Objective Questions

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

Solution:

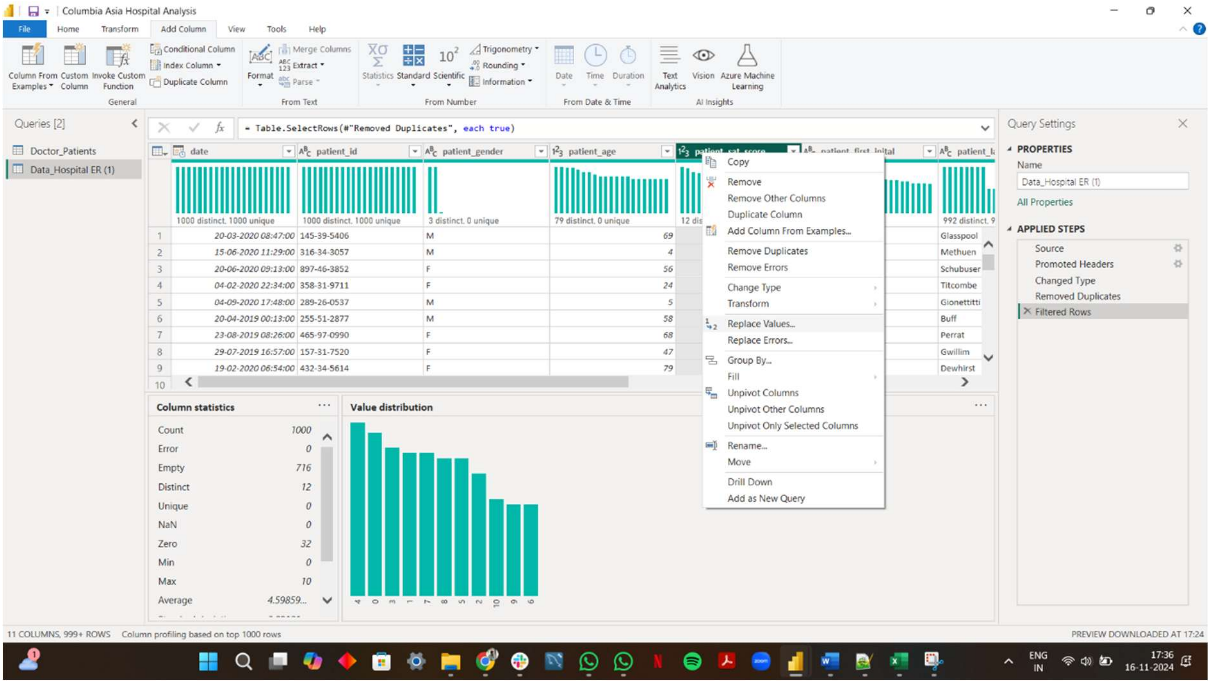
Data Cleaning can be performed with the help of Power Query Editor. We can identify the null values by clicking on the columns dropdown list and see if there is any null value.

Or,In Power Query Editor, we can go to View -> Data Preview. In this, we can find various options like Monospaced, Value Distribution etc. to check the distribution of null values in the dataset.

For numeric data type such as age, satisfication score, we can replace it with the average value of the parameters.

For text data types, we can replace it with placeholders like Unknown / NA.

For date datatype, we can either replace the null values with the minimum date or the maximum date.



```
= Table.SelectRows(#"Removed Duplicates", each true)
```

	date	patient_id	patient_gender	patient_age	patient_sat_score	patient_first_inital	patient_l
	1000 distinct, 1000 unique	1000 distinct, 1000 unique	3 distinct, 0 unique	79 distinct, 0 unique	12 dis		992 distinct, 9
1	20-03-2020 08:47:00	145-39-5406	M	69			Glasspool
2	15-06-2020 11:29:00	316-34-3057	M	4			Methuen
3	20-06-2020 09:13:00	897-46-3852	F	56			Schubuser
4	04-02-2020 22:34:00	358-31-9711	F	24			Titcombe
5	04-09-2020 17:48:00	289-26-0537	M	5			Gionettitti
6	20-04-2019 00:13:00	255-51-2877	M	58			Buff
7	23-08-2019 08:26:00	465-97-0990	F	68			Perrat
8	29-07-2019 16:57:00	157-31-7520	F	47			Gwillim
9	19-02-2020 06:54:00	432-34-5614	F	79			Dewhirst
10							

Column statistics

Count	1000
Error	0
Empty	716
Distinct	12
Unique	0
NaN	0
Zero	32
Min	0
Max	10
Average	4.59859...

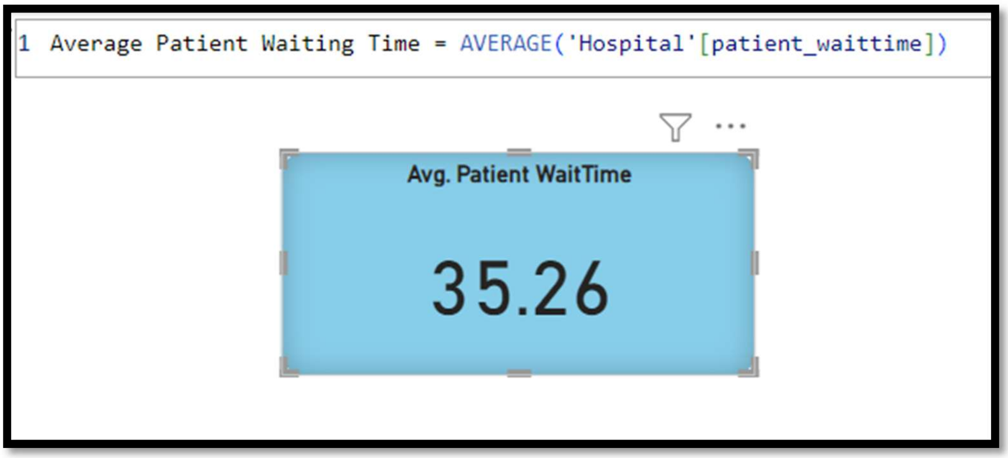
11 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 17:24

---

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

Solution:

Average Patient Waiting Time **= AVERAGE('Hospital'[patient_waittime])**



```
1  Average Patient Waiting Time = AVERAGE('Hospital'[patient_waittime])
```

Avg. Patient WaitTime

35.26

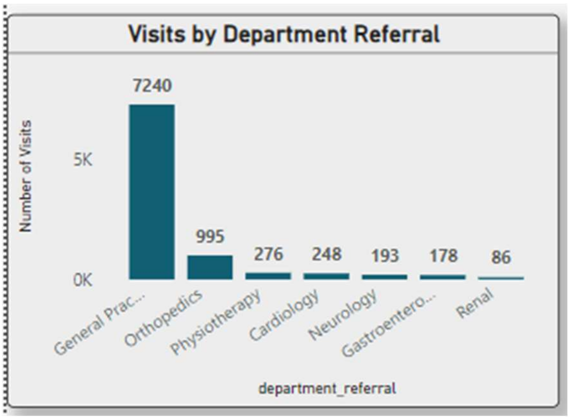
---

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.

Solution:

Total Number of Patients: **9216**

We can see that General Practice department has the highest number of visitors at 7240, followed by Orthopedics with 995 visitors. The departments with the lowest number of visitors are Renal and Gastroenterology at 86 and 178 respectively.



---

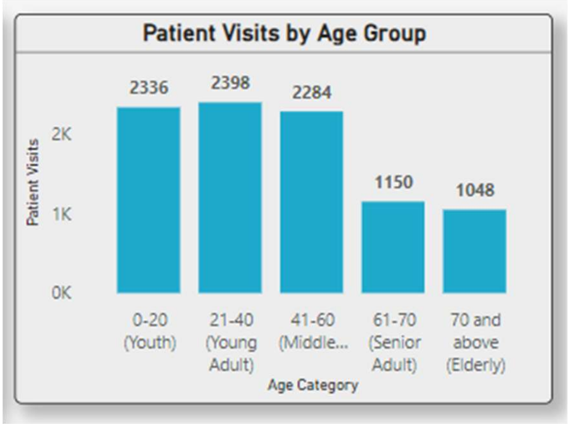
4. **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

Solution:

We make use of Conditional Columns under the Add column tab in Power Query Editor to add a column based on conditions.

- Patients with age <= 20 are categorised as **Youth**.
- Patients with age in the range of 21 – 40 are categorised as **Young Adult.**
- Patients with age in the range of 41 – 60 are categorised as **Middle-Aged.**
- Patients with age in the range of 61 – 70 are categorised as **Senior Adult.**
- Patients with age >= 70 are categorised as **Elderly.**

```
= Table.AddColumn(#"Replaced Value", "Age Category", each if [patient_age] <= 20 then "0-20 (Youth)" else if [patient_age] <= 40 then "21-40 (Young Adult)" else if [patient_age] <= 60 then "41-60 (Middle-Aged)" else if [patient_age] <= 70 then "61-70 (Senior Adult)" else if [patient_age] > 70 then "70 and above (Elderly)" else null)
```



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?

Solution:

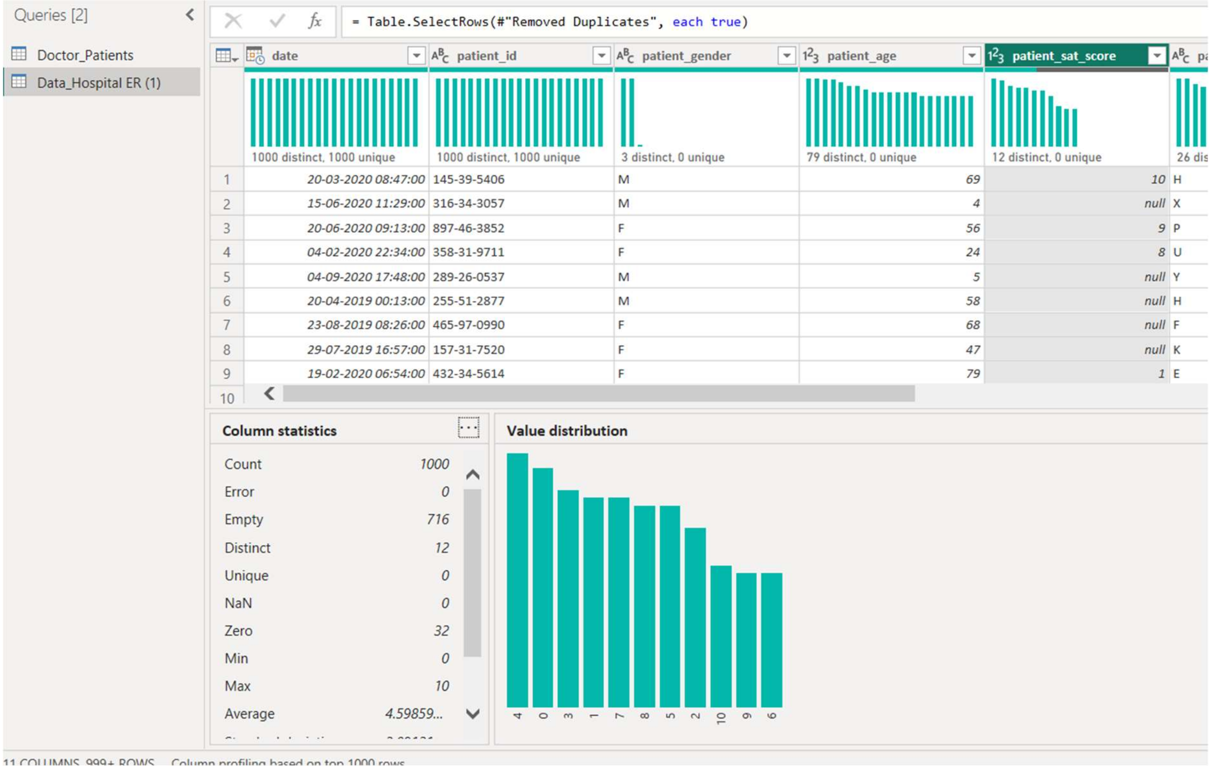
The patient_sat_score column in hospital table had the highest number of null values. We can replace the null values by calculating the average of the patient satisfaction scores and replace the null values with the average of the patient satisfaction score.

To find the average of the satisfaction score, we can create a measure as follows: **Average Satisfaction Score = AVERAGE(Hospital[patient_sat_score]),** which gives us a value of 4.9978298611111107 which can be rounded off to 5.

Then, in the power query editor, right click on patient_sat_score column header and click on Replace Values. In the Value to Find box, we won't be entering any value since we have to find the NULL values. In the Replace With box, enter 5 and click on OK.

We can see that the NULL values has been replaced with the average of satisfaction scores which is 5.

