

Heart Disease Data Analysis using SQL

Presented by Jobyar Ahmed



Project Objective

The objective of this project is to analyze the Heart Disease dataset using SQL to identify significant medical factors influencing the presence of heart disease. By applying structured SQL queries, the project aims to extract meaningful patterns, summarize patient data, and evaluate key relationships among attributes such as age, gender, cholesterol, blood pressure, chest pain type, and other health indicators.

Through this analysis, the project seeks to uncover insights that can help in understanding how various risk factors contribute to heart disease and demonstrate how SQL can be effectively used for data-driven healthcare analysis.

Project Goal

The main goal of this project is to perform an in-depth analysis of the Heart Disease dataset using SQL to understand trends, correlations, and key predictors of heart disease. By leveraging SQL's data querying and aggregation capabilities, the project aims to transform raw medical data into meaningful insights that can assist in identifying high-risk patients and common contributing factors.



Scope of the Project

- The project uses a Heart Disease dataset containing 1,190 patient records and 12 medical attributes.
- SQL is used to perform data analysis, including data extraction, filtering, grouping, and aggregation.
- The analysis focuses on identifying patterns and relationships among factors such as age, sex, cholesterol, blood pressure, and chest pain type.
- Queries are categorized into beginner, intermediate, and advanced levels for structured understanding.
- The project is limited to data analysis only – no machine learning or visualization tools are applied.
- The outcome highlights key insights and risk trends related to heart disease using SQL-based analytical techniques.

Dataset Overview

	age	sex	chest_pain_type	resting_bp_s	cholesterol	fasting_blood_sugar	resting_ecg	max_heart_rate	exercise_angina	oldpeak	st_slope	target
	40	1	2	140	289	0	0	172	0	0	1	0
	49	0	3	160	180	0	0	156	0	1	2	1
	37	1	2	130	283	0	1	98	0	0	1	0
	48	0	4	138	214	0	0	108	1	1.5	2	1
	54	1	3	150	195	0	0	122	0	0	1	0
	39	1	3	120	339	0	0	170	0	0	1	0
	45	0	2	130	237	0	0	170	0	0	1	0
	54	1	2	110	208	0	0	142	0	0	1	0
	37	1	4	140	207	0	0	130	1	1.5	2	1
	48	0	2	120	284	0	0	120	0	0	1	0
	37	0	3	130	211	0	0	142	0	0	1	0
	58	1	2	136	164	0	1	99	1	2	2	1
	39	1	2	120	204	0	0	145	0	0	1	0

Dataset Analysis using SQL

- Count the total number of patients.

```
SELECT COUNT(*) AS total_patients FROM heart_disease;
```

	total_patients
▶	1190

- Find the number of male and female patients.

```
SELECT  
CASE WHEN SEX=1 THEN 'MALE' ELSE 'FEMALE' END AS GENDER,COUNT(*) AS  
TOTAL_PATIENT  
FROM heart_disease  
GROUP BY SEX;
```

	GENDER	TOTAL_PATIENT
▶	MALE	909
	FEMALE	281

- Find patients with fasting blood sugar greater than 120 mg/dl.

```
SELECT COUNT(*) as fasting_blood_sugar_patients  
FROM heart_disease  
WHERE fasting_blood_sugar=1;  
##(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
```

	fasting_blood_sugar_patients
▶	254

- Find the average age of all patients.

```
SELECT FLOOR(AVG(AGE)) as Avg_Age FROM heart_disease;
```

	Avg_Age
▶	53

- Count the total number of patients who have exercise-induced angina.

```
SELECT COUNT(*) as exercise_angina_patients  
FROM heart_disease  
WHERE exercise_angina=1;  
##1 = yes; 0 = no
```

	exercise_angina_patients
▶	461

- List distinct chest pain types found in the dataset.

```
SELECT DISTINCT CHEST_PAIN_TYPE,
CASE WHEN CHEST_PAIN_TYPE=1 THEN 'typical angina'
WHEN CHEST_PAIN_TYPE=2 THEN 'atypical angina '
WHEN CHEST_PAIN_TYPE=3 THEN 'non-anginal pain '
ELSE 'asymptomatic' END AS CHEST_PAIN
FROM heart_disease
ORDER BY CHEST_PAIN_TYPE
```

	CHEST_PAIN_TYPE	CHEST_PAIN
▶	1	typical angina
	2	atypical angina
	3	non-anginal pain
	4	asymptomatic

