**Chinook Project**

**Objective Questions:**

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

To determine if any table in your dataset has missing values or duplicates, you would typically follow these steps:

A screenshot of a computer

Description automatically generated

* **Null Values Check**
  + **Table:** Album
  + **Findings:** There are no missing values for the album ID; the data is complete.
* **Duplicates Check**
  + **Method:** Employed COUNT(\*) with grouping by album ID.
  + **Findings:** No duplicate rows were found in the album table.
* **Overall Data Integrity**
  + **Check:** Conducted similar assessments for all other tables.
  + **Findings:** No missing values or duplicates were detected across any of the tables.

.

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

* Here first I have found top selling tracks in the USA

A screenshot of a computer program

Description automatically generated

* Then I found Top artist in the USA

A screenshot of a computer program

Description automatically generated

**Top Artist in the USA**

* **Artist**: Van Halen
* **Total Sales**: 42.57 million

A screenshot of a computer code

Description automatically generated

**A screenshot of a computer

Description automatically generated**

 The **Top 5 Genres in the USA** are:

1. **Rock**
2. **Alternative & Punk**
3. **Metal**
4. **R&B/Soul**
5. **Blues**

 **Rock** is the leading genre in the USA, with total sales of **555.39 million**.

 Additionally, I identified the famous genres associated with the **Top Artist** in the USA

A screenshot of a computer

Description automatically generated

 For the top artist **Van Halen**, the top genre **Rock** has total sales of **42.57 million.**

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

A white background with black text

Description automatically generated

**Output:**

A screenshot of a computer

Description automatically generated

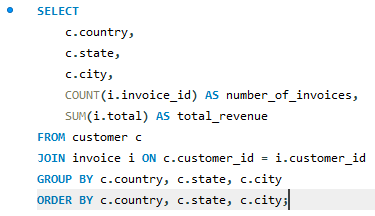
.

 Since age and gender data were not available in the Chinook customer base, we are unable to categorize customers by these factors. Instead, we can consider categories such as country and city.

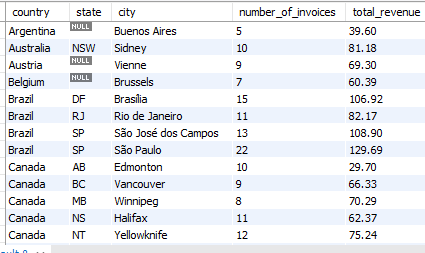
 I have identified the number of customers from each country by city.

The USA has the highest customer count, followed by Canada, Brazil, and France, with customer counts of 13, 8, 5, and 5, respectively.

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

To calculate the total revenue and the number of invoices for each country, state, and city:****

**Output:**

****

 The table above displays the total revenue and the number of invoices for each country, state, and city.

 The first graph illustrates the total revenue by country, while the second graph shows the total number of invoices by country. It’s evident that the USA has the highest total revenue of **1040.49** and the highest total number of invoices at **131**.