Before the transformation:



```
= Table.SelectRows(#"Removed Duplicates", each true)
```

	date	patient_id	patient_gender	patient_age	patient_sat_score	pa
	1000 distinct, 1000 unique	1000 distinct, 1000 unique	3 distinct, 0 unique	79 distinct, 0 unique	12 distinct, 0 unique	26 dis
1	20-03-2020 08:47:00	145-39-5406	M	69	10	H
2	15-06-2020 11:29:00	316-34-3057	M	4	null	X
3	20-06-2020 09:13:00	897-46-3852	F	56	9	P
4	04-02-2020 22:34:00	358-31-9711	F	24	8	U
5	04-09-2020 17:48:00	289-26-0537	M	5	null	Y
6	20-04-2019 00:13:00	255-51-2877	M	58	null	H
7	23-08-2019 08:26:00	465-97-0990	F	68	null	F
8	29-07-2019 16:57:00	157-31-7520	F	47	null	K
9	19-02-2020 06:54:00	432-34-5614	F	79	1	E
10						

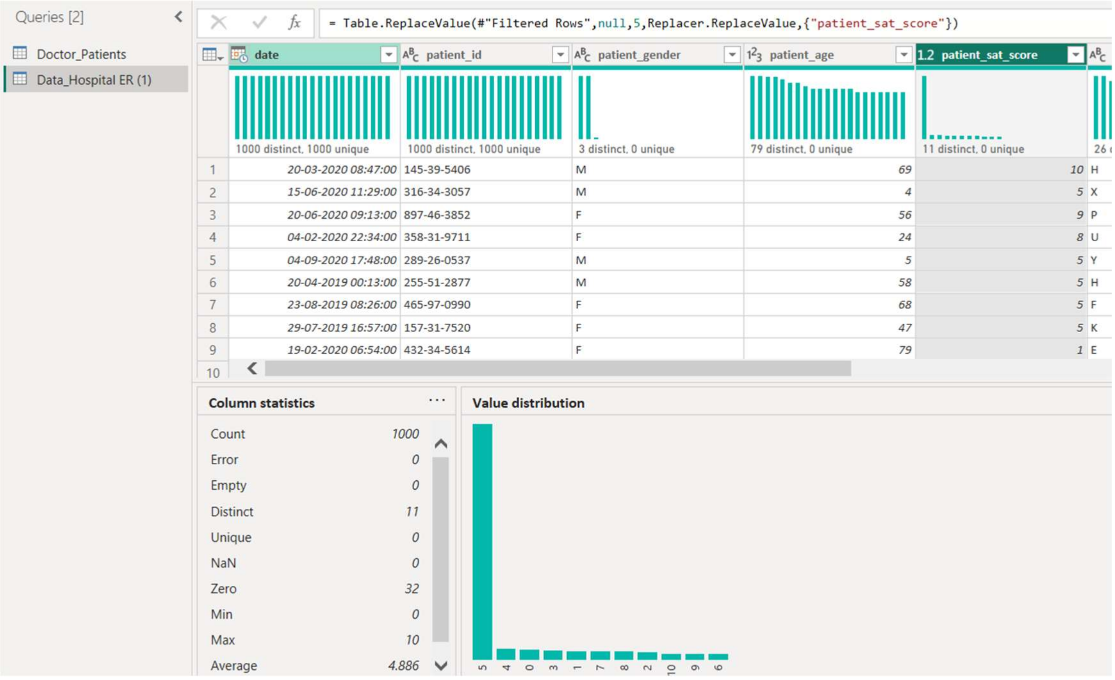
Column statistics

Count	1000
Error	0
Empty	716
Distinct	12
Unique	0
NaN	0
Zero	32
Min	0
Max	10
Average	4.59859...

Value distribution

11 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

After the Transformation:



```
= Table.ReplaceValue(#"Filtered Rows",null,5,Replacer.ReplaceValue,{"patient_sat_score"})
```

	date	patient_id	patient_gender	patient_age	patient_sat_score	
	1000 distinct, 1000 unique	1000 distinct, 1000 unique	3 distinct, 0 unique	79 distinct, 0 unique	11 distinct, 0 unique	26
1	20-03-2020 08:47:00	145-39-5406	M	69	10	H
2	15-06-2020 11:29:00	316-34-3057	M	4	5	X
3	20-06-2020 09:13:00	897-46-3852	F	56	9	P
4	04-02-2020 22:34:00	358-31-9711	F	24	8	U
5	04-09-2020 17:48:00	289-26-0537	M	5	5	Y
6	20-04-2019 00:13:00	255-51-2877	M	58	5	H
7	23-08-2019 08:26:00	465-97-0990	F	68	5	F
8	29-07-2019 16:57:00	157-31-7520	F	47	5	K
9	19-02-2020 06:54:00	432-34-5614	F	79	1	E
10						

Column statistics

Count	1000
Error	0
Empty	0
Distinct	11
Unique	0
NaN	0
Zero	32
Min	0
Max	10
Average	4.886

---

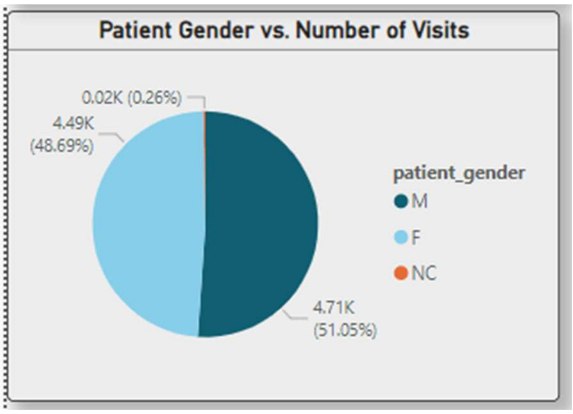
6. Is there any relation between the number of visits and the Gender of the patients?

Solution:

The following graph reveals the patient's gender distribution:

- Male: 4.71 K **(51.05%)**
- Female: 4.49 K **(48.69%)**
- NC (Not Categorized): 0.02 K **(0.26%)**

We can see a slight difference between male and female patients with male patients being higher in number compared to the female patients.



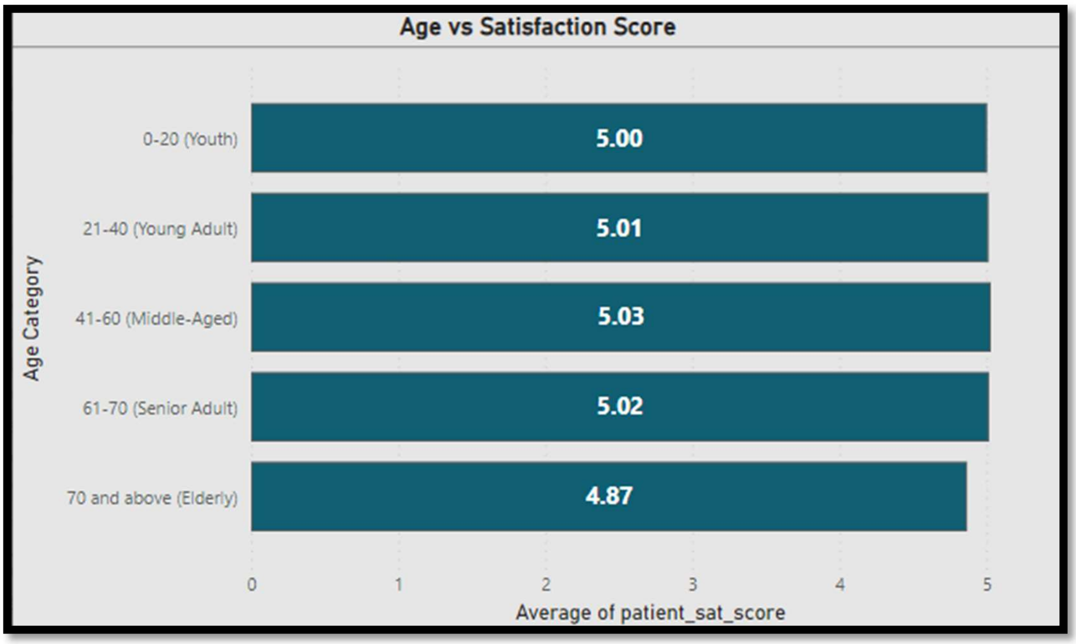
---

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.

Solution:

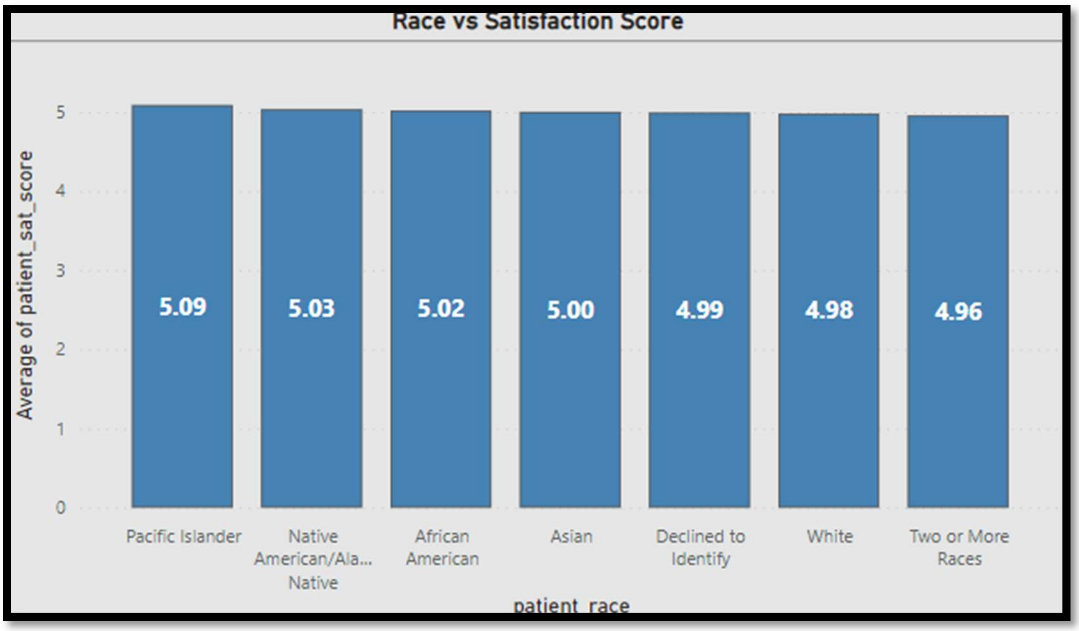
Age vs Satisfaction Score:

- The visual below shows that the satisfaction score is on an average higher in the Youth age group (age <=20) whereas the Elderly people (age >=70) have the least average satisfaction score of 4.87.
- This helps us to analyse the satisfaction score based on the age group and tailor the healthcare services accordingly.



Race vs Satisfaction Score:

- The visual below shows that the satisfaction score is higher amongst the race of Pacific Islanders at 5.09 followed by Native Americans at 5.03.
- The patients with two or more races have the least satisfaction score of 4.96 which gives us insights on ways to improvise the satisfaction score of the particular race.



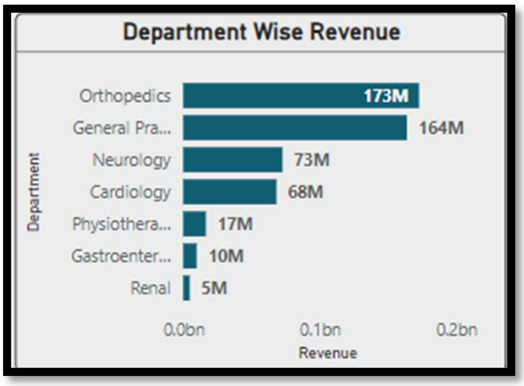
---

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

Solution:

The following visual depicts the revenue generated by each department in the hospital.

- Orthopedics has generated the highest revenue ($173 million) followed by General Practice ($ 164 million).
- The Renal department has generated the least revenue at $5 million.

