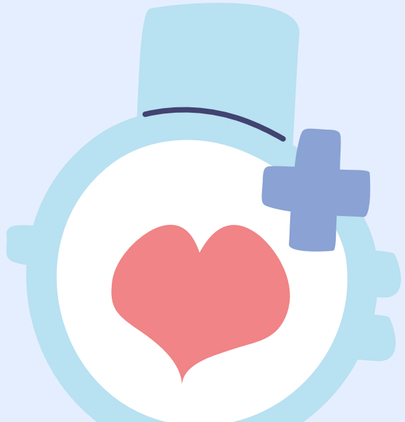
SQL PROJECT ON HEART DISEASE

Arianty Siahaan



Download dataset here

Heart Diseases:
(Dataset Kaggle)
https://www.kaggle.com/datasets/ketangangal/heart-disease-dataset-uci

Data analysis using PostgreSQL

About Dataset

Age :
numeric

Sex :
male : 1; female : 0

chest pain type
-- Value 1: typical angina
-- Value 2: atypical angina
-- Value 3: non-anginal pain
-- Value 4: asymptomatic

resting blood pressure
numeric (in mm Hg on admission to the hospital serum

cholestoral
in mg/dl

fasting blood sugar (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

resting electrocardiographic results

-- Value 0: normal
-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
-- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

thalach: **maximum heart rate** achieved

exercise induced angina
(1 = yes; 0 = no). Angina is chest pain or discomfort caused whe your heart muscle doesn't get enough oxygen-rich blood.
It may feel like pressure or squeezing in your chest.

oldpeak = ST depression induced by exercise relative to rest

slope: the slope of the peak exercise ST segment
--Value 1: upsloping
-- Value 2: flat
-- Value 3: downsloping

vessels colored by flourosopy :
number of major vessels (0-3) colored by flourosopy

A blood disorder called **thalassemia**
(3 = normal; 6 = fixed defect; 7 = reversable defect)

Target :
0 No Heart disease
1 Heart disease

1. Displaying the Average Age of Patients with Heart Disease

```sql
select
        target,
        avg(age) as average_age,
        count(*) as totalpatients
from
        heartdiseasetrain_test ht
group by target;
```

123 target	123 average_age	123 totalpatients
0	56.5691382766	499
1	52.4087452471	526

Result:

The average age is 52.4 years, and the total number of patients is 526.

2. Displaying the Total Number of Patients by Gender and Heart Disease Status

```sql
select
        sex,
        count(*) as totalpatients,
        target
from
        heartdiseasetrain test ht
group by sex, target;
```

	A-Z sex	123 totalpatients	123 target
1	Female	226	1
2	Male	413	0
3	Male	300	1
4	Female	86	0

Result:
The total number of female patients with heart disease is 226, and the number of male patients is 300.

3. Displaying Patients with Cholesterol Greater than 250 mg/dl and Heart Disease

```sql
select
	age,
	sex,
	cholestoral ,
	target
from
	heartdiseasetrain test ht
where
	cholestoral >250 and "target" =1;
```

	123 age	A-Z sex	123 cholestoral	123 target
1	63	Female	252	1
2	63	Female	252	1
3	63	Female	252	1
4	47	Male	253	1
5	47	Male	253	1
6	47	Male	253	1
7	47	Male	253	1
8	50	Female	254	1

Result:

The total number of patients with high cholesterol (>250 mg/dl) and heart disease is 180.

4. Counting the Number of Heart Disease Patients Based on Chest Pain Type (CP)

```sql
select
	chest_pain_type,
	count(*) as Total_Patients_with_Heart_Diseases
from
	heartdiseasetrain_test ht
where
	"target" =1
group by
	chest_pain_type ;
```

	chest_pain_type	total_patients_with_heart_diseases
1	Non-anginal pain	219
2	Atypical angina	134
3	Asymptomatic	51
4	Typical angina	122

Result:

The largest number of heart disease patients are those with non-anginal pain (219 patients).

5. Displaying Patients with Old Peak > 0.56 and Angina

```sql
select
	age,
	oldpeak,
	ht. exercise_induced_angina,
	target
from
	heartdiseasetrain_test ht
where
	oldpeak > 0.56 and exercise_induced_angina ='Yes';
```

	123 age	123 oldpeak	A-Z exercise_induced	123 target
1	53	3.1	Yes	0
2	70	2.6	Yes	0
3	55	0.8	Yes	0
4	54	3.2	Yes	0
5	43	3	Yes	0
6	51	4.2	Yes	0
7	52	1	Yes	0
8	54	2.2	Yes	0

Result:

There are 262 patients with an old peak > 0.56 and angina.

