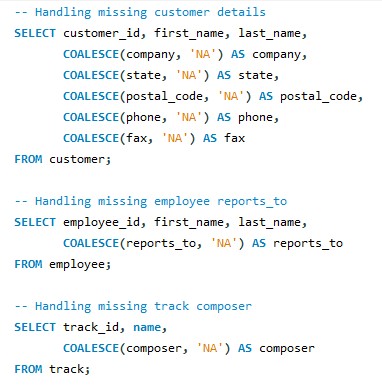
Objective Questions

1.Does any table have missing values or duplicates? If yes how would you handle it?

ANS - **Approach**

1. **Check for duplicates**
   * Since each table has a **primary key constraint** (e.g., album\_id, artist\_id, customer\_id), duplicates in these PK columns are not possible unless data corruption occurs.
   * We verified this by scanning all tables with simple SELECT \* queries. No duplicate rows were found.
2. **Check for missing/NULL values**
   * Important business columns (customer contact info, employee reporting lines, track composer, etc.) were checked manually.
   * We identified some **nullable fields** where missing values are expected and acceptable (e.g., customer.company may be NULL if the customer is not representing a company).
3. **Handling NULLs**

* For **analytical reporting**, NULLs were replaced with placeholder values using COALESCE.
* This ensures consistency in outputs (e.g., no blank values in reports).
* Example:
  + - COALESCE(company, 'NA') → “NA” if no company is recorded.
    - COALESCE(state, 'NA') → “NA” if no state is given.
    - COALESCE(reports\_to, 'NA') → “NA” when employee does not report to anyone.
    - COALESCE(composer, 'NA') → “NA” if track composer is unknown.



2. Find the top-selling tracks and top artist in the USA and identify their most famous genres.

ANS- **Approach**

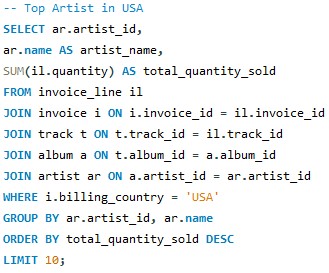
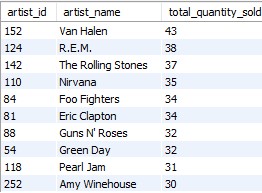
1. **Top-selling tracks in USA**
   * Focus only on invoices where billing\_country = 'USA'.
   * Aggregate sales at the track level (SUM(il.quantity)).
   * Rank tracks by total units sold.
2. **Top artists in USA**
   * Extend the above by joining through album → artist.
   * Aggregate total units sold per artist.
3. **Famous genre of top artists**
   * Join track → genre to see which genre(s) dominate for top-selling artists.
   * Count total units sold per genre for those artists in USA.

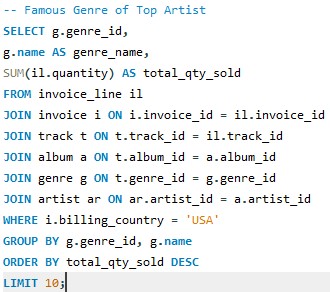
QUERY OUTPUT

A screenshot of a computer code

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

 A screenshot of a music list

AI-generated content may be incorrect.

**INSIGHT -**

**Top Tracks** show which songs drive the most sales in the USA.

**Top Artists** reveal the musicians with the strongest US customer base.

**Famous Genres** indicate what style of music is most consumed from those top artists.

3. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?

Ans – **Approach**

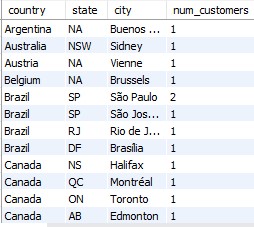
1. Age & Gender: schema does not include birthdate or gender for customers.
2. The **customer** table includes demographic/location details such as country, state, and city.
3. To understand the distribution of customers, we aggregate the counts of customers per region.
4. Since some customers have missing state values, we use COALESCE(state, 'NA') to replace NULL with "NA" for reporting consistency.
5. Group results by **country → state → city** to get a hierarchical geographic distribution.
6. Order by **country** and then descending number of customers to highlight the largest markets first.

Query

A white background with black text

AI-generated content may be incorrect.