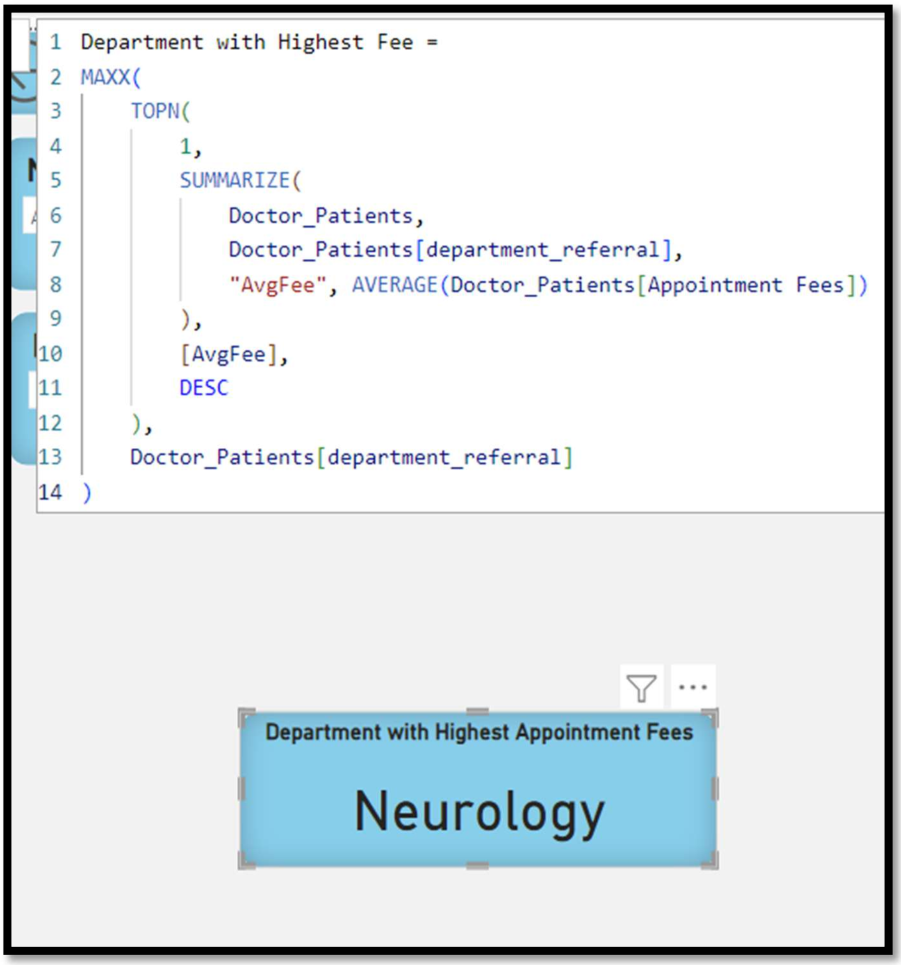


---

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

Solution:

```
Department with Highest Fee =
MAXX(
   TOPN(
      1,
      SUMMARIZE(
         Doctor_Patients,
         Doctor_Patients[department_referral],
         "AvgFee", AVERAGE(Doctor_Patients[Appointment Fees])
      ),
      [AvgFee],
      DESC
   ),
   Doctor_Patients[department_referral]
)
```



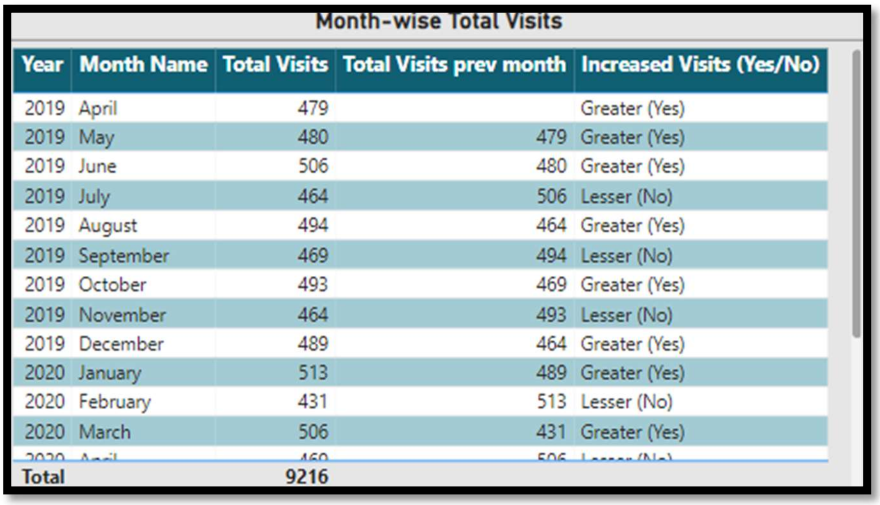
```
Department with Highest Fee =
MAXX(
    TOPN(
        1,
        SUMMARIZE(
            Doctor_Patients,
            Doctor_Patients[department_referral],
            "AvgFee", AVERAGE(Doctor_Patients[Appointment Fees])
        ),
        [AvgFee],
        DESC
    ),
    Doctor_Patients[department_referral]
)
```

Department with Highest Appointment Fees

Neurology

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.

Solution:



Month-wise Total Visits

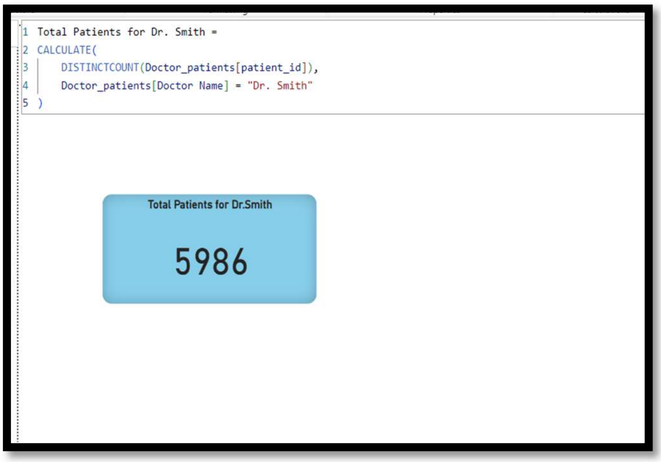
Year	Month Name	Total Visits	Total Visits prev month	Increased Visits (Yes/No)
2019	April	479		Greater (Yes)
2019	May	480	479	Greater (Yes)
2019	June	506	480	Greater (Yes)
2019	July	464	506	Lesser (No)
2019	August	494	464	Greater (Yes)
2019	September	469	494	Lesser (No)
2019	October	493	469	Greater (Yes)
2019	November	464	493	Lesser (No)
2019	December	489	464	Greater (Yes)
2020	January	513	489	Greater (Yes)
2020	February	431	513	Lesser (No)
2020	March	506	431	Greater (Yes)
2020	April	460	506	Lesser (No)
Total		9216		

---

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

Solution:

```
Total Patients for Dr. Smith =
CALCULATE(
   DISTINCTCOUNT(Doctor_patients[patient_id]),
   Doctor_patients[Doctor Name] = "Dr. Smith"
)
```



```
Total Patients for Dr. Smith =
CALCULATE(
    DISTINCTCOUNT(Doctor_patients[patient_id]),
    Doctor_patients[Doctor Name] = "Dr. Smith"
)
```

Total Patients for Dr.Smith

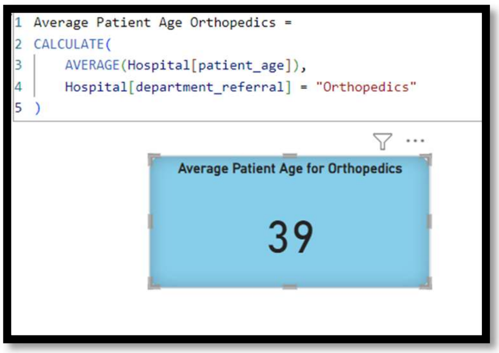
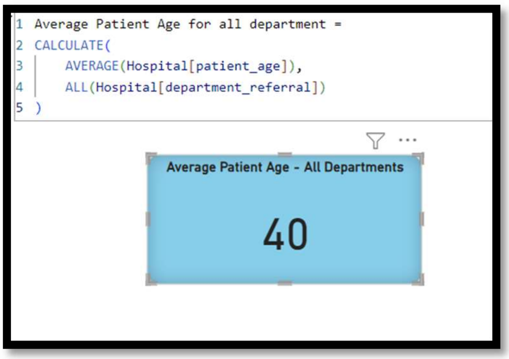
5986

---

12. 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?

Solution:

S.No	**Average Age for Orthopedics**	**Average Age for all Departments**
1	Average Patient Age Orthopedics = CALCULATE(AVERAGE(Hospital[patient_age]), Hospital[department_referral] = "Orthopedics")	Average Patient Age for all department = CALCULATE(AVERAGE(Hospital[patient_age]), ALL(Hospital[department_referral]))

2	
	`Average Patient Age Orthopedics =`
	`CALCULATE(`
	`AVERAGE(Hospital[patient_age]),`
	`Hospital[department_referral] = "Orthopedics"`
	`)`
	
		
		`Average Patient Age for all department =`
		`CALCULATE(`
		`AVERAGE(Hospital[patient_age]),`
		`ALL(Hospital[department_referral])`
		`)`

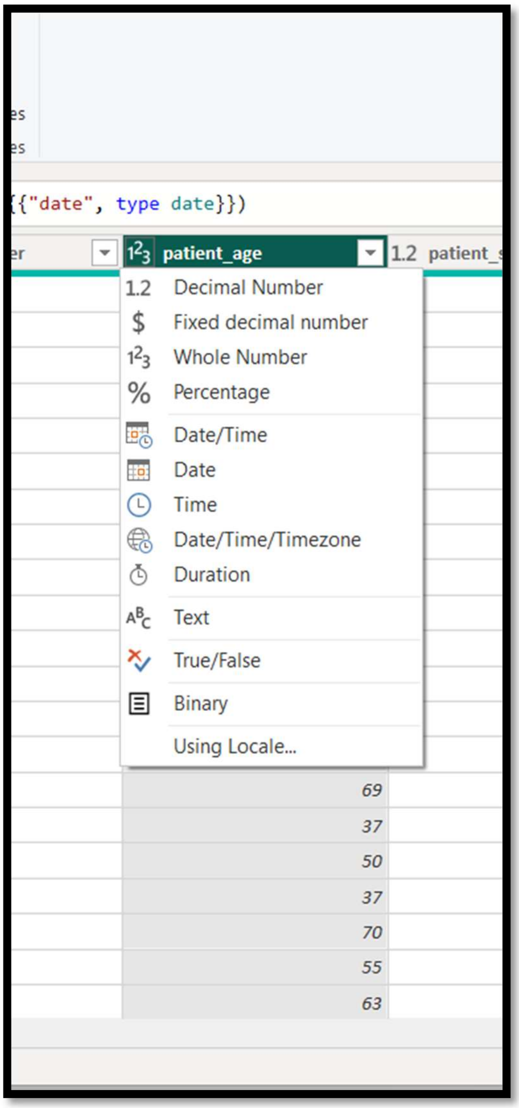
----

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

Solution:

Firstly, check the datatypes of the attributes in the given dataset. If they are not in the proper format, then do the following:

1. Go to Power Query Editor.
2. Select the column/attribute whose datatype is not in the proper format.
3. On the left side of the column header, there would be a symbol which specifies the data type. Click on it.
4. Based on the data in the column, change the datatype. For eg) if age is in text data type, change it to **Whole Number** datatype.

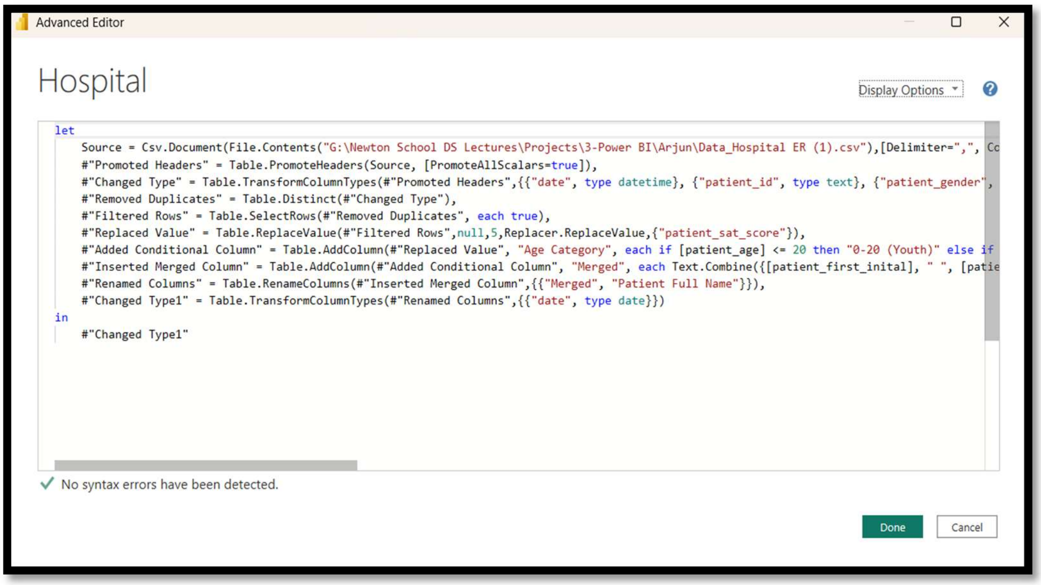


```
{{"date", type date}})
```

---

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?