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

**A screenshot of a computer program

Description automatically generated**

**Output:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **customer\_id** | **first\_name** | **last\_name** | **country** | **total revenue** |
| 56 | Diego | Gutiérrez | Argentina | 39.6 |
| 55 | Mark | Taylor | Australia | 81.18 |
| 7 | Astrid | Gruber | Austria | 69.3 |
| 8 | Daan | Peeters | Belgium | 60.39 |
| 1 | Luís | Gonçalves | Brazil | 108.9 |
| 13 | Fernanda | Ramos | Brazil | 106.92 |
| 12 | Roberto | Almeida | Brazil | 82.17 |
| 11 | Alexandre | Rocha | Brazil | 69.3 |
| 10 | Eduardo | Martins | Brazil | 60.39 |
| 3 | François | Tremblay | Canada | 99.99 |
| 30 | Edward | Francis | Canada | 91.08 |
| 33 | Ellie | Sullivan | Canada | 75.24 |
| 32 | Aaron | Mitchell | Canada | 70.29 |
| 15 | Jennifer | Peterson | Canada | 66.33 |
| 57 | Luis | Rojas | Chile | 97.02 |
| 5 | František | Wichterlová | Czech Republic | 144.54 |
| 6 | Helena | Holý | Czech Republic | 128.7 |
| 9 | Kara | Nielsen | Denmark | 37.62 |
| 44 | Terhi | Hämäläinen | Finland | 79.2 |
| 42 | Wyatt | Girard | France | 99.99 |
| 39 | Camille | Bernard | France | 79.2 |
| 43 | Isabelle | Mercier | France | 73.26 |
| 40 | Dominique | Lefebvre | France | 72.27 |
| 41 | Marc | Dubois | France | 64.35 |
| 37 | Fynn | Zimmermann | Germany | 94.05 |
| 36 | Hannah | Schneider | Germany | 85.14 |
| 2 | Leonie | Köhler | Germany | 82.17 |
| 38 | Niklas | Schröder | Germany | 73.26 |
| 45 | Ladislav | Kovács | Hungary | 78.21 |
| 58 | Manoj | Pareek | India | 111.87 |
| 59 | Puja | Srivastava | India | 71.28 |
| 46 | Hugh | O'Reilly | Ireland | 114.84 |
| 47 | Lucas | Mancini | Italy | 50.49 |
| 48 | Johannes | Van der Berg | Netherlands | 65.34 |
| 4 | Bjørn | Hansen | Norway | 72.27 |
| 49 | Stanisław | Wójcik | Poland | 76.23 |
| 34 | João | Fernandes | Portugal | 102.96 |
| 35 | Madalena | Sampaio | Portugal | 82.17 |
| 50 | Enrique | Muñoz | Spain | 98.01 |
| 51 | Joakim | Johansson | Sweden | 75.24 |
| 53 | Phil | Hughes | United Kingdom | 98.01 |
| 54 | Steve | Murray | United Kingdom | 79.2 |
| 52 | Emma | Jones | United Kingdom | 68.31 |
| 17 | Jack | Smith | USA | 98.01 |
| 20 | Dan | Miller | USA | 95.04 |
| 22 | Heather | Leacock | USA | 92.07 |
| 21 | Kathy | Chase | USA | 91.08 |
| 26 | Richard | Cunningham | USA | 86.13 |

 I used the RANK() function to identify the top 5 customers, partitioned by country.

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

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

* To identify the top-selling track for each customer, we need to find the track that generates the highest revenue.

A screenshot of a computer program

Description automatically generated

**Output:**  
A screenshot of a computer

Description automatically generated

 This is a small portion of the output table shown above.

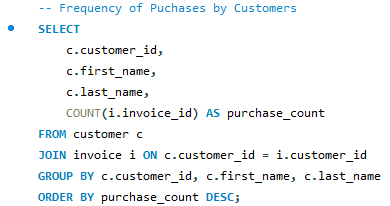
 The output displays a list of individual customers along with their respective top-selling tracks. Each customer is associated with one song they purchased, reflecting the highest recorded value for that track.

 Most tracks sold for **$0.99**, with a few exceptions where tracks were sold for a higher price, such as **$1.98**.

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

* Analyzing customer purchasing behavior for patterns or trends—such as purchase frequency, preferred payment methods, and average order value—can provide valuable insights.

1. **Frequency of Purchases by Customers**:
   * To determine purchase frequency, we can calculate how often each customer makes a purchase, which will help identify regular customers.



**Output:**

A screenshot of a computer

Description automatically generated

* From the output, we can see that **František Wichterlová** has the highest purchase count, with a total of **18**.