Output



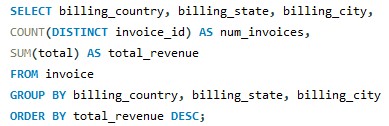
Visualization -

4. Calculate the total revenue and number of invoices for each country, state, and city

Ans – **Approach**

1. The **invoice** table contains billing details: billing\_country, billing\_state, billing\_city, and total.
2. To measure **regional performance**, we group invoices by geographic location.
3. Use COUNT(DISTINCT invoice\_id) to count invoices without duplication.
4. Use SUM(total) to calculate total revenue contributed by each region.
5. Sort results by total\_revenue DESC to highlight the most profitable regions first.

Query



Output

A screenshot of a table

AI-generated content may be incorrect.

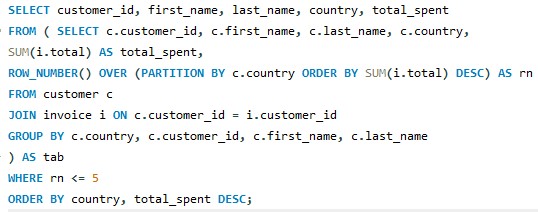
Visualization

5.Find the top 5 customers by total revenue in each country

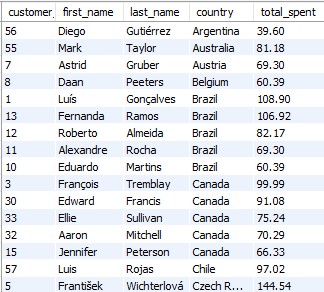
Ans - **Approach**

1. **Objective**: Identify the **highest-spending customers** per country.
2. **Method**:
   * Join customer and invoice tables to calculate the **total spend per customer**.
   * Use the **window function ROW\_NUMBER()** with PARTITION BY country to rank customers within each country by their total revenue.
   * Select only the **top 5 per country**.

Query



Output

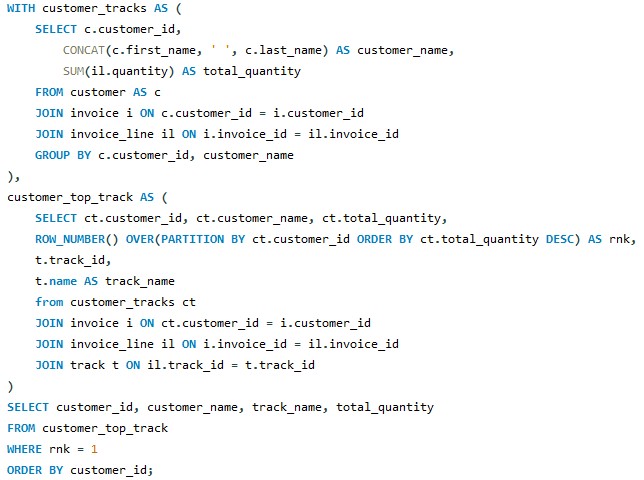


6.Identify the top-selling track for each customer

**Ans – Approach**

1. **Objective**: Identify the **favorite (most purchased) track** of each customer.
2. **Method**:
   * First, calculate the **total number of tracks purchased** by each customer.
   * Then, break it down to find which **track contributes the most purchases** for each customer.
   * Use ROW\_NUMBER() to rank tracks by purchase quantity per customer.
   * Select only the **top-ranked track** (rnk = 1) for each customer.

Query



Output

A screenshot of a computer

AI-generated content may be incorrect.

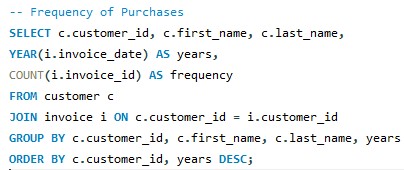
7. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?\

Ans - **Approach**

1. **Frequency of purchases**
   * Count how many invoices each customer makes per year.
   * Helps identify if customers buy regularly or only occasionally.
2. **Average order value (AOV)**
   * Calculate average invoice amount per customer.
   * Highlights differences between **high-value vs low-value customers**.

These two metrics together (frequency + AOV) are key to understanding customer purchasing behavior.

