Data Analytics in Hospital Emergency Department

Done by: Ben Tan, 08 May 2024

Contents

1. Problem Statement

With the data that we have, I imagined myself to be a manager in the hospital, and wondered what kind of questions might I have:

1. Who are our patients?
2. How satisfied are our patients now?
3. Are there any bottlenecks in the process
4. Do we have enough resources to accommodate for the volume of patients?
5. Are our staffs lacking in any competency that we can work on
6. Is the safety of our patients fair and accounted for
7. When are the peak periods through the year/week/day

2. Data Dictionary

We have obtained data from an emergency department from a fictional hospital:

- Date - YYYY-MM-DD HH:MM:SS
- Patient_ID – identifier with pattern xxx-xx-xxxx
- Patient_gender
- Patient_age
- Patient_sat_score – satisfaction score from patients on their stay at the Emergency Department
- Patient_first_initial
- Patient_last_name
- Patient_race
- Patient_admin_flag – Patients are flagged for some reason, could be special needs, critical condition, safety concerns, legal issues, or requires follow up care
- Patient_waittime – Assumed to be in minutes
- Department_referral

3. Understanding the Data

1. Seeing what each column's data look like

```
SELECT *
FROM hospital_er
LIMIT 10;
```

date	patient_id	patient_gender	patient_age	patient_sat_score	patient_first_inital	patient_last_name	patient_race	patient_admin_flag	patient_waittime	department_referral
2020-03-20 08:47:00	145-39-5406	M	69	10	H	Glasspool	White	FALSE	39	None
2020-06-15 11:29:00	316-34-3057	M	4		X	Methuen	Native American/Alaska Native	TRUE	27	None
2020-06-20 09:13:00	897-46-3852	F	56	9	P	Schubuser	African American	TRUE	55	General Practice
2020-02-04 22:34:00	358-31-9711	F	24	8	U	Titcombe	Native American/Alaska Native	TRUE	31	General Practice
2020-09-04 17:48:00	289-26-0537	M	5		Y	Gionettitti	African American	FALSE	10	Orthopedics
2019-04-20 00:13:00	255-51-2877	M	58		H	Buff	Asian	FALSE	59	None
2019-08-23 08:26:00	465-97-0990	F	68		F	Perrat	White	TRUE	43	None
2019-07-29 16:57:00	157-31-7520	F	47		K	Gwillim	Two or More Races	TRUE	23	None
2020-02-19 06:54:00	432-34-5614	F	79	1	E	Dewhirst	White	FALSE	42	None
2020-10-11 05:25:00	609-17-8678	M	62		M	Crebo	African American	FALSE	51	None

2. Column Datatype

```
-- Checking the datatypes of each column
SELECT
    COLUMN_NAME AS Column_name,
    DATA_TYPE AS Data_type
FROM INFORMATION_SCHEMA.COLUMNS
WHERE table_name = 'hospital_er';
```

Column_name	Data_type
date	datetime
patient_id	text
patient_gender	text
patient_age	int
patient_sat_score	text
patient_first_inital	text
patient_last_name	text
patient_race	text
patient_admin_flag	text
patient_waittime	int
department_referral	text

3. Checking the number of rows and columns in our data

```
15      -- Dimension of our data
16  •   WITH `rows` AS (
17          SELECT COUNT(*) AS num_rows
18          FROM hospital_er
19      ),
20      cols AS (
21          SELECT COUNT(*) AS num_cols
22          FROM information_schema.columns
23          WHERE table_name = 'hospital_er'
24      )
25      SELECT * FROM `rows`, cols;
```

	num_rows	num_cols
▶	9216	11

4. Number of null values in each column

```
28 •  SELECT
29        SUM(CASE WHEN date IS NULL THEN 1 ELSE 0 END) AS null_date,
30        SUM(CASE WHEN patient_id IS NULL THEN 1 ELSE 0 END) AS null_id,
31        SUM(CASE WHEN patient_gender IS NULL THEN 1 ELSE 0 END) AS null_gender,
32        SUM(CASE WHEN patient_age IS NULL THEN 1 ELSE 0 END) AS null_age,
33        SUM(CASE WHEN patient_sat_score IS NULL THEN 1 ELSE 0 END) AS null_SATscore,
34        SUM(CASE WHEN patient_first_inital IS NULL THEN 1 ELSE 0 END) AS null_initials,
35        SUM(CASE WHEN patient_last_name IS NULL THEN 1 ELSE 0 END) AS null_lastname,
36        SUM(CASE WHEN patient_race IS NULL THEN 1 ELSE 0 END) AS null_race,
37        SUM(CASE WHEN patient_admin_flag IS NULL THEN 1 ELSE 0 END) AS null_flag,
38        SUM(CASE WHEN patient_waittime IS NULL THEN 1 ELSE 0 END) AS null_waittime,
39        SUM(CASE WHEN department_referral IS NULL THEN 1 ELSE 0 END) AS null_referral
40    FROM
41        hospital_er;
```