- Find the maximum, minimum, and average cholesterol level.

```
SELECT
MAX(cholesterol) as Maximum_cholesterol,
MIN(cholesterol) as Minimum_cholesterol,
AVG(cholesterol) as Average_cholesterol
FROM heart_disease;
```

	Maximum_cholesterol	Minimum_cholesterol	Average_cholesterol
▶	603	0	210.3639

- Count how many patients have heart disease (target=1) vs normal (target=0).

```
SELECT target,
CASE WHEN target=1 then 'heart disease'
ELSE 'normal' end as class,
COUNT(*) as total_heart_disease_patient
FROM heart_disease
group by target;
```

target	class	total_heart_disease_patient
0	normal	561
1	heart disease	629

- Find the average resting blood pressure for patients with and without heart disease.

```
SELECT target,
CASE WHEN target=1 then 'heart disease'
ELSE 'normal'
end as class,
AVG(resting_bp_s) as Avg_Resting_blood_pressure
FROM heart_disease
group by target;
```

target	class	Avg_Resting_blood_pressure
0	normal	129.7932
1	heart disease	134.2591

- Group patients by chest pain type and find the average cholesterol for each group.

```
SELECT chest_pain_type,
CASE WHEN CHEST_PAIN_TYPE=1 THEN 'typical angina'
WHEN CHEST_PAIN_TYPE=2 THEN 'atypical angina '
WHEN CHEST_PAIN_TYPE=3 THEN 'non-anginal pain '
ELSE 'asymptomatic' END AS chest_pain,
avg(cholesterol) as avg_cholesterol
FROM heart_disease
group by chest_pain_type
order by chest_pain_type;
```

chest_pain_type	chest_pain	avg_cholesterol
1	typical angina	216.4242
2	atypical angina	235.7593
3	non-anginal pain	210.9647
4	asymptomatic	200.6752

- Find the average maximum heart rate by gender.

```
SELECT sex,
CASE WHEN SEX=1 THEN 'MALE' ELSE 'FEMALE'
END AS GENDER,
avg(max_heart_rate) as avg_max_heart_rate
FROM heart_disease
group by sex
order by sex;
```

sex	GENDER	avg_max_heart_rate
0	FEMALE	148.0747
1	MALE	137.1540

- Find how age affects heart disease: group patients by age ranges (e.g., 30-40, 41-50, etc.) and show average target value.

SELECT

```
CASE WHEN age BETWEEN 20 AND 30 THEN '20-30'
WHEN age BETWEEN 31 AND 40 THEN '31-40'
WHEN age BETWEEN 41 AND 50 THEN '41-50'
WHEN age BETWEEN 51 AND 60 THEN '51-60'
ELSE '61+' END AS age_group,
AVG(TARGET)*100 as heart_disease_percent
FROM heart_disease
group by age_group
order by age_group;
```

age_group	heart_disease_percent
20-30	0.0000
31-40	37.2549
41-50	38.3051
51-60	56.2753
61+	68.2594

- List top 10 patients with the highest cholesterol levels.

```
SELECT * FROM heart_disease
ORDER BY cholesterol
DESC LIMIT 10;
```

age	sex	chest_pain_type	resting_bp_s	cholesterol	fasting_blood_sugar	resting_ecg	max_heart_rate	exercise_angina	oldpeak	st_slope	target
54	1	4	130	603	1	0	125	1	1	2	1
67	0	3	115	564	0	2	160	0	1.6	2	0
67	0	3	115	564	0	2	160	0	1.6	2	0
32	1	4	118	529	0	0	130	0	0	2	1
53	1	3	145	518	0	0	130	0	0	2	1
44	1	4	135	491	0	0	135	0	0	2	1
53	0	2	113	468	0	0	127	0	0	1	0
40	1	4	120	466	1	0	152	1	1	2	1
58	1	4	132	458	1	0	69	0	1	3	0
65	0	3	140	417	1	2	157	0	0.8	1	0

- Find correlation trends: For example, check how many patients with high cholesterol (>240) also have heart disease.

```
SELECT COUNT(*) as High_cholesterol_hear_disease_patient  
FROM heart_disease  
WHERE cholesterol>240 AND TARGET=1;
```