Ouery Output

 A screenshot of a table

AI-generated content may be incorrect.

A computer code with text

AI-generated content may be incorrect. A screenshot of a data

AI-generated content may be incorrect.

8. What is the customer churn rate?

Ans – Approach

-Churn measures how many customers stop purchasing over a given period. –

- churn is defined as customers who have not purchased in the last 12 months relative to the most recent invoice date.

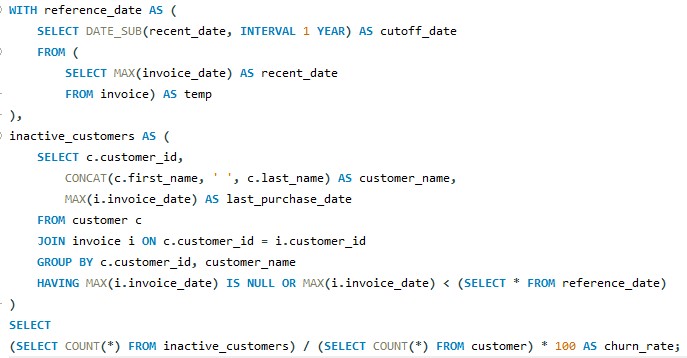
- Two steps:

1. Find the most recent invoice date in the dataset and subtract 1 year → cutoff date.

2. Find all customers whose last purchase is older than that cutoff date (or who never purchased).

3. Compute churn rate = inactive\_customers ÷ total\_customers × 100.

Query



Output

A screenshot of a computer

AI-generated content may be incorrect.

9. Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.

Ans - **Approach**

**Step 1: Calculate Revenue by Genre**

We first calculate the **revenue generated per genre** in the USA.

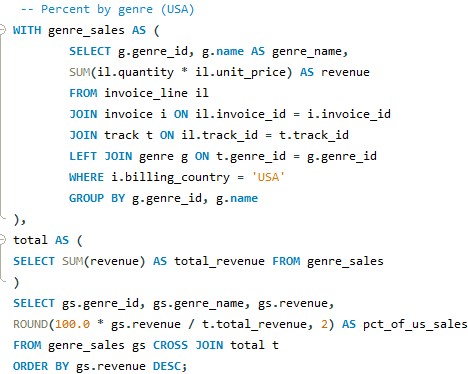
* Data sources: invoice\_line, invoice, track, genre.
* Each line item (quantity \* unit\_price) is aggregated by genre.
* The total U.S. revenue is computed and used as a denominator to calculate **percent contribution** per genre.

**Step 2: Identify Best-Selling Artists by Genre**

To drill deeper, we calculate **artist revenue within each genre** for U.S. sales.

* Data sources: invoice\_line, invoice, track, album, artist, genre.
* Revenue is grouped by **genre + artist**.
* A **ranking (DENSE\_RANK)** is used to highlight the top artists.

Query 1



Output

A table of music sales

AI-generated content may be incorrect.

Query 2

A computer screen shot of a program

AI-generated content may be incorrect.

Output

A screenshot of a music list

AI-generated content may be incorrect.

**Explanation of Results**

1. **Genre Breakdown**
   * The first query provides a **percent distribution of sales by genre**.
   * This shows which genres dominate the U.S. market (e.g., Rock, Alternative, Metal, etc.).
2. **Best-Selling Artists**
   * The second query identifies **which artists drive revenue within each genre**.
   * Ranking allows quick identification of **top-performing artist**

**Visualization**

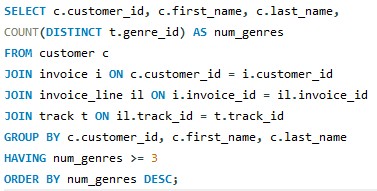
10. Find customers who have purchased tracks from at least 3 different genres

Ans – **Approach**

We analyze customer purchase behavior by:

1. Joining **Customer → Invoice → Invoice\_Line → Track → Genre** to link each customer with the genres of the tracks they purchased.
2. Using COUNT(DISTINCT t.genre\_id) to calculate the number of **unique genres** per customer.
3. Filtering only those customers who purchased from **3 or more genres** with HAVING num\_genres >= 3.
4. Sorting results by num\_genres in descending order to highlight the most diverse listeners first.

