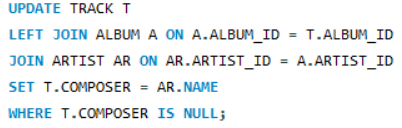
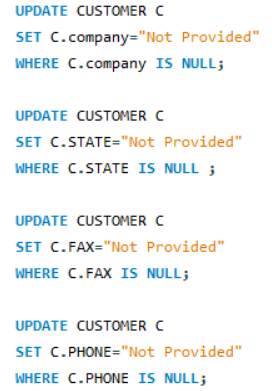
**Objective Questions**

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

**Guidelines :** Checked all tables for missing values and duplicates .Found missing values in TRACK and CUSTOMER tables. No exact duplicate rows found .

**Visualization :** Updated the table by using Update function.

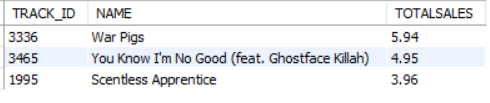
**** ****

**Observation :** Used UPDATE with JOIN and IS NULL to fill missing values directly in the table. Ensured updates target only NULLs to preserve original data.

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

**Guidelines :** Filtered sales data using billing\_country = 'USA' from the invoice table. Joined relevant tables: invoice, invoiceline, track, album, artist, and genre. Grouped by track, artist, and genre to identify top performers.

**Visualization :** Top selling tracks and top artist from USA

**** ** **

****

**Observation** : Identified the top 3 best-selling tracks by revenue in the USA.

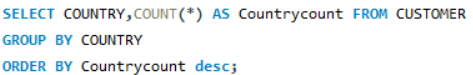
They are War Pigs ,You Know I'm No Good (feat. Ghostface Killah), Scentless

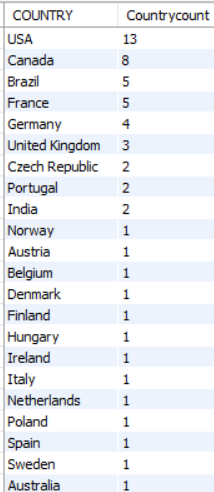
Apprentice .The top artist is Van Halen and his most famous genre is ROCK.

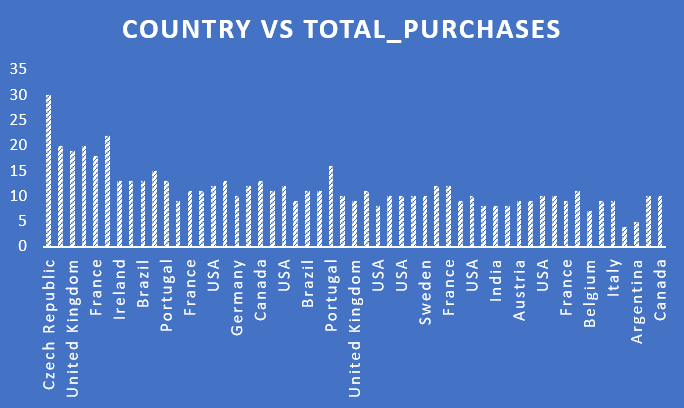
1. **Is the customer demographic breakdown (age, gender, location) of Chinook's customer base?**

**Guidelines :** Analysed customer demographics using available data in the customer table. Focused on location as the only demographic attribute available . Chinook dataset does not include age or gender, so the breakdown is limited to geography.

**Visualization :** Show casing customer count from every country





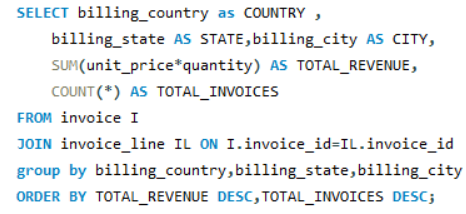
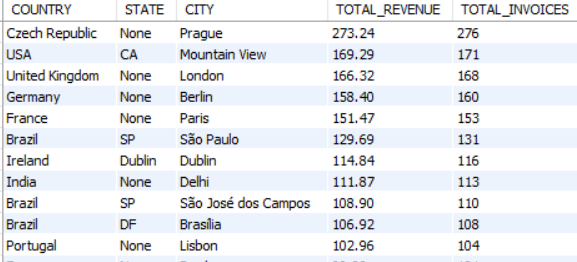


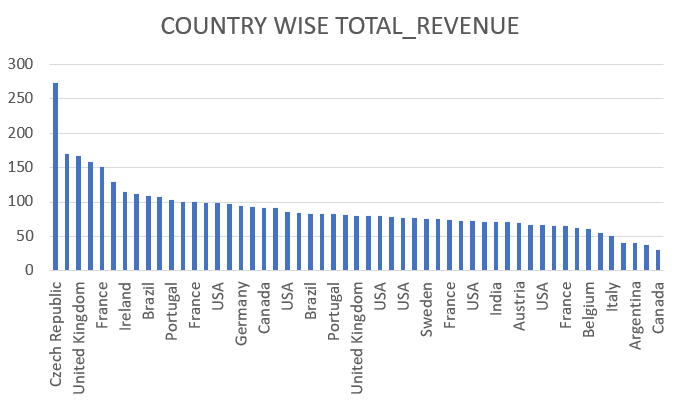
**Observation** : Majority of customers are concentrated in a few key countries like USA and CANADA .This breakdown helps identify the primary geographic markets.

1. **Calculate the total revenue and number of invoices for each country, state, and city:**

**Guidelines :** Joined invoice and invoice\_line tables to calculate Total Revenue . Grouped results by country, state, and city. Counted total invoice lines to measure sales activity.