**Key Insights**: This highlights which customers are the most frequent buyers.

1. **Preferred Payment Methods**:
   * Since the Chinook database does not include a column for payment methods, we cannot analyze which payment methods are used most frequently without additional information.
   * If there were a column representing payment methods, we could analyze which methods customers use most often.
   * **Hypothetical Analysis**: If there were a payment method column:

**SELECT**

**payment\_method,**

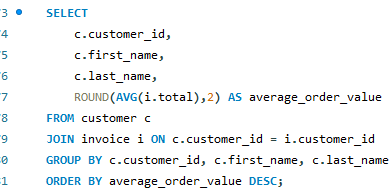
**COUNT(invoice\_id) AS count\_of\_invoices**

**FROM invoice**

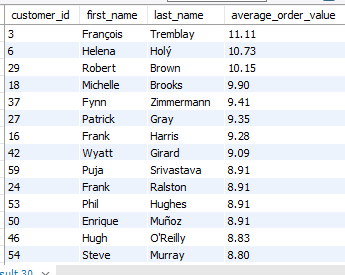
**GROUP BY payment\_method**

**ORDER BY count\_of\_invoices DESC;**

**Average Order Value**:

* The Average Order Value (AOV) is calculated by dividing the total revenue by the number of invoices. This metric can help identify trends in how much customers spend on average per purchase.
* order.

**Output:**



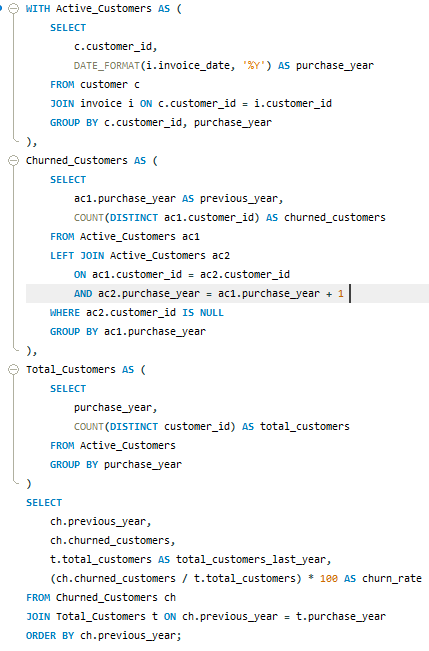
* Here, we can see that **François Tremblay** spends more per purchase.

**Key Insights**: This helps identify which customers typically spend more on each transaction.

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

* The customer churn rate is a key metric for understanding how many customers stop using a service over a specific period. To calculate the churn rate for Chinook’s customer base, we need to determine how many customers ceased making purchases compared to the total number of customers during that period.
* **Define Active Customers**: Active customers are those who made a purchase within the given period (e.g., month, quarter, or year).
* **Identify Churned Customers**: A churned customer is one who was active in a previous period but did not make any purchases in the current period.
* **Calculate the Churn Rate**: The churn rate can be calculated using the following formula:

Churn Rate=(Number of Churned CustomersTotal Number of Customers at the Beginning of the Period)×100\text{Churn Rate} = \left( \frac{\text{Number of Churned Customers}}{\text{Total Number of Customers at the Beginning of the Period}} \right) \times 100Churn Rate=(Total Number of Customers at the Beginning of the PeriodNumber of Churned Customers​)×100

* Here we are considering the customer churn rate on an annually basis.

**Output:**

A screenshot of a number

Description automatically generated

Here’s a refined version of your statements:

The results will show the churn rate for each year, including:

* **previous\_year**: The year in which churn was assessed.
* **churned\_customers**: The count of customers who churned that year.
* **total\_customers\_last\_year**: The total number of customers who made a purchase in the previous year.
* **churn\_rate**: The calculated churn rate for that year.
* In 2016, **4 customers** who made a purchase did not make any purchases in 2017.
* In 2017, **5 customers** who made a purchase did not return in 2018.
* Only **1 customer** from 2018 did not make a purchase again in 2019.
* All **58 customers** who purchased in 2019 did not return in 2020.

### Observations:

* **Gradual Increase in Churn**: Churn rates slightly increased in 2017 and 2018, with very low churn in 2019.
* **Major Churn in 2020**: In 2020, there was complete customer churn—no returning customers from the previous year. This could indicate a significant event in 2020 that caused all customers to stop purchasing.