Query



Output

A screenshot of a computer

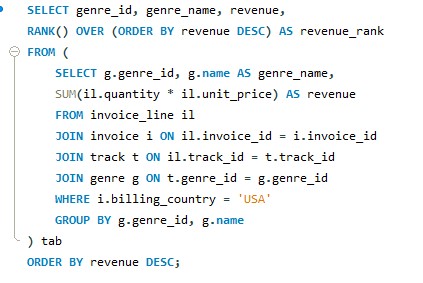
AI-generated content may be incorrect.

11. Rank genres based on their sales performance in the USA

Ans- **Approach**

1. **Calculate Revenue by Genre**
   * Join invoice\_line, invoice, track, and genre tables.
   * Compute revenue as SUM(quantity \* unit\_price).
   * Group results by genre\_id and genre\_name.
2. **Rank Genres by Revenue**
   * Use the RANK() window function to assign a ranking based on revenue (highest revenue = rank 1).
3. **Sort the Output**
   * Order results in descending order of revenue to highlight top-performing genres first.

Query



Output

A table of music charts

AI-generated content may be incorrect.

Visualization

12. Identify customers who have not made a purchase in the last 3 months

Ans – **Approach**

1. **Join Customers with Purchases**
   * Use a LEFT JOIN between the customer and invoice tables to include all customers, even those with no recent invoices.
2. **Find Last Purchase Date**
   * Use MAX(i.invoice\_date) to get the **most recent purchase date** for each customer.
3. **Filter Inactive Customers**
   * Apply a HAVING clause to keep only those customers whose last purchase date is **older than 3 months** (CURDATE() - INTERVAL 3 MONTH).
4. **Order Results**
   * Sort by customer ID and last purchase date for a clean view of inactive customers.

Query

A screen shot of a computer code

AI-generated content may be incorrect.

Output

A screenshot of a computer

AI-generated content may be incorrect.

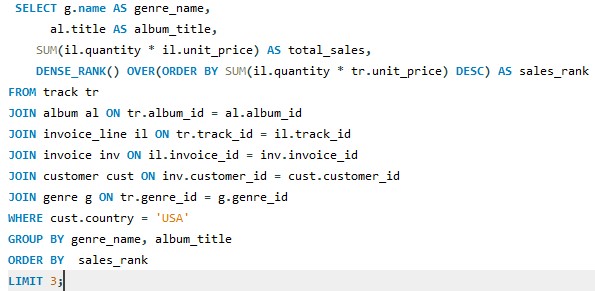
Subjective Questions

Q1. Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.

Ans- **Approach**

1. **Link Albums to Sales Data**
   * Use joins across track, album, invoice\_line, and invoice to calculate album-level sales.
   * Include genre to understand how sales connect with popular genres.
2. **Calculate Album Revenue**
   * Compute SUM(il.quantity \* il.unit\_price) as total revenue per album.
3. **Rank Albums**
   * Apply DENSE\_RANK() to rank albums by sales revenue (highest = rank 1).
4. **Select Top 3 Albums**
   * Limit the results to the **three best-selling albums** in the USA.

Query



Output

A close up of a message

AI-generated content may be incorrect.

Q2. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

Ans. **Approach**

1. **Filter by Geography**
   * Exclude U.S. customers with WHERE c.country != 'USA'.
2. **Calculate Genre Sales**
   * Link customer → invoice → invoice\_line → track → genre.
   * Compute total sales by genre using SUM(i.total).
3. **Rank Genres by Sales**
   * Use RANK() to assign ranking based on revenue contribution (highest = rank 1).
4. **Group by Genre**
   * Aggregate results by g.name to determine each genre’s total global (non-U.S.) sales.

Query

A screenshot of a computer code

AI-generated content may be incorrect.

