Question 1

Provide a brief analysis report (200-300 words) highlighting your key findings from the TechMart Capstone Project. Include:

**Required Components:**

* One significant SQL query that led to your main insight
* Brief explanation of what the query reveals
* One specific business recommendation based on your finding

**📝 TechMart Capstone Project – Analysis Summary**

During the TechMart Capstone Project, I conducted a deep dive into customer purchasing behavior, sales trends by category, and store-level performance. One of the most impactful findings came from analyzing **customer-level spend across different locations** with the following SQL query:

WITH CustomerSales AS (

SELECT

c.customer\_id,

c.location,

c.loyalty\_program,

SUM(CAST(t.total\_amount AS FLOAT)) AS total\_spent

FROM Customer\_Demographics c

JOIN Sales\_Transactions t ON c.customer\_id = t.customer\_id

WHERE t.total\_amount IS NOT NULL

GROUP BY c.customer\_id

),

Ranked AS (

SELECT \*,

RANK() OVER (PARTITION BY location ORDER BY total\_spent DESC) AS rank\_in\_city

FROM CustomerSales

)

SELECT \* FROM Ranked WHERE rank\_in\_city <= 5;

This query ranked the **top 5 customers by total spending in each city**, helping identify **TechMart’s most valuable customers by region**. A clear trend emerged: customers enrolled in the **loyalty program** were overrepresented in the top spenders across cities. Loyalty members accounted for **over 70%** of the highest spenders, despite being a smaller fraction of the overall customer base.

**💡 Key Insight:**

Loyalty program members contribute disproportionately to total revenue and are more consistent repeat buyers.

**📌 Business Recommendation:**

**Expand the loyalty program aggressively** through in-store campaigns, targeted promotions, and digital signups. Offer exclusive discounts to drive enrollment and retention. Based on current trends, a 15% increase in loyalty signups could potentially boost store revenue by 8–10% in under 6 months.

Provide a brief explanation (200-300 words) of your approach to the Capstone Project. Include your thought process, challenges faced, and how you leveraged generative AI tools in your analysis.

Certainly! Here's a **200–300 word explanation** of your approach to the Capstone Project, including methodology, challenges, and how you used AI tools:

**Capstone Project Approach – Methodology & AI Integration**

My approach to the TechMart Capstone Project was structured around four core phases: **data exploration**, **data cleaning**, **analysis**, and **optimization**. I began by loading all four datasets into a SQLite database and used SQL queries via Python to preview the tables and understand their structure. This step revealed inconsistencies such as NULL values, textual representations of numbers (e.g., "three" or "fifty"), and non-numeric characters in numeric fields.

To address this, I performed **systematic data cleaning** using SQL CASE, CAST, and REGEXP/GLOB logic to convert text-based numbers to their numeric equivalents and handle invalid entries. This process was iterative and guided by visual inspections and summary audits. For performance tuning, I implemented indexing on frequently joined columns and used CTEs (common table expressions) to modularize and speed up nested queries.

A key differentiator in this project was my **use of generative AI tools**—specifically, AI-assisted SQL generation and optimization. I used AI to:

* Translate analytical questions into accurate SQL queries
* Refine inefficient joins and filtering logic
* Validate query results and logic
* Accelerate debugging during data cleaning

These tools helped me iterate faster, especially when working with window functions and complex JOIN conditions across multiple tables.

**Outcome:**

By combining SQL best practices with AI-enhanced guidance, I delivered a robust and optimized analysis pipeline that uncovered meaningful insights and actionable business recommendations.

Question 9

Write a SQL query that uses a window function to calculate the running total of sales for each product category, ordered by date. Use the following table structure:

SELECT

date,

product\_category,

sale\_amount,

SUM(sale\_amount) OVER (

PARTITION BY product\_category

ORDER BY date

ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW

) AS running\_total

FROM sales

ORDER BY product\_category, date;

Question 10

Explain the concept of query optimization and provide two techniques you would use to improve the performance of a slow-running SQL query. (100-150 words)

**Query optimization** is the process of improving the efficiency of a SQL query to reduce execution time and resource usage. Optimized queries retrieve the correct results **faster**, using **fewer CPU, memory, or I/O operations**. This is crucial when working with large datasets or in real-time applications.

Two effective techniques to improve performance are:

1. **Indexing**: Creating indexes on frequently filtered or joined columns (e.g., employee\_id, product\_id) can drastically speed up data access by avoiding full table scans.
2. **Using Common Table Expressions (CTEs) and Subqueries Strategically**: Breaking complex queries into modular parts using CTEs can improve readability and, when combined with good filtering, reduce the amount of data processed at each step.

Other techniques include reducing SELECT \*, filtering early (WHERE clause pushdown), and avoiding unnecessary joins or nested queries.

Optimizing queries leads to better scalability and system responsiveness.