

# Heart Attack SQL Analysis

SQL Queries, Procedures, and Insights



# Table of Contents

S. No.	Section	Details
1	Introduction	Overview of the project and purpose of the analysis
2	Abstract	Summary of the dataset, approach, and key analytical objectives
3	About the Dataset	Source and structure of the South Africa Heart Dataset
4	Objectives of the Analysis	Goals and research questions driving the analysis
5	SQL Queries and Insights	Analysis by gender, cholesterol, stress, lifestyle, etc.
6	Stored Procedures	Reusable SQL procedures for targeted queries
7	Triggers for Data Validation	Ensuring data consistency before insertion
8	Key Insights and Visualizations	Summary of findings and graphical representations
9	Conclusion	Final observations and implications of the study

# HEART ATTACK DATASET

Patient_ID	Age	Gender	Cholesterol_Pressure	Pressure	Pressure	DSmoking	SAcohol	In Physical	AObesity	In Diabetes	Family_His	Diet_Qual	Stress_Lev	Heart_Attr	Medication	Triglycerid	LDL_Level	HDL_Level	Heart_Attack_Outcome
1	76	Female	156	94	79	No	High	Sedentary	35.2	No	No	Good	High	No	No	264	141	65	0
2	39	Female	160	185	88	No	Low	Sedentary	21.3	No	No	Average	Medium	No	Yes	150	81	42	0
3	85	Male	254	173	113	Yes	Moderate	Highly Active	32	Yes	No	Good	Low	No	No	220	124	70	1
4	45	Female	261	187	65	No	High	Sedentary	36.4	No	Yes	Average	Low	No	Yes	141	56	38	1
5	48	Male	206	189	84	No	Low	Active	26.6	No	No	Good	High	No	No	221	186	26	0
6	27	Female	228	136	100	No	Low	Active	28.7	No	No	Average	Medium	No	Yes	130	145	26	0
7	46	Male	278	171	76	No	Moderate	Sedentary	22.5	No	No	Average	High	Yes	No	257	156	45	0
8	77	Male	272	145	112	No	Low	Active	28.2	Yes	No	Average	Medium	No	No	221	99	44	1
9	26	Male	166	100	62	Yes	High	Sedentary	25.8	No	No	Poor	Medium	No	No	226	103	54	0
10	54	Male	215	168	72	Yes	Low	Active	25.7	No	No	Poor	Low	No	No	145	114	63	0
11	62	Female	253	179	61	No	Low	Sedentary	26.4	No	No	Average	High	No	Yes	263	142	28	1
12	26	Male	273	193	100	Yes	Low	Sedentary	23.6	No	No	Average	Low	No	Yes	51	106	65	1
13	88	Female	207	95	66	No	Low	Sedentary	33.2	No	No	Good	Medium	No	Yes	162	102	68	0
14	84	Female	197	138	89	Yes	High	Sedentary	35.9	No	Yes	Poor	Medium	Yes	Yes	116	170	46	1
15	45	Female	212	141	78	Yes	Moderate	Highly Active	24	No	Yes	Poor	Medium	Yes	No	65	144	62	1
16	57	Female	150	152	68	No	Low	Sedentary	25.6	No	No	Average	High	No	Yes	199	77	69	0
17	82	Female	188	146	67	No	Low	Sedentary	22.2	No	No	Average	High	No	No	236	77	49	0
18	46	Male	271	111	103	No	Moderate	Highly Active	39.5	No	Yes	Average	Medium	No	Yes	277	95	75	1
19	73	Female	294	110	106	No	High	Sedentary	31.4	Yes	No	Average	Low	No	Yes	270	140	69	1
20	83	Female	263	136	97	No	Low	Highly Active	24.3	No	No	Poor	Medium	No	No	151	166	58	1
21	66	Male	294	145	82	No	Moderate	Sedentary	21.1	No	No	Good	Medium	No	No	177	79	29	1
22	84	Female	179	176	64	No	Moderate	Sedentary	23.5	No	No	Poor	Medium	No	No	173	92	24	0
23	39	Male	252	198	103	Yes	Low	Sedentary	30.4	Yes	No	Average	Medium	No	No	237	114	50	1
24	86	Female	266	154	106	No	Moderate	Active	35.1	No	Yes	Average	Medium	No	Yes	129	175	65	1
25	86	Female	160	185	66	Yes	Moderate	Active	23.2	No	No	Poor	Medium	No	No	172	123	68	1