6. Displaying Patients Based on Age and Gender

```sql
select
        age,
        sex,
        count(*) as totalpatients
from
        heartdiseasetrain_test ht
group by
        age, sex
order by
        age;
```

select age, sex, count | Enter a SQL expression to filter resu

Grid | Text

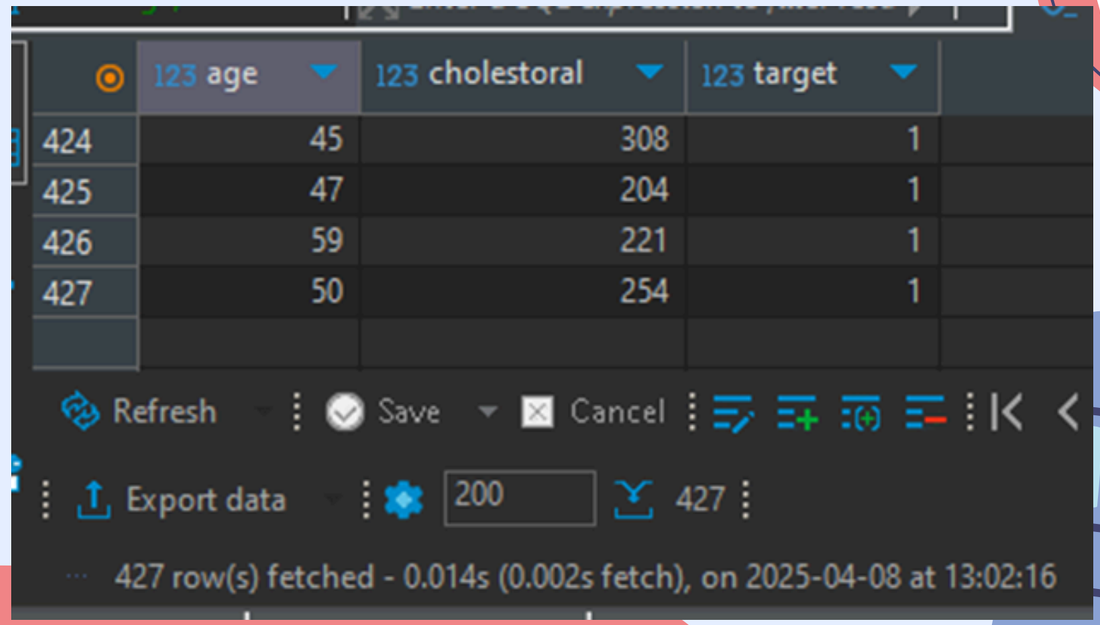
	123 age	A-Z sex	123 totalpatients
1	29	Male	4
2	34	Male	3
3	34	Female	3
4	35	Male	11
5	35	Female	4

Result:

There are 3 male and 3 female patients aged 34.

7. Displaying Patients with Heart Disease and Cholesterol Over 200

```sql
select
	age,
	ht.cholestoral ,
	"target"
from
	heartdiseasetrain_test ht
where
	ht.cholestoral >200 and target=1;
```



	123 age	123 cholestoral	123 target
424	45	308	1
425	47	204	1
426	59	221	1
427	50	254	1

Refresh | Save | Cancel | Export data | 200 | 427

427 row(s) fetched - 0.014s (0.002s fetch), on 2025-04-08 at 13:02:16

Result:

The total number of patients with heart disease and cholesterol > 200 is 427.

8. Displaying the Relationship Between Blood Pressure and Heart Disease

```sql
select
       avg(ht.resting_blood_pressure) as avg_resting_blood_pressure,
       "target"
from heartdiseasetrain_test ht
group by target;
```

	123 avg_resting_blood_pressure	123 target
1	134.1062124248	0
2	129.2452471483	1

Result:
Patients with heart disease have an average resting blood pressure of 129.2.