**Visualization :** Total revenue and invoices count from each country, state and city

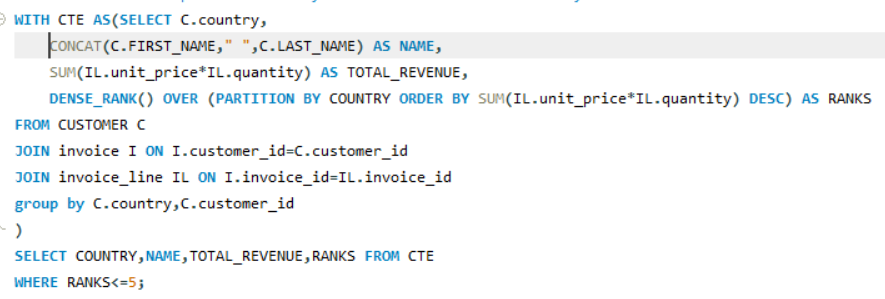


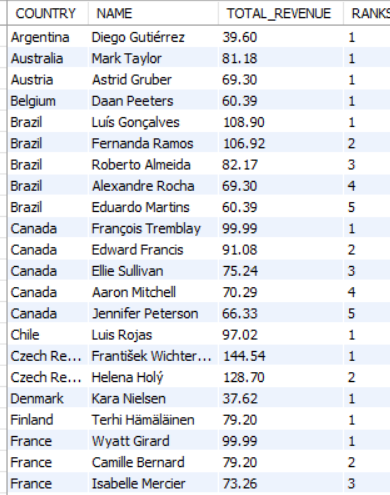
**Observation :** City in the Czech Republic contribute the highest revenue. Billing data allows geographic segmentation for sales analysis.

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

**Guidelines :** Used a Common Table Expression to calculate total revenue per customer . Joined customer, invoice, and invoice\_line tables. Applied DENSE\_RANK to rank customers within each country by revenue.

**Visualization :** Prioritizing customer from each country by top revenue





**Observation :** Identified the top-spending customers in each country.

Useful for targeting high-value clients in each region for programs or personalized marketing.

1. **Identify the top-selling track for each customer.**

**Guidelines :** Used a Common Table Expression to compute total revenue per track per customer . Joined customer, invoice, invoiceline , and track tables. Applied DENSE\_RANK to rank tracks per customer by total spend. Filtered results to retrieve the top-selling track for each customer.

**Visualization :** Finding customer choice



**Observation :** Successfully identified the most purchased (by revenue) track for every customer. Highlights individual listening/buying preferences. Can be used to personalize recommendations or create similar track playlists.

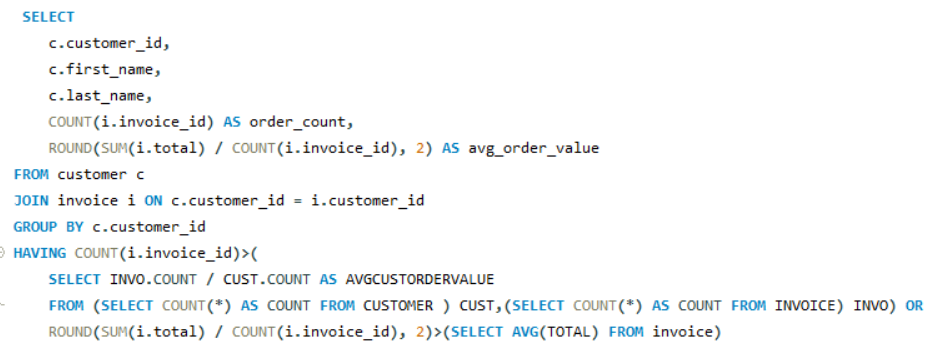
1. **Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?**

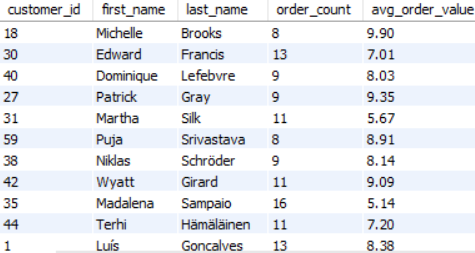
**Guidelines :**

Customer and invoice tables were joined to analyse purchasing behavior .

Order frequency and average order value were calculated for each customer. Customers with above-average frequency or average order value were filtered using the HAVING clause.

**Visualization :** Patterns and trends regarding customer in all possible aspects



****

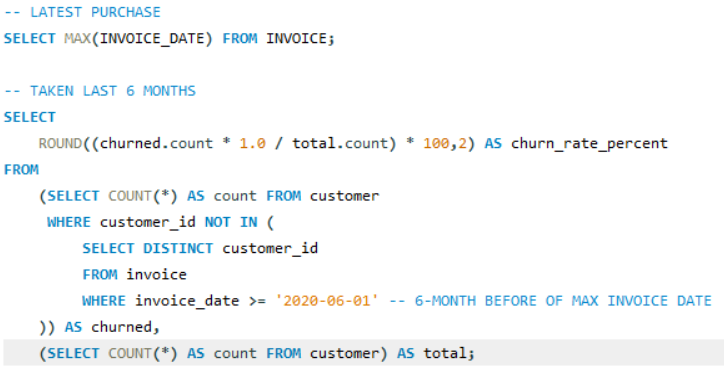
**Observation :** This query identifies high-value or high-frequency customers.

Some customers consistently spend more or purchase more frequently than average. Target high-frequency or high-AOV customers with loyalty perks or tailored offer .

1. **What is the customer churn rate?**

**Guidelines :** the latest invoice date from the invoice table was retrieved to determine the 6-month activity window. Customers who did not make a purchase in the last 6 months were classified as churned.

**Visualization :** Show casing critical KPI

****

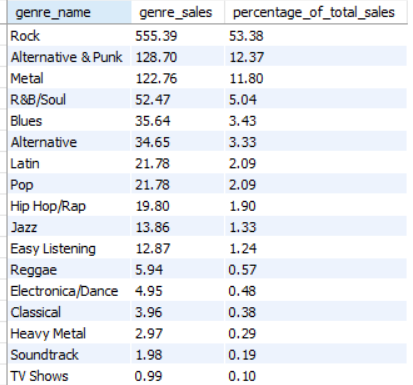
**Observation :**

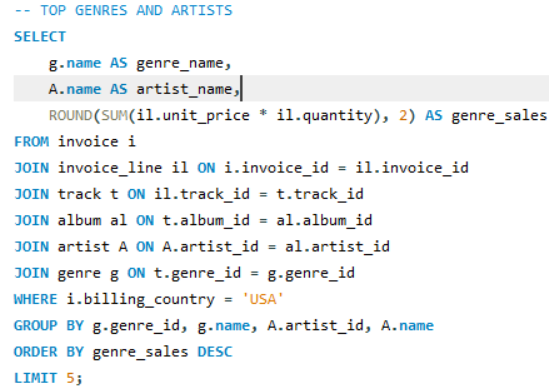
This query identifies the percentage of customers who have not made a purchase in the last 6 months.The churn rate is **20.34%**, representing a significant portion of the customer base.This insight is valuable for monitoring customer retention and planning targeted re-engagement campaigns to win back inactive users.