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

* To calculate the percentage of total sales contributed by each genre in the USA and to identify the best-selling genres and artists, we will use the provided tables, specifically focusing on the following:
  + **Customer Table**: To identify customers located in the USA.
  + **Invoice Table**: To retrieve total sales from the total field and filter for invoices related to USA-based customers.
  + **Invoice Line Table**: To obtain sales details per track.
  + **Track and Genre Tables**: To associate tracks with their respective genres.
  + **Artist Table**: To access artist details.

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

**Output:**

A screenshot of a data

Description automatically generated

* **Top-Selling Genre**: The genre **Rock** dominates sales in the USA, contributing **53.38%** of the total revenue, significantly ahead of other genres such as **Alternative & Punk** (12.37%) and **Metal** (11.80%).
* **Top Artist**: **Van Halen** consistently ranks as the top-selling artist across all genres, contributing **42.57 million** to the total sales.

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

* To find customers who have purchased tracks from at least **3 different genres**, we can use the following SQL query. This query joins the **customer**, **invoice**, **invoice\_line**, **track**, and **genre** tables to identify the distinct genres each customer has purchased tracks from, and then filters for customers who have purchased from at least **3 different genres**.

A screenshot of a computer program

Description automatically generated

**Output:**

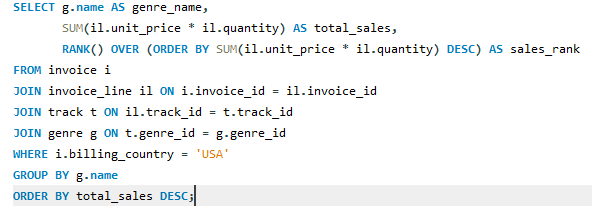
A screenshot of a computer

Description automatically generated

* **Top Customers by Genres Purchased**: **Leonie Köhler** (customer ID 2) is the top customer, having purchased tracks from **14 different genres**. She is followed by several customers, including **František Wichterlová**, **Terhi Hämäläinen**, and **Madalena Sampaio**, each purchasing from **13 genres**.
* **Distribution of Genre Purchases**: Most customers have purchased tracks from **8 to 12 different genres**, with only a few having purchased from fewer genres, such as **Robert Brown**, who purchased from **5 genres**.

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

* To rank genres based on their sales performance in the USA, we can use a SQL query to aggregate the sales data by genre and order them by total sales in descending order. We can utilize the RANK() function to assign rank numbers accordingly.



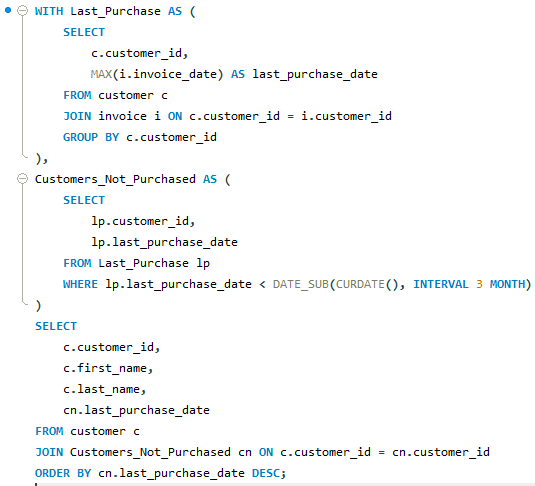
**Output:**

|  |  |  |
| --- | --- | --- |
| **genre\_name** | **total\_sales** | **sales\_rank** |
| Rock | 555.39 | 1 |
| Alternative & Punk | 128.7 | 2 |
| Metal | 122.76 | 3 |
| R&B/Soul | 52.47 | 4 |
| Blues | 35.64 | 5 |
| Alternative | 34.65 | 6 |
| Latin | 21.78 | 7 |
| Pop | 21.78 | 7 |
| Hip Hop/Rap | 19.8 | 9 |
| Jazz | 13.86 | 10 |
| Easy Listening | 12.87 | 11 |
| Reggae | 5.94 | 12 |
| Electronica/Dance | 4.95 | 13 |
| Classical | 3.96 | 14 |
| Heavy Metal | 2.97 | 15 |
| Soundtrack | 1.98 | 16 |
| TV Shows | 0.99 | 17 |

 **Top Genres**: **Rock** leads by a significant margin with total sales of **555.39 million**, followed by **Alternative & Punk** and **Metal**, which rank second and third with sales of **128.7 million** and **122.76 million**, respectively.