Solution:



```
let
    Source = Csv.Document(File.Contents("G:\Newton School DS Lectures\Projects\3-Power BI\Arjun\Data_Hospital ER (1).csv"),[Delimiter=",", Co
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"date", type datetime}, {"patient_id", type text}, {"patient_gender",
    #"Removed Duplicates" = Table.Distinct(#"Changed Type"),
    #"Filtered Rows" = Table.SelectRows(#"Removed Duplicates", each true),
    #"Replaced Value" = Table.ReplaceValue(#"Filtered Rows",null,5,Replacer.ReplaceValue,{"patient_sat_score"}),
    #"Added Conditional Column" = Table.AddColumn(#"Replaced Value", "Age Category", each if [patient_age] <= 20 then "0-20 (Youth)" else if
    #"Inserted Merged Column" = Table.AddColumn(#"Added Conditional Column", "Merged", each Text.Combine({[patient_first_inital], " ", [patie
    #"Renamed Columns" = Table.RenameColumns(#"Inserted Merged Column",{{"Merged", "Patient Full Name"}}),
    #"Changed Type1" = Table.TransformColumnTypes(#"Renamed Columns",{{"date", type date}})
in
    #"Changed Type1"
```

#"Added Conditional Column" = Table.AddColumn(#"Replaced Value", "Age Category", each if [patient_age] <= 20 then "0-20 (Youth)" else if [patient_age] <= 40 then "21-40 (Young Adult)" else if [patient_age] <= 60 then "41-60 (Middle-Aged)" else if [patient_age] <= 70 then "61-70 (Senior Adult)" else if [patient_age] > 70 then "70 and above (Elderly)" else null)

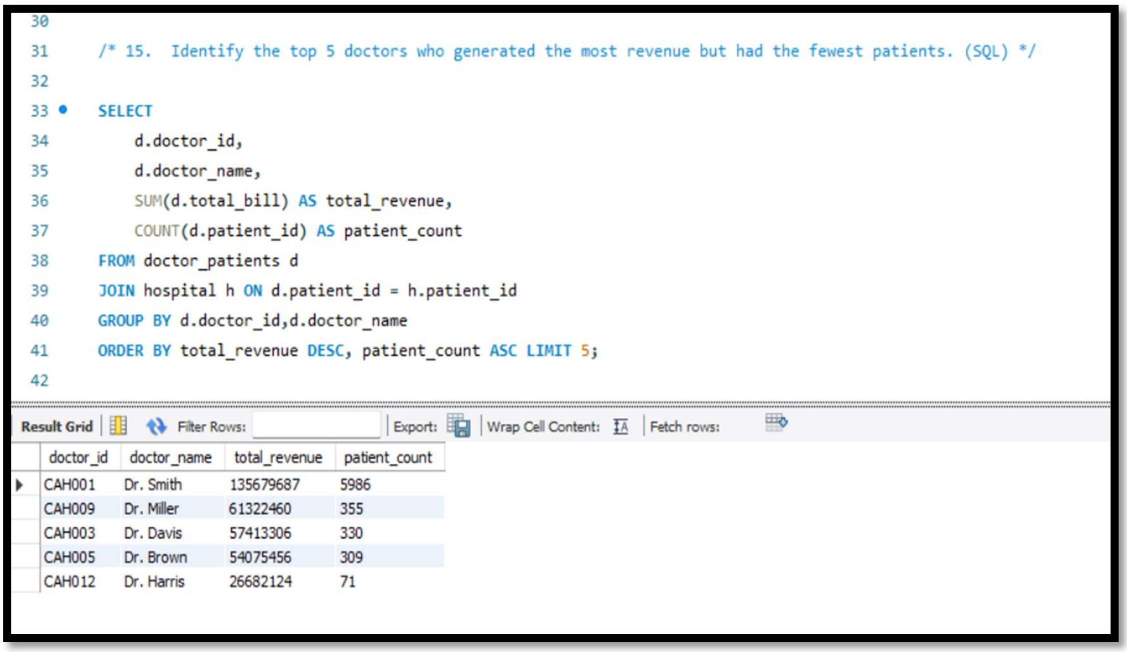
The above snippet is used to categorize age based on conditions.

----

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

Solution:

```
SELECT
        d.doctor_id,
        d.doctor_name,
        SUM(d.total_bill) AS total_revenue,
        COUNT(d.patient_id) AS patient_count
FROM doctor_patients d
JOIN hospital h ON d.patient_id = h.patient_id
GROUP BY d.doctor_id,d.doctor_name
ORDER BY total_revenue DESC, patient_count ASC LIMIT 5;
```



```
/* 15.  Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL) */

SELECT
    d.doctor_id,
    d.doctor_name,
    SUM(d.total_bill) AS total_revenue,
    COUNT(d.patient_id) AS patient_count
FROM doctor_patients d
JOIN hospital h ON d.patient_id = h.patient_id
GROUP BY d.doctor_id,d.doctor_name
ORDER BY total_revenue DESC, patient_count ASC LIMIT 5;
```

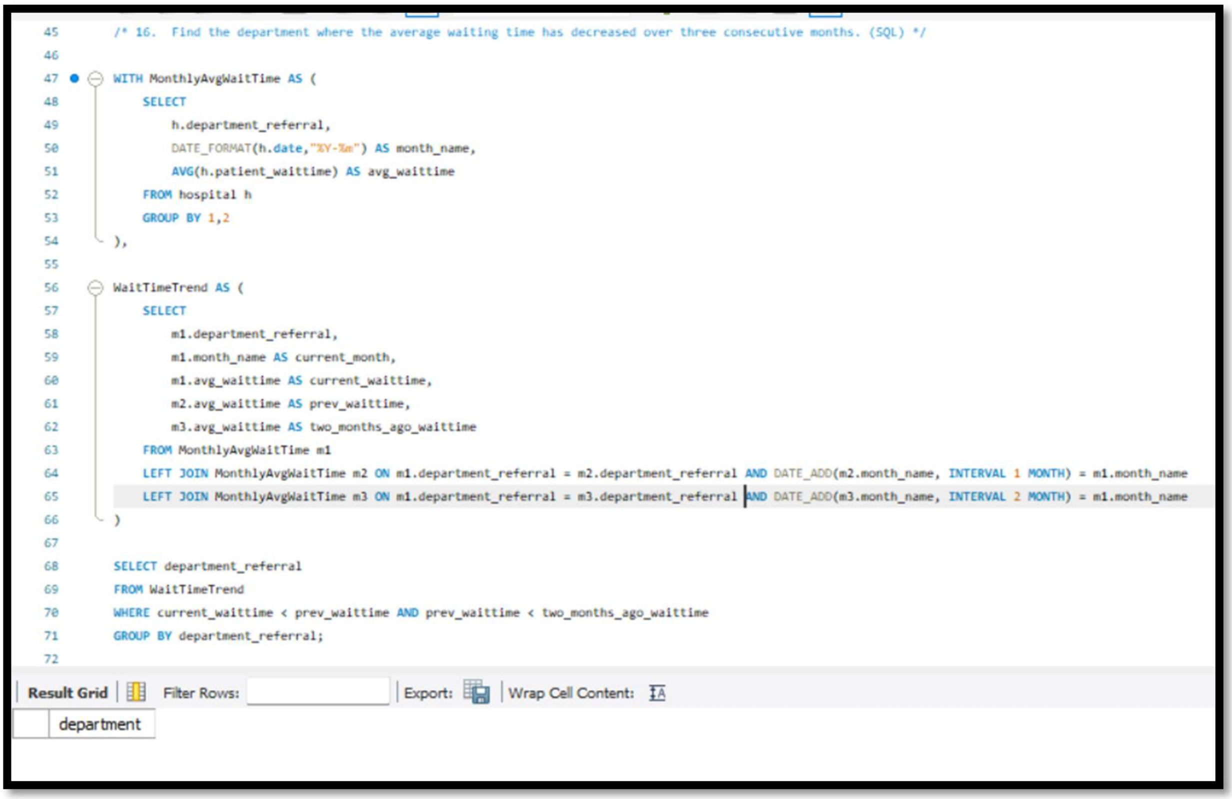
doctor_id	doctor_name	total_revenue	patient_count
CAH001	Dr. Smith	135679687	5986
CAH009	Dr. Miller	61322460	355
CAH003	Dr. Davis	57413306	330
CAH005	Dr. Brown	54075456	309
CAH012	Dr. Harris	26682124	71

----

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

Solution:

```
WITH MonthlyAvgWaitTime AS (
    SELECT
            h.department_referral,
            DATE_FORMAT(h.date,"%Y-%m") AS month_name,
            AVG(h.patient_waittime) AS avg_waittime
    FROM hospital h
    GROUP BY 1,2
),

WaitTimeTrend AS (
    SELECT
            m1.department_referral,
            m1.month_name AS current_month,
            m1.avg_waittime AS current_waittime,
            m2.avg_waittime AS prev_waittime,
            m3.avg_waittime AS two_months_ago_waittime
    FROM MonthlyAvgWaitTime m1
    LEFT JOIN MonthlyAvgWaitTime m2 ON m1.department_referral = m2.department_referral
    AND DATE_ADD(m2.month_name, INTERVAL 1 MONTH) = m1.month_name
    LEFT JOIN MonthlyAvgWaitTime m3 ON m1.department_referral = m3.department_referral
    AND DATE_ADD(m3.month_name, INTERVAL 2 MONTH) = m1.month_name
)

SELECT
    department_referral
FROM WaitTimeTrend
WHERE current_waittime < prev_waittime
AND prev_waittime < two_months_ago_waittime
GROUP BY department_referral;
```



```sql
/* 16.  Find the department where the average waiting time has decreased over three consecutive months. (SQL) */

WITH MonthlyAvgWaitTime AS (
    SELECT
        h.department_referral,
        DATE_FORMAT(h.date,"%Y-%m") AS month_name,
        AVG(h.patient_waittime) AS avg_waittime
    FROM hospital h
    GROUP BY 1,2
),

WaitTimeTrend AS (
    SELECT
        m1.department_referral,
        m1.month_name AS current_month,
        m1.avg_waittime AS current_waittime,
        m2.avg_waittime AS prev_waittime,
        m3.avg_waittime AS two_months_ago_waittime
    FROM MonthlyAvgWaitTime m1
    LEFT JOIN MonthlyAvgWaitTime m2 ON m1.department_referral = m2.department_referral AND DATE_ADD(m2.month_name, INTERVAL 1 MONTH) = m1.month_name
    LEFT JOIN MonthlyAvgWaitTime m3 ON m1.department_referral = m3.department_referral AND DATE_ADD(m3.month_name, INTERVAL 2 MONTH) = m1.month_name
)

SELECT department_referral
FROM WaitTimeTrend
WHERE current_waittime < prev_waittime AND prev_waittime < two_months_ago_waittime
GROUP BY department_referral;
```

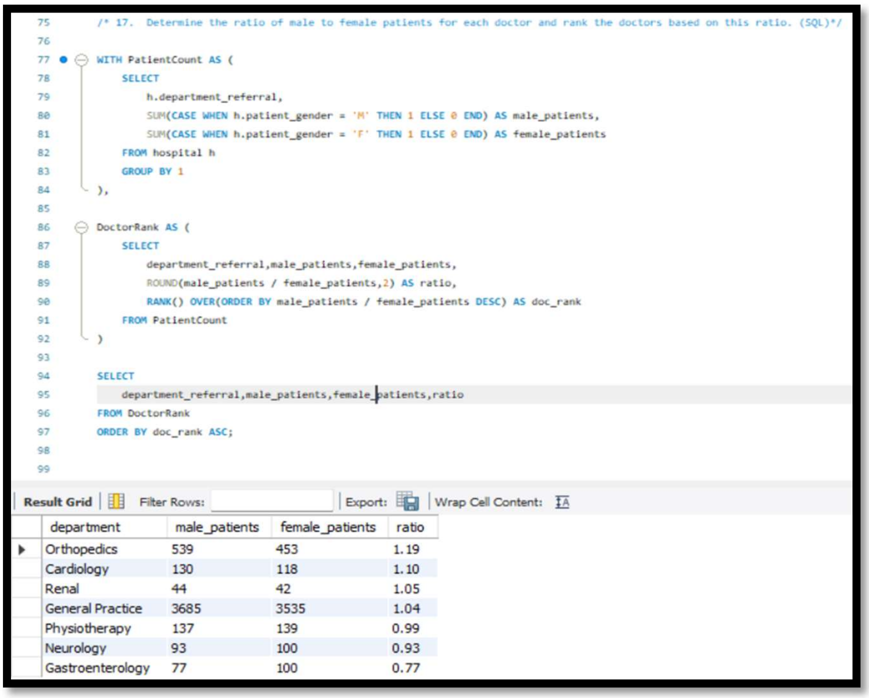
department

----

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

Solution:

```
WITH PatientCount AS (
    SELECT
        h.department_referral,
        SUM(CASE WHEN h.patient_gender = 'M' THEN 1 ELSE 0 END) AS male_patients,
        SUM(CASE WHEN h.patient_gender = 'F' THEN 1 ELSE 0 END) AS female_patients
    FROM hospital h
    GROUP BY 1
),

DoctorRank AS (
    SELECT
        department_referral,male_patients,female_patients,
        ROUND(male_patients / female_patients,2) AS ratio,
        RANK() OVER(ORDER BY male_patients / female_patients DESC) AS doc_rank
    FROM PatientCount
)

SELECT department_referral, male_patients, female_patients, ratio
FROM DoctorRank
ORDER BY doc_rank ASC;
```



