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.
9	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
6	Conclusion	Final observations and implications of the study

HEART ATTACK DATASET

utcome																									
_Attack_0																									
el Heart	0	0	1	1	0	0	0	1	0	0	1	1	0	1	1	0	0	1	1	1	1	0	1	1	_
 년	59	45	70	38	76	26	45	44	54	63	28	65	89	46	62	69	49	75	69	28	29	24	20	65	89
LDL_Leve	141	81	124	99	186	145	156	66	103	114	142	106	102	170	144	11	11	95	140	166	79	95	114	175	173
riglycerid	264	150	220	141	221	130	257	221	226	145	263	51	162	116	65	199	236	777	270	151	177	173	237	129	177
ledicatio T	9	Yes	%	Yes	N _o	Yes	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No	No	No	Yes	No
eart_AttaN	9	No	9	No	No	No	Yes	No	No	No	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	9	No
ress_Lev H	High	Medium	Low	Low	High	Medium	High	Medium	Medium	Low	High	Low	Medium	Medium	Medium	High	High	Medium	Low	Medium	Medium	Medium	Medium	Medium	Medium
t_Quali Str	ро	Average N	po	Average	po	Average	Average	Average		٥٢	Average	Average				Average	Average	Average	Average		Good		Average	Average	
ily_His Die	G00d	Ave	G00d		G00d	Ave	Ave	Ave	Poor	Poor	Ave	Ave	Good	Poor	Poor	Ave	Ave		Ave	Poor	9	Poor	Ave		Poor
es_(Fan	N _o	No	No	Yes	No	N	N	No	No	No	No	No	No	Yes	Yes	No	No	Yes	No	No	No	No	No	Yes	S
In Diabet	8	No	Yes	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	Yes	9	N
Obesity_	35.2	21.3	32	36.4	26.6	28.7	22.5	28.2	25.8	25.7	26.4	23.6	33.2	35.9	24	25.6	22.2	39.5	31.4	24.3	21.1	23.5	30.4	35.1	23.2
Physical_A	Sedentary	Sedentary	Highly Acti	Sedentary	Active	Active	Moderate Sedentary	Active	Sedentary	Active	Sedentary	Sedentary	Sedentary	Sedentary	Highly Acti	Sedentary	Sedentary	Highly Acti	Sedentary	Highly Acti	Moderate Sedentary	Moderate Sedentary	Sedentary	Active	Active
Alcohol_I	High	Low	Moderate Highly	High	Low	Low	Moderate	Low	High	Low	Low	Low	Low	High	Moderate Highly	Low	Low	Moderate Highly	High	Low	Moderate	Moderate	Low	Moderate Active	Moderate Active
Smoking_	No	No	Yes	No	No	No	No	No	Yes	Yes	No	Yes	No	Yes	Yes	No	No	No	No	No	No	No	Yes	No	Yes
essure_D	6/	88	113	65	84	100	9/	112	62	72	61	100	99	68	78	89	29	103	106	6	82	64	103	106	99
ressure_91	94	185	173	187	189	136	171	145	100	168	179	193	95	138	141	152	146	111	110	136	145	176	198	154	185
Cholesterc Pressure_Bressure_D Smoking_S Alcohol_In Physical_A Obesity_In Diabetes_(Family_His Diet_Quali Stress_Lev Heart_Atta Medicatio Triglycerid LDL_Level HDL_Level Heart_Attack_Outcome	156	160	254	261	500	228	278	272	166	215	253	273	207	197	212	150	188	271	294	263	294	179	252	566	160
Gender	Female	Female	Male	Female	Male	Female	Male	Male	Male	Male	Female	Male	Female	Female	Female	Female	Female	Male	Female	Female	Male	Female	Male	Female	Female
Age	9/	33	85	45	48	77	46	11	76	54	62	76	88	84	45	57	82	46	73	83	99	84	39	98	98
Patient_ID	Ţ	2	c c	4	5	9	7	∞	6	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	7.5

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 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;

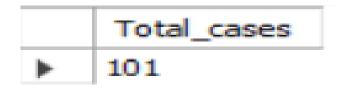
	Gender	No_of_Patients
>	Male	44
	Female	57

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";



Insight:

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

Q3: Average Cholesterol Level of Patients:

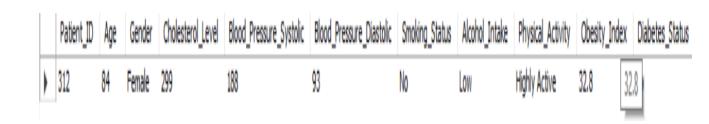
SELECT AVG(Cholesterol_Level) AS avg_cholestrol FROM heart;

<u>Insight</u>:

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;



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";



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";

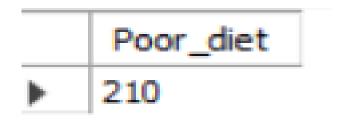
	Patient_ID	Age	Gender	Stress_Level
•	1	76	Female	High
	5	48	Male	High
	7	46	Male	High
	11	62	Female	High
	16	57	Female	High

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";



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

<u>Insight</u>:

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";

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";



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

<u>Insight</u>:

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 "Nill"
        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)



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
                                 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)
    BFGIN
    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):

	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 datadriven approaches in understanding cardiovascular risk factors and guiding healthcare interventions and preventive strategies.