 **Lower-Ranking Genres**: The **TV Shows** and **Soundtrack** genres have the lowest total sales, with **0.99 million** and **1.98 million**, respectively, ranking 16th and 17th.

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

To identify customers who have not made a purchase in the last three months, I have used:

**Output:**

A screenshot of a computer

Description automatically generated

* **Customers Who Haven't Made Purchases in Over 3 Months**: The most recent purchases recorded in the list are from late December 2020, indicating that many customers have not made purchases in a significant amount of time.
* **Longer Gaps in Activity**: Some customers haven't made a purchase since as early as September 2019, reflecting a longer gap in customer engagement for a subset of the list.

**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.**

Based on the genre sales analysis, the three albums that should be prioritized for advertising and promotion in the USA from the new record label should align with the top-performing genres. Here’s the recommendation:

* **Rock Genre**: As the highest-selling genre in the USA with total sales of **555.39 million**, a Rock album would have the strongest potential for mass appeal and high returns.
* **Alternative & Punk Genre**: With total sales of **128.7 million**, this genre ranks second in performance, making it a strong candidate for promotion. Fans of Alternative & Punk tend to be passionate, which could drive significant engagement.
* **Metal Genre**: With sales of **122.76 million**, Metal is another top-performing genre. Prioritizing a Metal album could attract a dedicated and loyal fan base, ensuring solid sales.

The query is shown below:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

**Output:**

A screenshot of a computer

Description automatically generated

 **Top 3 Rock Albums for Promotion**: Based on sales analysis, the three albums to prioritize for advertising and promotion in the USA are:

* **From The Muddy Banks Of The Wishkah [Live]** by Nirvana
* **Are You Experienced?** by Jimi Hendrix
* **The Doors** by The Doors

 **Genre Focus**: All three albums fall under the Rock genre, indicating that Rock is a high-performing genre in terms of sales, making it a key area for promotional efforts.

1. **Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.**
2. To determine the top-selling genres in countries other than the USA and identify commonalities or differences, I followed a step-by-step SQL approach using Common Table Expressions (CTEs):
   1. **Extract Invoice Data**: Gather invoice data from countries other than the USA.
   2. **Group Sales Data**: Group the sales data by genre and calculate total sales for each genre in non-USA countries.
   3. **Identify Top-Selling Genres**: Sort the sales totals for each genre in each country to identify the top-selling genres.
   4. **Analyze Commonalities or Differences**: Analyze the top genres across the dataset to identify any commonalities or differences among countries.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

**Output:**

A screenshot of a music list

Description automatically generated

* **Rock Dominates Globally**: Across most countries, Rock is the top-selling genre. Countries like Brazil, Canada, France, and the United Kingdom show particularly high sales in Rock, demonstrating its universal appeal outside of the USA.
* **Genre Variations by Country**: While Rock is dominant, Alternative & Punk and Metal are frequently in the top 3 genres in various countries, indicating their widespread popularity. Some countries also have unique preferences, like Jazz in Spain and Austria, or Latin music in Argentina and Ireland.

**3. 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?**

* To analyse customer purchasing behaviour--comparing long-term customers with new customers--we need to calculate the following for both groups:
* **Frequency**: The number of purchases made (i.e., the number of invoices).
* **Basket Size**: The average number of items per order.
* **Spending Amount**: The total and average spending per customer.



A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

**Output:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **customer\_type** | **avg\_frequency** | **avg\_basket\_size** | **avg\_total\_spending** | **avg\_spending\_per\_order** |
| Long-Term | 10.4068 | 7.83495593 | 805.138475 | 78.66075885 |

**Insights:**

* Long-term customers typically have a higher purchase frequency, indicating stronger loyalty.
* A larger basket size for long-term customers suggests they tend to purchase more items per order.
* Long-term customers are likely to spend more in total and have a higher average spending per order, which can be an indicator of trust and satisfaction with the service.

