* **Creating a database:**

create database healthcare\_db;

use healthcare\_db;

* **creating two tables called diabetic\_data and demographic out of the same data:**

show tables;

select \*from diabetic\_data;

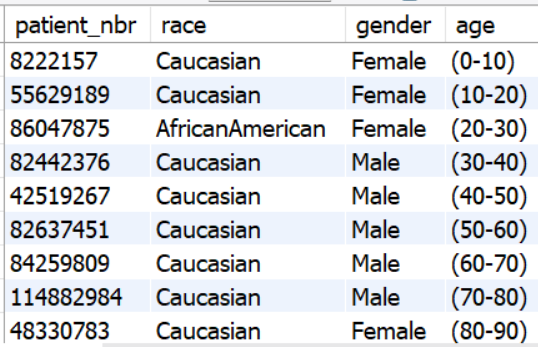
select count(\*) from diabetic\_data;

describe diabetic\_data;

* **updating age column in demographic table**

update demographic

set age=replace(age, '[', '(');



* **Just wanted to find the max and min of patient number:**

select max(patient\_nbr) from demographic;

select min(patient\_nbr) from demographic;

* **finding duplicates using row number windows function**

select \*

from (

select \*,

row\_number() over(partition by patient\_nbr order by patient\_nbr) as RowNum

from demographic

)d

where d.RowNum>1;

* **Another way of finding duplicates in the dataset:**

select count(\*)

from(

select \*

from (

select \*,

row\_number() over(partition by patient\_nbr order by patient\_nbr) as RowNum

from demographic

)d

where d.RowNum>1

)as dup;

* **Finding duplicates in the table diabetic\_data:**

select \*

from (

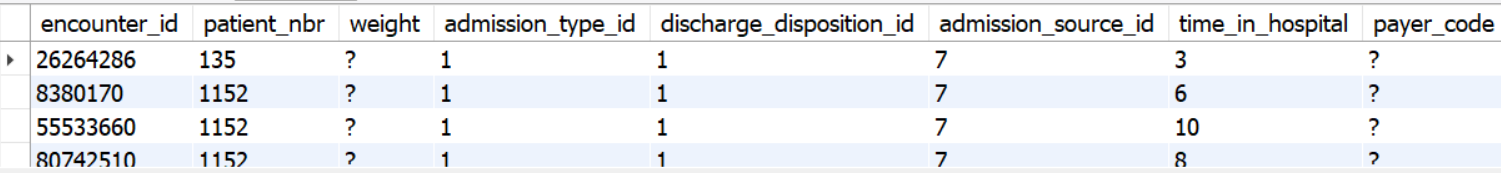
select \*,

row\_number() over(partition by patient\_nbr order by patient\_nbr) as RowNum

from diabetic\_data

)d

where d.RowNum >1;



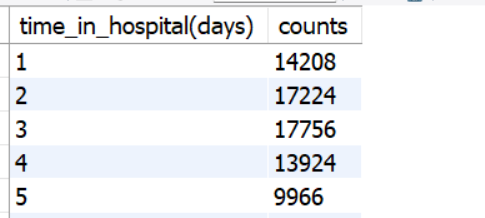
* **Finding duration of stay in hospital**

select time\_in\_hospital, count(\*) as counts

from diabetic\_data

group by time\_in\_hospital

order by time\_in\_hospital;



