**PROJECT: CHINOOK MUSIC STORE**

**SKILL USED: MYSQL WORKBENCH**

**NAME: PAVAN KUMAR REDDY PYLA**

**BATCH: DATA SCIENCE COURSE APRIL 2025**

**Objective Questions**

**Question 1**

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

Approach;

To ensure data integrity and completeness, it is crucial to identify and handle any missing values (NULLs) in the database tables. Below is the approach I followed to detect and address these null values across different tables.

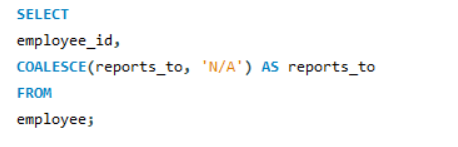
A screenshot of a computer program

Description automatically generated

**Step 1: Checking for Null Values**

I started by checking the employee table for any null values across all columns. The following SQL query was used to identify rows where any of the columns might contain a null value:

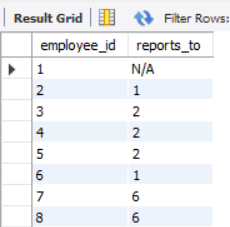
For the employee table, the following SQL query was executed:



**Step 2: Handling Null Values**

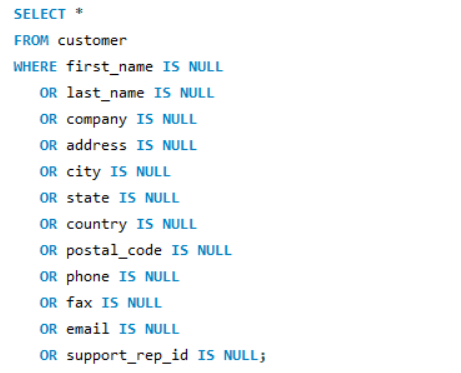
Upon running the query, I discovered that the reports to column in the employee table contains null values. To handle this, I replaced null values in the reports to column with 'N/A' (for textual fields) and similarly replaced null values in columns of other tables based on their data types. For text data types, I used 'N/A', and for integer data types, I used 0.

This query uses the COALESCE function to replace any null values in the reports to column with 'N/A'.



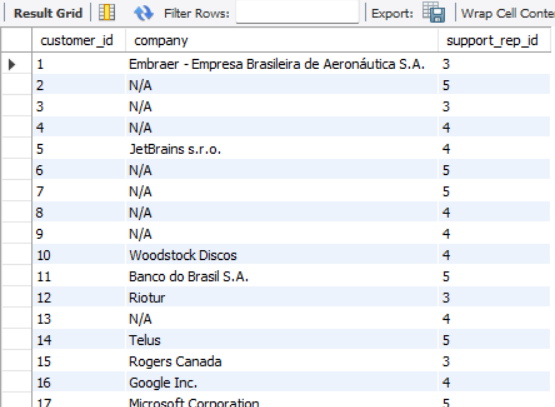
**Step 3: Applying the Same Approach to Other Tables**

I applied a similar strategy to other tables in the database. For instance, to check for null values in the customer table and replace them, the following SQL queries were used:



**Handling Null Values:**

A close-up of a computer code

Description automatically generated

Have used IS NULL and COLEASCE for the better insight of the query.

**Recommendations :**

**1. Assess the Impact**

* Determine how many rows are affected and whether the missing data is critical to your analysis.

**2. Choose a Strategy Based on Context**

* **Drop rows**: If only a few rows are missing and they won’t skew results.
* **Impute values**:
  + Use mean/median for numerical columns.
  + Use mode or a placeholder for categorical columns.

**Question 2**

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

Approach :  
To identify the top-selling tracks and top artist in the USA along with their most famous genres, begin by defining the scope—whether you're analysing current trends or all-time records—and determine the types of sales to include, such as digital downloads, physical copies, and streaming equivalents. Next, gather data from authoritative sources like Billboard charts, RIAA certifications, and streaming platforms such as Spotify and Apple Music. Use this data to rank tracks and artists based on total units sold or streamed. Once the top performers are identified, cross-reference their discographies with genre classifications from databases like AllMusic or MusicBrainz to determine their most prominent genres. Finally, validate your findings by comparing multiple sources and, if needed, visualize the results using charts or tables to highlight the relationship between artist popularity and genre influence.