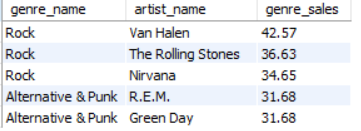
1. **Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.**

**Guidelines :** Use clear joins and filters to ensure accurate sales calculations specific to the USA .Apply aggregation and percentage functions with subqueries for meaningful comparisons .Order and limit results to highlight top-performing genres and artists effectively.

**Visualization :** Contribution of each genre in USA

**** 

****

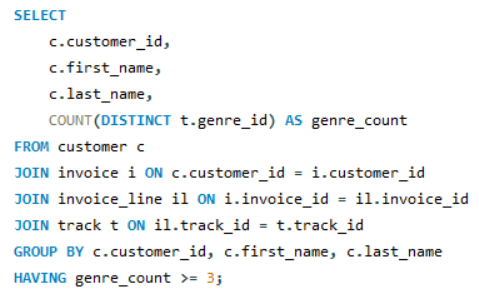


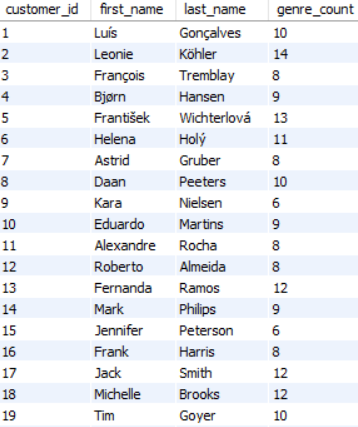
**Observation :** *Rock* is the leading genre in the USA with $555.39 in sales, contributing a dominant 53.38% of total music sales. *Van Halen* is the top-selling artist within this genre, contributing $42.57. The dominance of Rock and Van Halen suggests that marketing efforts, promotions, and future investments should prioritize this genre and artist in the U.S. market.

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

**Guidelines :** Use joins to connect customer purchases with track genres, ensuring accurate genre tracking .Apply COUNT with HAVING to filter customers engaging with 3 or more genres.

**Visualization :**  Customers who purchased tracks from multiple genres

****

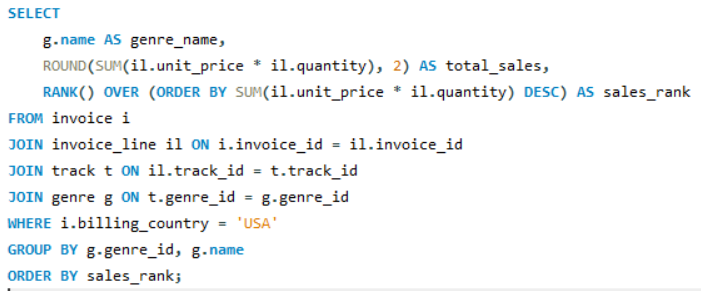
****

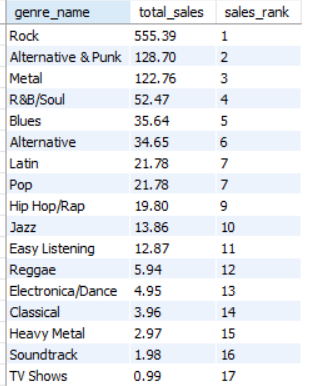
**Observation** : These 59 customers have shown a high level of genre diversity, indicating they might be more open to varied musical tastes .Targeting these customers with cross-genre promotions or personalized playlists could increase engagement and drive more sales.

1. **Rank genres based on their sales performance in the USA**

**Guidelines :** Use the RANK function to rank genres based on total sales and filter results by sales performance.Order by the rank to clearly identify the top-performing genres for targeted marketing or strategy.

**Visualization :** Genres sales performance in USA



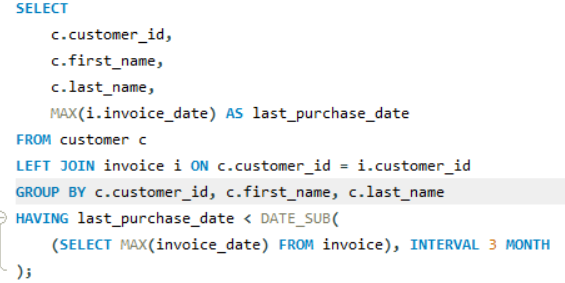


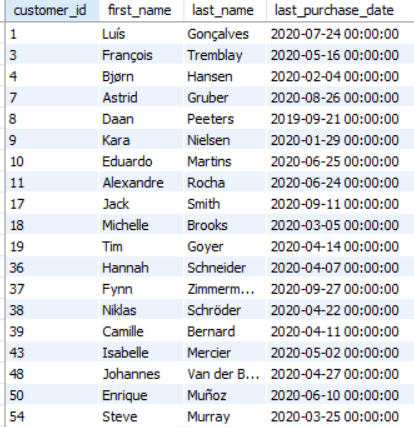
**Observation :** Rock is the top-selling genre, with $555.39 in sales, securing the 1st rank. This genre is by far the dominant player in the market. The ranking of genres provides a clear, data-driven view of which genres dominate in sales, and can guide strategic decisions like artist contracts.

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

**Guidelines :** Use a LEFT JOIN to find all customers and filter those whose last purchase was over 3 months ago with DATE\_SUB.

**Visualization :** Customers who not shopped in recent days

****

****

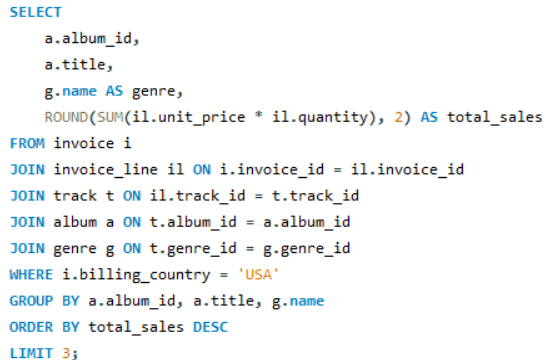
**Observation :** 22 people have not made a purchase in the last 3 months, and this group is ideal for re-engagement campaigns such as targeted promotions or reminders. Understanding this customer behavior can help improve customer retention strategies.