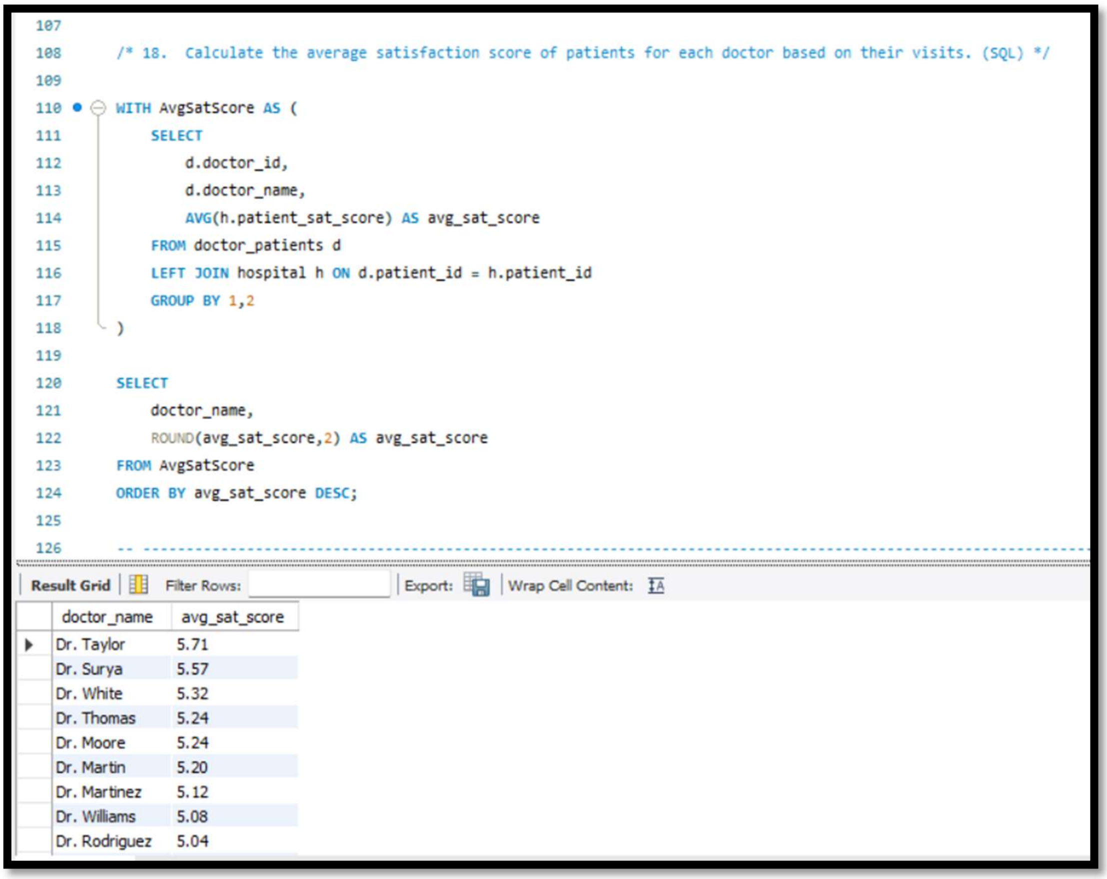
```
/* 17.  Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)*/

WITH PatientCount AS (
    SELECT
        h.department_referral,
        SUM(CASE WHEN h.patient_gender = 'M' THEN 1 ELSE 0 END) AS male_patients,
        SUM(CASE WHEN h.patient_gender = 'F' THEN 1 ELSE 0 END) AS female_patients
    FROM hospital h
    GROUP BY 1
),

DoctorRank AS (
    SELECT
        department_referral,male_patients,female_patients,
        ROUND(male_patients / female_patients,2) AS ratio,
        RANK() OVER(ORDER BY male_patients / female_patients DESC) AS doc_rank
    FROM PatientCount
)

SELECT
    department_referral,male_patients,female_patients,ratio
FROM DoctorRank
ORDER BY doc_rank ASC;
```

	department	male_patients	female_patients	ratio
▶	Orthopedics	539	453	1.19
	Cardiology	130	118	1.10
	Renal	44	42	1.05
	General Practice	3685	3535	1.04
	Physiotherapy	137	139	0.99
	Neurology	93	100	0.93
	Gastroenterology	77	100	0.77

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

Solution:

```
WITH AvgSatScore AS (
    SELECT
            d.doctor_id,
            d.doctor_name,
     AVG(h.patient_sat_score) AS avg_sat_score
  FROM doctor_patients d
    LEFT JOIN hospital h ON d.patient_id = h.patient_id
  GROUP BY 1,2
)

SELECT
    doctor_name,
  ROUND(avg_sat_score,2) AS avg_sat_score
FROM AvgSatScore
ORDER BY avg_sat_score DESC;
```



```

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

WITH AvgSatScore AS (
    SELECT
        d.doctor_id,
        d.doctor_name,
        AVG(h.patient_sat_score) AS avg_sat_score
    FROM doctor_patients d
    LEFT JOIN hospital h ON d.patient_id = h.patient_id
    GROUP BY 1,2
)

SELECT
    doctor_name,
    ROUND(avg_sat_score,2) AS avg_sat_score
FROM AvgSatScore
ORDER BY avg_sat_score DESC;

-- -----------------------------------------------------------------------------------------------------------
```

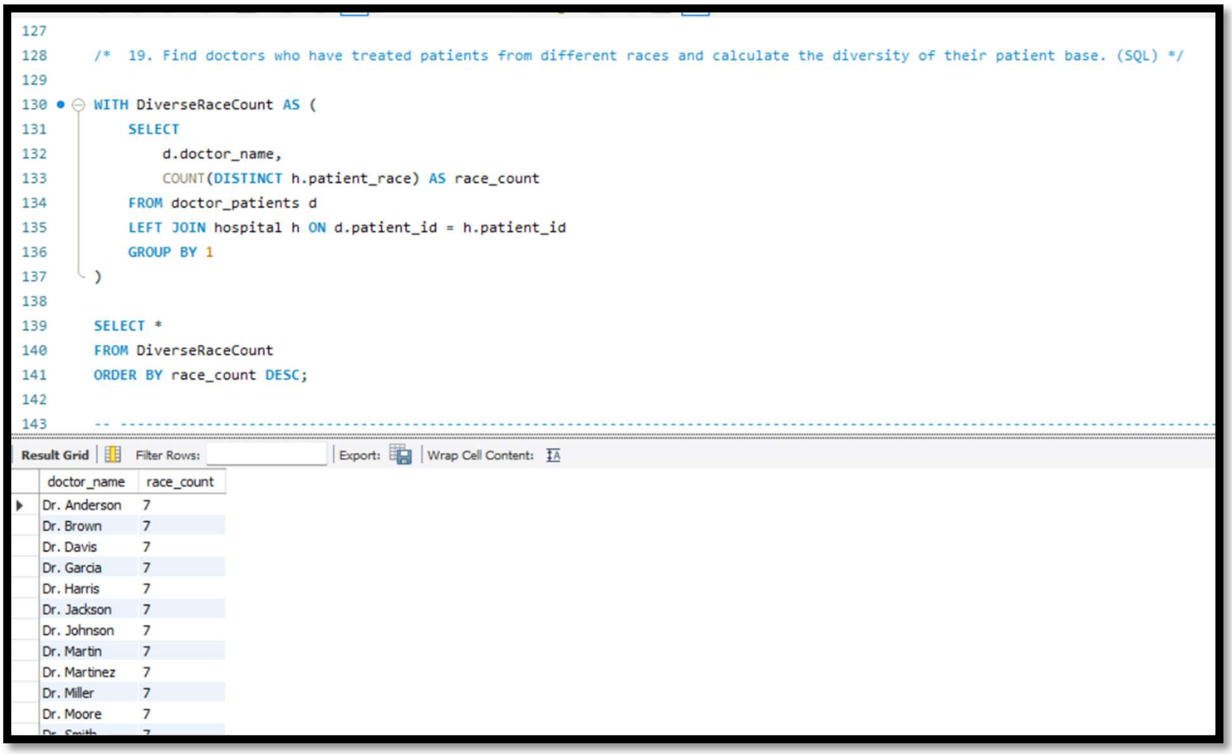
doctor_name	avg_sat_score
Dr. Taylor	5.71
Dr. Surya	5.57
Dr. White	5.32
Dr. Thomas	5.24
Dr. Moore	5.24
Dr. Martin	5.20
Dr. Martinez	5.12
Dr. Williams	5.08
Dr. Rodriguez	5.04

----

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

Solution:

```
WITH DiverseRaceCount AS (
    SELECT
        d.doctor_name,
        COUNT(DISTINCT h.patient_race) AS race_count
     FROM doctor_patients d
    LEFT JOIN hospital h ON d.patient_id = h.patient_id
  GROUP BY 1
)

SELECT *
FROM DiverseRaceCount
ORDER BY race_count DESC;
```



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

WITH DiverseRaceCount AS (
    SELECT
        d.doctor_name,
        COUNT(DISTINCT h.patient_race) AS race_count
    FROM doctor_patients d
    LEFT JOIN hospital h ON d.patient_id = h.patient_id
    GROUP BY 1
)

SELECT *
FROM DiverseRaceCount
ORDER BY race_count DESC;

-- ---------------------------------------------------------------------------------------------------------
```

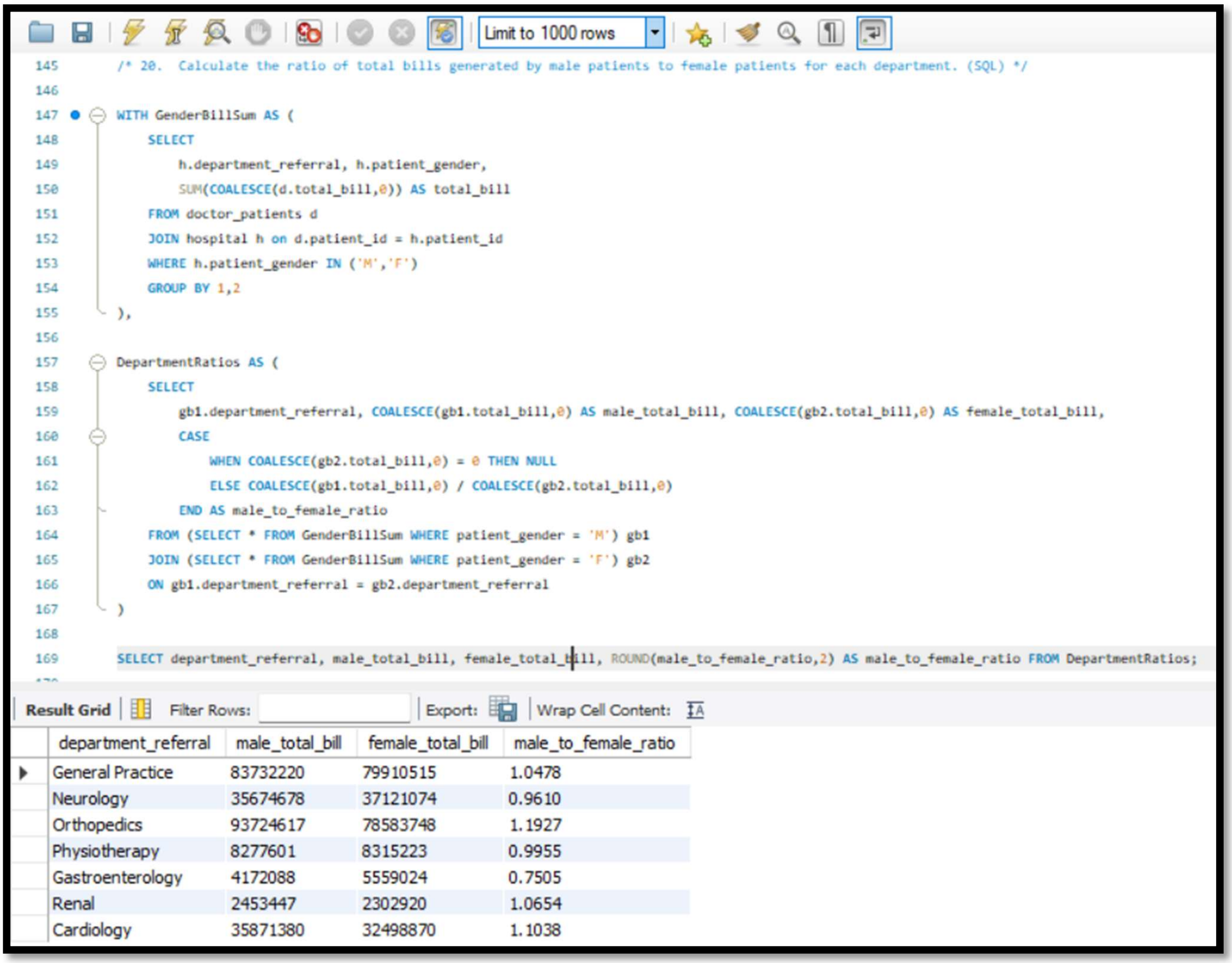
```
doctor_name   race_count
Dr. Anderson  7
Dr. Brown     7
Dr. Davis     7
Dr. Garcia    7
Dr. Harris    7
Dr. Jackson   7
Dr. Johnson   7
Dr. Martin    7
Dr. Martinez  7
Dr. Miller    7
Dr. Moore     7
Dr. Smith     7
```

----

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

Solution:

