### \*\*Cost and Usage Trends:\*\*

1. \*\*What are the monthly cost trends?\*\*

- Identify months where cloud costs spiked and analyze the reasons (e.g., increased traffic, inefficient usage).

2. \*\*Which services incur the highest costs?\*\*

- Break down costs by services such as EC2 instances, storage, and networking to determine which services are the most expensive.

3. \*\*Which regions or zones are the most cost-efficient?\*\*

- Analyze costs by geographical region to find out if using certain regions could be more cost-effective.

4. \*\*How does resource usage vary over time (daily, weekly, monthly)?\*\*

- Explore patterns in resource usage to identify whether the company can reduce resources during certain time periods (e.g., off-peak hours).

### \*\*Resource Optimization:\*\*

5. \*\*Are there underutilized resources that can be downsized or decommissioned?\*\*

- Identify any virtual machines or services that are underutilized but still running, helping to lower operational costs.

6. \*\*How can reserved instances save costs compared to on-demand instances?\*\*

- Evaluate whether shifting from on-demand instances to reserved instances would reduce costs for long-running processes.

7. \*\*Which services or resources have sudden increases in usage, and why?\*\*

- Identify services where the usage and costs have increased unexpectedly and investigate the causes (e.g., poor scaling configurations).

8. \*\*How does the storage usage grow over time, and are there any opportunities to reduce costs by optimizing storage tiers?\*\*

- Track storage growth and evaluate whether migrating to lower-cost storage tiers for rarely accessed data would be beneficial.

### \*\*Billing and User Behavior:\*\*

9. \*\*Which department or team generates the most cloud costs?\*\*

- Break down costs by department or project to allocate expenses and identify areas where cloud spend can be better controlled.

10. \*\*How does cost per unit (e.g., per user or per transaction) fluctuate?\*\*

- Determine whether the cost per user or transaction is increasing, helping businesses optimize for user-level efficiency.

### \*\*Predictive Insights:\*\*

11. \*\*What would future cloud costs look like based on historical data?\*\*

- Use historical trends to predict future cloud spending and budget accordingly.

12. \*\*What are potential cost savings with better scaling policies?\*\*

- Analyze whether auto-scaling policies could reduce costs by adjusting resources automatically during high and low traffic periods.

Each of these questions can drive impactful cost-saving strategies, helping businesses to optimize their cloud infrastructure expenses effectively.

### Business Problem Questions:

1. \*\*Resource Utilization Analysis:\*\*

- Which cloud resources (e.g., instances, volumes, or buckets) are being used the most?

- How many times does each resource type appear?

2. \*\*Cost Optimization:\*\*

- Which resources are being overused or duplicated (e.g., same instance ID used multiple times)?

- Are there underutilized resources that could be decommissioned to save costs?

3. \*\*Infrastructure Dependency:\*\*

- How many different types of cloud services (S3, RDS, EC2) are utilized in the infrastructure?

- What dependencies or relationships exist between resources (e.g., an instance with attached volumes)?

**Exploratory Data Analysis (EDA):**

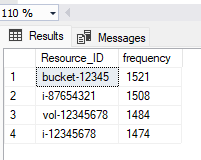
1. \*\*Frequency Distribution:\*\*

- Analyze the frequency of resource types (e.g., `i-` for instances, `vol-` for volumes).

select Resource\_ID, count(Resource\_ID) as frequency from cloud\_infrastructure\_cost\_optimization\_cleaned

group by Resource\_ID

order by 2 desc



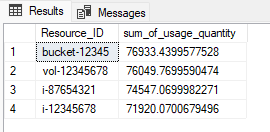
2. \*\*Grouping by Resource Type:\*\*

- Group the data by resource type (e.g., instances, volumes, buckets) to understand their distribution.

select Resource\_ID, sum(Usage\_Quantity) as sum\_of\_usage\_quantity from cloud\_infrastructure\_cost\_optimization\_cleaned

group by Resource\_ID

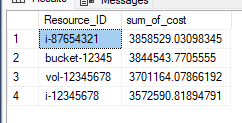
order by 2 desc



select Resource\_ID, sum(cost) as sum\_of\_cost from cloud\_infrastructure\_cost\_optimization\_cleaned

group by Resource\_ID

order by 2 desc



3. \*\*Duplicate Resources:\*\*

- Identify any duplicate resources that could indicate inefficiencies (e.g., duplicate instance IDs).

4. \*\*Trend Analysis:\*\*

- If there’s a timestamp or time-based information in the dataset, you could analyze trends in resource utilization over time.

To perform these analyses in Python, you can use libraries like pandas to group, aggregate, and visualize the data. Let me know if you'd like specific code examples for any of these analyses!