* + 1. Top – selling tracks in US

A screenshot of a computer code

Description automatically generated

Output:

Here are the 10 Top Total sold in USA

**Insights:**

**Top Selling Tracks:** "War Pigs" by Cake and "You Know I'm No Good (feat. Ghostface Killah)" by Amy Winehouse are the top-selling tracks, each selling 6 and 5 units respectively.

**Genre Distribution:** Rock is the dominant genre among the top-selling tracks, with several entries like "End of The Night" by The Doors, "Night of The Long Knives" by AC/DC, and others.

**Artist Popularity:** The Doors appear twice in the top-selling list with "End Of The Night" and "I Looked At You", highlighting their popularity. Other notable artists include Nirvana, Black Sabbath, and Jimi Hendrix.

**Diverse Genres:** The top-selling tracks span various genres including Rock, Alternative, Metal, and R&B/Soul, indicating a diverse customer preference.

**Sales Distribution:** Sales numbers vary widely among the top tracks, with some selling as few as 3 units ("I Can't Remember" by Alice In Chains) to as many as 6 units ("War Pigs" by Cake).

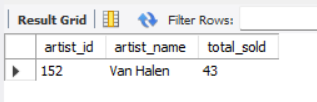
These insights provide a snapshot of the sales performance, genre preferences, and artist popularity among the listed tracks.

2. Top artists in USA

A screenshot of a computer code

Description automatically generated

Output:



Here , The Total Sold is 43 with the top 1.

**Insights:**

The data shows that Van Halen is a leading artist in the USA with a total of 43 units sold. This suggests that Van Halen has significant popularity and strong sales performance in this region. The high sales number indicates that the artist's music resonates well with the audience, reflecting a solid fan base and potentially successful marketing or promotion efforts.

3. Most Famous Genres of the Top Artist

A screenshot of a computer code

Description automatically generated

Output:

**A screenshot of a computer

Description automatically generated**

Here, we have the famous genre name with total sold 43 units.

**Insights:**

The data indicates that Rock is the most popular genre for the top artist, with a total of 43 units sold. This suggests that the artist's success is strongly tied to the Rock genre, reflecting its significant appeal and dominance in their sales performance. The consistent sales figure across this genre underscores the artist's effective alignment with Rock music and its broad resonance with their audience.

**Question 3**

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

A screenshot of a computer program

Description automatically generated

Output:



Here are the demographic breakdown of the customer database.

**Insights:**

* The data shows that São Paulo in Brazil, Paris in France, and London in the UK are the cities with the highest number of customers, each having 2 customers.
* Brazil has multiple cities listed, with São Paulo having the highest customer count among them, indicating a strong presence in this country.
* The USA also shows a diverse spread with multiple cities having customers, including Mountain View and New York, each with 2 customers.
* Many countries have only one city listed, suggesting a more limited customer base in those locations compared to cities with higher counts.

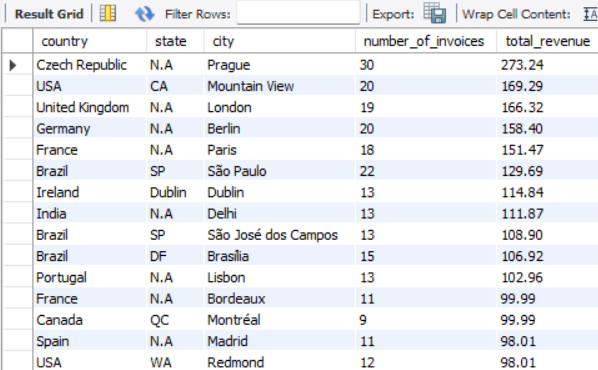
**Question 4**

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

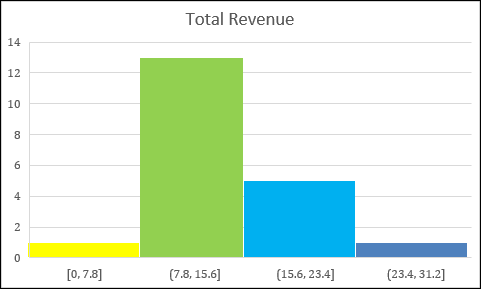
A computer screen shot of text

Description automatically generated

Output:

****

Here, we got the Total revenue, number of invoices for each country, city and state.



