



# Hospital Data Insights

## **An exploration of factors impacting hospitalization duration**

### **Overview**

As a nurse during coronavirus I saw first hand how truly valuable a hospital bed is amidst a pandemic. Hospital readmissions occurring within 30 days after a patient's discharge can result in high costs and suggest an insufficient treatment approach. Additionally, a prolonged hospital stay may have financial consequences for the facility, as insurance reimbursement systems such as Medicare's prospective payment system come into play. Therefore, striking a balance between length of stay and the risk of readmission is crucial for hospital administrators and policymakers.

### **The Data**

I used MySQL to analyze a data set containing 10 years (1999-2008) of data collected from 130 US hospitals. It's not a recent dataset, but it shows some good insights and can be used for future prediction. The dataset can be found on [Kaggle](#).

### **Objectives**

The overall objective of this project is to gather knowledge about patient demographics and their associations with hospitalization duration. By doing so I was able to address business concerns from management by providing insights on:

- Distribution of time for hospitalized patients
- Success rate of emergency cases
- Racial biases regarding women's health
- Relationship between length of hospital stay and laboratory procedures
- Average number of procedures performed by specialty
- Summary of patients hospital stay

## The Analysis

### 1. How many patients stay in the hospital for different lengths?

Answering this is a crucial element that aids in predicting the availability of new admissions.

I performed the SQL statement below to visualize the count number by a simple histogram.

```

1 • SELECT * FROM hospital;
2 • SELECT ROUND(time_in_hospital, 1) AS bucket,
3     COUNT(*) AS count,
4     RPAD(' ', COUNT(*)/100, '*') AS BAR
5 FROM Hospital
6 GROUP BY bucket
7 ORDER By bucket;

```

	bucket	count	BAR
▶ 1	1	14127	*****
2	2	17137	*****
3	3	17639	*****
4	4	13792	*****
5	5	9838	*****
6	6	7389	*****
7	7	5708	*****
8	8	4208	*****
9	9	2872	*****
10	10	2229	*****
11	11	1755	*****
12	12	1366	*****
13	13	1110	*****
14	14	930	*****

minimum, average, and maximum number of procedures performed within the

dataset.

```
1 • SELECT MIN(num_lab_procedures) as minimum, ROUND(AVG(num_lab_procedures),0) as average,  
2         MAX(num_lab_procedures) as maximum  
3 FROM Hospital;
```

	minimum	average	maximum
▶ 1	43	132	

Based on this information, I used the CASE WHEN function to create a column for the following categories:

- Less than 25 as Few
- 25-54 as Average
- Greater than 55 as Many

Then I investigated the correlation between the number of procedures versus the length of hospitalization in total by listing the average time in hospital and grouping them together.

```
1 • SELECT ROUND(AVG(time_in_hospital), 0) as days_stay,  
2         CASE WHEN num_lab_procedures >= 0 AND num_lab_procedures < 25 THEN "few"  
3             WHEN num_lab_procedures >= 25 AND num_lab_procedures < 55 THEN "average"  
4             WHEN num_lab_procedures >= 55 THEN "many" END AS procedure_frequency  
5  
6 FROM Hospital  
7 GROUP BY procedure_frequency  
8 ORDER BY days_stay;
```

days_stay	procedure_frequency
3	few
4	average
6	many

It's evident that patients who stay longer in the hospital require more lab procedures, leading to increased expenses for both the patients and hospitals.

### 3. The newly appointed hospital director has requested a comprehensive list of all specialties along with the average number of procedures performed at the hospital.

I filtered out N/A data like “?” by using the WHERE clause. I used GROUP BY and DISTINCT functions to extract specific data for specialties without repeating the same values. It also helps in counting the total specialties in the hospital. I ROUND the AVG function for the number of procedures with 1 decimal. Finally, I organized the data in descending order.

```

1 • SELECT DISTINCT medical_specialty, COUNT(medical_specialty) as total,
2       ROUND (AVG(num_procedures),1) as average_procedures
3 FROM Hospital
4 WHERE NOT medical_specialty = "?"
5 GROUP BY medical_specialty
6 ORDER BY average_procedures DESC;
```

There are a few of the specialties that only have 1 patient. This could be due recording or input error.

medical_specialty	total	average_procedures
► Proctology	1	4.0
Obsterics&Gynecology-GynecologicOnco	25	3.6
Surgery-Thoracic	109	3.5
Radiologist	1109	3.3
Surgery-Cardiovascular/Thoracic	650	3.3
Dermatology	1	3.0
Surgery-PlasticwithinHeadandNeck	1	3.0
Cardiology	5311	2.7
Surgery-Vascular	529	2.6
Radiology	52	2.5
Podiatry	100	2.4
Surgeon	45	2.2
Surgery-Plastic	39	2.2
Gynecology	57	2.1
Surgery-Cardiovascular	95	2.1
Neurology	201	2.1
Psychiatry-Addictive	1	2.0
Pediatrics-InfectiousDiseases	1	2.0

**4. To streamline the list of specialties, management then requested that only specialties with at least 50 patients and an average of more than 2.5 procedures on average be included.**

The HAVING function was used for conditioning aggregated groups and AND operators for procedures number required.

```
1 • SELECT DISTINCT medical_specialty, COUNT(medical_specialty) as total,
2     ROUND (AVG(num_procedures),1) as average_procedures
3 FROM Hospital
4 WHERE NOT medical_specialty = "?"
5 GROUP BY medical_specialty
6 HAVING total > 50 AND average_procedures > 2.5
7 ORDER BY average_procedures DESC;
```

