**Healthcare Management Analysis Project with SQL**

The number of new patients and re-admitted patients is raising year by year significantly. Therefore, the pressure for hospital management is also increased to be able to serve patients in all conditions. Some of the concerns from the management are:

1. The hospital's performance in managing the hospital capacity for all patients.
2. Budgeting the procedures and specialists for each patient case.
3. How patients are treated regardless of their race and identification for better care.
4. The success rate for emergency cases.

About the dataset: This project used data from 130 medical clinics for Diabetes in the United States from 1999 to 2008.

<https://www.kaggle.com/code/iabhishekofficial/prediction-on-hospital-readmission/data?select=diabetic_data.csv>

Diabetic patients in the United States (1999-2008).

**THE ANALYSIS**

In this article, I will explore the dataset by using multiple functions like SELECT, FROM , CONCAT, UNION, JOIN, GROUP BY, ORDER BY, CASE WHEN HAVING, DISTINCT, MIN, MAX ,SUM, AVG and ROUND functions and answer a few questions:

***Question 1:* The health care management wants to know the distribution of time spent in the hospital by a patient.**

To help the management learn about hospital's capacity for emergency and general patients.

Approach: I selected the total\_days of hospitalizes and count the number of patients based on their time in hospital by grouping them in the same total\_days with the order of total\_days sorted in ascending order.

SELECT ROUND (time\_in\_hospital,1) as total\_days, COUNT(\*) as count,   
FROM patient.health  
GROUP BY total\_days  
ORDER BY total\_days;

As the result, most patients stay less than 7 days in the hospital, mostly around 3 days or less.

***Question 2:* Budget management- A brand-new hospital director wants a list of all specialties and the average total of the number of procedures currently practiced at the hospital.**

Approach : Here I focused on the column named medical\_specialty and num\_procedures as they contain the information we needed. I also 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. Also added the filter for total more than 50 and with average\_procedures more than 2.5.

SELECT medical\_specialty, COUNT(medical\_specialty) as total,   
 ROUND(AVG(num\_procedures), 1) as average\_procedures  
FROM patient.health  
WHERE NOT medical\_specialty = "?"  
GROUP BY medical\_specialty  
HAVING total > 50 AND average\_procedures > 2.5  
ORDER BY average\_procedures DESC;

A screenshot of a computer

Description automatically generated

The result showed that only 5 specialties met the requirements including:

**Question 3: Integrity- The Chief of Nursing wants to know if the hospital seems to be treating patients of different races differently, specifically with the number of lab procedures done.**

Approach : I needed the information about race in the demographics table and the num\_lab\_procedures in the health table for the analysis. Therefore, I had to combine the 2 tables into one to be able to extract information from them by using the INNER JOIN function on the patient\_nbr column.

SELECT d.race, ROUND(AVG(h.num\_lab\_procedures),1) as average\_num\_lab\_procedures  
FROM patient.health h  
JOIN patient.demographics d  
ON h.patient\_nbr=d.patient\_nbr  
GROUP BY d.race  
ORDER BY average\_num\_lab\_procedures DESC;

The result showed there are 6 listings for race and the average number of lab procedures were quite close to each other with the highest result being for African Americans. Overall, there is no difference in treatment based on race at the hospital. However, the “?” and another listing also showed 44 lab procedures which could affect the result if they are identified.

**Question 4: Do people need more procedures if they stay longer in the hospital?**

Approach: First, I wanted to see the general measurement for the number of procedures at the hospital.

SELECT MIN(num\_lab\_procedures) as minimum, ROUND(AVG(num\_lab\_procedures),0) as average,   
 MAX(num\_lab\_procedures) as maximum  
FROM patient.health;

A close up of a number

Description automatically generated

Based on this information, I divided the number of procedures into 3 different categories: few (0–25 procedures), average (25–55 procedures), and many (>55 procedures). For this purpose, I created CASE.

SELECT ROUND(AVG(time\_in\_hospital), 0) as days\_stay,   
 CASE HEN num\_lab\_procedures >= 0 AND num\_lab\_procedures < 25 THEN "few"  
 WHEN num\_lab\_procedures >= 25 AND num\_lab\_procedures < 55 THEN "average"  
 WHEN num\_lab\_procedures >= 55 THEN "many" END AS procedure\_frequency  
   
FROM patient.health  
GROUP BY procedure\_frequency  
ORDER BY days\_stay;

A screenshot of a computer

Description automatically generated

It is now clear to see that the longer the patient stayed at the hospital, the more lab procedures were required for them and in reverse which would higher expenses for both patients and hospitals.

**Question 5: The Hospital Administrator wants to highlight some of the biggest success stories of the hospital. They are looking for opportunities when patients came into the hospital with an emergency (admission\_type\_id of 1) but stayed less than the average time in the hospital.**

**Approach :** For this challenge, I needed to filter again for emergency cases only but time in the hospital is less than average which needs to be calculated and I counted the total successful cases.

WITH average\_time\_hospital AS(  
 SELECT AVG(time\_in\_hospital) as average  
 FROM patient.health  
)  
SELECT COUNT(\*) as successful\_case  
FROM patient.health   
WHERE admission\_type\_id = 1  
AND time\_in\_hospital < (SELECT\* FROM averaqge\_time\_hospital);

The result was 33684 patients. I then compared this number with the total of patients.

SELECT DISTINCT COUNT(\*) as total\_patients  
FROM patient.health;

There was 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.

**Question 6: The requirement is to obtain 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.**

Approach: In this challenge, I needed the information from the health table including patient\_bnr, num\_lab\_procedures, num\_medications, readmitted (or not), and from the demographics table (race). Therefore, I needed to combine the 2 tables again by using the INNER JOIN function and LIMIT with 50 results.

The printing format included the aggregated data and string so I used CONCAT to return the values and CASE WHEN statements for filtering the “readmitted” condition for each patient.

SELECT CONCAT('Patient ', h.patient\_nbr,' was ',d.race,' and ',  
 CASE WHEN h.readmiited = "NO" THEN " was not readmitted. They had " ELSE   
 " was readmiited. They had " END,  
 h.num\_medications, " medications and", h.num\_lab\_procedures, " lab procedures." as Summary  
FROM patient.health h  
INNER JOIN patient.demographics d  
ON h.patient\_nbr=d.patient\_nbr  
ORDER BY h.num\_medications DESC, num\_lab\_procedures DESC  
LIMIT 10;

And the result is ready to report to the management.

A screenshot of a computer screen

Description automatically generated

I enjoyed working on this project as I reviewed the SQL knowledge that I learned during the summer and practiced analyzing the insights from the data.

Results : After the analysis, I explored:

1. Most patients stayed less than 7 days and a maximum of 14 days in the hospital.
2. There were several lab procedures and medications for each patient with different specialists which could be considered to reduce the cost if necessary.
3. As an immigrant country, the United States has people from many ethnic backgrounds. The analysis found that there was fair treatment at healthcare facilities with similar procedures and lab work.
4. The longer the patients stay at the hospital, the more lab procedures are required.
5. Around 33% of patients needed emergency services but stayed shorter at the hospital which reflects the good management system.

Future Scope: I would like to investigate further into other features such as:

1. How the ages of patients affect their level of diabetes.
2. Re-admitted rate and how to improve it.
3. How aftercare service works and the follow-up schedule for each patient.
4. Waiting time for service: emergency, regular check-up, etc.

Thank you.