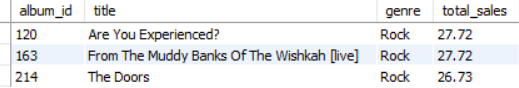
**Subjective Questions**

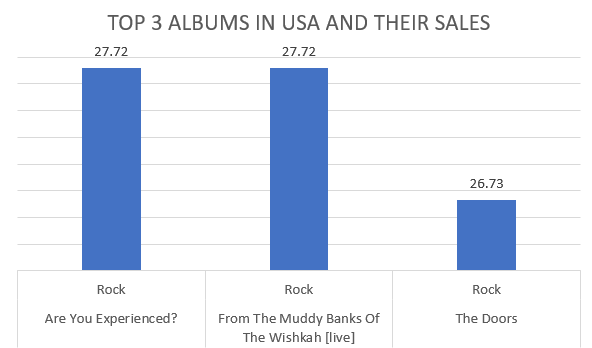
1. **Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.**

**Guidelines :** Use JOINs across invoice, invoice line, track, album, and genre to calculate total U.S. sales per album. Filter by country = 'USA' to focus only on U.S.-based purchases. Group by album id.

**Visualization :** Top 3 albums

****

****



**Insights :**

Rock dominates the U.S. market in this dataset, indicating strong listener loyalty and catalogue demand.The top 3 albums are rock, suggesting continued commercial value of legacy artists and nostalgia-driven consumption.Sales values are very close, showing competitive performance among classic rock albums a good sign for consistent audience engagement.

**Recommendations :**

Focus advertising and promotion on:

***Are You Experienced?***

***From The Muddy Banks Of The Wishkah [Live]***

***The Doors***

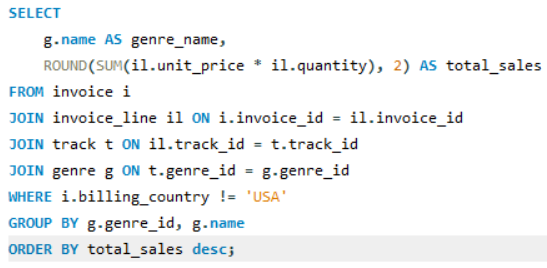
These albums should lead U.S. marketing efforts due to their proven sales

Performance , genre strength, and cultural legacy.

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

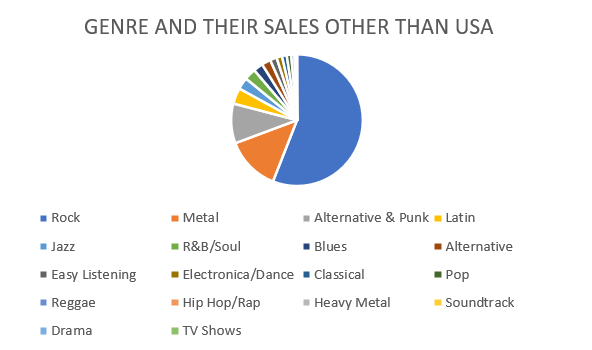
**Guidelines :** Join invoice, invoice line, track, and genre to analyse sales by genre .Use country != 'USA' for non-USA. Group by genre and compare total sales across both regions.

**Visualization :** Top-selling genres in countries other than the USA.

****

**OUTPUT :**





**Insights :**

Rock dominates both U.S. and non-U.S. sales, making it a strong global

genre . Alternative & Punk, Metal, R&B/Soul, and Blues appear in the top 5

of both datasets. Indicates a shared global preference .

**Recommendations :**

Prioritise Rock for global promotion, as it's the top-selling genre in both USA

and international markets . Increase international marketing for Metal and

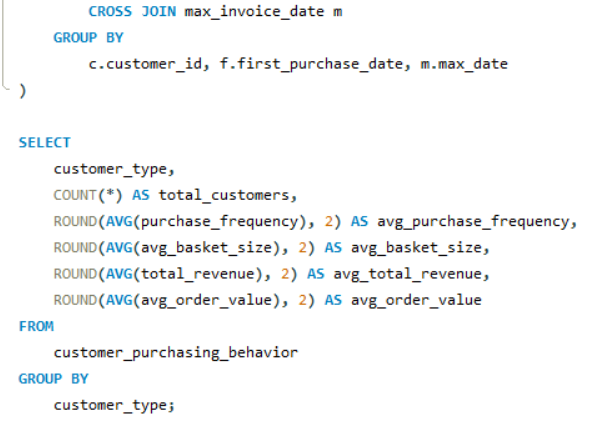
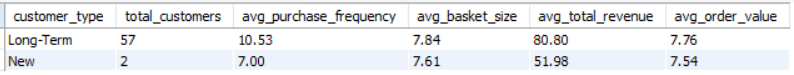
Alternative, which perform significantly better outside the USA.

1. **Customer Purchasing Behaviour 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?**

**Guidelines :** Join customer, invoice, and invoice line to analyse purchases.

Classify customers as Long-Term if their first purchase was over 3 years ago, else New .Calculate purchase frequency, average basket size, total revenue, and average order value per customer.Group by customer type to compare behaviour and identify loyalty trends.

**Visualization :** Customer Purchasing Behaviour Analysis .

**Insights :**

Long-Term customers have slightly larger basket sizes and higher overall spending than New customers.Despite fewer Long-Term customers (57 vs. 2), they purchase more frequently (10.53 vs. 7.00).

The average order value is marginally higher for Long-Term customers, indicating stronger engagement.New customers show potential with similar basket sizes but lower frequency and revenue.

**Recommendation** :

Focus on retaining and expanding the Long-Term segment, as they generate more revenue and purchase more often. Implement onboarding incentives to convert New customers into repeat buyers.

Encourage higher purchase frequency among New customers with targeted promotions or loyalty programs.

