Health Data Analysis

Summary

I was going through a portfolio which made me interested in doing this analysis using the health care data and understanding and the getting the right out come for every question which is been asked in this data analysis.

Tool Used: Azure Data Studio

Questions

What is the **distribution of time**spent in the hospital for patients, are most of them staying less than7 days?

Select time\_in\_hospital,count(patient\_nbr) as cnt

from diabetic\_data

group by time\_in\_hospital

order by time\_in\_hospital ASC

A table of numbers with numbers

Description automatically generated

Which departments perform the most average number of proceduresand surgeries?

select distinct(medical\_specialty),avg(num\_procedures) as

avg\_num\_procedures,count(\*) as cnt

from diabetic\_data

group by medical\_specialty

order by avg\_num\_procedures desc, cnt desc

A screenshot of a medical data

Description automatically generated

```Since the nursing director is looking for avg\_num\_procedures more than 2 and mdical specialities more than 50 patients```

select distinct(medical\_specialty),avg(num\_procedures) as avg\_num\_procedures,count(\*) as cnt

from diabetic\_data

group by medical\_specialty

having

avg(num\_procedures)>2 and count(\*)>50

order by avg\_num\_procedures desc, cnt desc

A screenshot of a computer

Description automatically generated

Do longer hospital stays **correlate** to more labs procedures being performed?

select distinct(time\_in\_hospital) as time\_in\_hospital,avg(num\_lab\_procedures) as avg\_lab\_procedures

from diabetic\_data

group by time\_in\_hospital

order by time\_in\_hospital desc,avg\_lab\_procedures desc

A screenshot of a computer

Description automatically generated

```So we want to derive procedures between 1-25 as few, 26-55 as average and anything above 55 as many with consideration of departments```

select medical\_specialty,num\_lab\_procedures,

case when num\_lab\_procedures between 1 and 25 then 'few'

when num\_lab\_procedures between 26 and 55 then 'Average'

when num\_lab\_procedures > 55 then 'Many' END,avg(time\_in\_hospital) as avg\_time\_in\_hospital

from diabetic\_data where medical\_specialty <>'?'

group by medical\_specialty,num\_lab\_procedures,

case when num\_lab\_procedures between 1 and 25 then 'few'

when num\_lab\_procedures between 26 and 55 then 'Average'

when num\_lab\_procedures > 55 then 'Many' END

A screenshot of a computer

Description automatically generated

```This query is returning the number of lab procedures with avg\_time\_in\_hospital. Now i want to find

avg\_number\_of\_lab\_procuders with avg\_time\_in\_hospital```

with cte as (select medical\_specialty,num\_lab\_procedures,

case when num\_lab\_procedures between 1 and 25 then 'few'

when num\_lab\_procedures between 26 and 55 then 'Average'

when num\_lab\_procedures > 55 then 'Many' END as category,avg(time\_in\_hospital) as avg\_time\_in\_hospital

from diabetic\_data where medical\_specialty <>'?'

group by medical\_specialty,num\_lab\_procedures,

case when num\_lab\_procedures between 1 and 25 then 'few'

when num\_lab\_procedures between 26 and 55 then 'Average'

when num\_lab\_procedures > 55 then 'Many' END)

Select distinct(avg\_time\_in\_hospital),avg(num\_lab\_procedures) as avg\_num\_lab\_procedures

from cte

where avg\_time\_in\_hospital>4

group by avg\_time\_in\_hospital

order by avg\_time\_in\_hospital desc

A screenshot of a data sheet

Description automatically generated

Is there a **racial bias** when it comes to evaluate the amount of lab procedures performed for patients, and does gender of patients has any effect?

A screenshot of a data sheet

Description automatically generated

How many people belong to **African American**race who are UP for 'Metformin' ?

A screenshot of a computer

Description automatically generated

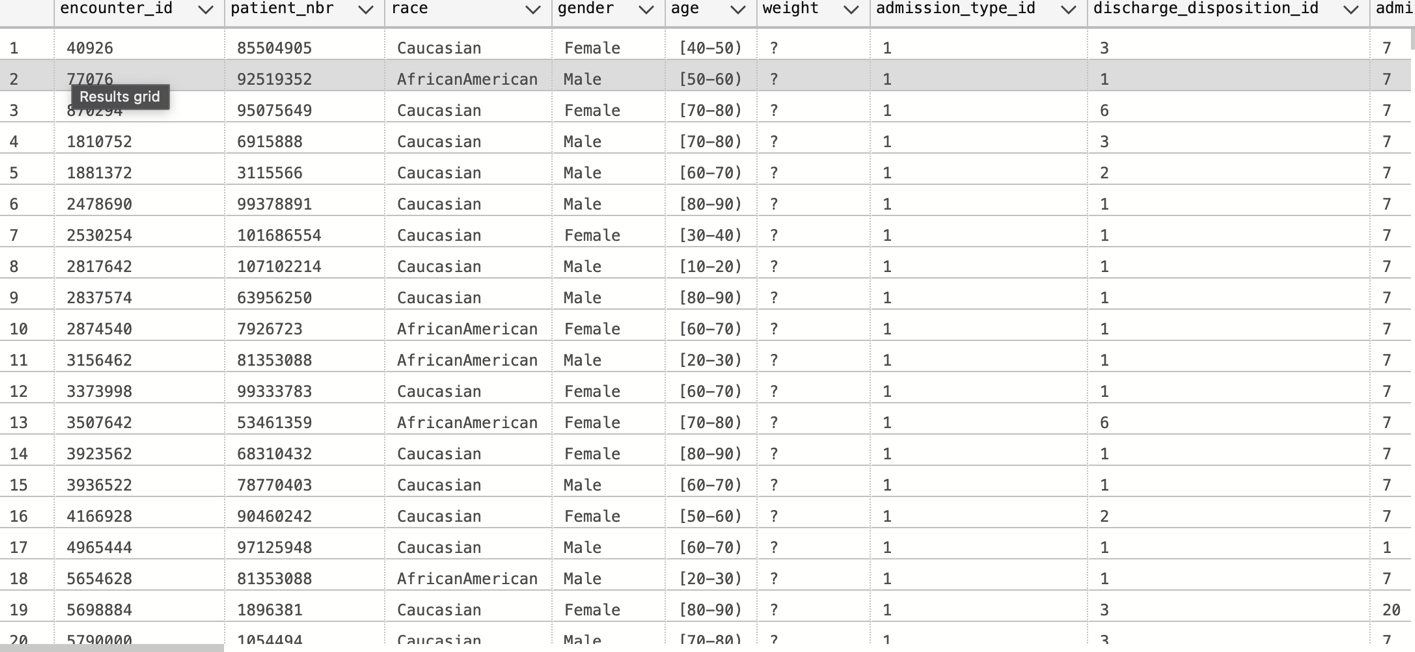
with cte as (select patient\_nbr,race,insulin

from diabetic\_data

where race='AfricanAmerican' and insulin=upper('up'))

select distinct(patient\_nbr) from cte

When patients are discharged, do they readmit to the hospital in **under 30 days**and are these returning patients coming through emergency room?



select \*

from diabetic\_data

where readmitted='<30' and admission\_type\_id=1

A screenshot of a computer

Description automatically generated

Select \*

from diabetic\_data

where readmitted='<30' and admission\_type\_id=1 and time\_in\_hospital<(select avg(time\_in\_hospital) from diabetic\_data)

A summary of **top 50 patients** with highest number of lab procedures.

select TOP 50 patient\_nbr,num\_lab\_procedures

from diabetic\_data

order by num\_lab\_procedures DESC

A screenshot of a table

Description automatically generated