The result showed a clean list of 5 specialties that met the requirements. This information can help the facility direct attention toward the specialties that are utilizing the most resources.

medical_specialty	total	average_procedures
Surgery-Thoracic	109	3.5
Surgery-Cardiovascular/Thoracic	650	3.3
Radiologist	1109	3.3
Cardiology	5311	2.7
Surgery-Vascular	529	2.6

**5. The Chief of Nursing wants to know if there is any subconscious racial bias specifically regarding women's health with the number of lab procedures done.**

```

SELECT race, (medical_specialty),
       ROUND(AVG(num_procedures),1)AS Avg_procedures,
COUNT(number_inpatient) AS Num_patients
FROM Hospital
WHERE NOT race = '?' AND NOT medical_specialty = '?'
      AND medical_specialty = 'ObstetricsandGynecology'
      AND gender = 'Female'
GROUP BY medical_specialty, race
ORDER BY avg_procedures DESC;

```

race	medical_specialty	Avg_procedures	Num_patients
Caucasian	ObstetricsandGynecology	2.1	416
Asian	ObstetricsandGynecology	2.0	5
Hispanic	ObstetricsandGynecology	1.9	34
Other	ObstetricsandGynecology	1.8	5
AfricanAmerican	ObstetricsandGynecology	1.6	187

The results showed an “other” group. It would be good to identify this group and reanalyze.

## 6. The hospital administrator wants to highlight cases where the hospital provided exceptional care.

I will provide a list of all patients who had an emergency but stayed less than the average time in the hospital.

A WHERE filter and CTE would be the most appropriate for this as shown below. Emergency is admission\_type\_id of 1.

```
WITH avg_time AS (SELECT AVG(time_in_hospital) FROM Hospital)
SELECT * FROM Hospital
WHERE admission_type_id = '1'
AND time_in_hospital < (SELECT * FROM avg_time)
ORDER BY time_in_hospital ASC;
```

The result was 33684 patients.

Then compare this number with the total of patients.

```
SELECT DISTINCT COUNT(*) as total_patients
FROM Hospital;
```

There were a total of 101766 patients reported. From this number, I can easily calculate the percentage of success rate was 33% which is reasonable for emergency and short-term stays only at the hospital.

**7. Lastly, management needs a summary for the top 50 medication patients, and break any ties with the number of lab procedures (highest at the top) by following the hospital's format.**

```
• SELECT CONCAT('Patient ', patient_nbr, ' was ', race, ' and ',
  (CASE WHEN readmitted = 'NO' THEN ' was not readmitted.'
    ELSE 'was readmitted.'
  END),
  ' They had ', num_medications, ' medications and ',
  num_lab_procedures, ' lab procedures.') AS Summary
FROM Hospital
ORDER BY num_medications DESC, num_lab_procedures DESC
LIMIT 50;
```

## Summary

Patient 24189597	was Caucasian and was readmitted. They had 81 medications and 57 lab procedures.
Patient 25112691	was Caucasian and was not readmitted. They had 79 medications and 57 lab procedures.
Patient 43503210	was Caucasian and was not readmitted. They had 75 medications and 76 lab procedures.
Patient 24526629	was Caucasian and was not readmitted. They had 75 medications and 61 lab procedures.
Patient 25450911	was Caucasian and was not readmitted. They had 74 medications and 62 lab procedures.
Patient 42147990	was Caucasian and was readmitted. They had 72 medications and 85 lab procedures.
Patient 42522309	was Caucasian and was readmitted. They had 72 medications and 73 lab procedures.
Patient 43515927	was Caucasian and was readmitted. They had 72 medications and 68 lab procedures.
Patient 25162767	was AfricanAmerican and was readmitted. They had 70 medications and 59 lab procedures.
Patient 23581485	was Caucasian and was readmitted. They had 70 medications and 52 lab procedures.
Patient 23233167	was Caucasian and was readmitted. They had 69 medications and 59 lab procedures.
Patient 24420438	was Caucasian and was readmitted. They had 69 medications and 57 lab procedures.
Patient 24218658	was Caucasian and was readmitted. They had 69 medications and 51 lab procedures.
Patient 23237658	was Caucasian and was readmitted. They had 69 medications and 51 lab procedures.
Patient 24686586	was AfricanAmerican and was not readmitted. They had 69 medications and 47 lab procedures.
Patient 41720949	was Caucasian and was readmitted. They had 68 medications and 88 lab procedures.
Patient 41931207	was Caucasian and was not readmitted. They had 68 medications and 85 lab procedures.

## Results

- The average patient stay is less than 7 days with the most common length of admission being 3 days.
- Approximately 33% of patients who required emergency services had a shorter hospital stay, indicating an effective management system.
- The number of procedures performed increases the longer a patient is admitted.
- The average number of procedures performed by a medical specialty does not necessarily correspond to the number of patients they treat.
- There does not appear to be any significant gaps between races and procedures ordered.

## Conclusion and Recommendation

By analyzing data, facilities can gain valuable insights into the factors that affect patient care outcomes and financial impacts, as demonstrated in this project. These insights can help inform policy, guide strategy, and provide opportunities for staff education. Regular analysis is essential to monitor the effects of changes and evolving health trends, which may require updates to policies and strategies. Additionally, further research is necessary to comprehend potential disparities in care among patients of different races and demographics.



Thank you for taking the time to review this project. To further discuss or connect, please reach out to me on [LinkedIn](#)