null_date	null_id	null_gender	null_age	null_SATscore	null_initials	null_lastname	null_race	null_flag	null_waittime	null_referral
0	0	0	0	0	0	0	0	0	0	0

Comments:

This is strange, as we saw empty values in the dataset for the patient_SAT_score column. This means that those data were not missing, but simply blank data. So we count the number of empty values this time.

```
44 •  SELECT
45        SUM(CASE WHEN patient_id = '' THEN 1 ELSE 0 END) AS null_id,
46        SUM(CASE WHEN patient_gender = '' THEN 1 ELSE 0 END) AS null_gender,
47        SUM(CASE WHEN patient_age = '' THEN 1 ELSE 0 END) AS null_age,
48        SUM(CASE WHEN patient_sat_score = '' THEN 1 ELSE 0 END) AS null_SATscore,
49        SUM(CASE WHEN patient_first_inital = '' THEN 1 ELSE 0 END) AS null_initials,
50        SUM(CASE WHEN patient_last_name = '' THEN 1 ELSE 0 END) AS null_lastname,
51        SUM(CASE WHEN patient_race = '' THEN 1 ELSE 0 END) AS null_race,
52        SUM(CASE WHEN patient_admin_flag = '' THEN 1 ELSE 0 END) AS null_flag,
53        SUM(CASE WHEN patient_waittime = '' THEN 1 ELSE 0 END) AS null_waittime,
54        SUM(CASE WHEN department_referral = '' THEN 1 ELSE 0 END) AS null_referral
55    FROM
56        hospital_er;
```

null_id	null_gender	null_age	null_SATscore	null_initials	null_lastname	null_race	null_flag	null_waittime	null_referral
0	0	0	6699	0	0	0	0	0	0

```
59 •  SELECT
60        ROUND(SUM(CASE WHEN patient_sat_score = '' THEN 1 ELSE 0 END)/(SELECT COUNT(*) FROM hospital_er)*100,2) AS percent_null_SAT
61    FROM hospital_er;
```

percent_null_SAT
72.69

5. Observations:

- There are 6699 null values in patient_SAT_score, which is 73% of the data.

- Our data has 11 columns, with 9216 rows of data
- Data is in datetime format, age and waittingtime is in integer format, while the rest are all in text format.

4. Exploratory Data Analysis

1. Date Range of our Data

```
64 •    SELECT
65          MIN(date) AS start_date,
66          MAX(date) AS end_date
67      FROM hospital er;
```

	start_date	end_date
▶	2019-04-01 01:13:00	2020-10-30 23:44:00

Comments:

It seems like the dataset is only 1.5 years long, with April to October data being sampled twice.

2. How satisfied are our patients now?

```
70 •    SELECT ROUND(AVG(patient_sat_score),2) AS Avg_score
71      FROM hospital_er
72      WHERE patient_sat_score <> '';
```

	Avg_score
▶	4.99

```
75 •    SELECT
76          patient_sat_score AS Score,
77          COUNT(patient_sat_score) AS count
78      FROM hospital_er
79      WHERE patient_sat_score <> ''
80      GROUP BY patient_sat_score
81      ORDER BY LENGTH(Score), Score;
```

	Score	count
▶	0	222
	1	246
	2	204
	3	228
	4	248
	5	221
	6	231
	7	256
	8	218
	9	222
	10	221

Comments:

- The null values were excluded to calculate this average score
- The score is very close to the middle point of 5 out of 10.
- There are more patients giving a score of 1, 4 and 7.