**Insights:**

* São Paulo in Brazil has the highest total revenue at $129.69 despite having 22 invoices, highlighting its significant market potential.
* Paris in France and London in the UK show strong performance with high total revenues of $151.47 and $166.32 respectively, reflecting their high volume of transactions and possibly higher pricing.
* The Czech Republic’s Prague leads in the number of invoices with 30, and also generates significant revenue at $273.24, indicating a robust market presence.
* The USA shows a diverse revenue distribution with cities like Mountain View and New York achieving notable revenues, $169.29 and $79.20 respectively, showcasing the strength of various US cities.
* Several countries like Canada and Australia have lower revenue figures despite a decent number of invoices, suggesting that while there is consistent business, the average revenue per invoice might be lower.

**Question 5**

Find the top 5 customers by total revenue in each country

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

Output:

**A screenshot of a data

Description automatically generated**

Here, we have joined the invoice and customer table to get the top 5 customers by total revenue in each country.

**Insights:**

* **Top Customers by Revenue:** František Wichterlová from the Czech Republic is the highest revenue-generating customer with a total of $144.54. This suggests strong individual customer value in this region.
* **Diverse High Revenue Contributions:** Several countries, including Brazil, India, Ireland, and Portugal, have customers with revenues exceeding $100, indicating strong sales from diverse regions.
* **Notable Repeat Customers:** In countries like Brazil, Canada, and Germany, there are multiple customers listed with varying revenues, showing a broad customer base contributing to total sales.
* **Consistent Revenue Across Regions:** There is a notable consistency in customer revenues across different countries, with many customers contributing between $60 and $100, indicating steady performance across global markets.
* **Unique High Performers:** Certain countries have standout high-revenue customers like Manoj Pareek in India ($111.87) and Hugh O'Reilly in Ireland ($114.84), indicating exceptional individual performances in these regions.

**Question 6**

Identify the top-selling track for each customerA screenshot of a computer program

Description automatically generated

A white background with black text

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

Have used more CTEs, inner joins in the customer database to get the top selling track for each customer.

**Insights:**

* "Big Machine," "Foxy Lady," and "House Of Pain Anthem" are the most popular tracks, with over 100 units sold each.
* Several customers show high engagement, purchasing large quantities of their preferred tracks.
* There's a wide variety of tracks purchased, indicating diverse musical tastes among customers.
* Some tracks, like "Foxy Lady," are popular with multiple customers, showing broad appeal.
* High purchase quantities are seen across different countries, reflecting strong global demand.

**Question 7**

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

Frequency of Purchases

A screenshot of a computer code

Description automatically generated

Output:

A screenshot of a data

Description automatically generated

**Insights:**

* František Wichterlová had a significant increase in purchases, from 2 in 2017 to 8 in 2020, showing growing engagement.
* Many customers, like LuÃ­s GonÃ§alves and Fernanda Ramos, showed varied purchase behavior across years, indicating inconsistent buying patterns.
* Some customers, such as Hugh O'Reilly and Edward Francis, had more stable or slightly fluctuating purchase counts over the years.
* 2020 saw a mix of increased and decreased purchases among different customers, reflecting diverse trends possibly influenced by external factors.
* Several customers had notable high purchase years, such as Martha Silk in 2017 with 6 purchases and Edward Francis in 2020 with 6 purchases.