* **SQL command creating Histogram**

select

round(time\_in\_hospital, 1) as bins,

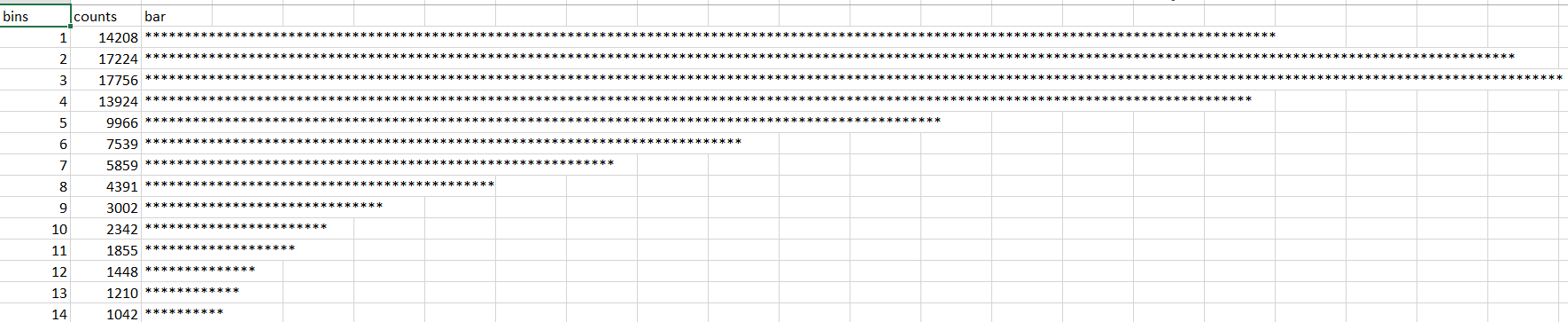
count(\*) as counts,

rpad('', count(\*)/100, '\*') as bar

from diabetic\_data

group by bins

order by bins;



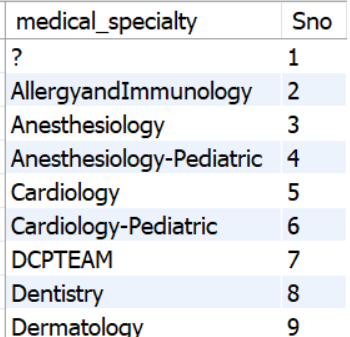
* **Just to find out how many medical specialties in the hospital**

select distinct(medical\_specialty),

dense\_rank() over (order by medical\_specialty) as serial\_no

from diabetic\_data

order by medical\_specialty;



* **Finding the average number of procedures done in each speciality:**

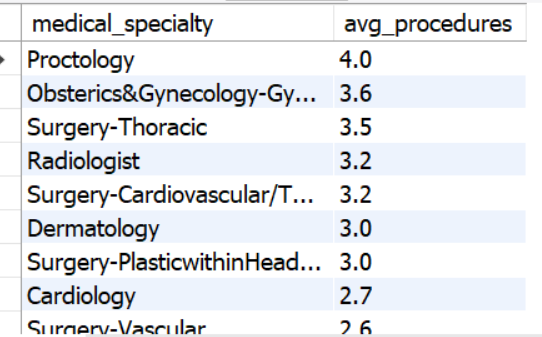
select medical\_specialty,

round(avg(num\_procedures),1)avg\_procedures

from diabetic\_data

group by medical\_specialty

order by avg\_procedures desc;



select medical\_specialty,

round(avg(num\_procedures),1)avg\_procedures,

count(\*) total\_number\_procedures\_done

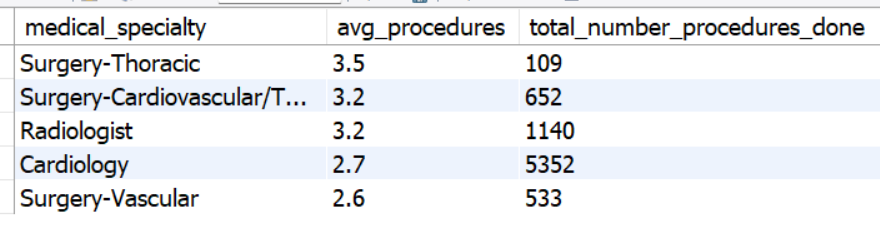
from diabetic\_data

group by medical\_specialty

having count(\*) > 50

and avg\_procedures >2.5

order by avg\_procedures desc;



select \* from diabetic\_data

join demographic on diabetic\_data.patient\_nbr = demographic.patient\_nbr;

