Unlocking Insights: A Deep Dive into Diabetic Patient Data Across 130 Hospitals Using SQL Analysis

I embarked on this project driven by a sense of curiosity. First, with **data from 130 different hospitals** at my disposal, I saw an opportunity to delve into a **vast and diverse dataset**, uncovering insights that could potentially **inform and improve healthcare practices**, sparking a desire to understand the trends and patterns within the data. Second, by leveraging **SQL for analysis**, I aimed to not only gain a comprehensive understanding of **patient demographics, treatments, and outcomes** but also to showcase the power of **data-driven decision-making in healthcare management**. Ultimately, this project represents a convergence of personal interest, professional development, and a commitment to **harnessing data for positive change in healthcare**.

Why you should keep reading:

Continuing to delve into this write-up promises a journey into the intricate landscape of **patient data analysis**, offering valuable insights and implications for both healthcare professionals and individuals. By staying engaged, readers will uncover not only **the trends and patterns within the data** but also the **potential implications for healthcare practices and patient outcomes**. Moreover, the utilization of SQL analysis showcases the power of data-driven decision-making in healthcare, providing a compelling case for the importance of **leveraging technology to drive positive change**. Whether you're a healthcare professional seeking to **optimize patient care** or a researcher interested in the **complexities of patient data**, this exploration offers a wealth of knowledge and perspectives that are sure to captivate and enlighten. So, let's embark together on this journey of discovery and insight, with the promise **of gaining a deeper understanding of how data can improve healthcare.**

The Data:

Step into the world of **data-driven healthcare exploration** as I unlock a treasure trove of **hospital records from 130 institutions across the United States, spanning from 1999 to 2008**. Within these records lie the stories of patients who traversed the corridors of medical care, undergoing **laboratory procedures**, medication regimens, and **hospital stays** of up to 14 days. This invaluable dataset, meticulously curated and available through multiple sources, https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008 and https://www.kaggle.com/code/iabhishekofficial/prediction-on-hospital-readmission/data?select=diabetic_data.csv, beckons us to unravel its mysteries and glean insights that could **reshape the landscape of healthcare**.

Using this data I looked to investigate captivating trends as they pertain to:

- Exploring the average duration of **hospital stays** and the diverse spectrum of lengths, spanning from **1 to 14 days**.
- Compiling a roster of **medical specialties attending to these patients** and examining the **frequency of procedures** they prescribed.

- Identifying the **medical specialties with the highest average procedure orders**, utilizing a criterion of an average exceeding 2.5 procedures and a cumulative procedure count surpassing 50.
- Investigating the **correlation between the quantity of procedures** administered to a patient and the **duration of their hospital stay**.
- Investigating **whether race influences the level of care received** by analyzing the average number of procedures aggregated by racial demographics.

Analysis of the Data:

To determine both the **average hospital stay duration** and its distribution, I employed a dual approach **utilizing SQL queries**. Initially, I sought to ascertain the **average length of a patient's hospital stay**, recognizing its pivotal significance for healthcare facilities. Understanding this metric is crucial for hospitals as it directly **impacts their capacity to accommodate patients**, which is inherently tied to the number of available beds. A shorter average stay not only enables hospitals to serve more patients but also signifies **efficiency in resource utilization**. However, it's imperative to strike a balance, ensuring that **patients receive optimal care** and that **longer stays correlate with higher acuity levels**. I used the following SQL query to determine the average length of stay:

```sql
SELECT ROUND(AVG(time_in_hospital),1) AS 'AVG Time In Hospital'
FROM health;
```

	AVG Time In Hospital
▶	4.4

I found that the **average hospital stay** duration was **4.4 days**, a reasonable figure, especially considering that the maximum stay recorded in the dataset is **14 days**.

Subsequently, my aim was to establish a **breakdown of hospital stays by day**. While SQL isn't inherently a data visualization tool, through **data manipulation** and crafting specific queries, I managed to **construct a basic histogram**. This rudimentary visualization method facilitated the depiction of stay distributions by categorizing patient stays into buckets corresponding to the number of days spent in the hospital.