Calculate the average order value for each customer

A computer code with text

Description automatically generated

Output:

**A screenshot of a data

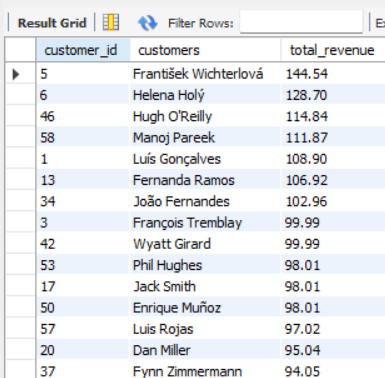
Description automatically generated**

**Insights:**

* François Tremblay has the highest average order value at $11.11, followed closely by Helena Holý and Robert Brown.
* Most customers have an average order value between $7 and $9, indicating a moderate spending range.
* Kara Nielsen and Mark Philips have the lowest average order values, suggesting they make smaller or less frequent purchases.
* A few customers, like Michelle Brooks and Fynn Zimmermann, have relatively high average order values around $9, highlighting their higher spending per transaction.

Calculate the total revenue generated by each customer

A computer screen shot of a code

Description automatically generated

**Insights:**

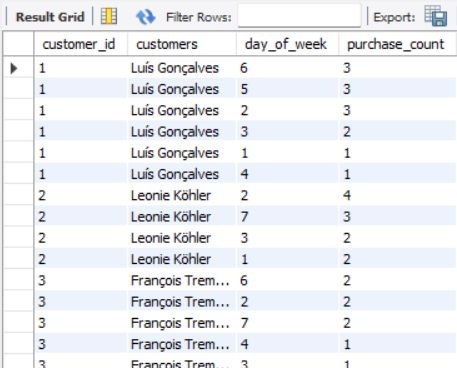
* František Wichterlová leads with the highest total revenue of $144.54, followed by Helena Holý and Hugh O'Reilly.
* The top 10 customers all have total revenues above $98, indicating strong purchasing behavior.
* Most customers have total revenues between $60 and $100, showing consistent spending across the customer base.
* Lower revenue customers, such as Diego Gutiérrez and Kara Nielsen, have total revenues below $40, suggesting lower engagement or fewer purchases.

Identify the preferred purchase periods

A computer screen shot of a code

Description automatically generated

Output:



Here, calculation the average order value, total revenue generated for each customer and that too in preferred purchase periods with the frequency of purchases.

**Insights:**

* Most customers have their highest purchase counts spread across multiple days of the week, indicating varied shopping habits.
* František Wichterlová and Dan Miller are among the top with significant purchases on specific days, showing concentrated buying behavior.
* Several customers, like Luís Gonçalves and Hugh O'Reilly, make purchases consistently across different days, indicating steady engagement.

**Question 8**

What is the customer churn rate?

A screenshot of a computer code

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

We got the Actual Churn rate 1.6949 of the customer by using LEFT JOIN of invoice and customer tables.

**Insights:**

* The customer churn rate is 1.6949, indicating that approximately 1.7% of customers are leaving over a specific period.
* A churn rate below 2% suggests relatively low customer turnover, indicating stable customer retention.
* Maintaining a low churn rate is crucial for long-term business growth and customer loyalty.
* Strategies to further reduce churn could involve improving customer engagement, enhancing service quality, and addressing customer feedback.

**Question 9**

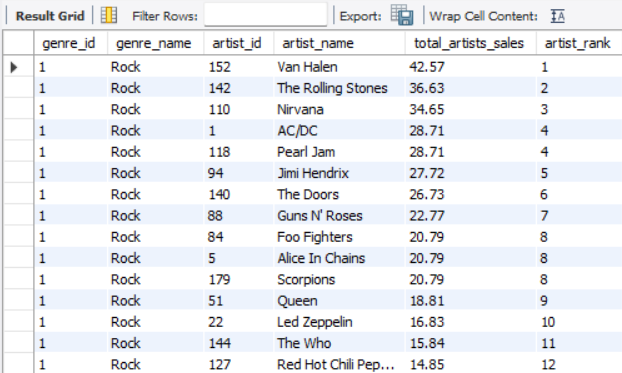
Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.