**Retention Strategies**:

* **For new customers**: Focus on increasing purchase frequency and encouraging larger basket sizes through targeted promotions, discounts, or personalized recommendations.
* **For long-term customers**: Continue nurturing loyalty with rewards programs, personalized offers, or exclusive deals. Their higher spending and frequency show strong engagement, so retention efforts can focus on deepening relationships and encouraging referrals.
* **Higher Engagement from Long-Term Customers**: Long-term customers, on average, make over 10 purchases, demonstrating consistent engagement and loyalty over time. Their average basket size is around 7.83 items per purchase, indicating that they tend to purchase more per transaction compared to what might be expected from new customers.
* **Substantial Spending**: Long-term customers show strong spending behaviour, with an average total spending of $805.14, and they spend approximately $78.66 per order. This suggests that long-term customers not only return frequently but also contribute significantly to revenue, making them crucial for sustaining business growth.

**4. 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?**

* To conduct a Product Affinity Analysis in the context of music purchases, we can explore which music genres, artists, or albums are frequently bought together.

**Query to Analyze Genre Affinities:**

**A screenshot of a computer code

Description automatically generatedA screenshot of a music list

Description automatically generated**

* **Rock Dominance:** The genre Rock shows a strong connection with several other genres, especially Metal and Alternative & Punk, as indicated by high purchase numbers (**1622**). This suggests that fans of Rock music are likely to enjoy and buy albums from these related genres.
* **Diverse Genre Interests:** Customers who purchase Rock also buy music from other genres like Latin and R&B/Soul, highlighting a varied interest in different music styles. This indicates opportunities for promoting albums across these genres to attract a broader audience.

**Query to Analyze Artist Affinities:**

A screenshot of a computer program

Description automatically generatedA screenshot of a music list

Description automatically generated

* **Popularity:** Many customers who buy **Green Day** albums also purchase music from **Led Zeppelin** and **Foo Fighters**, suggesting opportunities for cross-promotion.
* **Artist Connections**: There are strong links between **Nirvana**, **Eric Clapton**, and **The Rolling Stones**, indicating that fans of these artists often enjoy each other's music, which can be useful for targeted marketing.

**Query to Analyze Album Affinities:**

**A screenshot of a computer program

Description automatically generatedA screenshot of a music album

Description automatically generated**

* **Strong Connections with "Mezmerize":** The album "**Mezmerize**" is frequently purchased alongside several other albums, including "**Are You Experienced?"** and "**Vault: Def Leppard's Greatest Hits**," suggesting it is a popular choice among customers who enjoy rock music.
* **Mutual Popularity:** Albums like "**The Singles**" and "**Dark Side Of The Moon**" are often bought together, indicating that customers who appreciate one of these albums are likely to be interested in the other, showing overlapping fan bases among these artists.

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?**

* To perform a **Regional Market Analysis** that examines customer purchasing behaviours and churn rates across different geographic regions, I have used the following SQL query:

A screenshot of a computer code

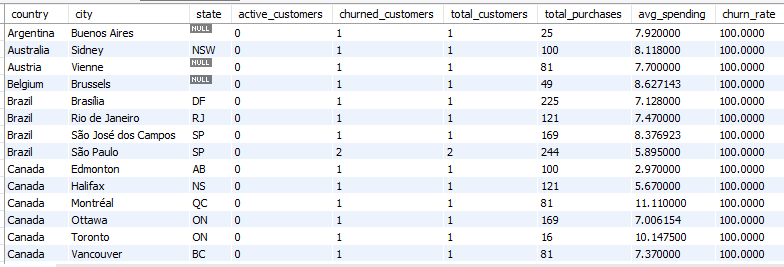
Description automatically generated

A screenshot of a computer program

Description automatically generated

* In summary, this query provides a comprehensive view of customer activity by region, counting how many customers are active versus churned, summing total purchases, and calculating the average spending along with the churn rate. It highlights differences in customer purchasing behaviour across various geographic locations.

