the visualization:

- **White** and **African American** patients account for the **highest total visits (2571 and 1951)**, indicating strong engagement and consistent hospital usage.
- **Two or More Races** and **Asian** groups show **moderate engagement (1557 and 1060 visits)**.
- **Pacific Islander** and **Native American/Alaska Native** groups have **the lowest visit counts (549 and 498)**, suggesting potential barriers or lower participation rates.
- The **Declined to Identify** group (1030 visits) remains neutral, showing average engagement.

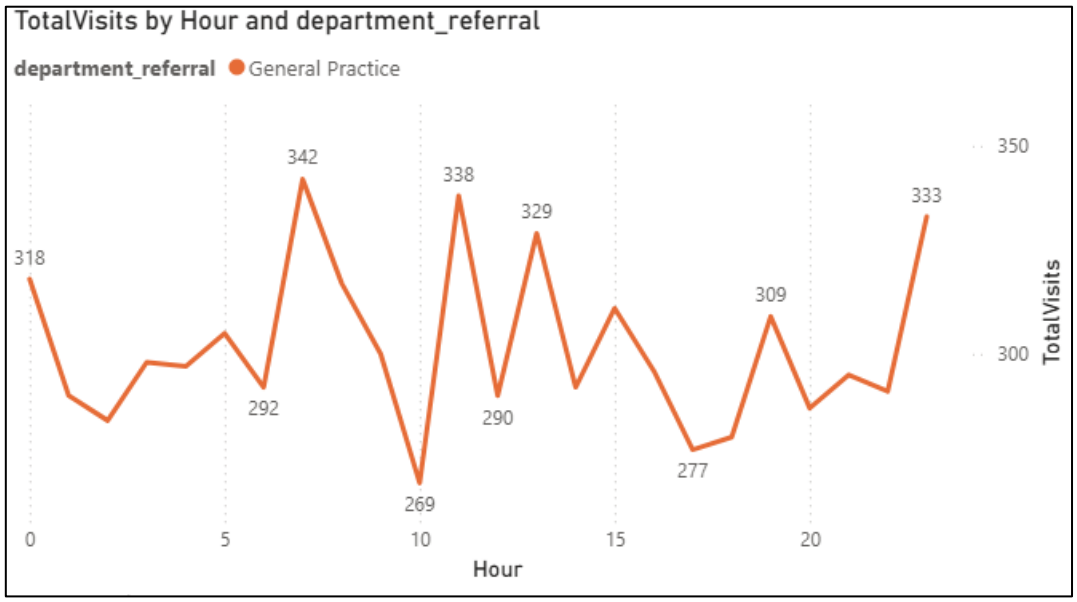
Recommendations / Discount Strategy

1. **High Discount (15–20%) → Native American/Alaska Native & Pacific Islander**
    - Reason: These groups have **lowest total visits**, suggesting lower hospital engagement.
    - Goal: Encourage healthcare access and improve inclusivity among underrepresented patient groups.
2. **Medium Discount (10%) → Asian & Two or More Races**
    - Reason: Moderate visit frequency can benefit from moderate incentives to boost engagement further.
    - Goal: Strengthen connection with moderately engaged groups.
3. **Low or No Discount → White & African American**
    - Reason: These groups already show **high visit frequency** and contribute significantly to hospital utilization.
    - Goal: Maintain retention and satisfaction rather than offering additional incentives.

Q11.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: Approach:

- Used **Hour** (extracted from the DateTime column) on the **X-axis** and **Total Visits** on the **Y-axis**.
- Applied a **filter** to show only the **General Practice** department.
- Created a **Line Chart** to visualize patient visit trends across different hours of the day.
- The chart reveals fluctuations in patient visits throughout the day, helping identify **peak and low-traffic hours** for shift planning.



Insights:

- **Peak Hours:** High patient inflow observed around **6 AM, 11 AM, and 1 PM** (342, 338, and 329 visits respectively).
- **Low Hours:** Noticeably lower patient visits around **10 AM (269 visits)** and between **5 PM–8 PM (277–300 visits)**.
- The variation indicates **two distinct busy periods** one in the **morning** and another around **midday**.
- Early morning and mid-afternoon dips suggest transition windows suitable for shift handovers.

Recommendations:

- **Shift 1 (Morning Shift): 6 AM – 2 PM**
  Covers early and mid-morning peaks (6 AM, 11 AM) ensuring sufficient doctor availability when patient inflow is high.
- **Shift 2 (Afternoon–Evening Shift): 2 PM – 10 PM**
  Covers midday and evening consultations while balancing workload after the peak hours.
- **Doctor Allocation:**
  - Allocate a **slightly higher number of doctors** in **Shift 1** due to higher visit volume in morning and midday hours.
  - Maintain **fewer doctors** in **Shift 2**, with potential backup or on-call doctors during occasional evening spikes.

Will the Two-Shift Policy Help?

Yes.
Implementing two well-balanced shifts will:

- Ensure **adequate doctor availability** during patient peak times.
- Reduce **doctor fatigue** by distributing workload more evenly.
- Improve **patient wait times** and **overall satisfaction**.

Q12.What do you understand by PowerBI gateway? What are its use cases?

Ans: Power BI Gateway is a **bridge (connector)** that enables **secure data transfer** between **on-premises data sources** (like SQL Server, Oracle, Excel files, etc.) and **Microsoft Power BI cloud services**.
It ensures that data in Power BI reports and dashboards stays **up-to-date** without manually uploading files each time.