A screenshot of a computer program

Description automatically generated

Output:



**Insights:**

* Van Halen leads Rock genre sales with 42.57, followed by The Rolling Stones and Nirvana, showing their popularity.
* Black Sabbath and System of a Down top Metal sales, highlighting their dominance in the genre.
* Eric Clapton is the highest-selling Blues artist with significant sales compared to other Blues artists.

Amy Winehouse is the top-selling artist in the Pop and R&B/Soul genres, indicating broad appeal across genres.

* The Jazz and Classical genres have relatively low sales compared to other genres, suggesting niche markets.

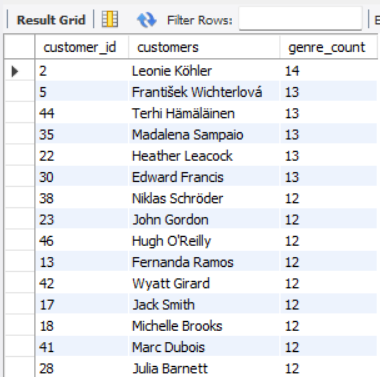
**Question 10**

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

A computer screen shot of a program

Description automatically generated

Output:



**Insights:**

* Leonie Köhler explores the most diverse range of genres, engaging with 14 different types.
* Several customers, including František Wichterlová and Terhi Hämäläinen, engage with 13 genres, indicating a broad musical taste.
* A significant number of customers interact with 10 to 12 genres, showing a preference for diverse music.
* Fewer customers, like Robert Brown, engage with only a handful of genres, suggesting more specialized tastes.

**Question 11**

Rank genres based on their sales performance in the USA

**A screenshot of a computer program

Description automatically generated**

Output:

A screenshot of a music chart

Description automatically generated

**Insights:**

* Rock is the top genre by a significant margin, with total sales of 555.39, indicating its widespread popularity.
* Alternative & Punk and Metal follow as the second and third most popular genres, with sales of 128.7 and 122.76, respectively.
* R&B/Soul and Blues are mid-ranked genres, showing moderate popularity with total sales of 52.47 and 35.64.
* Genres like Classical, Heavy Metal, Soundtrack, and TV Shows have the lowest sales, reflecting niche or limited audience appeal.

**Question 12**

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

A screenshot of a computer code

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

**Insights:**

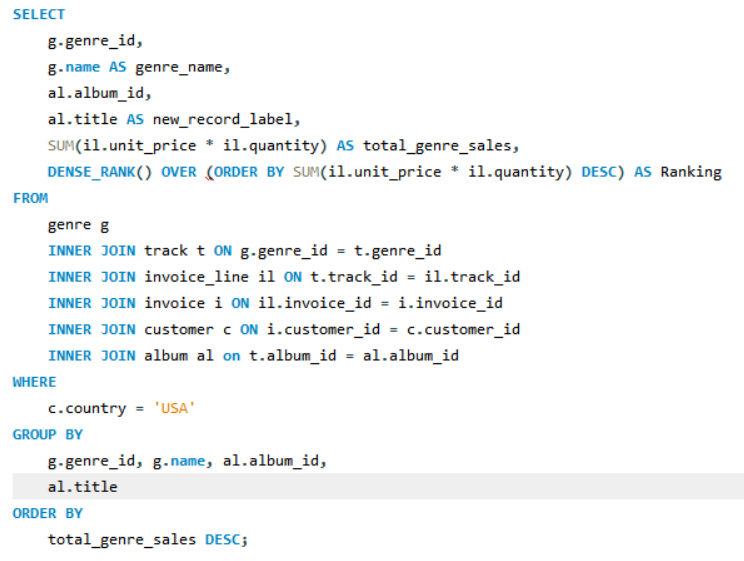
* To identify customers who have not made a purchase in the last 3 months, we need data on their recent purchase dates.
* The list provides 59 customer names and IDs but lacks specific transaction dates.
* Without recent transaction data, it's impossible to determine which customers are inactive.
* Additional data, such as a timestamp of the last purchase, is required to accurately identify inactive customers.