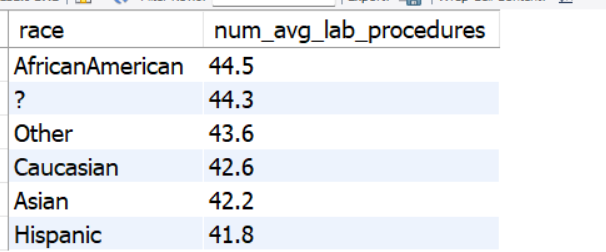
select race, round(avg(num\_lab\_procedures),1) num\_avg\_lab\_procedures

from diabetic\_data

join demographic on diabetic\_data.patient\_nbr = demographic.patient\_nbr

group by race

order by num\_avg\_lab\_procedures desc;



* **Finding average stay in the hospital and the number of procedure carried out.**

select round(avg(time\_in\_hospital),1) avg\_stay\_hospital,

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"

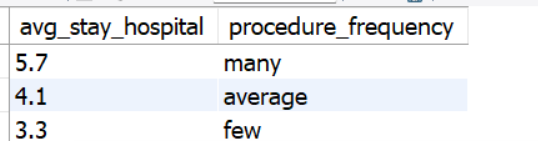
else "many"

end as procedure\_frequency

from diabetic\_data

group by procedure\_frequency

Order by avg\_stay\_hospital desc;



Select count(\*) as total\_count

From

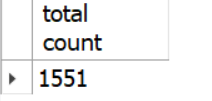
(

select patient\_nbr from demographic where race = "Asian"

union

select patient\_nbr from diabetic\_data where metformin = "up"

) as combined\_total;



select concat('Patient',' ', diabetic\_data.patient\_nbr,' ', 'was',' ', race,' ', 'and',

(case

when readmitted like '%<30%' Then "was readmitted <30 days after discharge"

-- when readmitted = 'no' Then "was not"

-- when readmitted = 'No' Then "was not"

else ' was not readmitted.'

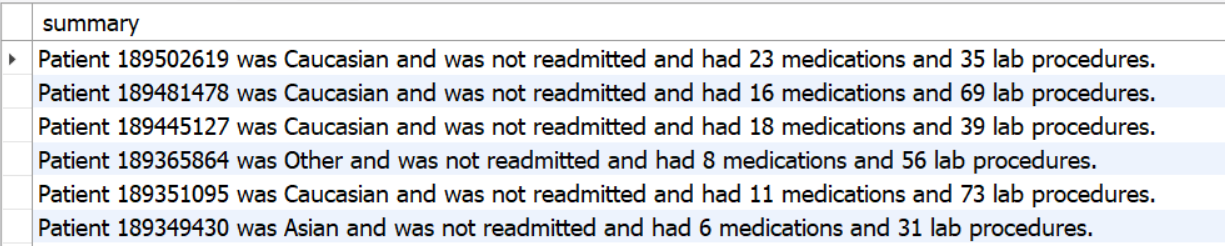
end),' ',

'and had',' ', num\_medications,' ', 'medications and ', num\_lab\_procedures,' ', 'lab procedures.') summary

from diabetic\_data

join demographic on diabetic\_data.patient\_nbr = demographic.patient\_nbr

order by diabetic\_data.patient\_nbr desc;



with cte as

(

select concat('Patient',' ', diabetic\_data.patient\_nbr,' ', 'was',' ', race,' ', 'and',' ',

(case

when readmitted like '%<30%' Then "was readmitted <30 days after discharge"

when readmitted like '%>30%' Then "was readmitted >30 days after discharge"

when not readmitted like "%NO%" then "was readmitted"

when readmitted like "%NO%" then "was not readmitted"

end),' ',

'and had',' ', num\_medications,' ', 'medications and ', num\_lab\_procedures,' ', 'lab procedures.') summary

from diabetic\_data

join demographic on diabetic\_data.patient\_nbr = demographic.patient\_nbr

)

select \*from cte where summary like "%<30%";

