

# Cardiovascular Disease Analysis Using SQL

## Dataset

- <https://www.kaggle.com/code/jocelyndumlao/cardiovascular-health-analysis/notebook>

## About

This project aims to explore and analyze the cardiovascular dataset by using SQL to extract meaningful insights about the factors contributing to cardiovascular diseases.

## Create the table for analysis


- Created a database and then a table in mysql workbench.
- Used “insert into” statement to import data used in analysis

```
3 • CREATE TABLE IF NOT EXISTS cardio(  
4     patientid      INTEGER NOT NULL PRIMARY KEY  
5     ,age            INTEGER NOT NULL  
6     ,gender         BIT NOT NULL  
7     ,chestpain      INTEGER NOT NULL  
8     ,restingBP      INTEGER NOT NULL  
9     ,serumcholesterol INTEGER NOT NULL  
10    ,fastingbloodsugar BIT NOT NULL  
11    ,restingelectro  INTEGER NOT NULL  
12    ,maxheartrate    INTEGER NOT NULL  
13    ,exerciseangia   BIT NOT NULL  
14    ,oldpeak         NUMERIC(3,1) NOT NULL  
15    ,slope           INTEGER NOT NULL  
16    ,noofmajorvessels INTEGER NOT NULL  
17    ,target          BIT NOT NULL  
18 );  
19 • INSERT INTO cardio(patientid,age,gender,chestpain,restingBP,serumcholesterol,fastingbloodsugar,restingelectro,maxheartrate,exerciseangia,oldpeak,slope,noofmajorvessels,target)  
20 • INSERT INTO cardio(patientid,age,gender,chestpain,restingBP,serumcholesterol,fastingbloodsugar,restingelectro,maxheartrate,exerciseangia,oldpeak,slope,noofmajorvessels,target)  
21 • INSERT INTO cardio(patientid,age,gender,chestpain,restingBP,serumcholesterol,fastingbloodsugar,restingelectro,maxheartrate,exerciseangia,oldpeak,slope,noofmajorvessels,target)
```

**Analysis:** Does gender affect the chance of having a cardiovascular disease?

- Answer: No, the data shows the percentage of people with heart disease were close to equal with 42% of males having a heart disease, and 43% of females having a heart disease.

```
-- 1) Which gender has the highest number of cardiovascular disease?  
-- Male or Female? (1 male, 0 female)  
  
select  
    gender,  
    count(*) as total_people,  
    count(case when target = 0 then 1 end) as no_heart_disease,  
    round(avg(case when target = 0 then 1 else 0 end), 2) as percentage_no_heart_disease  
from cardio  
GROUP BY gender;  
-- Conclusion: Men and Women have nearly equal odds of having heart-disease
```

Result Grid				
Filter Rows:		Export:  Wrap Cell Co		
	gender	total_people	no_heart_disease	percentage_no_heart_disease
▶	1	765	318	0.42
	0	235	102	0.43

**Analysis:** Does age group have an effect on whether a person has a cardiovascular disease?

- Answer: The results suggest people in the age group of 40-50 have the highest chance of having a heart disease. There is a pattern where the further you are from this age range, the lower the chance an individual has a heart disease.

```
-- 2) Which age group has the highest percentage of people with heart disease

select
  case
    when age < 20 then "<20"
    when age between 20 and 30 then "20-30"
    when age between 30 and 40 then "30-40"
    when age between 40 and 50 then "40-50"
    when age between 50 and 60 then "50-60"
    else "60+"
  end as age_group,
  count(*) as people_in_age_group,
  sum(case when target = 1 then 1 end) as count_with_heart_disease,
  avg(case when target = 1 then 1 else 0 end) as percentage_with_heart_disease
from cardio
group by age_group
order by age_group;

-- Conclusion: People in the 40-50 age group have the highest percentage with heart_disease,
-- There is a pattern where the further you are away from this age-group the lower the percentage
-- of people with heart_disease
```

	age_group	people_in_age_group	count_with_heart_disease	percentage_with_heart_disease
►	20-30	196	102	0.5204
	30-40	164	98	0.5976
	40-50	162	105	0.6481
	50-60	160	94	0.5875
	60+	318	181	0.5692

**Analysis:** Does chest pain have an effect on whether a person has a cardiovascular disease?

- Answer: Yes patients without chest pain (Chest Pain type is 0) Have a significantly lower chance of having a heart disease.

```
-- 3) Count the number of patients with each chest_pain type, do chest_pain types affect the likely hood of having
-- a cardiovascular disease?

select
    chestpain,
    count(target) as count_with_chestpain,
    sum(case when target = 1 then 1 end) as count_with_heart_disease,
    avg(case when target = 1 then 1 else 0 end) as percentage_with_heart_disease
from cardio
group by chestpain
order by chestpain;
```

	chestpain	count_with_chestpain	count_with_heart_disease	percentage_with_heart_disease
▶	0	420	108	0.2571
	1	224	154	0.6875
	2	312	279	0.8942
	3	44	39	0.8864

**Analysis:** Does an individual's cholesterol level affect their likelihood of having cardiovascular disease?

- Answer: It does have an effect, patients with an extremely high amount of cholesterol have a significantly higher chance of having a hard disease. Those with the highest cholesterol levels have a 83% chance to have a heart disease. However scrutiny must be applied here since cholesterol is grouped under 200 as healthy, 200-239 as at risk and 240+ as Dangerous. Since the maximum value far exceeds this the data is likely inaccurate. (<https://my.clevelandclinic.org/health/articles/11920-cholesterol-numbers-what-do-they-mean>)

cholesterol_group	patient_count	count_with_heart_disease	percentage_with_heart_disease
1-Very Low Cholesterol (85-149)	65	7	0.1077
2-Low Cholesterol (150-199)	58	36	0.6207
3-Borderline High Cholesterol (200-239)	85	47	0.5529
4-High Cholesterol (240-299)	200	82	0.4100
5-Very High Cholesterol (300-399)	276	136	0.4928
6-Extremely High Cholesterol (400-602)	263	219	0.8327

## Final Thoughts

From this cardiovascular disease analysis using SQL, several insights were gained about the contributing factors to heart disease. The data suggests that:

- Gender is not a major factor, as the percentages of males and females with cardiovascular disease are nearly identical.
- Age plays a significant role, with individuals aged 40-50 being at the highest risk. This suggests that preventive measures should target individuals as they approach middle age.
- Chest Pain is a strong indicator, as patients without chest pain have a much lower chance of heart disease.
- Cholesterol Levels also strongly correlate with cardiovascular disease, but care must be taken when interpreting the data. Extremely high cholesterol levels are associated with heart disease, but the data's upper limits appear inaccurate, which could skew results.

Knowing this, further action can be taken:

- Data Accuracy: Clean and verify outliers or inaccuracies in cholesterol levels to ensure the analysis is reliable.
- In hospitals using this information can be used to make a decision to add more prevention strategies for those in the 40-50 age group. Such as more checkups on cholesterol level since they are at risk more.

This analysis provides a strong foundation for understanding cardiovascular risk factors, but further refinement and deeper investigation into outliers are necessary for more accurate conclusions.