3. Demographic of our patients?

Age Demographic

```
85      -- Average age of our patients
86      SELECT AVG(patient_age) AS Age FROM hospital_er;
```

Age
39.8551

Age group distribution:

```
89      CREATE OR REPLACE VIEW age_separation AS
90          SELECT
91              CASE
92                  WHEN patient_age BETWEEN 0 AND 12 THEN 'Child'
93                  WHEN patient_age BETWEEN 13 AND 19 THEN 'Teenager'
94                  WHEN patient_age BETWEEN 20 AND 39 THEN 'Adult'
95                  WHEN patient_age BETWEEN 40 AND 59 THEN 'Middle Age'
96                  WHEN patient_age >= 60 THEN 'Senior'
97                  ELSE 'unknown'
98              END AS age_group,
99              date,
100             patient_sat_score AS score,
101             patient_waittime
102         FROM hospital_er;
103
104     -- Distribution of patients' age group
105     SELECT
106         age_group,
107         COUNT(*) AS count
108     FROM age_separation
109     GROUP BY age_group
110     ORDER BY count DESC;
```

age_group	count
Adult	2388
Senior	2307
Middle Age	2286
Child	1413
Teenager	822

Gender Demographic

```
112     -- Gender and Age Distribution of our patients
113 •   SELECT
114         patient_gender AS gender,
115         COUNT(patient_gender) AS gender_count,
116         AVG(patient_age) AS Avg_age
117     FROM hospital_er
118     GROUP BY patient_gender;
```

	gender	gender_count	Avg_age
▶	M	4705	39.7299
	F	4487	39.9608
	NC	24	44.6667

Race demographic

```
112     -- Checking the distribution of race amongst our patients
113 •   SELECT
114         patient_race,
115         count(*) AS Count
116     FROM hospital_er
117     GROUP BY patient_race;
```

	patient_race	Count
▶	White	2571
	Native American/Alaska Native	498
	African American	1951
	Asian	1060
	Two or More Races	1557
	Pacific Islander	549
	Declined to Identify	1030

Flagged Patients

```
119     -- Percentage of flagged patients
120 •   SELECT
121         patient_race,
122         COUNT(*) AS flagged_count,
123         ROUND(COUNT(*)/(SELECT COUNT(*) FROM hospital_er)*100,2) AS flagged_percentage
124     FROM hospital_er
125     WHERE patient_admin_flag = 'TRUE'
126     GROUP BY patient_race;
```

	patient_race	flagged_count	flagged_percentage
▶	Native American/Alaska Native	251	2.72
	African American	995	10.80
	White	1289	13.99
	Two or More Races	763	8.28
	Declined to Identify	508	5.51
	Asian	541	5.87
	Pacific Islander	265	2.88

Seeing the number of flagged patients from each age group

```sql
-- Agegroup of flagged patients, and percentage of the agegroup flagged.
WITH try AS(
    SELECT
        CASE
            WHEN patient_age BETWEEN 0 AND 12 THEN 'Child'
            WHEN patient_age BETWEEN 13 AND 19 THEN 'Teenager'
            WHEN patient_age BETWEEN 20 AND 39 THEN 'Adult'
            WHEN patient_age BETWEEN 40 AND 59 THEN 'Middle Age'
            WHEN patient_age >= 60 THEN 'Senior'
            ELSE 'unknown'
        END AS age_group,
        patient_admin_flag
    FROM hospital_er
)
SELECT
    age_group,
    SUM(CASE WHEN patient_admin_flag = 'TRUE' THEN 1 ELSE 0 END) AS flagged_count,
    SUM(CASE WHEN patient_admin_flag = 'FALSE' THEN 1 ELSE 0 END) AS unflagged_count,
    ROUND(SUM(CASE WHEN patient_admin_flag = 'TRUE' THEN 1 ELSE 0 END) / (SUM(CASE WHEN patient_admin_flag = 'TRUE' THEN 1 ELSE 0 END) + SUM(CASE WHEN patient_admin_flag = 'FALSE' THEN 1 ELSE 0 END)) * 100,2) AS percentage
FROM try
GROUP BY age_group;

-- Count of Referral Types
SELECT
    department_referral AS 'Referral Types',
    COUNT(*) AS Count
FROM hospital_er
GROUP BY department_referral;
```

patient_race	flagged_count	flagged_percentage
Native American/Alaska Native	251	2.72
African American	995	10.80
White	1289	13.99
Two or More Races	763	8.28
Declined to Identify	508	5.51
Asian	541	5.87
Pacific Islander	265	2.88

Referral Types

```sql
-- Count of Referral Types
SELECT
    department_referral AS 'Referral Types',
    COUNT(*) AS Count
FROM hospital_er
GROUP BY department_referral;
```