```
WITH GenderBillSum AS (
    SELECT
        h.department_referral,
        h.patient_gender,
       SUM(COALESCE(d.total_bill,0)) AS total_bill
    FROM doctor_patients d
  JOIN hospital h on d.patient_id = h.patient_id
  WHERE h.patient_gender IN ('M','F')
  GROUP BY 1,2
),

DepartmentRatios AS (
   SELECT
       gb1.department_referral,
       COALESCE(gb1.total_bill,0) AS male_total_bill,
       COALESCE(gb2.total_bill,0) AS female_total_bill,
       CASE
           WHEN COALESCE(gb2.total_bill,0) = 0 THEN NULL
           ELSE COALESCE(gb1.total_bill,0) / COALESCE(gb2.total_bill,0)
       END AS male_to_female_ratio
   FROM (SELECT * FROM GenderBillSum WHERE patient_gender = 'M') gb1
   JOIN (SELECT * FROM GenderBillSum WHERE patient_gender = 'F') gb2
   ON gb1.department_referral = gb2.department_referral
)

SELECT
    department_referral,
    male_total_bill,
    female_total_bill,
    ROUND(male_to_female_ratio,2) AS male_to_female_ratio
FROM DepartmentRatios;
```



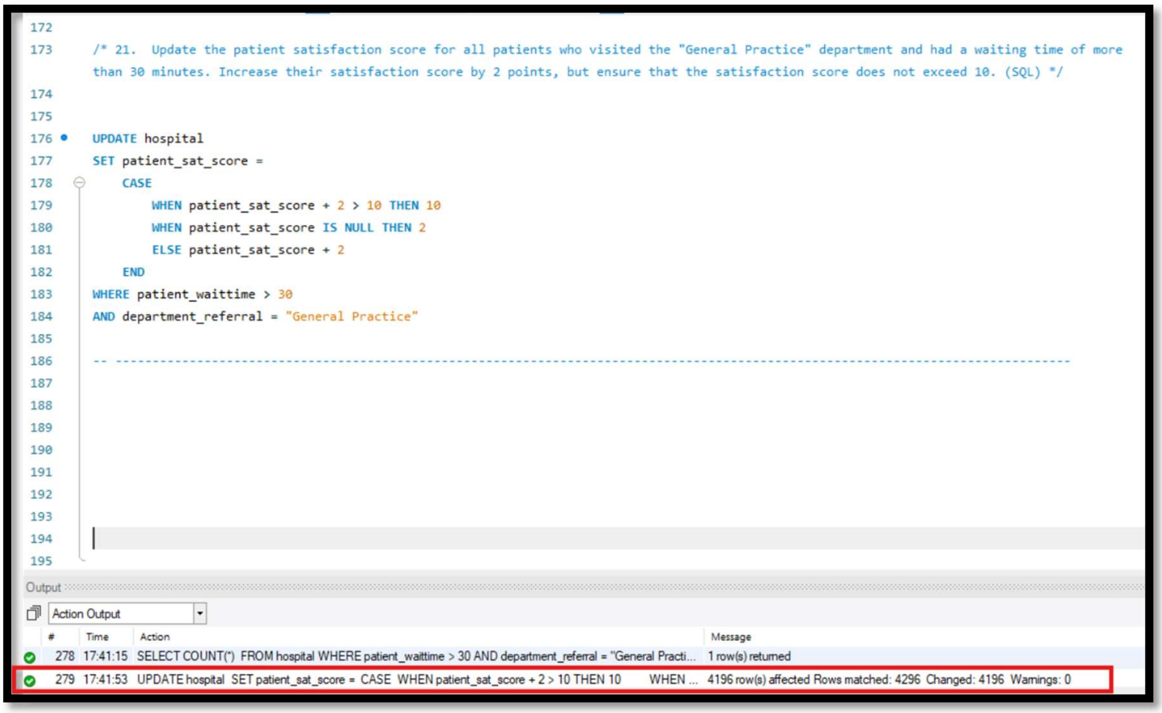
```
/* 20.  Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL) */

WITH GenderBillSum AS (
    SELECT
        h.department_referral, h.patient_gender,
        SUM(COALESCE(d.total_bill,0)) AS total_bill
    FROM doctor_patients d
    JOIN hospital h on d.patient_id = h.patient_id
    WHERE h.patient_gender IN ('M','F')
    GROUP BY 1,2
),

DepartmentRatios AS (
    SELECT
        gb1.department_referral, COALESCE(gb1.total_bill,0) AS male_total_bill, COALESCE(gb2.total_bill,0) AS female_total_bill,
        CASE
            WHEN COALESCE(gb2.total_bill,0) = 0 THEN NULL
            ELSE COALESCE(gb1.total_bill,0) / COALESCE(gb2.total_bill,0)
        END AS male_to_female_ratio
    FROM (SELECT * FROM GenderBillSum WHERE patient_gender = 'M') gb1
    JOIN (SELECT * FROM GenderBillSum WHERE patient_gender = 'F') gb2
    ON gb1.department_referral = gb2.department_referral
)

SELECT department_referral, male_total_bill, female_total_bill, ROUND(male_to_female_ratio,2) AS male_to_female_ratio FROM DepartmentRatios;
```

department_referral	male_total_bill	female_total_bill	male_to_female_ratio
General Practice	83732220	79910515	1.0478
Neurology	35674678	37121074	0.9610
Orthopedics	93724617	78583748	1.1927
Physiotherapy	8277601	8315223	0.9955
Gastroenterology	4172088	5559024	0.7505
Renal	2453447	2302920	1.0654
Cardiology	35871380	32498870	1.1038

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)

Solution:

```
UPDATE hospital
SET patient_sat_score =
    CASE
        WHEN patient_sat_score + 2 > 10 THEN 10
        WHEN patient_sat_score IS NULL THEN 2
        ELSE patient_sat_score + 2
    END
WHERE patient_waittime > 30
AND department_referral = "General Practice";
```



```
/* 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) */


UPDATE hospital
SET patient_sat_score =
    CASE
        WHEN patient_sat_score + 2 > 10 THEN 10
        WHEN patient_sat_score IS NULL THEN 2
        ELSE patient_sat_score + 2
    END
WHERE patient_waittime > 30
AND department_referral = "General Practice"

-- ----------------------------------------------------------------------------------------------------------------------------
```

```
#    Time      Action                                                                                     Message
278  17:41:15  SELECT COUNT(*)  FROM hospital WHERE patient_waittime > 30 AND department_referral = "General Practi...  1 row(s) returned
279  17:41:53  UPDATE hospital  SET patient_sat_score =  CASE  WHEN patient_sat_score + 2 > 10 THEN 10      WHEN ...  4196 row(s) affected Rows matched: 4296  Changed: 4196  Warnings: 0
```

Subjective Questions

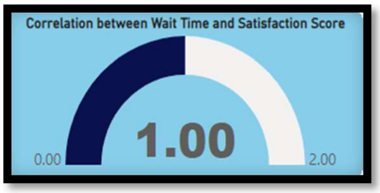
1. What is the relation between patient wait time and satisfaction scores?

Solution:

To find the correlation between wait time and satisfaction scores, we make use of a gauge chart to display the correlation co-efficient.

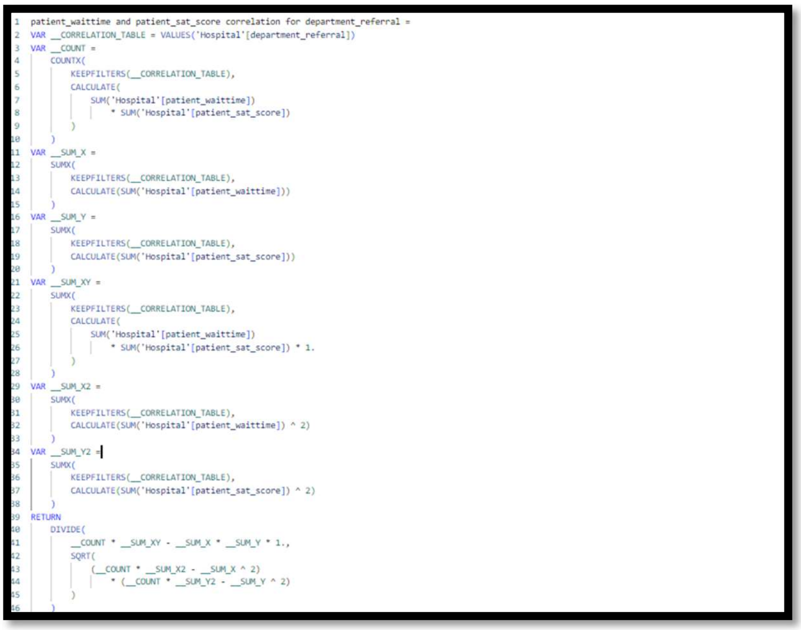
To get this correlation co-efficient parameter, we make use of Quick Measure and assign the following values:

- Category **– department_referral**
- Measure **X – patient_waittime**
- Measure **Y – patient_sat_score**



Then, click on a gauge chart visual and add this quick measure under value in the visualization tab.

- We can see that the correlation co-efficient is 1 which implies that the wait time of patients does have an impact on the satisfaction scores of the patients.
- Departments with lower waiting times have a higher satisfaction score but other factors such as time management and overall service quality might vary based on the patients point of view.