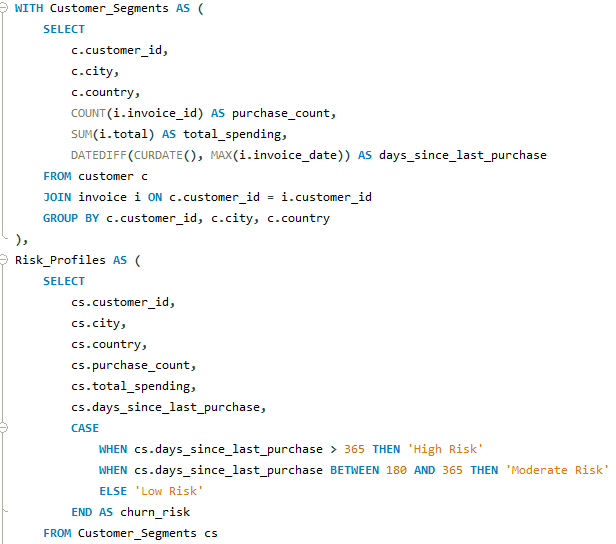
**Output:**

****

* **Customer Engagement:** Each row represents a city in different countries, showing that there are both active and churned customers. The counts for active customers and churned customers are mostly 1 or 2, indicating low customer engagement in these locations.
* **Average Spending:** The average spending per customer varies across locations, with most average spending amounts around 7 to 9, highlighting differences in customer purchasing behaviour based on geographic factors.

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?**

* The customer segments that are more likely to churn or pose a higher risk of reduced spending are primarily those classified as High Risk. These customers exhibit the following patterns, which contribute to their risk:

****

**A computer code with text

Description automatically generated**

**Output:**

****

**Factors Contributing to Churn Risk:**

* **High Risk Segment**: These **59** customers haven't made a purchase in over a year, which is a clear indicator of potential churn. Even though their average spending (**$79.82**) and average purchase count (**10.41**) are not particularly low, the fact that they haven't engaged in over 365 days signals that they are at high risk of stopping their relationship with the company.
* **Days Since Last Purchase:** Since they haven’t made a purchase in more than a year, they are highly likely to churn unless re-engaged.
* **Purchase Count:** Their purchase count, while not low, may reflect past loyalty or activity. However, the recent inactivity suggests disengagement.
* **Spending:** These customers once spent around $80 on average, but the absence of recent transactions highlights the need for intervention to prevent churn.

**Recommendation:**

* By focusing on re-engaging this high-risk segment, the company could potentially recover some of these customers and prevent further revenue loss.

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?**

* To predict **Customer Lifetime Value (CLV)**, we can leverage historical customer data such as tenure, purchase history, and engagement patterns.
* **Tenure** generally refers to the length of time that someone holds a particular position, job, or role within an organization.

A screenshot of a computer code

Description automatically generated

A computer code with text

Description automatically generated

**Output:**

**A screenshot of a computer

Description automatically generated**

* **Customer Segment**: Most customers fall into the "Active" category, meaning they are still purchasing regularly, with a churn risk of "Low."
* **Spending and Predicted CLV**: Active customers have an average current spending of approximately **$79.82**, with a projected customer lifetime value (CLV) of around **$53.26**, indicating good engagement and future potential for revenue.

**Insights and Targeted Strategies:**

1. **Loyalty Programs**: Focus on high-CLV customers by offering rewards to keep them engaged and extend their lifetime value.
2. **Re-Engagement Campaigns**: Target at-risk customers with incentives like discounts or personalized offers to prevent churn.
3. **Churn Prevention**: Analyze common characteristics (e.g., low engagement, high days since last purchase) among customers who churn to improve retention strategies.
4. **If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?**

If data on promotional campaigns is available, I would measure their impact on customer acquisition, retention, and overall sales by following steps;

* **Customer Acquistion:** I would count the number of new customers during each campaign.