# INTRODUCTION

This project presents a comprehensive analysis of heart attack data using Structured Query Language (SQL). The objective is to uncover significant health patterns and risk factors contributing to heart disease by leveraging SQL queries, procedures, and triggers. The analysis focuses on key variables such as cholesterol levels, stress, diet quality, obesity index, physical activity, and genetic predisposition. It also includes demographic breakdowns (e.g., gender and diabetes status) to evaluate how different factors impact heart attack prevalence across various groups. To ensure robust and actionable insights, stored procedures were implemented to automate repetitive queries, and triggers were used to maintain data integrity. Each SQL query is followed by its output and a brief insight, making the findings easy to interpret and use for healthcare decision-making or preventive care planning. This project ultimately serves as a practical demonstration of using SQL in medical data analytics, offering valuable guidance for future data-driven health initiatives.

# ABSTRACT

This project explores the use of SQL for analyzing heart attack-related data, aiming to identify key health risk factors and trends. Through a series of structured queries, procedures, and triggers, the analysis provides insights into cholesterol levels, stress, obesity, lifestyle habits, and demographic influences such as gender and diabetes status. By grouping and filtering patient data, the project highlights the prevalence of heart attacks across different health and lifestyle categories. Stored procedures were used to streamline repeated analyses, while triggers ensured the accuracy of data entries. The results offer a data-driven understanding of heart attack risks, supporting both preventive strategies and clinical research in cardiovascular health.

# About the Dataset

The South Africa Heart Attack Dataset is a medical dataset that contains patient-level information used to study the risk factors associated with heart disease. It includes various attributes such as age, gender, cholesterol levels, LDL and HDL levels, obesity index, diabetes status, smoking and alcohol habits, physical activity, diet quality, stress levels, and family history of heart disease. The dataset also records whether a patient has a history of heart attacks, serving as the target variable for analysis. Widely used in academic and analytical projects, this dataset supports the exploration of correlations between lifestyle, health conditions, and cardiovascular risk, particularly in a South African population.

## Q1: Genderwise Number of Patients who had Heart Attack :

- SELECT Gender, COUNT(\*) AS No\_of\_Patients FROM heart WHERE Heart\_Attack\_History="Yes" GROUP BY Gender;

	Gender	No_of_Patients
▶	Male	44
	Female	57

### Insight :

The query shows that there are 44 Males and 57 Females who had Heart attack.



Q2: Total No. of Heart Attack Cases Reported :

- `SELECT COUNT(*) AS Total_cases FROM heart WHERE Heart_Attack_History="Yes";`

	Total_cases
▶	101

Insight :

The Query shows that there are 101 heart attack cases reported in total.

Q3: Average Cholesterol Level of Patients :

- SELECT AVG(Cholesterol\_Level) AS avg\_cholesterol  
FROM heart;

	avg_cholesterol
▶	226.4347

Insight :

The Query shows that the average cholesterol level of patients is 226.4347

Q4: Patient with Highest Cholesterol Level :

- `SELECT * FROM heart ORDER BY Cholesterol_Level DESC LIMIT 1;`

	Patient_ID	Age	Gender	Cholesterol_Level	Blood_Pressure_Systolic	Blood_Pressure_Diastolic	Smoking_Status	Alcohol_Intake	Physical_Activity	Obesity_Index	Diabetes_Status
▶	312	84	Female	299	188	93	No	Low	Highly Active	32.8	32.8 ▶

Insight :

The Query shows that Female Patient with ID 312 of 84 years is the one with highest cholesterol level 299.