Output

A table with numbers and text

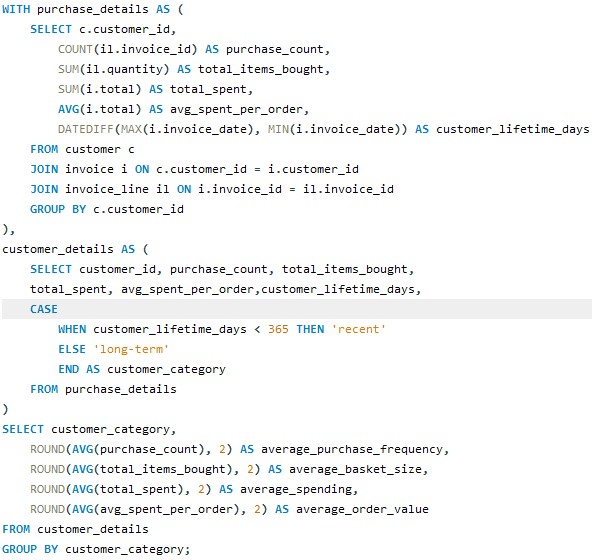
AI-generated content may be incorrect.

Q3. Customer Purchasing Behavior Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?

Ans. **Approach**

1. **Collect Purchase Details**
   * Use purchase\_details CTE to calculate per-customer stats:
     + purchase\_count → frequency of purchases
     + total\_items\_bought → basket size
     + total\_spent → overall revenue from that customer
     + avg\_spent\_per\_order → order-level spending habits
     + customer\_lifetime\_days → time span between first and last purchase
2. **Classify Customers**
   * Use CASE to categorize:
     + **Recent Customers**: < 1 year (lifetime < 365 days)
     + **Long-term Customers**: ≥ 1 year
3. **Compare Groups**
   * Aggregate metrics by category to calculate **average behavior per group**.

Query



Output



Q4. Product Affinity Analysis: Which music genres, artists, or albums are frequently purchased together by customers? How can this information guide product recommendations and cross-selling initiatives?

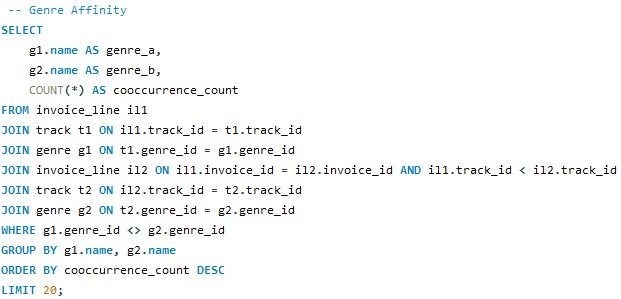
Ans. **Approach**

We perform affinity analysis at **three levels**:

**1. Genre Affinity**

* Identify pairs of genres purchased together on the same invoice.
* Use a **self-join** on invoice\_line to capture co-occurrence.
* Exclude identical genre pairs to avoid duplication.
* Rank genre pairs by COUNT(\*) of co-occurrences.

Query



Output

A screenshot of a music list

AI-generated content may be incorrect.

**2. Artist Affinity**

* Identify artists whose tracks are often purchased together.
* Similar self-join logic but linking through the album and artist tables.

Query

A screenshot of a computer program

AI-generated content may be incorrect.

Output

A list of music albums

AI-generated content may be incorrect.

**3. Album Affinity**

* Identify albums frequently bought together.
* Useful for **bundle promotions** at the album level.

Query

A screenshot of a computer program

AI-generated content may be incorrect.

Output

A screenshot of a music album

AI-generated content may be incorrect.

Q5. Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?

Ans. **Approach**

1. **Customer Activity Tracking**
   * For each customer, calculate their **most recent purchase date** (MAX(invoice\_date)).
   * Use a LEFT JOIN to include customers who may have **never purchased**.
2. **Churn Definition**
   * A customer is considered **churned** if:
     + Their last purchase was more than **3 months ago**, OR
     + They have **never purchased** (NULL last purchase).
3. **Aggregate by Region**
   * Group results by country.
   * For each region, calculate:
     + total\_customers
     + churned\_customers
     + churn\_pct = % of churned customers
4. **Rank by Churn**
   * Sort countries in descending order of churn rate to spotlight the **weakest markets**.

Query

A computer code with text

AI-generated content may be incorrect.