9. Displaying Patients Based on Gender and Blood Pressure Category (Normal Blood Pressure ⩽ 120)

```sql
select
	sex,
	case
		when ht.resting_blood_pressure <=120 then 'Normal'
		when ht.resting_blood_pressure>120 then 'High Blood Pressure'
	end as blood_pressure_category,
	count(*) as TotalPatients
from heartdiseasetrain_test ht
group by sex, blood_pressure_category;
```

	sex	blood_pressure_category	totalpatients
1	Female	Normal	90
2	Female	High Blood Pressure	222
3	Male	Normal	239
4	Male	High Blood Pressure	474

Result:

There are 222 female patients with high blood pressure, and 474 male patients with high blood pressure.

10. Using a Subquery to Find Patients with Cholesterol Higher than the Average

```sql
select
	age,
	ht.cholestoral ,
	"target"
from
	heartdiseasetrain test ht
where
	ht.cholestoral > (select avg(ht2.cholestoral ) from heartdiseasetrain test
ht2) ;
```

	123 age	123 cholestoral	123 target
1	62	294	
2	58	248	
3	58	318	
4	55	289	
5	46	249	
6	54	286	
7	43	341	
8	51	298	

Result:
There are 456 patients with cholesterol higher than the average.

11. Calculating the Average Age and Cholesterol Based on Heart Disease Target and Displaying Patients with Age and Cholesterol Above the Averages

```sql
with average_stats as (
	select
		target,
		avg(age) as avg_age,
		avg(cholestoral) as avg_chol
	from
		heartdiseasetrain_test ht
	group by
		target
)
select
	ht.age ,
	ht.cholestoral ,
	ht."target" ,
	a.avg_age,
	a.avg_chol
from
	heartdiseasetrain_test ht
join
	average_stats a on ht."target" = a.target
where
	ht.age >a.avg_age and ht.cholestoral >a.avg_chol;
```

	123 age	123 cholestoral	123 target	123 avg_age	123 avg_chol
1	62	294	0	56.5691382766	251.2925851703
2	58	248	1	52.4087452471	240.9790874525
3	58	318	0	56.5691382766	251.2925851703
4	63	252	1	52.4087452471	240.9790874525
5	61	307	0	56.5691382766	251.2925851703
6	58	319	0	56.5691382766	251.2925851703
7	70	269	0	56.5691382766	251.2925851703
8	59	271	1	52.4087452471	240.9790874525

Result:

There are 277 patients with both age and cholesterol above average.

12. Using CTE to Calculate the Average Cholesterol for Patients with Exercise Induced Angina and Displaying Patients with Cholesterol Higher Than This Average

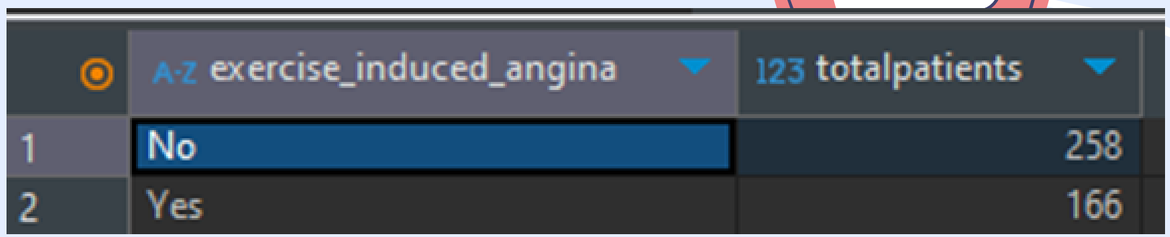
```sql
with average_chol as (
	select
		avg(ht.cholestoral) as avg_chol
	from
		heartdiseasetrain_test ht
	where
		ht.exercise_induced_angina='Yes'
)
select
	age,
	cholestoral,
	exercise_induced_angina,
	target
from
	heartdiseasetrain_test ht2
where
	cholestoral>(select a.avg_chol from average_chol a);
```

	123 age	123 cholestoral	A-Z exercise_induced_angina	123 target
1	62	294	No	0
2	58	318	No	0
3	55	289	Yes	0
4	54	286	Yes	0
5	43	341	Yes	0
6	51	298	Yes	0
7	51	308	No	1
8	54	266	Yes	0