Types of Power BI Gateways:

1. **Personal Mode Gateway:**
   - Installed and used by **individual users**.
   - Connects to data sources for **personal use** in Power BI Service.
   - Does **not support multiple users** or **scheduled refresh for shared datasets**.

2. **Standard Mode (Enterprise Gateway):**
    - Used in **organizational or enterprise environments**.
    - Allows **multiple users** to connect securely to **on-premises data sources**.
    - Supports **scheduled refreshes, DirectQuery**, and **live connections**.
    - Managed centrally by IT teams for better **security and governance**.

Use Cases of Power BI Gateway:

1. **Scheduled Data Refresh:**
    - Automatically refresh Power BI reports with the latest data from on-premises databases.
2. **DirectQuery and Live Connection:**
    - Enables real-time analytics without moving data to the cloud useful for systems like **SQL Server**, **SAP HANA**, or **Analysis Services**.
3. **Hybrid Data Environment:**
    - Integrates both **cloud-based** and **on-premises** data sources for unified reporting.
4. **Data Security & Compliance:**
    - Keeps sensitive data **within the organization's network**, transmitting only query results securely to the Power BI Service.
5. **Multi-user Collaboration:**
    - Enterprise Gateway allows different departments to access and refresh shared datasets securely.

Q13.How would you approach this problem, if the objective and subjective questions weren't given?

Ans: **Approach if Objective and Subjective Questions Were Not Provided**

If no predefined questions were available, my approach would focus on **exploring the datasets in depth** to uncover patterns, relationships, and insights that could guide hospital operations and decision-making.

1. Understanding the Dataset

- **Initial Review:** I would begin by studying the available tables (doctor_patients_data and hospital_er) to understand their structure, columns, and data types.
- **Data Profiling:** Next, I'd perform data profiling to detect missing values, data inconsistencies, and distribution patterns for key fields such as patient age, gender, race, satisfaction score, doctor name, and department referral. This helps assess the dataset's quality and readiness for analysis.

2. Identifying Key Analysis Areas

- **Patient Demographics:** Examine how different age groups and genders influence visit frequency, satisfaction, and wait times.
- **Department Insights:** Analyze each department's patient load, satisfaction, and performance to find high and low-performing areas.
- **Doctor Evaluation:** Track individual doctor performance using metrics like satisfaction scores, visit counts, and generated revenue.
- **Time-Based Trends:** Study temporal patterns (daily, weekly, monthly) in visits and satisfaction to understand service demand over time.

3. Building Metrics and KPIs

- **Wait Time vs. Satisfaction:** Identify if longer wait times negatively affect patient satisfaction.
- **Revenue Analysis:** Evaluate department-wise and doctor-wise revenue to understand profitability.
- **Patient Diversity:** Measure demographic diversity within each doctor's patient group.
- **Performance Trends:** Compare average satisfaction scores across doctors and departments to identify best practices.

4. Visual Representation in Power BI

To bring clarity to the findings, I'd use Power BI visualizations such as:

- **Line Charts** for time-based trends.
- **Bar/Column Charts** for department or demographic comparisons.
- **Scatter Plots** to analyze correlations like wait time vs. satisfaction.
- **KPI Cards and Gauges** for key metrics (e.g., average satisfaction score, total revenue).

5. Insights and Recommendations

After analyzing the visuals, I'd extract actionable insights.
For example:

- If longer wait times lower satisfaction, recommend optimizing appointment scheduling.
- If certain doctors or departments underperform, suggest targeted training or better resource management.

6. Reporting the Findings

Finally, I'd document all results and recommendations in a clear report format, highlighting how the insights align with hospital goals such as **enhancing patient satisfaction, optimizing resources, and boosting revenue**.

Q14.Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?

Ans Approach:

To analyze the relationship between **Doctor ID** and **Department**, I created a DAX summary table:

```
DoctorsRelationWithDepartment =
SUMMARIZE(
    'Sheet1',
     Sheet1[Doctor Name],
    "TotalNoOfDepartments", DISTINCTCOUNT(Sheet1[department_referral])
)
```

This helped identify how many **unique departments** each doctor is associated with.

Doctor Name	Sum of TotalNoOfDepartments
Dr. Surya	3
Dr. Anderson	1
Dr. Brown	1
Dr. Clark	1
Dr. Davis	1
Dr. Garcia	1
Dr. Harris	1
Dr. Jackson	1
Dr. Johnson	1
Dr. Martin	1
Dr. Martinez	1
Dr. Miller	1
Dr. Moore	1
Dr. Robinson	1
Dr. Rodriguez	1
Dr. Smith	1
Dr. Taylor	1
Dr. Thomas	1
Dr. Thompson	1
Dr. White	1
Dr. Williams	1
Dr. Wilson	1
Total	**24**

Insights:

- From the visualization, it is observed that **almost all doctors** are associated with **only one department**.
- **Dr. Surya** is the **only doctor** who works across **multiple departments (3)**, indicating a cross-department role or shared specialization.
- The rest of the doctors have a **1:1 relationship** with departments meaning each doctor belongs to one specific department

Conclusion:

- The overall relationship between **Doctor ID** and **Department** is **one-to-one**.
- However, there is **one exception (Dr. Surya)**, where the relationship becomes **one-to-many**, suggesting involvement in multiple specialties or departments.
- This mixed pattern indicates that while the hospital generally assigns doctors to a single department, some doctors may support multiple areas to balance workloads or share expertise.