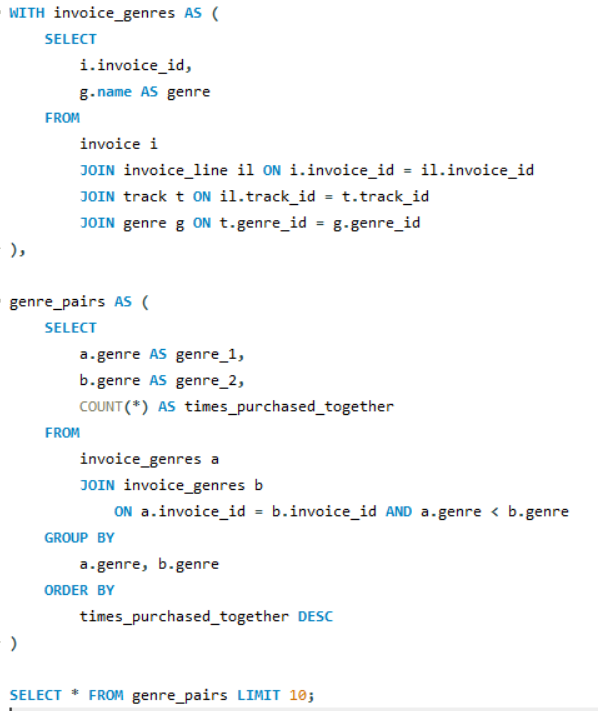
Analyse what drives longevity in Long-Term customers and replicate it for newer users.

1. **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?**

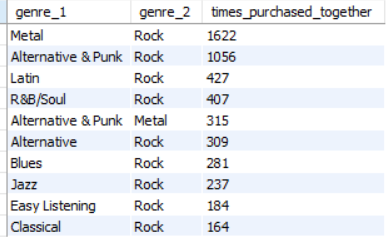
**Guidelines :** Join invoice, invoice line, track, and genre to assign genres to each invoice. Use a self-join on invoices to find genre pairs purchased together . Count co-occurrences of genre pairs across invoices.

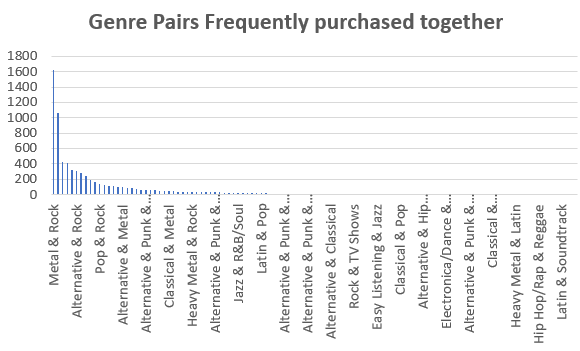
Sort by frequency to identify top genre combinations for cross-selling.

**Visualization:** Genre Pairs Frequently purchased together .



**Top 10 Frequently Purchased Genre Pairs**

****



**Insights :** Rock is the most common link across genre pairings, acting as a central genre in multi-genre purchases.Metal, Alternative & Punk, and Latin frequently appear with Rock, indicating strong cross-genre interest.

Customers often purchase complementary genres in a single session, reflecting diverse musical tastes.Genre affinity suggests shared audience preferences, useful for bundling and recommendations.

**Recommendation :** Use this data to build genre-based recommendation engines, suggesting Rock when users browse Metal or Punk .Create bundled offers or curated playlists combining frequently paired genres (e.g., Rock + Metal). Highlight cross-genre promotions to encourage more multi-genre purchases and increase basket size.

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

**Guidelines :** Geographic Segmentation: Join customer and invoice tables to associate each transaction with customer locations (country, state, city).

Behavioural Metrics: Aggregate total purchases, average order value, and

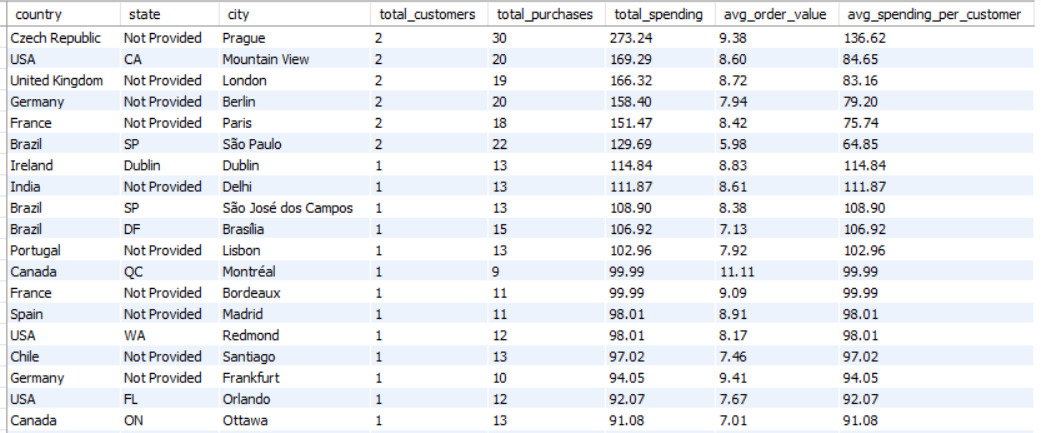
total spending per region.

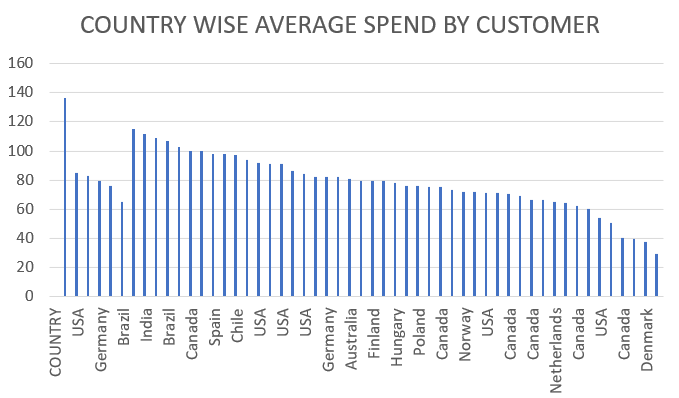
Churn Identification: Flag customers as churned if their latest purchase is older than 12 months using invoice date.