```
patient_waittime and patient_sat_score correlation for department_referral =
VAR __CORRELATION_TABLE = VALUES('Hospital'[department_referral])
VAR __COUNT =
    COUNTX(
        KEEPFILTERS(__CORRELATION_TABLE),
        CALCULATE(
            SUM('Hospital'[patient_waittime])
                * SUM('Hospital'[patient_sat_score])
        )
    )
VAR __SUM_X =
    SUMX(
        KEEPFILTERS(__CORRELATION_TABLE),
        CALCULATE(SUM('Hospital'[patient_waittime]))
    )
VAR __SUM_Y =
    SUMX(
        KEEPFILTERS(__CORRELATION_TABLE),
        CALCULATE(SUM('Hospital'[patient_sat_score]))
    )
VAR __SUM_XY =
    SUMX(
        KEEPFILTERS(__CORRELATION_TABLE),
        CALCULATE(
            SUM('Hospital'[patient_waittime])
                * SUM('Hospital'[patient_sat_score]) * 1.
        )
    )
VAR __SUM_X2 =
    SUMX(
        KEEPFILTERS(__CORRELATION_TABLE),
        CALCULATE(SUM('Hospital'[patient_waittime]) ^ 2)
    )
VAR __SUM_Y2 =
    SUMX(
        KEEPFILTERS(__CORRELATION_TABLE),
        CALCULATE(SUM('Hospital'[patient_sat_score]) ^ 2)
    )
RETURN
    DIVIDE(
        __COUNT * __SUM_XY - __SUM_X * __SUM_Y * 1.,
        SQRT(
            (__COUNT * __SUM_X2 - __SUM_X ^ 2)
                * (__COUNT * __SUM_Y2 - __SUM_Y ^ 2)
        )
    )
```

----

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

Solution:

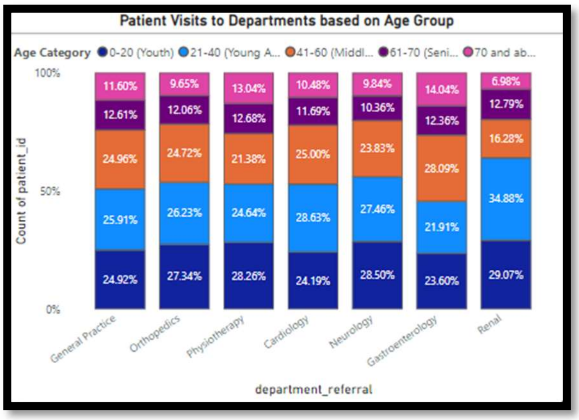
To find out how patient demographics affect the frequency of visits across departments, we make use of **100% stacked column chart** and consider the demographics such as:

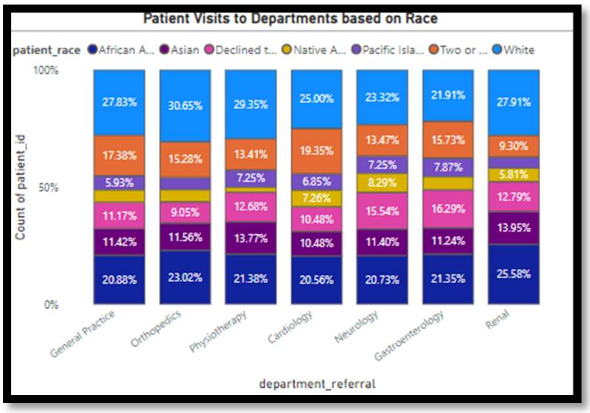
- Age Group
- Gender
- Race

To visualize all the demographics in the same visual, we make use of Field Parameters in the Modelling ribbon.

- Go to Modelling -> Parameters -> New Parameters -> Fields
- A dialog box appears. Here, we have to select the required columns for our analysis which is **Age Category**, **patient_gender** and **patient_race.**
- Then, click Ok. We can see that a table named **Age Group-Gender-Race** has been created.

The Field Parameters allows us to switch between different columns all under the same visualization. In this case, we can view the impact of patients demographics (Age Group, Gender and Race) with regards to the visits across departments.





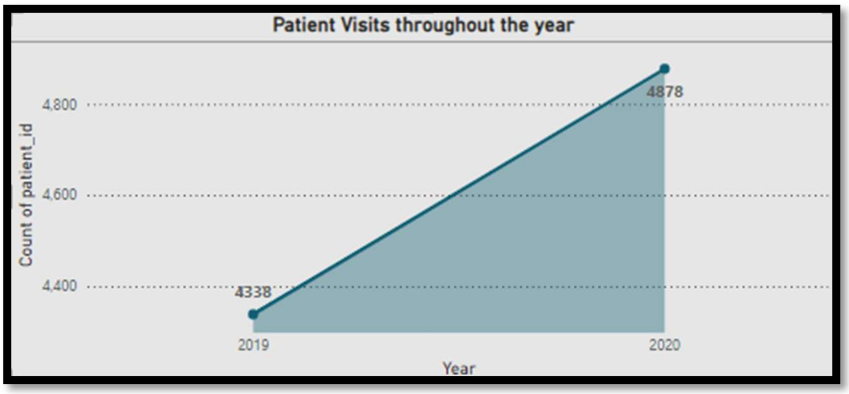
---

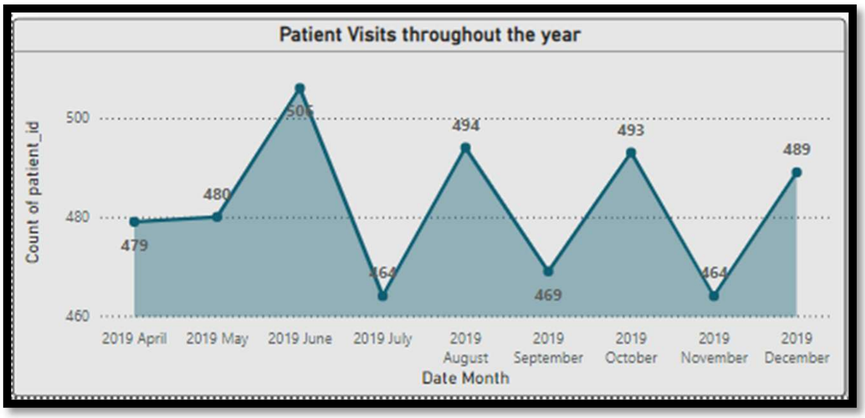
3. Is there a noticeable trend in the volume of patient visits throughout the year?

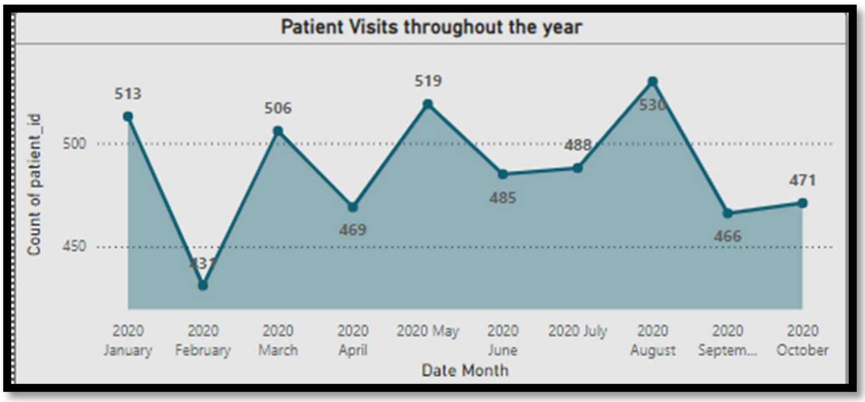
Solution:

We can observe that there is a 7% increase in the number of patients in the year 2020 (**4878**) when compared to 2019 (**4338**).

- In comparison of YoY, we can see that there is an increase in the number of patients in the months of May and August.
  - May 2019 – **480**, August 2019 - **494**
  - May 2020 – **519**, August 2020 - **530**





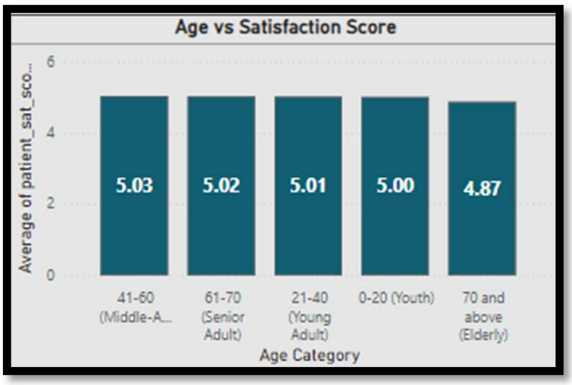


----

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

Solution:

- The age group of Middle-aged adults (41-60) had the **highest** satisfaction score of **5.03** on average
- The elderly age group (70 and above) had the **least** satisfaction score of **4.87** on average.



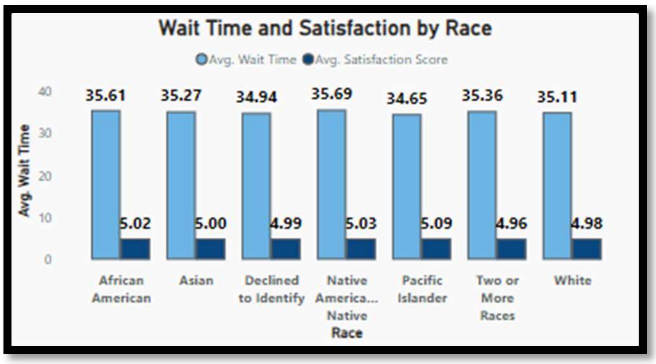
----

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?

Solution:

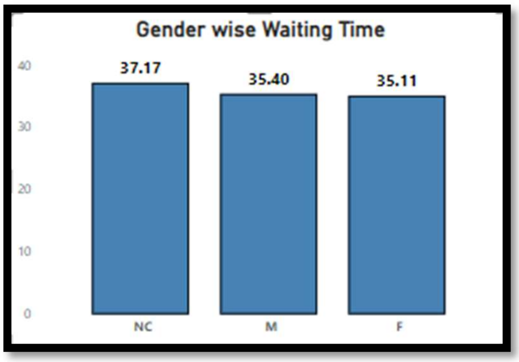
On the basis of race,

- Average waiting time for a patient ranges from **34 - 36** minutes.
- Satisfaction scores are consistent, ranging from **4.96 - 5.03**.



On the basis of gender,

- The waiting times for gender usually ranges from **35 – 37** mins.



Based on both the visuals, we can conclude that there is no discrimination on the basis of gender or race at the hospital.

----

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

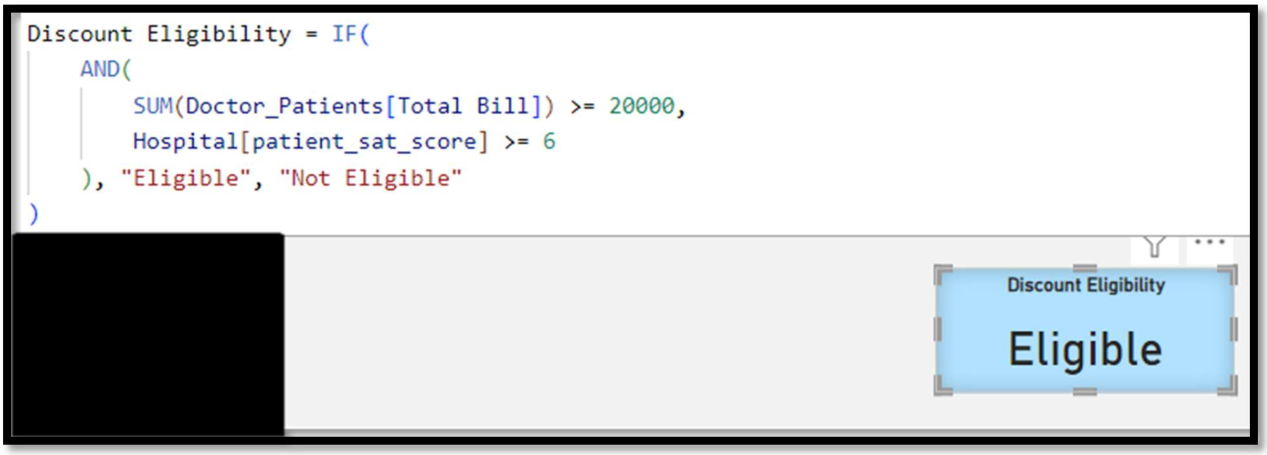
Solution:

```
Discount Eligibility = IF(
        AND(
                SUM(Doctor_Patients[Total Bill]) >= 20000,
                Hospital[patient_sat_score] >= 6
        ), "Eligible", "Not Eligible"
)
```

The discounts can be offered to patients who satisfy the following conditions:

- Total bill amount should be >= $20,000.
- Satisfaction score should be >= 6

The patients satisfying the above conditions will be eligible for discounts.