Referral Types	Count
None	5400
General Practice	1840
Orthopedics	995
Gastroenterology	178
Physiotherapy	276
Neurology	193
Cardiology	248
Renal	86

4. Is there a particular subgroup of patients that are less satisfied?

Age

```
166     -- Seeing if different agegroups have different average scores
167 •   SELECT
168         age_group,
169         ROUND(AVG(score),2) AS Avg_Score
170     FROM age_separation
171     WHERE score <> ''
172     GROUP BY age_group;
```

age_group	Avg_Score
Senior	4.77
Middle Age	5.17
Adult	5.07
Child	5.09
Teenager	4.71

Observations:

- Teens and Seniors give poorer rating

Gender

```
174     -- Ratings from different gender
175 •    SELECT
176         patient_gender AS gender,
177         ROUND(AVG(patient_sat_score),2) AS Avg_score,
178         COUNT(*) AS Count
179     FROM hospital_er
180     WHERE patient_sat_score <> ''
181     GROUP BY gender;
```

gender	Avg_score	Count
M	5.03	1306
F	4.96	1206
NC	3.2	5

Observations:

- NC has the lowest scores
- M and F almost same, but F have lower scores.
- There are also very little patients that are NC and scored

Race

```
183     -- Rating from each race
184 •   SELECT
185         patient_race,
186         ROUND(AVG(patient_sat_score),2) AS avg_score,
187         COUNT(*) AS Count
188     FROM hospital_er
189     WHERE patient_sat_score <> ''
190     GROUP BY patient_race;
```

patient_race	avg_score	Count
White	4.94	734
African American	5.07	514
Native American/Alaska Native	5.12	138
Asian	5.01	293
Declined to Identify	4.97	275
Two or More Races	4.83	416
Pacific Islander	5.33	147

Observations:

- Multirace people have lower rating
- Pacific Islander has highest rating

Flagged

```
192     -- Do flagged patients have higher or lower ratings?
193 •   SELECT
194         patient_admin_flag AS 'Flagged Patients',
195         ROUND(AVG(patient_sat_score),2) AS 'Average Score'
196     FROM hospital_er
197     WHERE patient_sat_score <> ''
198     GROUP BY patient_admin_flag;
```

Flagged Patients	Average Score
FALSE	4.91
TRUE	5.08

Observations:

- a bit difference, flagged patients give better ratings

Department referral

```
200     -- Referral type vs avg rating
201 •   SELECT
202         department_referral,
203         ROUND(AVG(patient_sat_score),2) AS Avg_score,
204         ROUND(AVG(PATIENT_WAITTIME),2) AS Avg_waittime
205     FROM hospital_er
206     WHERE patient_sat_score <> ''
207     GROUP BY department_referral
208     ORDER BY avg_score DESC;
```

department_referral	Avg_score	Avg_waittime
Gastroenterology	5.8	33.43
Neurology	5.28	37.02
Cardiology	5.14	33.25
General Practice	5.06	34.84
Physiotherapy	4.99	37.04
None	4.95	35.66
Orthopedics	4.86	34.62
Renal	4.57	37.43

Observations:

- Renal referrals have the lowest score of 4.5
- Gastroenterology has the highest score of 5.8

Waiting Time

```
210     -- Min and Max waiting time?
211 •   SELECT
212         MIN(patient_waittime) AS Min,
213         MAX(patient_waittime) AS Max
214     FROM hospital_er;
```

Min	Max
10	60

Pearson Correlation between waiting time and score:

Correlation
-0.0036

Separating waiting time into different buckets:

```sql
-- Rating of patients from different waiting time
CREATE OR REPLACE VIEW time_separation AS
SELECT
    CASE
        WHEN patient_waittime <= 20 THEN 'Fastest'
        WHEN patient_waittime BETWEEN 20 AND 30 THEN 'Fast'
        WHEN patient_waittime BETWEEN 30 AND 40 THEN 'Normal'
        WHEN patient_waittime BETWEEN 40 AND 50 THEN 'Slow'
        WHEN patient_waittime >= 50 THEN 'Slowest'
    END AS time_group,
    patient_waittime,
    patient_sat_score
FROM hospital_er;
```

Getting the average rating from each bucket:

	Waiting Time	Avg_score	Count
▶	Fastest	5.21	542
	Fast	4.84	467
	Normal	4.82	483
	Slow	5.09	509
	Slowest	4.97	516

Observations:

- The faster the wait time, the higher the rating.
- But longer wait time does not necessarily mean lower rating

Particular timing

```sql
-- Preferred timing rating
SELECT
    HOUR(date) AS hours,
    ROUND(AVG(patient_sat_score),2) AS Avg_score
FROM hospital_er
WHERE patient_sat_score <> ''
GROUP BY hours
ORDER BY LENGTH(hours), hours;
```

hours	Avg_score
0	4.68
1	4.32
2	5.42
3	5.33
4	4.77
5	5.14
6	4.59
7	4.83
8	5.28
9	4.84
10	5.02
11	5.28
12	4.81
13	5.34
14	5.03
15	5.08
16	5.11
17	5.22
18	5.17
19	4.77
20	5.11
21	5.24
22	4.42
23	4.88

Observations:

- 10pm to 1am have lower avg_rating.
- Rating is all higher than 5 from 1pm to 6pm

Conclusion

- Teenagers and Seniors have the lowest rating
- Non-Conforming patients have the lowest rating
- Patients referred from Renal department have the lowest rating
- A shorter waiting time will lead to higher rating, but a longer waiting time may not necessarily lead to lower rating.
- Ratings are generally lower from 10pm to 1am, and higher from 1pm to 6pm

5. When are the peak periods through the year

```
281     -- Number of visits and average waiting time of patients each year
282 •   SELECT
283         YEAR(date) AS Year,
284         COUNT(*) AS Num_of_visits,
285         AVG(patient_waittime) AS Average_Waiting_Time
286     FROM hospital_er
287     GROUP BY Year;
```

Year	Num_of_visits	Average_Waiting_Time
2020	4878	35.4651
2019	4338	35.0290

```sql
-- Number of visits and average waiting time of patients throughout the year?
SELECT
    monthname(date) AS month,
    COUNT(*) AS Num_of_visits,
    AVG(patient_waittime) AS Average_Waiting_Time
FROM hospital_er
WHERE date < '2020-04-02 00:00:00'
GROUP BY month
ORDER BY Average_Waiting_Time DESC;
```

month	Num_of_visits	Average_Waiting_Time
February	431	36.6705
August	494	36.3927
January	513	36.3236
March	506	35.8794
June	506	35.5810
November	464	35.1853
October	493	34.9331
April	491	34.8269
December	489	34.7648
July	464	34.7198
May	480	34.4292
September	469	34.2687

Observations:

- There is a slightly longer waiting time from January to March, as compared to the other months.
- It can also be seen that January has the highest number of cases. However, the following month on February, has the lowest number of cases.
- There is also longer waiting time in August
- There are more cases in year 2020 because there are 2 more months in 2020 than 2019 in the dataset.

6. When are the peak periods through the week

```sql
--  Number of visits and average waiting time of patients throughout the week?
SELECT
    DAYNAME(date) AS Day,
    COUNT(*) AS Num_of_visits,
    AVG(patient_waittime) AS Average_Waiting_Time
FROM hospital_er
GROUP BY Day
ORDER BY Num_of_visits DESC;
```

Day	Num_of_visits	Average_Waiting_Time
Monday	1377	35.6500
Saturday	1332	34.8791
Tuesday	1318	35.2473
Wednesday	1314	35.6682
Sunday	1310	35.0756
Thursday	1305	35.0935
Friday	1260	35.1873

Observations:

- Mondays have the most visits through the week
- Friday has the least
- They are more or less quite similar, around 1.3k per day.
- The average waiting time throughout the week is also very similar, very close to 35 minutes.

7. When are the peak periods through the day

```
308     -- Number of visits and average Waiting time of patients throughout the day?
309 •   SELECT
310         HOUR(date) AS Hour,
311         COUNT(*) AS Num_of_visits,
312         AVG(patient_waittime) AS Average_Waiting_Time
313     FROM hospital_er
314     GROUP BY Hour
315     ORDER BY Hour;
```

Hour	Num_of_visits	Average_Waiting_Time
0	406	34.2512
1	372	34.0645
2	376	35.3644
3	385	37.2182
4	384	34.9297
5	393	36.3868
6	375	35.1333
7	415	35.2627
8	386	35.3212
9	388	36.3608
10	349	35.6160
11	403	36.4615
12	366	33.8798
13	410	35.5610
14	368	34.7418
15	394	34.3858
16	378	34.5847
17	359	34.6936
18	370	34.7784
19	383	34.2663
20	372	34.9274
21	376	34.8830
22	372	36.6022
23	436	36.2500

