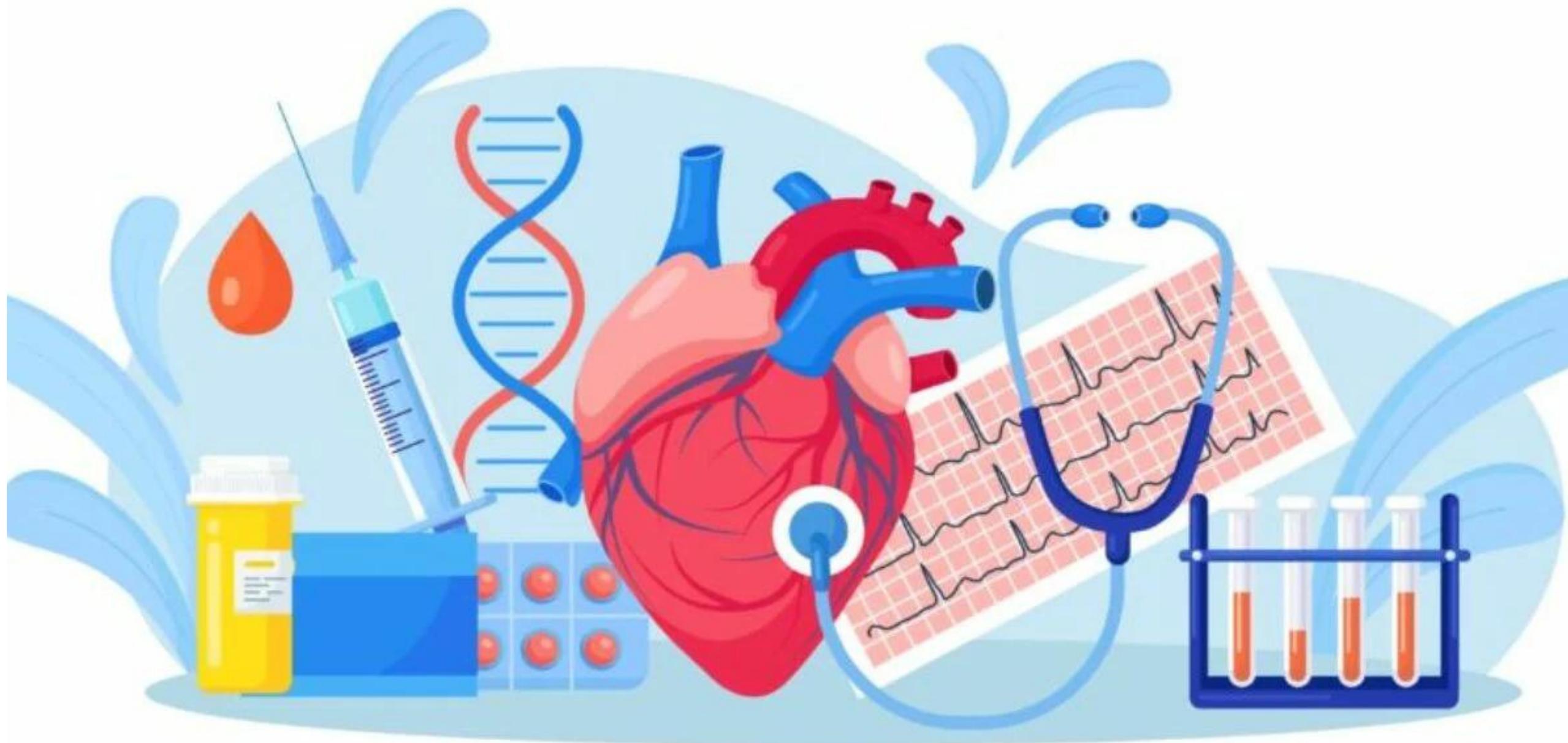


CARDIOVASCULAR DISEASE SQL ANALYSIS



PRESENTED BY ANUNAY DUBEY

PROJECT OVERVIEW

Using SQL queries, this project analyzes a cardiovascular dataset to uncover patterns and correlations in patient demographics and health metrics. The insights gained contribute to a deeper understanding of cardiovascular health, aiding medical research and decision-making.