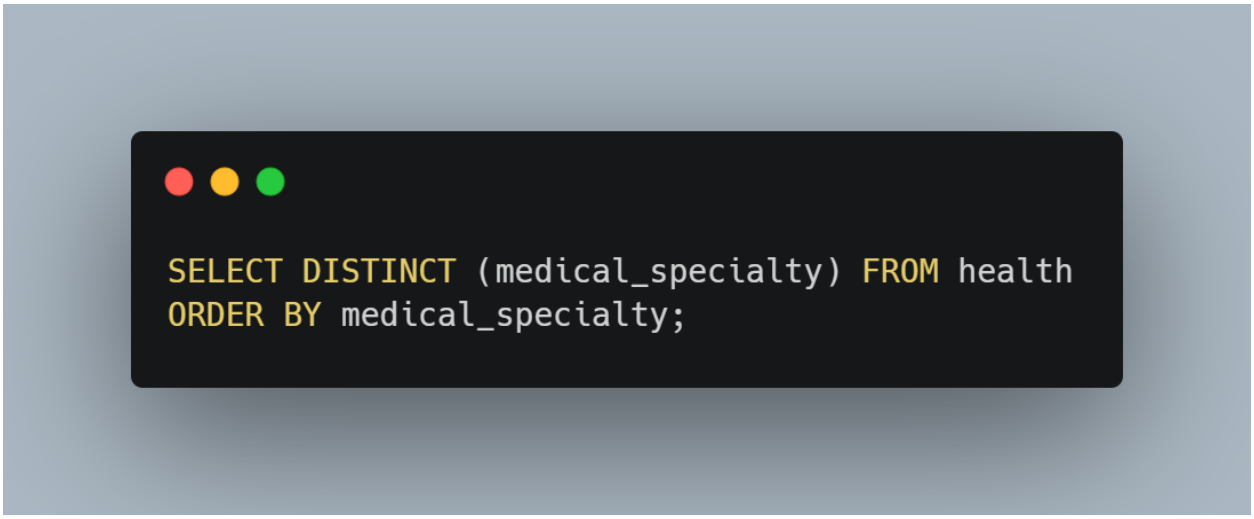
```sql
USE patient;
SELECT ROUND(time_in_hospital, 1) AS bucket,
COUNT(*) AS count,
RPAD('', COUNT(*)/100, '*') AS bar
FROM health
GROUP BY bucket
ORDER BY bucket;
```

bucket	count	bar
1	14208	**********
2	17224	**********
3	17756	**********
4	13924	**********
5	9966	**********
6	7539	**********
7	5859	**********
8	4391	**********
9	3002	**********
10	2342	**********
11	1855	*********
12	1448	*******
13	1210	******
14	1042	*****

As depicted by this distribution, **the bulk of patients spend between 1 to 4 days** in the hospital, with the highest concentration falling within the 2 to 3-day range. This distribution

likely brings contentment to hospitals, as it signifies a **positive influence on their ability to assist a larger number of patients**.

Procedures represent one of the most significant expenses for a hospital, prompting my investigation into **which medical specialties were responsible for the highest volume of procedures**. Initially, I aimed to determine the number of distinct medical specialties present in the dataset and the corresponding count of procedures ordered by each. Leveraging the **HAVING and DISTINCT functions** facilitated this inquiry. Through this analysis, I uncovered a total of **72 distinct specialties** within the dataset, with one specialty listed as unknown.



```
SELECT DISTINCT (medical_specialty) FROM health
ORDER BY medical_specialty;
```

medical_specialty
?
AllergyandImmunology
Anesthesiology
Anesthesiology-Pediatric
Cardiology
Cardiology-Pediatric
DCPTEAM
Dentistry
Dermatology
Emergency/Trauma
Endocrinology
Endocrinology-Metabolism
Family/GeneralPractice
Gastroenterology

224 15:14:44 SELECT DISTINCT (medical_specialty) FROM health ORDER BY medical... 73 row(s) returned

