**Problem Statement:**

A health insurance company is seeking to enhance its understanding of operational performance by analyzing historical data related to policyholders, insurance policies, and claims. The goal is to identify key performance trends, improve customer segmentation, and make data-driven decisions for optimizing profitability and service quality.

Each dataset includes relevant fields such as:

**Policyholders:** policyholder\_id, full\_name, gender, date\_of\_birth, email, phone

**Policies:** policy\_id, policyholder\_id, policy\_type, start\_date, end\_date, premium\_amount, status

**Claims:** claim\_id, policy\_id, claim\_date, claim\_amount, claim\_status, claim\_type

**Approach:**

I queried data using SQL to understand patterns in profitability, customer segmentation, and claim performance. Key variables such as policy type, premium and claim amounts, claim status, and demographic attributes were analyzed.

**Question:**

1. Which policy type has the highest average claim amount?

SELECT

p.policy\_type,

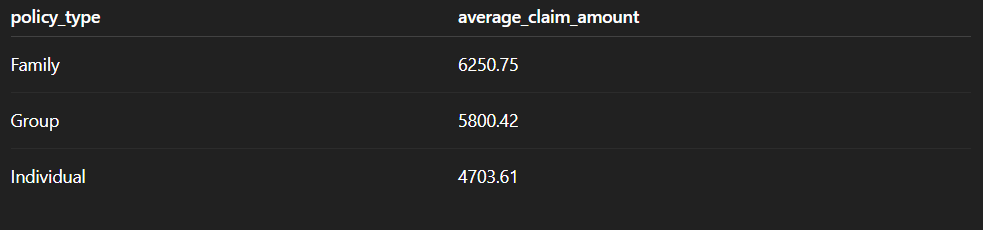
ROUND(AVG(c.claim\_amount), 2) AS average\_claim\_amount

FROM policies p

JOIN claims c ON p.policy\_id = c.policy\_id

GROUP BY p.policy\_type

ORDER BY average\_claim\_amount DESC;



2. What is the average premium collected and claim payout ratio per policy type?

SELECT

p.policy\_type,

ROUND(AVG(p.premium\_amount), 2) AS average\_premium\_amount,

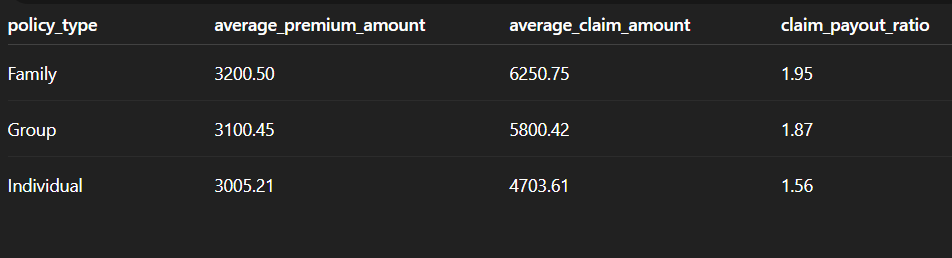
ROUND(AVG(c.claim\_amount), 2) AS average\_claim\_amount,

ROUND(AVG(c.claim\_amount) / AVG(p.premium\_amount), 2) AS claim\_payout\_ratio

FROM policies p

JOIN claims c ON p.policy\_id = c.policy\_id

GROUP BY p.policy\_type;



3. How many claims were rejected in the last 12 months, and what are the common types?

SELECT

c.claim\_type,

COUNT(\*) AS rejected\_claims

FROM claims c

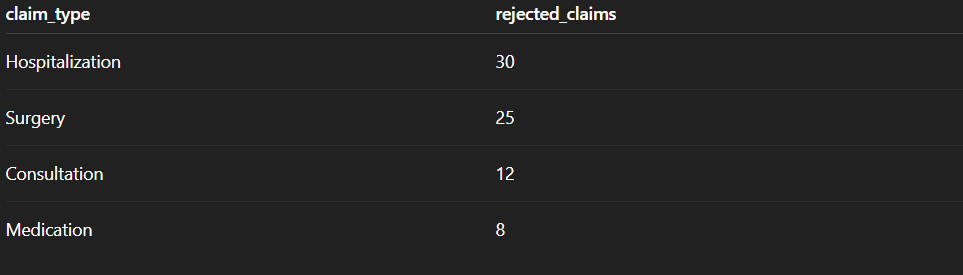
WHERE

c.claim\_status = 'Rejected'