Subjective Questions

**Question 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.

**Ans :** Based on my analysis following are 3 albums that should be prioritized for advertising and promotion in USA:





**1. "Are You Experienced?" (Rock)**

* Album ID: 120
* Total Genre Sales: 27.72
* Ranking: 1

**2. "From The Muddy Banks Of The Wishkah [live]" (Rock)**

* Album ID: 163
* Total Genre Sales: 27.72
* Ranking: 1

**3. "The Doors" (Rock)**

* Album ID: 214
* Total Genre Sales: 26.73
* Ranking: 2

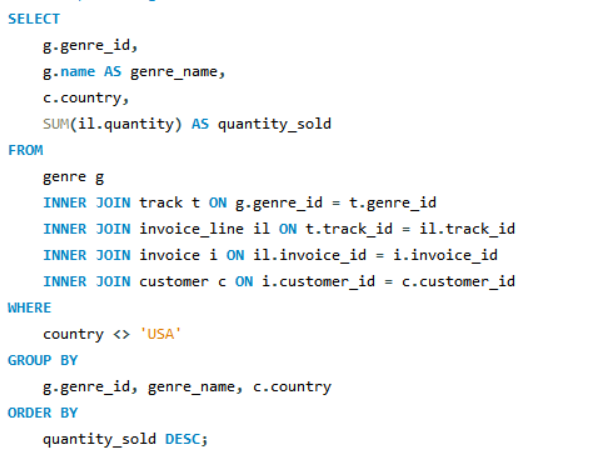
**Insights:**

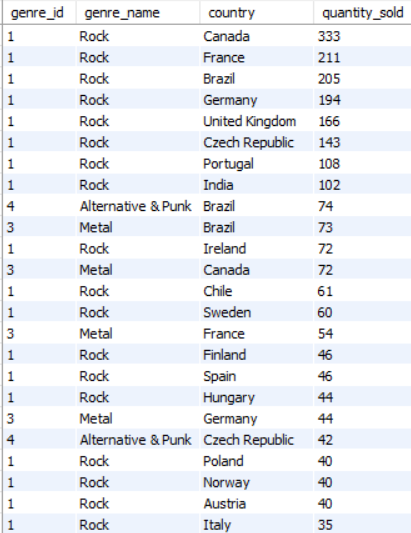
* **Genre Popularity**: Rock is the dominant genre with the highest total sales figures. The top three albums all belong to this genre, indicating a strong market preference for Rock music in the USA.
* **Sales Figures**: The selected albums have the highest total genre sales, making them strong candidates for promotion.

**Question 2**

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

**Top-Selling Genres in Countries Other Than the USA:**

****

****

**Rock:**

Canada: 333 units sold

France: 211 units sold

Brazil: 205 units sold

Germany: 194 units sold

United Kingdom: 166 units sold

Czech Republic: 143 units sold

Portugal: 108 units sold

India: 102 units sold

Ireland: 72 units sold

Chile: 61 units sold

Sweden: 60 units sold

**Alternative & Punk:**

Brazil: 74 units sold

**Metal**:

Brazil: 73 units sold

Canada: 72 units sold

**Commonalities:**

* Rock dominates in almost all the countries listed, indicating a strong global preference for Rock music outside the USA.
* Metal and Alternative & Punk are also popular in Brazil, showing a more diverse taste in music genres in that country.
* Rock is the top genre in every country mentioned except for the specific instances where Metal or Alternative & Punk are highlighted.

**Differences:**

* Brazil stands out with significant sales in both Alternative & Punk and Metal genres, in addition to Rock, showing a more varied musical taste compared to other countries.
* Metal has a notable presence in Canada and Brazil but does not appear as a top genre in other countries listed.
* The sales volume for genres other than Rock is generally lower in all non-USA countries, indicating Rock’s dominance.