Q1.Count the number of male and female patients:



```
SELECT gender, COUNT(*) AS count  
FROM disease  
GROUP BY gender;
```

Output

	gender character varying (6) 	count bigint 
1	Female	235
2	Male	765

Q2.Calculate the average serum cholesterol level for patients with and without fasting blood sugar above a certain threshold:



```
SELECT fastingsugar, AVG(serumcholesterol) AS average_cholesterol  
FROM disease  
GROUP BY fastingsugar;
```

Output

	fastingbloodsugar bit	average_cholesterol numeric
1	0	287.55539772727273
2	1	368.2702702702703

Q3.Find the age distribution of patients with chest pain, ordered by the count in descending order



```
SELECT age, COUNT(*) AS count
FROM disease
WHERE chestpain IS NOT NULL
GROUP BY age
ORDER BY count DESC;
```

Output

	age integer	lock	count bigint	lock
1	20		28	
2	58		23	
3	76		22	
4	24		22	
5	73		21	
6	45		21	
7	43		21	
8	46		21	
9	44		21	
10	25		20	
11	51		20	
12	74		20	
13	62		20	
14	22		19	
15	34		19	

Q4. Identify the top 5 patients with the highest serum cholesterol levels:

```
● ● ●  
SELECT *  
FROM disease  
ORDER BY serumcholesterol DESC  
LIMIT 5;
```

Output

	patientid [PK] integer	age integer	gender character varying (6)	chestpain integer	restingbp integer	serumcholestrol integer
1	3142651	39	Male	2	156	602
2	1932417	30	Male	1	193	602
3	7797411	52	Male	2	149	601
4	4503480	77	Male	2	169	601
5	8104279	51	Male	2	163	601

Q5. Identify patients with a higher than average serum cholesterol level in their respective gender and age groups:



```
WITH AvgCholesterol AS (
    SELECT gender, age, AVG(serumcholesterol) AS avg_cholesterol
    FROM disease
    GROUP BY gender, age
)
SELECT t.*
FROM disease t
JOIN AvgCholesterol a ON t.gender = a.gender AND t.age = a.age
WHERE t.serumcholesterol > a.avg_cholesterol;
```

Output

	patientid [PK] integer	age integer	gender character varying (6)	chestpain integer	restingbp integer	serumcholestrol integer
1	132514	43	Male	0	138	295
2	168686	79	Male	2	130	240
3	188225	62	Male	0	121	357
4	201030	59	Female	1	190	529
5	208877	58	Male	2	192	409
6	229445	58	Male	0	170	354
7	236763	42	Female	3	137	350
8	260870	35	Male	0	127	269
9	322287	72	Male	1	177	397
10	384606	40	Male	1	191	349
11	417218	72	Female	2	126	478
12	422067	49	Female	1	192	521

Q6.Calculate the cumulative count of patients by age group, ordered by age group and patient age:



```
SELECT age, patientid,  
       SUM(1) OVER (PARTITION BY age ORDER BY age, patientid) AS cumulative_count  
FROM disease;
```

Output

	age integer	patientid [PK] integer	cumulative_count bigint
1	20	463851	1
2	20	690148	2
3	20	1604440	3
4	20	1666741	4
5	20	1889122	5
6	20	2060446	6
7	20	2300065	7
8	20	2301297	8
9	20	2356338	9
10	20	3178004	10
11	20	3188178	11

Q7.Calculate the correlation coefficient between age and serum cholesterol levels for all patients:



```
SELECT CORR(age, serumcholesterol) AS correlation_coefficient FROM disease;
```

Thank you

Anunay Dubey