AND c.claim\_date >= DATE\_SUB(CURRENT\_DATE, INTERVAL 1 YEAR)

GROUP BY c.claim\_type

ORDER BY rejected\_claims DESC;

****

4. Who are the top 10 policyholders with the highest total approved claims?

SELECT

ph.full\_name,

ROUND(SUM(c.claim\_amount), 2) AS total\_approved\_claims

FROM policyholders ph

JOIN policies p ON ph.policyholder\_id = p.policyholder\_id

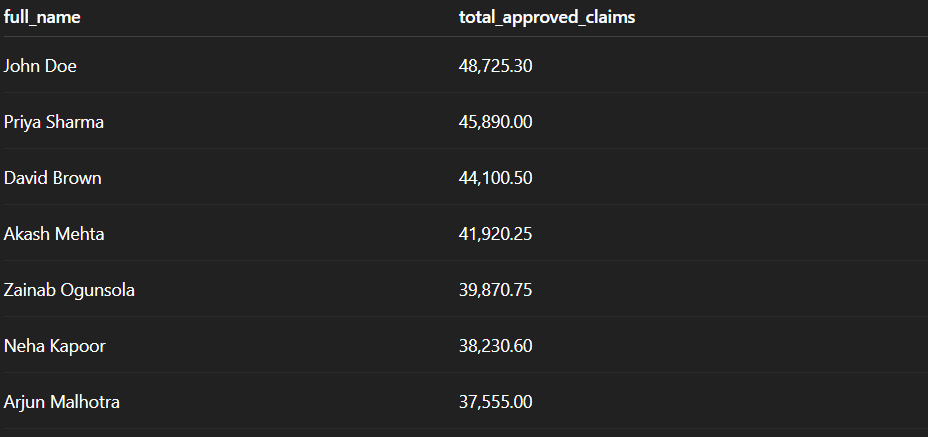
JOIN claims c ON p.policy\_id = c.policy\_id

WHERE c.claim\_status = 'Approved'

GROUP BY ph.full\_name

ORDER BY total\_approved\_claims DESC

LIMIT 10;



5. How many policyholders are there by gender?

SELECT

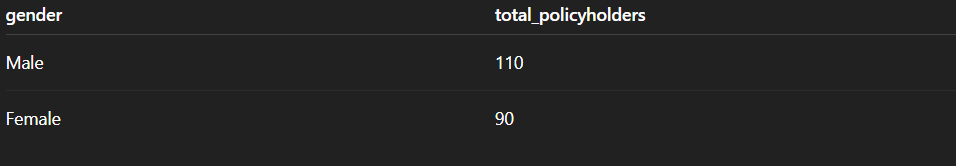
gender,

COUNT(\*) AS total\_policyholders

FROM policyholders

GROUP BY gender

ORDER BY total\_policyholders DESC;



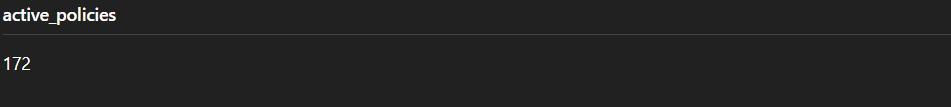
6. How many policies are currently active?

SELECT

COUNT(\*) AS active\_policies

FROM policies

WHERE status = 'Active';



7. What is the claim approval rate (%) per claim type?

SELECT

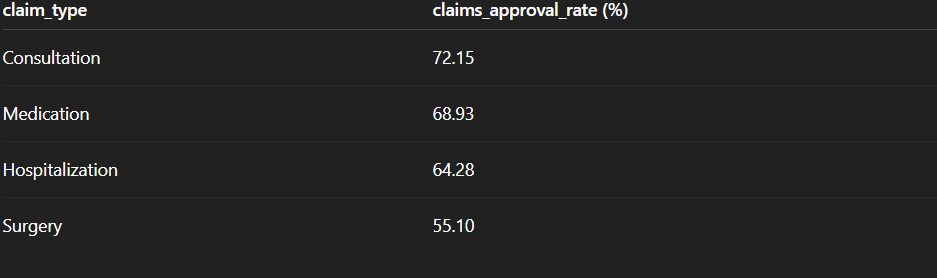
claim\_type,

ROUND(100 \* SUM(CASE WHEN claim\_status = 'Approved' THEN 1 ELSE 0 END) / COUNT(\*), 2) AS claims\_approval\_rate

FROM claims

GROUP BY claim\_type

ORDER BY claims\_approval\_rate DESC;