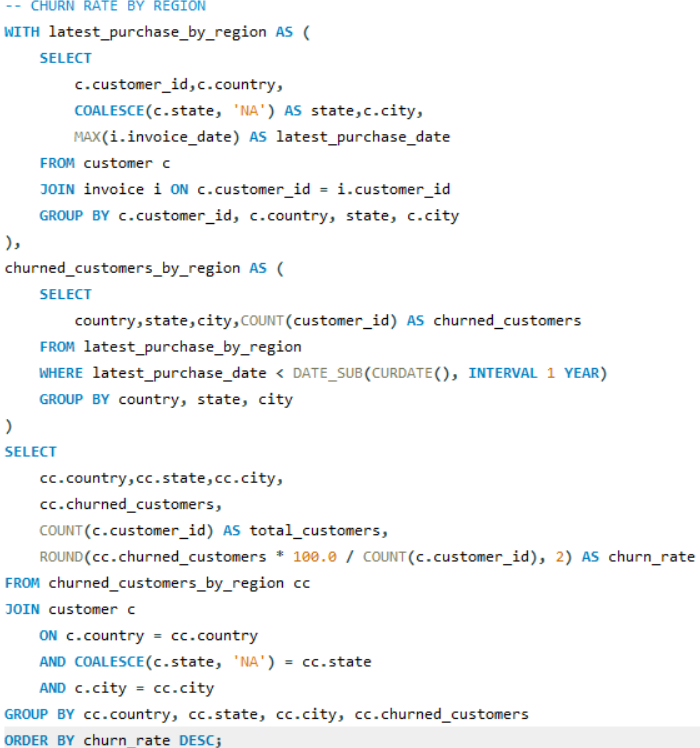
Regional Aggregation: Group data by region to compare customer activity, engagement, and churn.

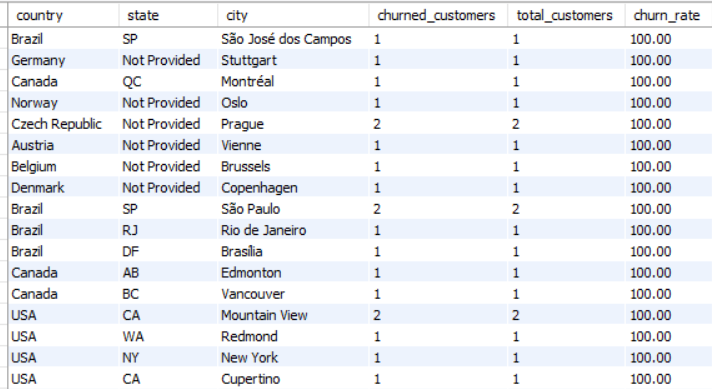
**Visualization:** Regional Purchasing Behaviour





**Churn Rate By Region :**



****

**Insights :**

High-Value Urban cities : Cities like Prague, London, Mountain View, and Berlin have high purchase frequencies and strong customer value.

Low Engagement Risk Zones: Cities like Edmonton, Buenos Aires, and Rome show low frequency and high churn risk.

Promising Mid-Tier Cities: Locations such as Delhi, Sao Paulo, and Dublin show moderate performance and good growth potential.

**Recommendation :**

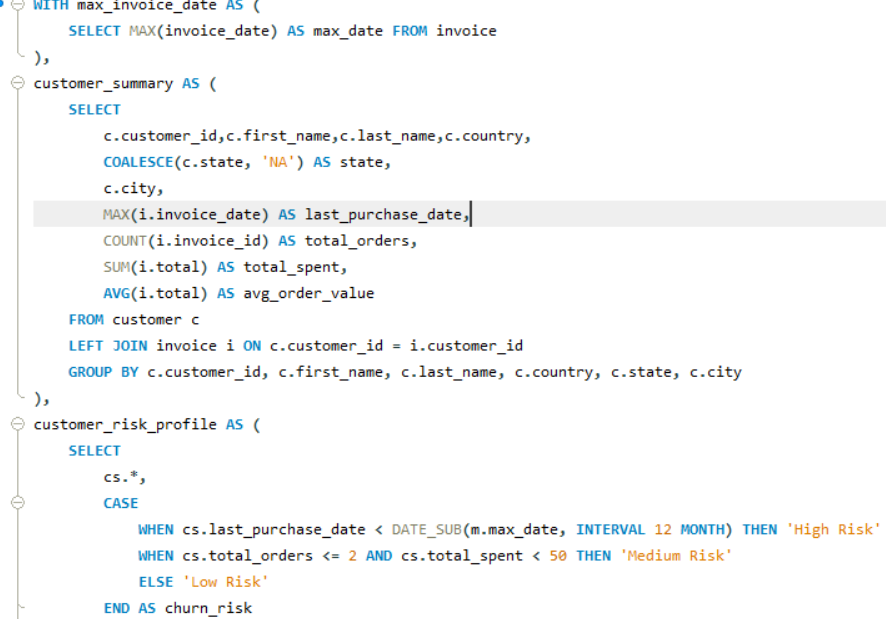
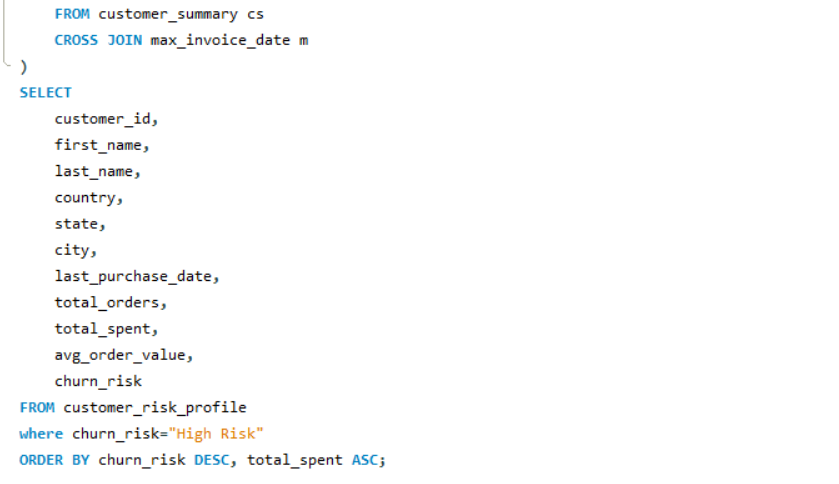
Retention & Growth: Re-engage churned users in cities like Rome and Buenos Aires; expand premium offerings in high-value cities like London and Prague.

Localized Strategy: Launch regional bundles in Delhi and Sao Paulo; use freemium or student pricing in low-engagement, price-sensitive areas.