Q5: Average LDL Level of Heart Attack Patients :

- SELECT AVG(LDL\_Level) AS avg\_LDL FROM heart WHERE Heart\_Attack\_History="Yes";

	avg_LDL
▶	118.2475

Insight :

The Query shows that the average Low-Density Lipoprotein(LDL) of patients is 118.2475

Q6 : Patients with High Stress Levels :

- `SELECT Patient_ID, Age, Gender, Stress_Level FROM Heart WHERE Stress_Level="High";`

	Patient_ID	Age	Gender	Stress_Level
▶	1	76	Female	High
	5	48	Male	High
	7	46	Male	High
	11	62	Female	High
	16	57	Female	High

Insight :

The Query shows that the 5 patients with high stress levels are the ones in the above figure

Q7: Number of Patients with Poor Diet Quality :

- `SELECT COUNT(*) AS Poor_diet FROM heart WHERE Diet_Quality="Poor";`

	Poor_diet
▶	210

Insight :

The Query shows that there are 210 patients with poor diet quality.

## Q8: Genderwise Count of Patients who Smoke and are Alcohol Addicted :

- `SELECT Gender, COUNT(*) AS No_of_patients FROM heart WHERE Smoking_Status="Yes" AND Alcohol_Intake="High" GROUP BY Gender;`

	Gender	No_of_patients
▶	Male	23
	Female	19

### Insight :

The Query shows that there are 23 Males and 19 Females who smoke and are alcohol addicted.



Q9: Number of Patients with a Family History of Heart Disease :

- `SELECT COUNT(*) AS No_of_patients FROM heart WHERE Family_History_Heart_Disease="Yes";`

	No_of_patients
▶	173

Insight :

The Query shows that there are 173 patients who become heart patients by genetic.

Q10 : Number of Patients Taking Medication :

- SELECT COUNT(\*) AS Intaking\_medication FROM heart WHERE Medication\_Usage="Yes";

	Intaking_medication
▶	322

Insight :

The Query shows that there are 322 patients Intaking medication.

## Q11: Physically Inactive Patients Based on Gender :

- `SELECT Gender, COUNT(*) AS Physically_Inactive  
FROM heart WHERE Physical_Activity="Sedentary"  
GROUP BY Gender;`

	Gender	Physically_Inactive
▶	Female	161
	Male	170

### Insight :

The Query shows that there are 161 Females and 170 Males who are physically Inactive

Q12: Patient with Lowest HDL Level :

- select Patient\_ID, Age, Gender, HDL\_Level from heart order by HDL\_Level asc limit 1;

Patient_ID	Age	Gender	HDL_Level
598	58	Male	20

Insight :

The Query shows that the Male patient with ID 598 of 58 years old is the one with lowest High\_Density Lipoprotein (HDL) level 20.

## Q13: Number of Patients Grouped by Physical Activity Level :

- `SELECT Physical_Activity, COUNT(*) AS patient_count  
FROM heart GROUP BY Physical_Activity;`

	Physical_Activity	patient_count
▶	Sedentary	331
	Highly Active	140
	Active	203

### Insight :

The Query shows that there are 331 patients who are physically Inactive, 140 patients who are Highly active and 203 patients who are in between.

## Q14: Number of Patients Based on Medication Usage :

- SELECT Medication\_Usage, COUNT(\*) AS patient\_count FROM heart GROUP BY Medication\_Usage;

	Medication_Usage	patient_count
▶	No	352
	Yes	322

### Insight :

The Query shows that there are 352 patients who are not using medication and 322 patients who are using medication.