```
Discount Eligibility = IF(
    AND(
        SUM(Doctor_Patients[Total Bill]) >= 20000,
        Hospital[patient_sat_score] >= 6
    ), "Eligible", "Not Eligible"
)
```

[redacted]

Discount Eligibility

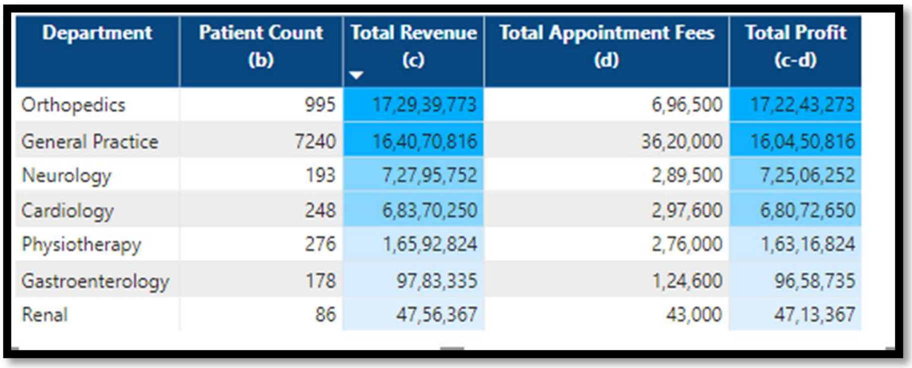
Eligible

---

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

Solution:

- Based on the visualization below, we can see that Orthopedics and General Practice departments have the highest revenue as well as the highest profit.
- Relatively, Neurology and Cardiology departments also show a promising revenue in spite of the low patient count.
- In conclusion, the new hires can be hired for Orthopedics, General Practice and (Neurology/Cardiology).



Department	Patient Count (b)	Total Revenue (c)	Total Appointment Fees (d)	Total Profit (c-d)
Orthopedics	995	17,29,39,773	6,96,500	17,22,43,273
General Practice	7240	16,40,70,816	36,20,000	16,04,50,816
Neurology	193	7,27,95,752	2,89,500	7,25,06,252
Cardiology	248	6,83,70,250	2,97,600	6,80,72,650
Physiotherapy	276	1,65,92,824	2,76,000	1,63,16,824
Gastroenterology	178	97,83,335	1,24,600	96,58,735
Renal	86	47,56,367	43,000	47,13,367

----

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

Solution:

To determine if the hospital is profitable or not, we make use of the following metrics for analysis such as:

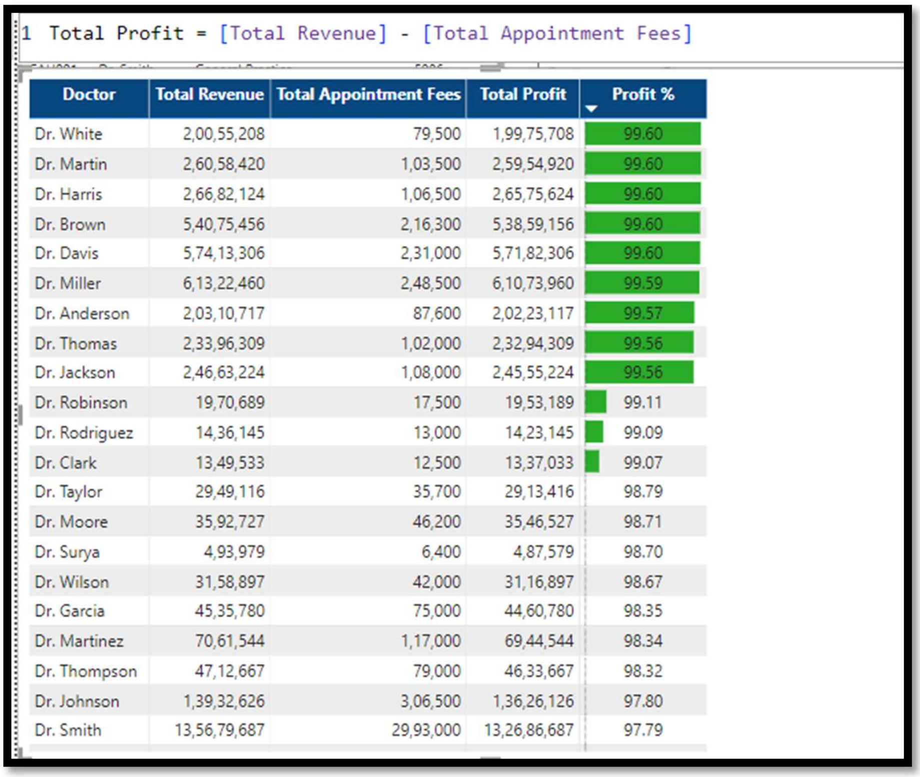
- Total Revenue (Total Bills)
- Total Appointment Fees
- Total Profit = Total Revenue – Total Appointment Fees
- Profit % = (Total Profit / Total Revenue) * 100

We make use of DAX to create the above metrics using Measures as follows:

- Total Revenue = SUM(Doctor_Patients[Total Bill])
- Total Appointment Fees = SUM(Doctor_Patients[Appointment Fees])
- Total Profit = [Total Revenue] - [Total Appointment Fees]
- Profit % = ([Total Profit] / [Total Revenue]) * 100

From the visual below, we can see that the profitability percent ranges from 97.79% to 99.6% which shows that the profitability is consistent.

The overall profit % is above 95% with which we can clearly say that the hospital is profitable with respect to the generated profit and revenue.



```
Total Profit = [Total Revenue] - [Total Appointment Fees]
```

Doctor	Total Revenue	Total Appointment Fees	Total Profit	Profit %
Dr. White	2,00,55,208	79,500	1,99,75,708	99.60
Dr. Martin	2,60,58,420	1,03,500	2,59,54,920	99.60
Dr. Harris	2,66,82,124	1,06,500	2,65,75,624	99.60
Dr. Brown	5,40,75,456	2,16,300	5,38,59,156	99.60
Dr. Davis	5,74,13,306	2,31,000	5,71,82,306	99.60
Dr. Miller	6,13,22,460	2,48,500	6,10,73,960	99.59
Dr. Anderson	2,03,10,717	87,600	2,02,23,117	99.57
Dr. Thomas	2,33,96,309	1,02,000	2,32,94,309	99.56
Dr. Jackson	2,46,63,224	1,08,000	2,45,55,224	99.56
Dr. Robinson	19,70,689	17,500	19,53,189	99.11
Dr. Rodriguez	14,36,145	13,000	14,23,145	99.09
Dr. Clark	13,49,533	12,500	13,37,033	99.07
Dr. Taylor	29,49,116	35,700	29,13,416	98.79
Dr. Moore	35,92,727	46,200	35,46,527	98.71
Dr. Surya	4,93,979	6,400	4,87,579	98.70
Dr. Wilson	31,58,897	42,000	31,16,897	98.67
Dr. Garcia	45,35,780	75,000	44,60,780	98.35
Dr. Martinez	70,61,544	1,17,000	69,44,544	98.34
Dr. Thompson	47,12,667	79,000	46,33,667	98.32
Dr. Johnson	1,39,32,626	3,06,500	1,36,26,126	97.80
Dr. Smith	13,56,79,687	29,93,000	13,26,86,687	97.79

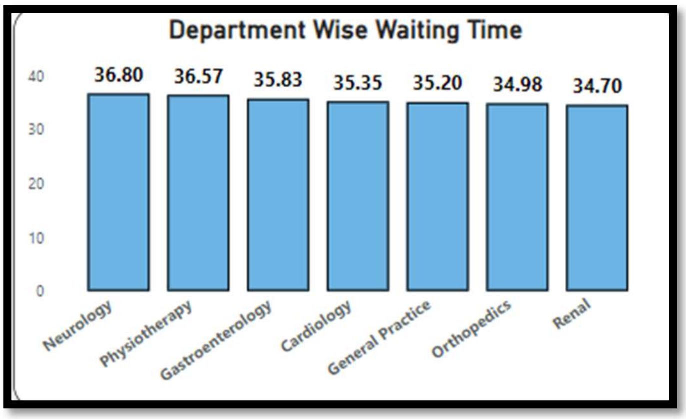
----

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

Solution:

To find if there is any oddly large waiting time among the departments, we make use of the column chart which has department and waiting time as parameters.

- We can observe that the waiting time lies in the range of 34 – 36 across all the departments.
- Neurology has the highest waiting time with 36.8 minutes while the department with the least waiting time is Renal with 34.7 minutes.
- The difference in highest and the least waiting time is approximately 2 minutes which is not much of a difference.
- Hence, we can conclude that there is no oddly large waiting time across the departments.



----

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

Solution:

The discounts could be offered to patients based on criteria such as total bill paid, satisfaction score etc.

- Provide discount to patients with bills greater than 15,000 and satisfaction score greater than or equal to 6, thus promoting customer loyalty.
- Referral discounts to the patients who refer their family / friends through word of mouth.
- Patients can be categorized based on criteria such as people who spend more than 25,000 can be offered 5% discount, 10% discount to the people whose bills exceed 50,000 etc.
- Tier Discounts can be applied to the patients based on the age groups. Like 7% on ages 70+, 5-6% for people with ages 51-70 etc.
- Patients with oddly large waiting times can be provided with some concession in bills due to the inconvenience caused as this can improve their experience to an extent.
- Seasonal or promotional discounts when there is low patient count to attract more patients and provide utmost care.

----

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?

Solution:

The following DAX formula is used to assign the shift to the doctors across all departments. With reference to the question, we have filtered out the doctors belonging to the **General Practice** department.

DAX Formula:

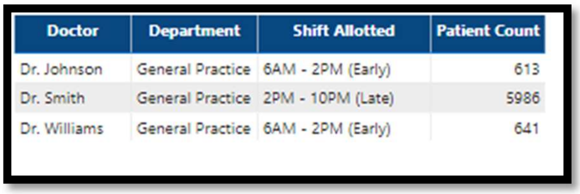
```
Shift Allotted =
    SWITCH(
            TRUE(),
            MOD(RANKX(ALL('Doctor_Patients'), Doctor_Patients[Doctor Name], , ASC), 2) = 0,
            "6AM - 2PM (Early)", "2PM - 10PM (Late)"
    )
```

- The above formula ensures that doctors are alternately assigned to one of the two shifts based on their rank in the sorted list.

- As we can see from the visual below, Dr. Johnson and Dr. Williams have been assigned with the **Early shift** (6 AM – 2 PM) since the patient count for the doctors is relatively low.

- Dr.Smith has been assigned with the **Late Shift** (2 PM – 10 PM) since the doctor has been visited frequently by the patients than the remaining 2 doctors combined.

- The shift allocation ensures that the workload is distributed evenly across the doctors, by improving satisfaction scores and reducing wait times.



Doctor	Department	Shift Allotted	Patient Count
Dr. Johnson	General Practice	6AM - 2PM (Early)	613
Dr. Smith	General Practice	2PM - 10PM (Late)	5986
Dr. Williams	General Practice	6AM - 2PM (Early)	641

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

Solution:

A Power BI Gateway is a bridge that facilitates secure data transfer between on-premises data sources (such as SQL Server, Excel, or Oracle) and Power BI service (cloud). It ensures data stays secure while being accessed, refreshed, or updated in the Power BI cloud reports and dashboards.

Types of Power BI Gateways

1. **Personal Mode Gateway**: Designed for individual users, ideal for small-scale projects, supporting only single-user access and scheduled refresh without direct query.
2. **Standard (Enterprise) Gateway**: Supports live connections, scheduled refreshes, and multiple data source connections, making it ideal for organizations requiring centralized data access management.

Use Cases

1. **Scheduled Data Refresh**: Automatically updates dashboards and reports with the latest data from on-premises systems.

2. **Live Connections**: Enables real-time data access from on-premises sources for analytics and dashboards.

3. **Hybrid Environments**: Seamlessly connects cloud and on-premises data systems for businesses using both.

4. **Secure and Centralized Management**: Ensures encrypted data transfer and centralized control of on-premises data access.

5. **Integration and Scalability**: Supports multiple data sources, Power Platform tools, and enterprise setups with shared access.

---

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

Solution:

- **Problem Statement:** Identify the real-time problems such as new hires, shift allocation for doctors, if the hospital is profitable or not and identify the stakeholders like the hospital management, Chief of Staff, Patient Care Chief etc.

- **Data Extraction:** Importing data into Power BI to begin the analysis of data.

- **Transformation of data:** Data cleaning and optimization to make sure that the data is well prepared before starting the analysis.

- **Data Loading:** Load the data into Power BI and start the analysis.

- **Identify Key Metrics:** Identify key metrics required for the dashboard / report where report is a collection of dashboards spanning across different pages.

- **Visualization:** To design the dashboard with different types of visuals such as Line Chart, Bar Chart, Pie Chart, Column Chart, Gauge Chart etc.

- **Dashboard Flexibility:** Usage of slicers for key filters (e.g., Date, Doctor Name, Patient Name) and drill-through functionality for deeper insights.

- **Validate with Stakeholders:** Present the draft dashboard to stakeholders to verify if it meets their expectations and gather feedback for refining out the layout, metrics, and interactivity.

---

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

Solution:

- The relationship between doctor_id and department is **one-to-many** since a doctor can be assigned to one department
- The relationship between department and doctors is **many-to-one** since many doctors can belong to one department.

Report

The hospital has asked for a report with three tabs:

- Main Tab
- Doctors' Tab
- Patients' Tab

❖ **Using the Main tab in the report,** the hospital should be able to look at the overall metrics like the number of daily visits, revenue produced on that day, customer satisfaction, how busy are different departments on that day, and general waiting time on that day. This tab should have a slicer of date.

❖ **Using the Doctors' Tab,** the Chief Of Staff at the hospital should be able to look at the individual doctor's performance metrics like customer satisfaction, the number of patients he was visited by, the revenue he has generated, and his appointment fees. This tab should have a slicer of the Doctor's Name or ID.

❖ **Using the Patients' Tab,** the Patient's Care Chief at the hospital wants to look at a customer's profile which would involve metrics like the most frequently visited department, their age, their race, their waiting time, number of visits, the total amount that they have paid to the hospital, etc. All the metrics using which they can address the patient very carefully in their visits. This tab should have a slicer of the Patient's Name or ID.

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

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