Observations:

- 3 am is the longest waiting time
- 12pm is the shortest waiting time
- We got the most number of visits at 11pm
- And least number of visits during 10am

8. Are there any bottle necks in the flow of patients?

Different patient age groups

```
318     -- See if it takes longer to see different agegroups
319 •   SELECT
320         age_group,
321         ROUND(AVG(patient_waittime),2) AS Avg_waitingtime
322     FROM age_separation
323     GROUP BY age_group;
```

age_group	Avg_waitingtime
Senior	35.07
Child	35.30
Middle Age	35.07
Adult	35.61
Teenager	35.20

Observations:

- Not much difference in average waiting time.

Gender

```
325     -- See if it takes longer to see different genders
326 •   SELECT
327         patient_gender AS gender,
328         ROUND(AVG(patient_waittime),2) AS Avg_waitingtime
329     FROM hospital_er
330     GROUP BY gender;
```

gender	Avg_waitingtime
M	35.40
F	35.11
NC	37.17

Observations:

- Non-conforming patients have longer average waiting time. Could there be some difficulty when dealing with non-conforming patients?

Flagged Patients

```
332     -- Does it take longer to see flagged patients?
333 •  SELECT
334         patient_admin_flag AS flagged,
335         ROUND(AVG(patient_waittime),2) AS Avg_waitingtime
336     FROM hospital_er
337     GROUP BY flagged;
```

	flagged	Avg_waitingtime
▶	FALSE	35.55
	TRUE	34.97

Observations:

- No big difference between flagged and unflagged patients.

Referral types

```
325     -- Referral Type vs Waiting Time
326 •  SELECT
327         department_referral,
328         ROUND(AVG(PATIENT_WAITTIME),2) AS Avg_waittime
329     FROM hospital_er
330     WHERE patient_sat_score <> ''
331     GROUP BY department_referral
332     ORDER BY Avg_waittime DESC;
```

	department_referral	Avg_waittime
▶	Renal	37.43
	Physiotherapy	37.04
	Neurology	37.02
	None	35.66
	General Practice	34.84
	Orthopedics	34.62
	Gastroenterology	33.43
	Cardiology	33.25

Observations:

- Seems like Renal, Physio and Neurology has significantly longer average waiting times than the other referral types. Perhaps they require some sort of additional processing, which can be considered a bottleneck.

Race

```
349     -- Average waiting time per race
350 •   SELECT
351         ROUND(AVG(WhiteWait),2) AS 'Average White Waittime',
352         ROUND(AVG(NA_AlaskaWait),2) AS 'Average NAA Waittime',
353         ROUND(AVG(AAWait),2) AS 'Average AA Waittime',
354         ROUND(AVG(AWait),2) AS 'Average Asian Waittime',
355         ROUND(AVG(BiracialWait),2) AS 'Average Biracial Waittime',
356         ROUND(AVG(PIWait),2) AS 'Average PI Waittime',
357         ROUND(AVG(RestWait),2) AS 'Average Others Waittime'
358     FROM racewait;
```

Average White Waittime	Average NAA Waittime	Average AA Waittime	Average Asian Waittime	Average Biracial Waittime	Average PI Waittime	Average Others Waittime
35.09	35.59	35.59	35.32	35.35	34.54	34.95

Observations:

- Not much difference in the average waiting time between different races.

Conclusions

- Non-conforming patients have a longer waiting time, there could be a bottleneck in the process of caring for these patients.
- Physio, Renal and Neurology patients have a longer waiting time as compared to other referral types. This could be a sign that there are bottlenecks in the process of caring for these patients.
- No significant bottleneck in different age groups, flagged, and race of patients.

9. Do we have enough resources to accommodate for the volume of patients?

We have previously found from other queries that the volume of patients do not significantly affect the patient_waittime. This could be a hint that the hospital is generally able to handle the volume of patients. To investigate further, we take a look at the relationship between patient volume and patient waiting time.

```
376     -- What is the average waiting time?
377 •   SELECT AVG(patient_waittime) FROM hospital_er;
```

AVG(patient_waittime)
35.2599

```
379     -- Finding the hours with the most patients
380 •   SELECT
381         DATE(date),
382         HOUR(date),
383         COUNT(*) AS 'Total Cases',
384         AVG(patient_waittime) AS 'Average Waiting Time'
385     FROM hospital_er
386     GROUP BY DATE(date), HOUR(date)
387     ORDER BY COUNT(*) DESC
388     LIMIT 10;
```