## Q15: Heart Attack Cases Grouped by Family History of Heart Disease :

- `SELECT Family_History_Heart_Disease, COUNT(*) AS heart_attack_count FROM heart WHERE Heart_Attack_History="Yes" GROUP BY Family_History_Heart_Disease;`

	Family_History_Heart_Disease	heart_attack_count
▶	No	69
	Yes	32

### Insight :

The Query shows that there are 69 patients who become heart patients by lifestyle and 32 patients who become heart patients by genetic.



## Q16: Number of Patients Based on LDL Cholesterol Level :

```
• SELECT
    CASE
        WHEN LDL_Level < 100 THEN "Low"
        WHEN LDL_Level BETWEEN 100 AND 160
        THEN "Moderate"
        ELSE "High"
    END AS LDL_Status,
    COUNT(*) AS patient_count FROM heart GROUP BY
LDL_Status;
```

	LDL_Status	patient_count
▶	Moderate	269
	Low	228
	High	178

### Insight :

The Query shows that there are 269 patients with Moderate LDL level, 228 with Low LDL level and 178 with High LDL level.

## Q17: Number of Patients Based on Gender and Diabetes Status :

- `SELECT Gender, Diabetes_Status, COUNT(*) AS patient_count FROM heart GROUP BY Gender, Diabetes_Status;`

	Gender	Diabetes_Status	patient_count
▶	Female	No	269
	Male	Yes	65
	Male	No	268
	Female	Yes	72

### Insight :

- The Query shows that there are 269 females with no diabetes and 72 with diabetes.
- Also there are 65 males with diabetes and 268 with no diabetes.

## Q18: Number of Patients Based on Diet Quality :

- `SELECT Diet_Quality, COUNT(*) AS patient_count FROM heart GROUP BY Diet_Quality;`

	Diet_Quality	patient_count
▶	Good	128
	Average	336
	Poor	210

### Insight :

The Query shows that there are 128 patients with Good diet quality, 336 with Average and 210 with Poor diet quality.

## Q19: Number of Heart Attack Cases Based on Obesity Index :

```
• SELECT
    CASE
        WHEN Obesity_Index < 20 THEN "Underweight"
        WHEN Obesity_Index BETWEEN 20 AND 30 THEN
            "Normal weight"
        WHEN Obesity_Index > 30 THEN "Overweight"
        ELSE "Nil"
    END AS Obesity_status ,
    COUNT(*) AS heart_attack_count FROM heart WHERE
    Heart_Attack_History="Yes" GROUP BY Obesity_status ;
```

	Obesity_status	heart_attack_count
►	Normal weight	48
	Over weight	42
	Under weight	11

### Insight :

The Query shows that there are 48 patients with normal weight had heart attack, 42 with over weight and also 11 patients with under weight had.

## Q20: Number of Patients Grouped by Stress Level :

- `SELECT Stress_Level, COUNT(*) AS No_of_patients  
FROM heart GROUP BY Stress_Level;`

	Stress_Level	No_of_patients
►	High	123
	Medium	333
	Low	218

### Insight :

The Query shows that there are 123 patients with High stress level, 333 with Medium and 218 with Low stress level.

Q21: No.of Patients with Higher Cholesterol than Average Cholesterol level :

```
•SELECT Cholesterol_Level, COUNT(*) AS No_of_patients
FROM heart WHERE Cholesterol_Level > (SELECT
AVG(Cholesterol_Level) FROM heart) GROUP BY
Cholesterol_Level;
```

	Cholesterol_Level	No_of_patients
▶	254	4
	261	5
	228	3
	278	4
	272	6

Insight :

The Query shows that there are 22 patients having higher cholesterol than average cholesterol level.

Q22: No.of Patients having similar ages more than 15 times :

- `SELECT age,COUNT(*) AS patient_count FROM HEART GROUP BY age HAVING COUNT(*) >15;`

	age	patient_count
▶	57	18
	86	19
	25	16

Insight :

The Query shows that there are 18 patients with the age of 57 years, 19 patients with the age of 86 years and 16 patients with the age of 25 years.



