SQL and Power BI Report for Cardiovascular disease analysis

Contents

- I used SQL and Power BI to perform an analysis on a cardiovascular disease dataset.
- The dataset in question:
  https://www.kaggle.com/code/jocelyndumlao/cardiovascular-health-analysis/notebook

The process I used was the following:

1) First of all I had to import the data into mysql. I created a table in a database and afterwards added all the rows using insert statements.

```
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        ,serumcholestrol   INTEGER  NOT NULL
10       ,fastingbloodsugar BIT   NOT NULL
11       ,restingrelectro   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,serumcholestrol,fastingbloodsugar,restingrelectro,maxheartrate,exerciseangia,oldpeak,slope,noofmajorvessels,target)
20 ●  INSERT INTO cardio(patientid,age,gender,chestpain,restingBP,serumcholestrol,fastingbloodsugar,restingrelectro,maxheartrate,exerciseangia,oldpeak,slope,noofmajorvessels,target)
21 ●  INSERT INTO cardio(patientid,age,gender,chestpain,restingBP,serumcholestrol,fastingbloodsugar,restingrelectro,maxheartrate,exerciseangia,oldpeak,slope,noofmajorvessels,target)
```

2) After importing all the data I began analysis.

- I investigated if gender makes a difference in likelihood of getting a cardiovascular disease.

```
-- 1) Which gender has the highest number of cardiovascular diease?
-- 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

- I also Investigated if different age groups can cause a difference in likelihood of getting a cardiovascular 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

- I also investigated if different types of chest pain affect the likelihood 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

Conclusion: Patients with chest pain had a higher chance of heart disease than those without.

- Lastly I investigated if cholesterol levels Increase the likelihood of heart_disease.

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

Power BI
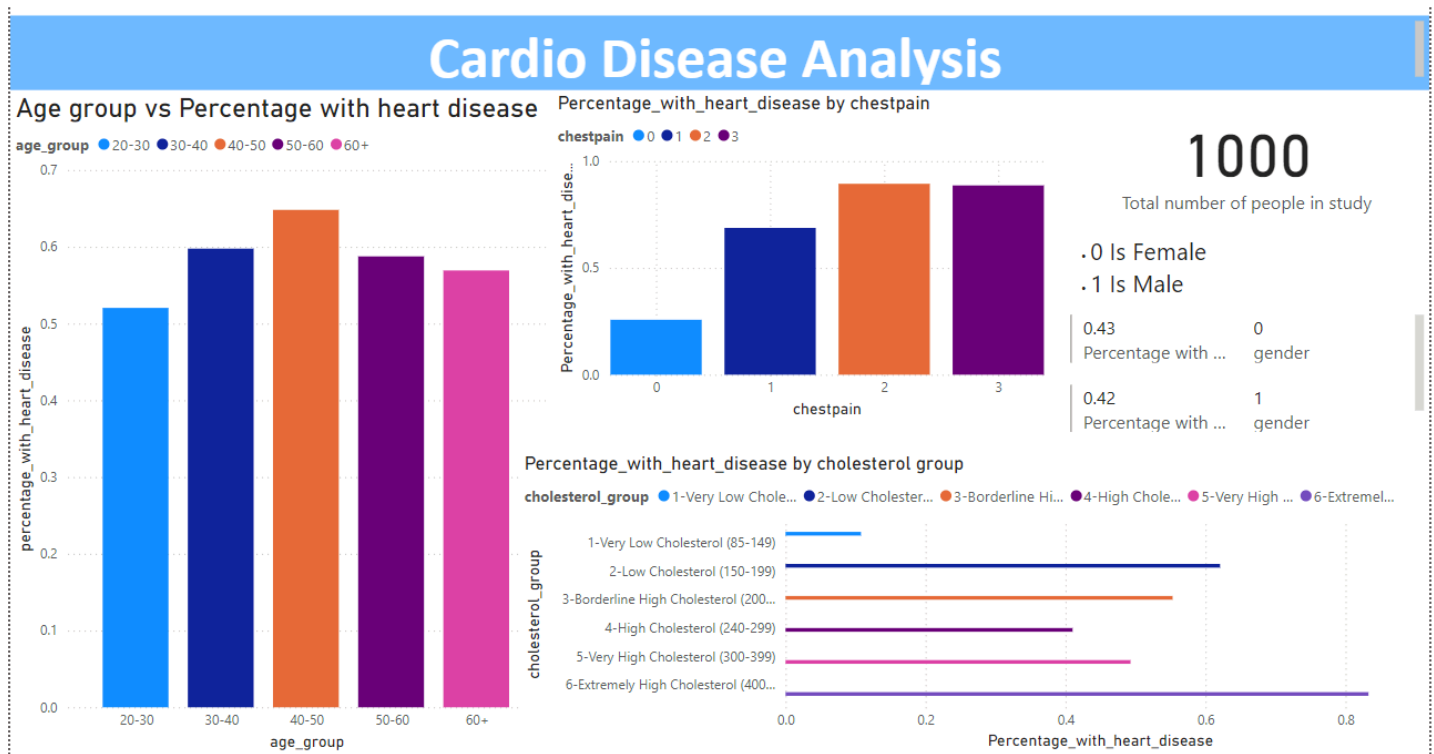
I made a dashboard with power BI based on the queries I created to show the results in a more visual way for results to be understood quickly.