	DATE(date)	HOUR(date)	Total Cases	Average Waiting Time
▶	2019-06-25	7	6	31.5000
	2019-09-19	16	6	28.1667
	2020-02-22	16	5	30.4000
	2020-08-17	11	5	28.8000
	2019-09-26	7	5	34.4000
	2019-10-09	23	5	44.8000
	2019-07-04	13	5	29.6000
	2019-08-27	0	5	40.2000
	2020-05-10	11	5	36.0000
	2019-12-15	11	5	35.4000

From the above query, we can see that the busiest hour was on 25th June 2019, 7am, with 6 patients within an hour. Even with the highest number of patients, the average waiting time is still 31.5 minutes, lower than the average waiting time of 35.26 minutes. This means that **the hospital is able to handle the patient volume**, and the increased waiting time is due to other factors such as complex cases, or additional processes.

Additionally, it can be observed that even with the same patient volume within an hour, the average waiting time of patients outside of office hour is generally higher. So let's take a look at the average score and waiting time outside of office hours.

```
444     -- Finding out if rating and waiting time is different outside of office hours
445 •   SELECT
446         CASE WHEN HOUR(date) >= 9 AND HOUR(date) <= 18 THEN 'Office Hour' ELSE 'Non-Office Hour' END AS Shift,
447         ROUND(AVG(patient_sat_score),2) AS Average_Score,
448         ROUND(AVG(patient_waittime),2) AS Average_Time
449     FROM hospital_er
450     WHERE patient_sat_score <> ''
451     GROUP BY Shift;
```

	Shift	Average_Score	Average_Time
▶	Non-Office Hour	4.92	35.52
	Office Hour	5.09	35.12

Based on the query, it can be seen that the **average SAT score is lower and average waiting time is longer outside of office hours**.

10. Are our staffs lacking in any competency that we can work on

From the previous queries, we have found that particular groups of patients have lower average SAT score:

- Teenagers and elderly patients have lower SAT score than the other age groups
- Non-Conforming gendered patients have lower SAT score compared to conforming patients.

Each age group of patients will require their own skillsets to handle, perhaps a lower SAT score coming from teenagers and elderly patients mean that there are space for improvement when dealing with these age groups.

Patients with a non-conforming gender are giving significantly lower scores, it is crucial for the hospital to look into the particular reasons why this group of patients are giving low SAT scores. It could be due to staffs requiring training, or hospital systems needing an improvement to include patients with non-conforming genders.

5. Conclusions

1. Satisfaction Score

- Overall, our hospital has an average SAT score of 4.99
- Teenagers and Seniors have the lower rating than other age groups
- Non-Conforming gendered patients gave the lowest rating
- Patients referred from Renal department have the lowest rating
- A shorter waiting time will lead to higher rating, but a longer waiting time may not necessarily lead to lower rating.
- Ratings are generally lower from 10pm to 1am, and higher from 1pm to 6pm

2. Peak Periods

- Throughout the year, January has the most number of patients.
- Throughout the week, Mondays have the most number of cases
- Throughout the day,
    - During midnight 11pm to 12am
    - Before the start of office hours, at 8am
    - And Before and after lunch at 11am and 1pm.
- ER department is able to handle the patient volume. During the busiest hour, there were 6 patients in an hour, but there was no impact to the waiting time. This means that bottlenecks were not caused by patient volume.

3. Bottle Necks

- Patients that were referred from Renal, Physiotherapy and Neurology have longer waiting time.
    - It could be due to additional processes or equipment required to diagnose and treat.
    - These departments may have more complex conditions that are harder to diagnose or treat.
- Non-conforming patients have longer waiting time.
    - The hospital’s internal job management system might not have accommodated for NC patients traditionally, and needs to be revised. For

example, a drop down menu might only have male or female genders. So we need to improve the system to add other options.

- It could also be due to the facilities or staffs being unfamiliar in handling NC patients as they appear in much smaller numbers.

6. Recommendations

1. Collect feedbacks from teenagers, elderlies and NC patients to further understand why their SAT score is lower.
2. Provide training for staffs to handle NC patients
3. Investigate into whether the hospital's system accommodates for NC patients
4. Collect feedbacks from staffs in ER, Renal, Physiotherapy and Neurology to understand where the bottlenecks are at when referrals come from these departments.