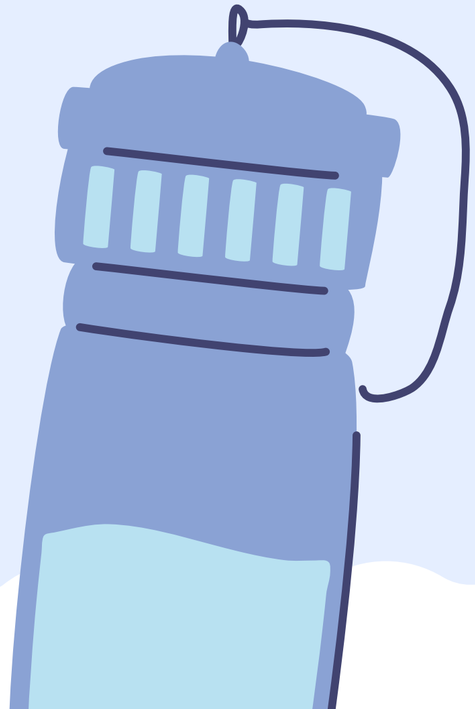
Result:
There are 424 patients with cholesterol above the average.

13. Displaying the Distribution of Angina from the Previous Query

```sql
with cte_angina as(
		with average_chol as (
			select
				avg(ht.cholestoral) as avg_chol
			from
				heartdiseasetrain_test ht
			where
				ht.exercise_induced_angina='Yes'
		)
		select
			age,
			cholestoral,
			exercise_induced_angina,
			target
		from
			heartdiseasetrain_test ht2
		where
			cholestoral>(select a.avg_chol from average_chol a)
)
select c.exercise_induced_angina, count(*) as totalpatients
from cte_angina c
group by c.exercise_induced_angina;
```



	exercise_induced_angina	totalpatients
1	No	258
2	Yes	166



Result:

There are 166 patients with angina.

14. Displaying Patients with Cholesterol Over 250 and Age Above the Average for Heart Disease Patients as High Risk

```sql
select
	age,
	cholestoral,
	target,
	case
		when cholestoral>250 and age>(select avg(age) from heartdiseasetrain_test ht where target=1) then 'high risk'
		else 'low risk'
	end as risk_category
from
	heartdiseasetrain_test ht
where
	target=1;
```

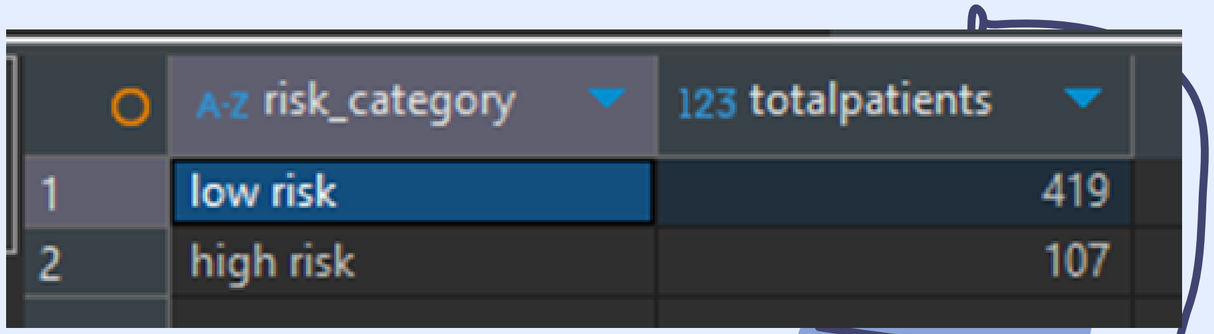
	123 age	123 cholestoral	123 target	A-Z risk_category
1	58	248	1	low risk
2	71	149	1	low risk
3	34	210	1	low risk
4	34	210	1	low risk
5	51	308	1	low risk
6	50	244	1	low risk
7	58	211	1	low risk
8	67	223	1	low risk
9	45	208	1	low risk

Result:

There are 526 patients in this category.

15. Displaying the Number of Patients by Risk Category from the Previous Query

```sql
with cte_1 as
(select
	age,
	cholestoral,
	target,
	case
		when cholestoral>250 and age>(select avg(age) from heartdiseasetrain test ht where target=1) then 'high risk'
		else 'low risk'
	end as risk_category
from
	heartdiseasetrain test ht
where
	target=1
)
select c. risk_category, count (*) as totalpatients
from cte_1 c
group by c.risk_category;
```



	risk_category	totalpatients
1	low risk	419
2	high risk	107

Result:

There are 107 patients in the high-risk category.

Thank You