# Procedure ( IN & OUT )

❑ Procedure (IN) : How many males and females had Heart attack ?

• DELIMITER //

```
CREATE PROCEDURE Heart_attack_gender (IN  
Patient_Gender VARCHAR(10))
```

```
BEGIN
```

```
SELECT COUNT(*) AS Heart_attack_count FROM  
heart
```

```
WHERE Heart_Attack_History = "Yes" AND Gender =  
Patient_Gender;
```

```
END //
```

```
DELIMITER ;
```

```
CALL Heart_attack_gender ("Male");
```

```
CALL Heart_attack_gender ("Female");
```

	Heart_attack_count
▶	44

	Heart_attack_count
▶	86

Insight :

The Query shows that there are 44 males and 86 females who had heart attack .

❑ Procedure (IN) : How many patients with the history of heart attack also have diabetes & don't have ?

```
DELIMITER //
```

```
CREATE PROCEDURE heart_attack_diabetes (OUT  
Diabetes_Condition VARCHAR(25))
```

```
BEGIN
```

```
    SELECT COUNT(*) AS Heart_attack_count FROM  
heart
```

```
    WHERE Heart_Attack_History="Yes" AND  
Diabetes_Status=Diabetes_Condition;
```

```
END //
```

```
DELIMITER ;
```

```
CALL heart_attack_diabetes("Yes");
```

```
CALL heart_attack_diabetes("No");
```

	Heart_attack_count
▶	15

	Heart_attack_count
▶	86

Insight :

The Query shows that there are 15 patients who had heart attack and also are diabetic patients. Also there are 86 patients who had heart attack but not are diabetic patients.

❑ Procedure (OUT) : How many Patients have a cholesterol level higher than 200 ?

- DELIMITER //

```
CREATE PROCEDURE high_cstl_patients (OUT  
cholesterol_count INT)  
BEGIN  
    SELECT COUNT(*) INTO cholesterol_count FROM  
heart WHERE  
    Cholesterol_Level >200;  
END //  
DELIMITER ;
```

```
CALL high_cstl_patients (@C) ;  
SELECT @ C AS High_cstl_count;
```

	High_cstl_count
▶	453

Insight :

The Query shows that there are 453 patients with High cholesterol level.

# Trigger (Before Insert):

- DELIMITER //

```
CREATE TRIGGER before_stress BEFORE INSERT  
ON heart FOR EACH ROW
```

```
BEGIN
```

```
IF NEW.Stress_Level NOT IN (  
'Low','Medium','High') THEN
```

```
SET NEW.Stress_Level="Unknown";
```

```
END IF;
```

```
END //
```

```
DELIMITER ;
```

```
INSERT INTO heart (Patient_ID, Stress_Level) VALUES  
(675, "Extreme");
```

```
SELECT Patient_ID, Stress_Level FROM heart WHERE  
Patient_ID = 675;
```

	Patient_ID	Stress_Level
▶	675	Unknown

## Insight :

The trigger Prevented Invalid stress value from being Inserted and assigned "Unknown" for the case where stress level "Extreme" inserted ,which is Invalid based on the given condition.

# CONCLUSION

- This SQL-based analysis of heart attack data revealed critical patterns and correlations between various health indicators and heart disease.
- The study found strong associations between high cholesterol, poor diet quality, physical inactivity, obesity, and increased heart attack risk.
- Gender- and diabetes-specific trends further enriched the insights, highlighting vulnerable groups.
- The implementation of stored procedures enabled efficient, reusable analysis for common queries, while triggers ensured data integrity by preventing invalid entries. Together, these techniques demonstrated how SQL can be effectively used not only for data retrieval but also for automation and quality control.
- Overall, this project emphasizes the value of data-driven approaches in understanding cardiovascular risk factors and guiding healthcare interventions and preventive strategies.