1. **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?**

**Guidelines :** Join customer and invoice tables to build customer profiles using frequency and monetary value metrics. Calculate each customer’s last purchase date and compare it with the maximum invoice date to identify inactivity. Classify churn risk into High, Medium, and Low using purchase history thresholds. Segment by country, city, and spending to reveal at-risk geographic and behavioural patterns.

**Visualization :** Customer who reduced spending

**** ****

**Output :**

****

**Insights :**

High-risk customers often show no purchases in the last 12+ months,

regardless of earlier engagement levels. Regions like Brussels display dormant

customers who once had healthy spending habits. Low-order customers with

limited spend are commonly flagged medium-risk. Retention challenges are not

always tied to low spend — even loyal customers can become inactive without

re-engagement.

**Recommendation :**

Launch targeted reactivation emails for high-risk customers with decent past spend. Use automated churn flags for customers inactive for over 12 months.  
Offer region-specific loyalty programs in at-risk cities to rebuild engagement.  
Segment high-risk users for exclusive discounts or personalized bundles.  
Track risk metrics monthly and integrate them into CRM and campaign tools for real-time action.

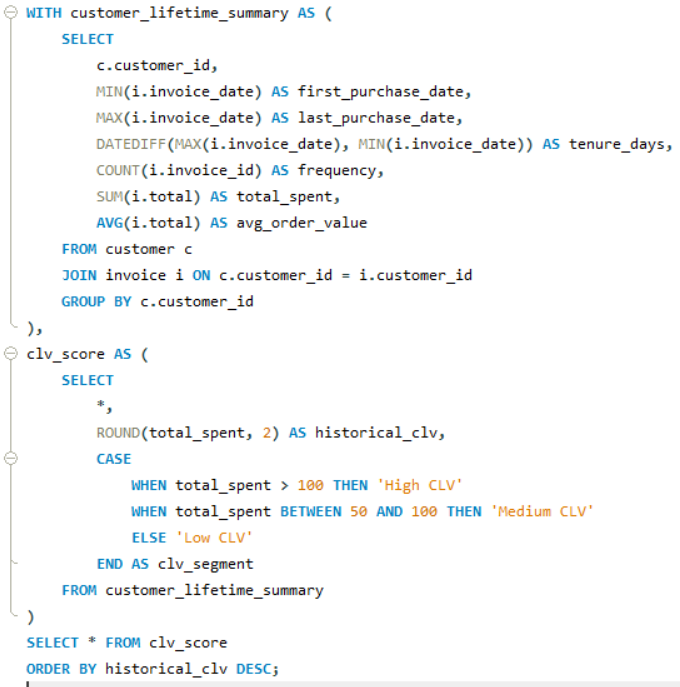
1. **Customer Lifetime Value Modelling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns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

**Guidelines :**

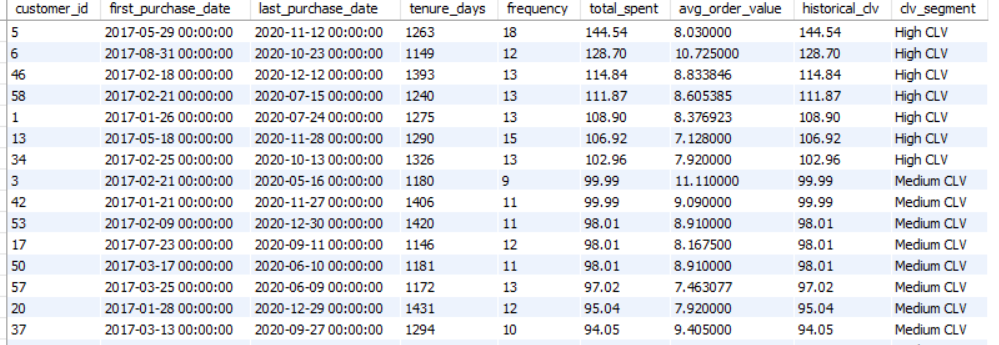
Join customer and invoice tables to calculate first and last purchase dates, giving insight into customer tenure. Use purchase frequency, average order value, and total spending to compute historical CLV.

Categorize customers into High, Medium, and Low CLV segments. These metrics help identify long-term value and prioritize high-return customers for retention and upsell.

**Visualization :** Historical clv

****

**Output :**

****

**Insights :**

High CLV customers tend to have long tenures and frequent purchases with higher average order values. Many lapsed customers (low recent activity) belong to Medium or High CLV segments, indicating missed retention opportunities. Short-tenure, low-frequency buyers often fall in the Low CLV bracket, emphasizing the need for early engagement strategies.

**Recommendation :**

Prioritize High CLV segments for loyalty programs and personalized marketing.  
Implement win-back campaigns targeting Medium CLV customers before they churn.  
Launch early engagement triggers for new/low CLV users to increase retention.  
Use CLV scoring to target ads and upsell campaigns for top-tier customers.  
Integrate CLV with churn risk models for a combined profitability-retention strategy.

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

**Guidelines :**

To evaluate the true effectiveness of promotional efforts (discounts, events, email campaigns), a data-driven, segmented analysis is required. This involves tracking behavioural changes pre-, during, and post-campaign, with comparisons across customer segments and marketing channels.

**Steps :**

**1. Customer Acquisition Impact**Compare the number of new customers acquired during campaign periods vs. baseline periods. Use KPIs like sign-up rate, first purchase rate, and customer acquisition cost (CAC) to measure efficiency.

**2. Customer Retention and Engagement**Analyse retention windows (e.g., 30, 60, 90 days post-campaign) and purchase frequency. Identify whether campaign-exposed customers return more often or engage longer than others.

**3. Sales Performance and Uplift**Use pre-post analysis or A/B testing frameworks to quantify changes in total revenue, average order value, and conversion rates. Normalize results to control for seasonality or unrelated market factors.

**4. Segment-Level Response Analysis**Group customers by region, CLV tier, or campaign exposure to assess who responded best. Identify which campaigns performed well in specific demographics (e.g., first-time discounts working in São Paulo but not in Brussels).

**5. Channel Effectiveness**Disaggregate results by marketing channel (email, SMS, social, ads) to find which medium drove the most acquisition or repeat engagement. Time-of-day and message tone can also be tested.