High_cholesterol_hear_disease_patient

266

- Find the percentage of patients with exercise-induced angina who also have heart disease.

```
SELECT  
CONCAT(ROUND((SUM(CASE WHEN exercise_angina=1  
AND target=1 THEN 1 ELSE 0 END)*100.0 / COUNT(*)),2),'%')  
AS percentage_of_patients  
FROM heart_disease;
```

percentage_of_patients

32.18%

- Find the first 5 patients whose cholesterol levels are higher than the average cholesterol of all patients in the dataset.

```
SELECT * FROM heart_disease
WHERE cholesterol > (SELECT AVG(cholesterol) AS Avg_cholesterol FROM heart_disease)
limit 5;
```

age	sex	chest_pain_type	resting_bp_s	cholesterol	fasting_blood_sugar	resting_ecg	max_heart_rate	exercise_angina	oldpeak	st_slope	target
40	1	2	140	289	0	0	172	0	0	1	0
37	1	2	130	283	0	1	98	0	0	1	0
48	0	4	138	214	0	0	108	1	1.5	2	1
39	1	3	120	339	0	0	170	0	0	1	0
45	0	2	130	237	0	0	170	0	0	1	0

- Find the most common combination of symptoms (e.g., chest pain type + fasting blood sugar + target).

```
SELECT chest_pain_type, fasting_blood_sugar, target,
COUNT(*) AS freq
FROM heart_disease
GROUP BY chest_pain_type, fasting_blood_sugar, target
ORDER BY freq DESC
LIMIT 5;
```

chest_pain_type	fasting_blood_sugar	target	freq
4	0	1	338
2	0	0	169
3	0	0	160
4	1	1	145
4	0	0	133

- Calculate the risk rate of heart disease for males vs females (percentage).

```
SELECT sex,
CASE WHEN SEX=1 THEN 'MALE' ELSE 'FEMALE' END AS GENDER,
ROUND((SUM(CASE WHEN TARGET=1 THEN 1 ELSE 0 END)/COUNT(*))*100,2) as
risk_rate
FROM heart_disease
GROUP BY sex;
```

sex	GENDER	risk_rate
1	MALE	61.50
0	FEMALE	24.91

- Find the Top 5 most at-risk patients who show multiple high-risk conditions – including:High cholesterol (>240 mg/dl),High resting blood pressure (>140 mm Hg),Exercise-induced angina (exercise_angina = 1),Confirmed heart disease (target = 1).

```
SELECT age,sex,cholesterol,resting_bp_s,exercise_angina,max_heart_rate,oldpeak,target
FROM heart_disease
WHERE cholesterol > 240 AND resting_bp_s > 140
AND exercise_angina = 1 AND target = 1
ORDER BY cholesterol DESC, resting_bp_s DESC
LIMIT 5;
```

age	sex	cholesterol	resting_bp_s	exercise_angina	max_heart_rate	oldpeak	target
58	0	393	180	1	110	1	1
56	1	388	170	1	122	2	1
67	1	384	160	1	130	0	1
67	1	369	146	1	110	1.9	1
48	1	355	160	1	99	2	1

Key Insights

- **Age and Gender Impact:** Heart disease occurrence increases significantly after age 40, with men showing a higher risk rate compared to women.
- **Cholesterol Levels:** Patients with cholesterol levels above 240 mg/dl are highly prone to heart disease, indicating a strong link between cholesterol and cardiac risk.
- **Blood Pressure Influence:** Those with resting blood pressure greater than 140 mm Hg are more likely to have heart disease, highlighting hypertension as a major contributing factor.
- **Exercise-Induced Angina:** A noticeable percentage of patients with exercise-induced angina also have heart disease, emphasizing the importance of physical activity monitoring.
- **Chest Pain Type Correlation:** The asymptomatic and atypical angina categories show the highest heart disease incidence among all chest pain types.
- **Combined Risk Factors:** Individuals with high cholesterol, high blood pressure, and angina together represent the most at-risk group for heart disease.

Conclusion

This project successfully demonstrates how SQL-based data analysis can extract meaningful medical insights from raw patient data.

Through systematic querying, filtering, and aggregation, it reveals critical patterns connecting age, gender, cholesterol, blood pressure, and lifestyle factors to heart disease risk.

The study highlights that data-driven healthcare analysis using SQL is not only effective for understanding patient trends but also a valuable tool for early identification of high-risk individuals.

In conclusion, this analysis supports the idea that preventive care and lifestyle management, guided by data insights, can play a key role in reducing heart disease prevalence.



The End