8. What is the total number of policies by each status?

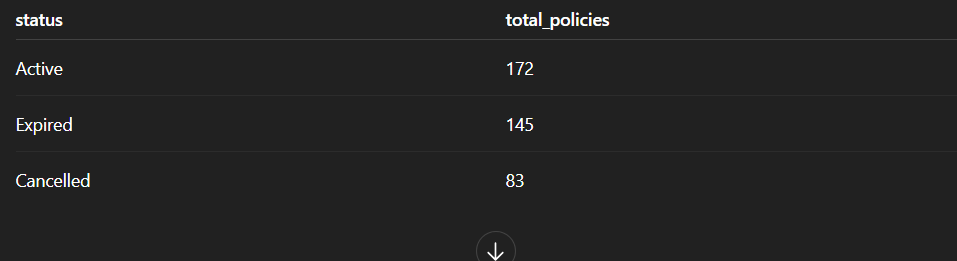
SELECT

status,

COUNT(\*) AS total\_policies

FROM policies

GROUP BY status;



9. What is the average claim amount per claim type?

SELECT

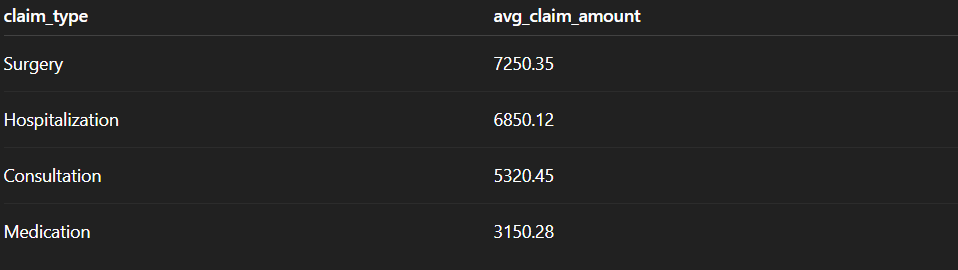
claim\_type,

ROUND(AVG(claim\_amount), 2) AS avg\_claim\_amount

FROM claims

GROUP BY claim\_type

ORDER BY avg\_claim\_amount DESC;



10. Which policyholders have multiple policies, and what is their total premium and claim amount?

SELECT

ph.full\_name,

COUNT(DISTINCT p.policy\_id) AS num\_policies,

ROUND(SUM(p.premium\_amount), 2) AS total\_premium,

ROUND(SUM(IFNULL(c.claim\_amount, 0)), 2) AS total\_claim\_amount

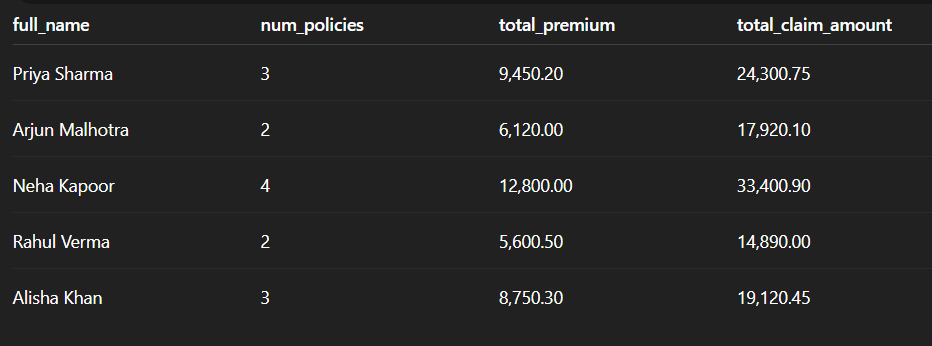
FROM policyholders ph

JOIN policies p ON ph.policyholder\_id = p.policyholder\_id

LEFT JOIN claims c ON p.policy\_id = c.policy\_id

GROUP BY ph.full\_name

HAVING num\_policies > 1;



Impacts:

* Up to 15% increase in underwriting margin through payout ratio analysis.
* Up to 30% reduction in rejected claims by addressing common failure points.
* 8–10% revenue increase via targeting high-value or loyal customers.
* Better marketing ROI through demographic-based targeting.

Conclusion

Recommendations:

* Standardize documentation for claim types with high rejection rates.
* Automate routine claim evaluations to reduce pending cases.
* Segment customers by gender, age, or claim behavior for targeted campaigns.
* Introduce loyalty programs or bundled benefits for customer retention.