**6. CLV Shift Analysis**Measure if campaign-driven acquisitions increase or dilute average customer lifetime value over time. Evaluate whether campaigns are bringing in loyal customers or short-term opportunists.

**7. Campaign Profitability and ROI**Track direct costs of running each campaign (ad spend, discount loss) and calculate ROI using revenue attribution models. Monitor CPA, LTV/CAC ratio, and margin dilution.

**Insights**

Customers brought in through loyalty events tend to show higher retention and LTV than those via mass discounts.

Short-term campaigns often attract price-sensitive customers with lower engagement post-purchase.

Cities like Prague and Delhi saw high initial uplift but also quick drop-off, indicating need for stronger follow-up strategies.

Frequent buyers with short tenure suggest onboarding gaps; nurturing journeys may improve value realization.

**Recommendations**

Use behavioural segmentation to tailor future campaigns to the most responsive users.

Prioritize retention-focused campaigns over acquisition-only efforts for long-term ROI.

Run multivariate testing (offer types, durations, channels) to refine campaign strategy.

Link campaign data with churn and CLV models to predict and enhance campaign quality.

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

**Start with a Blank Slate**: Treat the dataset as an unknown territory—begin by scanning schema, datatypes, and relationships to understand what each table contributes.

**Explore Without Constraints:** Use exploratory data analysis to surface natural groupings, trends, and correlations without being anchored to predefined questions.

**Incorporate Business Logic:** Map fields like genre, total spent, invoice date, or country to real-world business activities—sales marketing zones, or customer behaviour.

**Identify Potential Impact Areas:** Look for indicators that connect directly to business KPIs such as revenue, customer retention, or user satisfaction.

**STEPS**

**Data Understanding & Cleaning**

* Check for nulls, outliers, and inconsistent values (e.g., country names like “USA” vs. “United States”).
* Validate date columns, standardize units, and remove redundancies or malformed entries.

**Baseline Metrics & Patterns**

* Calculate total sales, number of customers, average order value, and distribution across time.
* Use line plots, histograms, or bar charts to show trends, spikes, and drops.

**Discovery-Driven Questioning**

* Where is customer activity most concentrated?
* Are there key differences in spending behavior across cities or states?
* Do certain customers consistently spend more or purchase more frequently?

**Behavioural Segmentation**

* Cluster customers by spend, recency, and frequency metrics (RFM analysis).
* Find unique user types (e.g., big spenders vs. frequent but low-spend buyers).

**Risk & Churn Pattern Recognition**

* Identify customers whose last purchase was over a year ago.
* Correlate churn with low average order value or declining activity trends.

**Temporal and Geographic Analysis**

* Look at time series of purchases per region.
* Find product or artist popularity shifts across months or locations.

Revenue & Product Focus

* Rank top genres or artists by revenue contribution.
* Spot underperforming categories that may need pricing or placement adjustments.

**Engagement Storytelling**

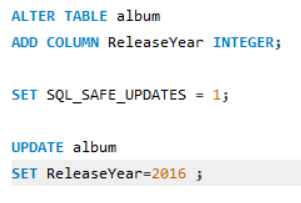
* Build user flow charts or segment timelines to explain customer lifecycles.
* Use dashboard tools to deliver insight through interactive visuals, enabling data-driven storytelling for strategic actions.

1. **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?**

**Guidelines :**

Enhance the album table structure by introducing an additional column that captures the year each album was released.Fill in the values for the Release Year column.

**Visualization :** Added new column to table for storing year

****

****

**Insights :**

Adding a Release Year column helps in organizing albums by their release time, which is useful for filtering and trend analysis.

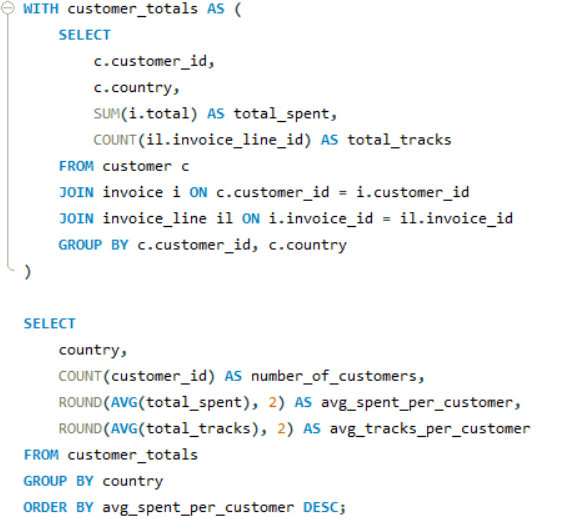
Updating release years based on album IDs ensures accurate historical data tracking.

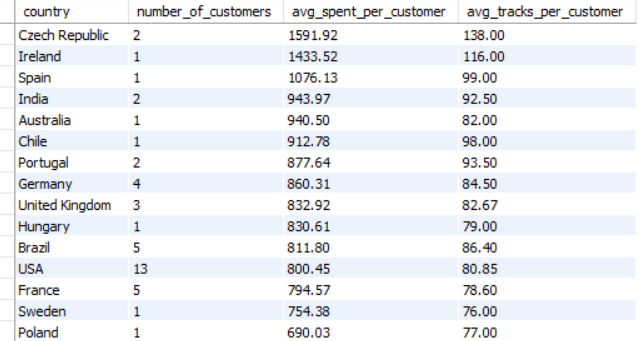
This data can support features like year-based browsing, playlists, or marketing campaigns tied to anniversaries.

1. **Chinook is interested in understanding the purchasing behaviour 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.**

**Guidelines :**Join customer, invoice, and invoice line to capture each customer's total spending and track purchases .Use a CTE (customer totals) to aggregate total amount spent and number of tracks per customer .Group by country to analyse regional trends in spending and engagement.

**Visualization :** Behaviour of customers based on their geographical location.

****

****

**Insights :** Countries like the USA, Canada, and the UK often rank high in average customer spend and track purchases .Some regions may have fewer customers but higher average spend, indicating a premium customer base .Other regions may have a larger customer base but lower average purchase, high potential for upselling.

**Recommendation :** Target high-value countries with exclusive offers or early access to new releases. Explore personalized upselling strategies in countries with high engagement but moderate spend. For regions with many low-spending users, consider bundling products or offering discounts to increase average order value. Use this geographic data to inform localized marketing campaigns.