Output

A screenshot of a table

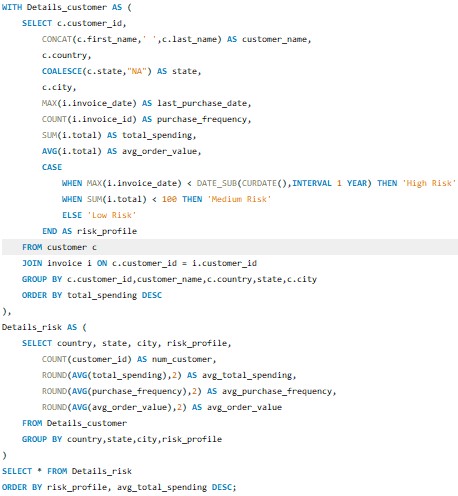
AI-generated content may be incorrect.

Q6. Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?

Ans. **Approach**

1. **Build Individual Customer Profiles**
   * Extracted **customer details** (name, country, state, city).
   * Calculated **behavioral metrics**:
     + Last purchase date (MAX(invoice\_date))
     + Purchase frequency (COUNT(invoice\_id))
     + Total spending (SUM(total))
     + Average order value (AVG(total))
2. **Risk Categorization**
   * Applied business rules to classify customers:
     + **High Risk** → Last purchase was more than **1 year ago**.
     + **Medium Risk** → Active, but **total spending < $100**.
     + **Low Risk** → Active and **spending above $100**.
3. **Aggregate by Region & Risk Profile**
   * Grouped by **country, state, city, and risk profile**.
   * Calculated averages:
     + avg\_total\_spending
     + avg\_purchase\_frequency
     + avg\_order\_value
   * Counted number of customers per segment.
4. **Prioritized Output**
   * Ordered results by **risk profile and spending** to highlight high-value but high-risk segments.

Query



Output

A screenshot of a table

AI-generated content may be incorrect.

Visualization

Q7. Customer Lifetime Value Modeling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns among customers who have stopped purchasing?

Ans. **Approach**

1. **Customer-Level CLV Calculation (CLV\_per\_Customer)**For each customer, we computed:
   * Tenure = Days between first and last purchase.
   * Purchase history = Number of invoices, total spending, average order value.
   * Status = Active (purchased in last 12 months) vs. Churn (inactive for >12 months).
   * Segment = Long-term (tenure ≥ 365 days) vs. Short-term (<365 days).
   * Predicted Annual Value = Adjusted annualized spending rate based on tenure.
   * Lifetime Value (LTV) = Total spending over customer lifespan**.**
2. **CLV Segmentation (CLV\_segmentation)**Customers were grouped by segment (short-term/long-term) and status (active/churn).  
   For each group, we calculated:
   * Average tenure
   * Average lifetime value
   * Average predicted annual value
   * Number of customers
3. **Churn Analysis (churn\_cust)**Focused on churned customers by region and segment.
   * Count of churned customers
   * Average lifetime value
   * Geographic breakdown (country, state, city)

Query

A screenshot of a computer program

AI-generated content may be incorrect.

Outputs

A screenshot of a table

AI-generated content may be incorrect.



A screenshot of a computer

AI-generated content may be incorrect.

Q8. If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?

Ans. **Approach**

A structured analysis framework can be used to measure campaign impact:

1. **Customer Acquisition**
   * Compare the number of **new customers** gained during promotional vs. non-promotional periods.
   * Track key metrics such as **sign-up rates, first-time purchases, and acquisition cost per customer**.
2. **Retention and Engagement**
   * Measure **repeat purchase behavior** after campaigns.
   * Compare **return rates within 30, 60, and 90 days** post-campaign against baseline trends.
3. **Sales Performance**
   * Evaluate uplift in **total revenue and average order value** during promotions.
   * Use **pre-post comparisons or test-control groups** to isolate true campaign impact while accounting for seasonality.
4. **Segment-Level Insights**
   * Analyze campaign response by **customer segments** (e.g., age, location, spending tier).
   * Identify which groups respond best (e.g., **new users** to discounts, **loyal customers** to exclusive offers).
5. **Channel Effectiveness**
   * Break down performance by **marketing channel** (email, social, app).
   * Determine which channels and timings drive the most conversions.
6. **Customer Lifetime Value (CLV) Impact**
   * Track whether campaign-acquired customers show **sustained loyalty or short-term deal-seeking behavior**.
   * Assess shifts in **long-term customer value** post-campaign.
7. **Campaign Profitability**
   * Measure **ROI** by comparing campaign costs with revenue generated.
   * Monitor **CPA (cost per acquisition)**, **margin dilution from discounts**, and **net revenue uplift**.