**Sample query will be:**

SELECT campaign\_id, COUNT(DISTINCT customer\_id) AS new\_customers

FROM orders

WHERE first\_purchase\_date BETWEEN campaign\_start\_date AND campaign\_end\_date

GROUP BY campaign\_id;

* **Customer Retention:** I would calculate how many customers made repeat purchases after participating in a campaign.

**Sample query will be:**

SELECT campaign\_id,

COUNT(DISTINCT customer\_id) AS total\_customers,

COUNT(DISTINCT CASE WHEN last\_purchase\_date > first\_purchase\_date THEN customer\_id END) AS retained\_customers

FROM customer\_purchases

WHERE first\_purchase\_date BETWEEN campaign\_start\_date AND campaign\_end\_date

GROUP BY campaign\_id;

* **Overall Sales:** I will find the total revenue, number of orders, and average order value for each campaign.

**Sample query will be:**

SELECT campaign\_id,

SUM(order\_value) AS total\_revenue,

COUNT(order\_id) AS total\_orders,

AVG(order\_value) AS average\_order\_value

FROM orders

WHERE order\_date BETWEEN campaign\_start\_date AND campaign\_end\_date

GROUP BY campaign\_id;

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

To approach the problem of analyzing music record sales data at Chinook without any objective and subjective questions, I would follow an approach focused on discovering meaningful insights to inform strategy that are:

* Firstly, I will try to understand the Chinook’s Model, their Market, and their specific goals within the physical music market.
* So, by reviewing their past sales data, understanding the product offerings, for example: music genres, albums, playlist etc.
* Identifying their key challenges or opportunities, for example how much sales they are expecting to sale for each artist, each genre etc.
* Identifying the performance of artists and their sales etc.
* Which type of customers love different genres, artist by categorizing their ages.
* Finding which region is in top sales for each genre.
* Finding which artist is having highest sales in each region.
* Also comparing which genre is highest in each region so that we can focus more on that genre and supply new albums to each and respective countries.
* Analyzing which all customers are long-term and short-term by seeing their past purchases.
* Based on this analyzing result, I would recommend the changes be made to increase more profit accordingly.
* Also I would suggest to provide new type of genre and new languages which can make a trending sector in some regions for example : Indian songs which are having many different languages and different type of taste to the music.

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?**

* To add a new column named ReleaseYear of type INTEGER to the Albums table, I have used the following SQL ALTER TABLE statement:

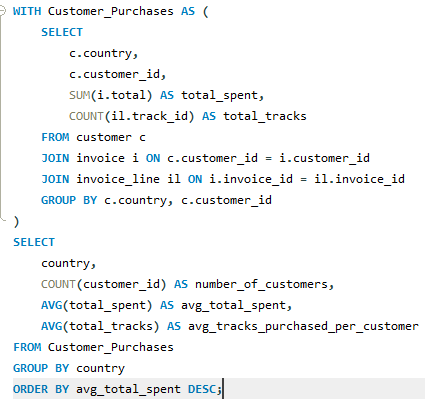
A close up of a sign

Description automatically generated

1. **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.**

* Firstly, calculated the total amount spent and total tracks purchased for each customer, grouped by their country. Then aggregated this data to find:

1. Number of customers per country.
2. Average total spent per customer.
3. Average number of tracks purchased per customer. It uses a WITH clause (CTE) for easier grouping and joins the necessary tables (customer, invoice, and invoice\_line) to gather purchase data.



**Output:**

**A screenshot of a data

Description automatically generated**

* **Top Spenders:** Countries like the **Czech Republic, Ireland**, and **Spain** have the highest average total spending per customer, with totals exceeding **$1,000** and more than **90** tracks purchased per customer.
* **Moderate to Low Spending:** The **USA** has the largest customer base (**13**), but the average spending per customer is lower at **$800** compared to top spenders. Other countries like **Denmark** and **Argentina** show much lower spending, with totals below **$400**.