Following this, I executed SQL queries to ascertain the **quantity of procedures ordered by each medical specialty**. Armed with this data, I delved further, aiming to **identify specialties with notably high average procedure counts**. I established the criteria at an **average of 2.5 procedures per patient** and a **minimum total count of 50 procedures** ordered per specialty. Through this inquiry, I pinpointed five medical specialties exhibiting elevated levels of procedure orders. Hospitals would be wise to **closely monitor these specialties**, as their practices could **potentially incur substantial costs**.

```
SELECT medical_specialty, ROUND (AVG(num_procedures), 1) AS avg_procedures,
COUNT(*) AS count
FROM health
GROUP BY medical_specialty
HAVING count>50 AND avg_procedures>2.5
ORDER BY avg_procedures DESC;
```

medical_specialty	avg_procedures	count
Surgery-Thoracic	3.5	109
Surgery-Cardiovascular/Thoracic	3.2	652
Radiologist	3.2	1140
Cardiology	2.7	5352
Surgery-Vascular	2.6	533

Now, my focus shifted to investigating the **potential correlation between the number of procedures a patient undergoes and the length of their hospital stay**. Particularly, I sought to understand whether **patients undergoing a greater number of procedures** tend to have **longer hospital stays**. To explore this, I employed the **CASE WHEN function** to query the dataset. This **analysis revealed a direct relationship**: as the number of procedures per patient increases, so does the length of their hospital stay.

```sql
SELECT AVG(time_in_hospital) AS avg_time,
CASE
WHEN num_lab_procedures >= 0 AND num_lab_procedures < 25 THEN "few"
WHEN num_lab_procedures >= 25 AND num_lab_procedures < 55 THEN "average"
ELSE "many"
END AS procedure_frequency
FROM health
GROUP BY procedure_frequency
ORDER BY avg_time DESC;
```

avg_time	procedure_frequency
5.657308841801269	many
4.065429992110237	average
3.27822257806245	few

Finally, I delved into the data to **investigate whether race influenced patient treatment**. Addressing issues of **equity and fairness in healthcare delivery** is crucial for hospitals committed to **providing quality care to all patients**, regardless of race or ethnicity. Identifying any **discrepancies in treatment based on race** allows hospitals to take proactive steps to rectify these disparities, ensuring that **all patients receive equitable care**. Additionally, addressing these issues can enhance patient trust and satisfaction, leading to improved health outcomes and fostering a positive reputation for the hospital within the community. Overall, investigating and addressing **potential racial disparities** in patient care aligns with the hospital's mission of providing compassionate and **inclusive healthcare services to all individuals**. Employing a JOIN function to amalgamate data tables, I established parameters centered on determining if race significantly impacted the care provided, focusing on the average number of procedures ordered by race. Following the execution of this query, **I found no evident disparities in the treatment of individuals across different racial demographics**.

```sql
SELECT race, ROUND (AVG(num_lab_procedures),3) AS num_avg_lab_procedures FROM health
JOIN demographics ON health.patient_nbr = demographics.patient_nbr
GROUP BY race
ORDER BY num_avg_lab_procedures DESC;
```

race	num_avg_lab_procedures
AfricanAmerican	44.078
?	43.997
Other	43.673
Caucasian	42.838
Hispanic	42.714
Asian	40.859

Main Takeaways:

Through SQL queries, I uncovered several key insights from the data. The **average patient stay was found to be 4.4 days**, with the majority enduring stays of 1-4 days, although the dataset spanned stays ranging from 1 to 14 days. It encompassed a total of **72 medical specialties**, including another one of which remained unidentified. Notably, the top five specialties ordering the highest number of procedures were Surgery-Thoracic, Surgery-Cardiovascular/Thoracic, Radiology, Cardiology, and Surgery-Vascular, prompting hospitals to **scrutinize these areas for potential cost implications**. Moreover, patients undergoing a **higher volume of procedures** tended to have **prolonged hospital stays**, indicating a direct correlation between the number of procedures and length of stay. Despite thorough analysis, **no discernible correlation** between a **patient's race and the quality of care received** was identified, affirming hospitals' commitment to compassionate and inclusive healthcare delivery.

Call to Action:

Thank you for dedicating your time to peruse my analysis of healthcare data utilizing SQL. Should you have any additional insights or ideas for further exploration within this dataset, or if you wish to share comments or suggestions on the analysis presented, please don't hesitate to reach out or leave a comment. Additionally, feel free to explore some of my other projects for further insights and inspirations. Your engagement and feedback are greatly appreciated!

Connect using the links below:

kendun1981@yahoo.com LinkedIn Portfolio