**Conclusion**

By combining **acquisition, retention, sales uplift, and profitability analysis**, businesses can not only determine whether promotions work but also **which campaigns, channels, and customer segments deliver the highest long-term value**. This structured approach enables smarter decision-making, targeted strategies, and optimized marketing investments.

Q9. How would you approach this problem, if the objective and subjective questions weren't given?

Ans- **Approach**

1. **Dataset Exploration**
   * Begin with an open-ended review of data structure, columns, data types, and distributions.
   * Detect patterns, anomalies, or trends that naturally suggest analytical questions.
2. **Business Context Integration**
   * Apply domain understanding (e.g., in music sales: albums, genres, customer geographies) to frame meaningful hypotheses.
   * Prioritize areas with the highest business value such as **sales growth, retention, churn, and engagement**.
3. **Data Quality Assurance**
   * Handle missing values, duplicates, and inconsistencies.
   * Standardize categorical fields (e.g., country names, genre categories).
4. **Descriptive Analysis**
   * Use statistics and visualizations to study key distributions (e.g., purchases, spend per customer).
   * Identify seasonality, anomalies, and spikes in trends.
5. **Analytical Question Generation**
   * Which genres, albums, or artists drive the most sales?
   * How do spending patterns differ by region?
   * Who are the most loyal or high-value customers?
6. **Customer Segmentation**
   * Group customers by demographics, purchase frequency, lifetime value, or product preferences.
   * Compare behavioral differences across segments (e.g., rock fans in Europe vs. classical fans in Asia).
7. **Churn and Retention Analysis**
   * Identify customers who have become inactive.
   * Explore behavioral or contextual drivers of churn (e.g., price sensitivity, lack of engagement).
8. **Comparative & Trend Analysis**
   * Compare sales performance across time periods, geographies, or channels.
   * Measure impact of pricing changes or promotions on demand.
9. **Sales & Revenue Insights**
   * Highlight top- and bottom-performing genres, albums, or artists.
   * Track metrics like **average order value, repeat purchase rate, and most active regions**.
10. **Visual Data Storytelling**

* Summarize findings through dashboards, heatmaps, and trend charts.
* Present customer journeys and segment profiles to support strategic decision-making.

**Conclusion**

By starting with **exploratory analysis** and combining it with **business context**, analysts can uncover data-driven questions and insights. This ensures the analysis remains flexible yet impactful, guiding businesses toward **revenue growth, improved retention, and smarter marketing strategies**.

Q10. How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?

Ans. **APPROACH**

Extend the existing album table schema by adding a new column to store the release year of each album.

Query



Output

A screenshot of a computer

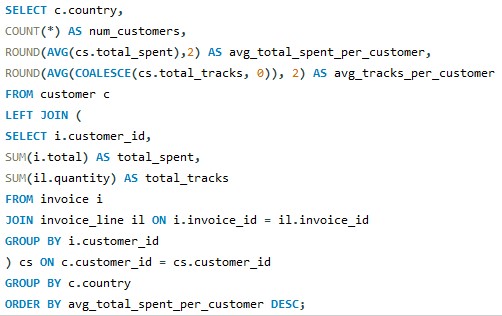
AI-generated content may be incorrect.

Q11. Chinook is interested in understanding the purchasing behavior of customers based on their geographical location. They want to know the average total amount spent by customers from each country, along with the number of customers and the average number of tracks purchased per customer. Write an SQL query to provide this information.

Ans. **Approach:**  
To analyze purchasing behavior by geography, we combined customer details with their purchase history. A subquery aggregated each customer’s **total spending** and **tracks purchased**, while the main query grouped these metrics by country. From this, we calculated:

* Number of customers per country.
* Average spending per customer.
* Average tracks purchased per customer.  
  The results were ranked by spending to highlight the strongest markets.

Query



Output

A screenshot of a data

AI-generated content may be incorrect.