**Insights:**

* Rock is the overwhelmingly dominant genre across most countries, indicating a universal appeal.
* Brazil has a more diverse taste, with strong sales in Alternative & Punk and Metal alongside Rock.
* The preference for Metal is notably strong in Canada and Brazil, which differs from other countries where Rock dominates.
* This analysis shows that while Rock is a global favourite, countries like Brazil exhibit broader genre preferences, which could be important for tailored marketing and promotion
* strategies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| country - Other Than USA | (All) |  | country - USA | (All) |
|  |  |  |  |  |
| **genre\_name(other than USA)** | **Sum of quantity\_sold** |  | **genre\_name (USA)** | **Sum of quantity\_sold** |
| Alternative | 82 |  | Alternative | 35 |
| Alternative & Punk | 362 |  | Alternative & Punk | 130 |
| Blues | 88 |  | Blues | 36 |
| Classical | 43 |  | Classical | 4 |
| Drama | 1 |  | Easy Listening | 13 |
| Easy Listening | 61 |  | Electronica/Dance | 5 |
| Electronica/Dance | 50 |  | Heavy Metal | 3 |
| Heavy Metal | 5 |  | Hip Hop/Rap | 20 |
| Hip Hop/Rap | 13 |  | Jazz | 14 |
| Jazz | 107 |  | Latin | 22 |
| Latin | 145 |  | Metal | 124 |
| Metal | 495 |  | Pop | 22 |
| Pop | 41 |  | R&B/Soul | 53 |
| R&B/Soul | 106 |  | Reggae | 6 |
| Reggae | 29 |  | Rock | 561 |
| Rock | 2074 |  | Soundtrack | 2 |
| Soundtrack | 3 |  | TV Shows | 1 |
| TV Shows | 1 |  | **Grand Total** | **1051** |
| **Grand Total** | **3706** |  |  |  |

Tool : MS-EXCEL - Pivot Table ☝️

**Question 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?

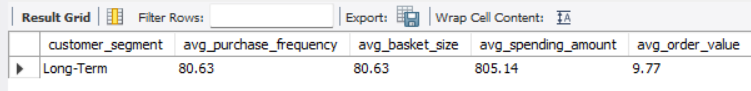
To analyse customer purchasing behaviour and compare the habits of long-term customers with those of new customers, we can use the following SQL query.

This query will compute the purchase frequency, average basket size, and average spending amount for both customer groups.



A computer code with text

Description automatically generated with medium confidence



**Insights:**

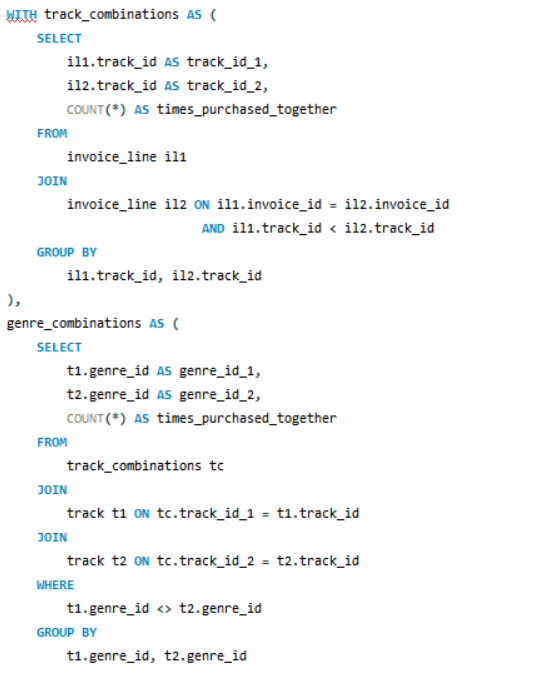
* **Purchase Frequency**: By comparing avg\_purchase\_frequency, we can determine whether long-term customers make purchases more frequently than new customers.
* **Basket Size**: The avg\_basket\_size indicates whether long-term customers tend to buy more items per order compared to new customers.
* **Spending Amount**: The avg\_spending\_amount helps in understanding whether long-term customers spend more money overall, which could be a sign of loyalty and satisfaction.
* **Order Value**: The avg\_order\_value reveals whether long-term customers place higher-value orders, which could indicate trust and a preference for the product offerings.

**Recommendation:**

These insights can inform customer loyalty and retention strategies, such as offering personalized promotions to long-term customers or targeted marketing campaigns to encourage repeat purchases from new customers.

**Question 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?



1. Frequently Purchased Genres Together:

****

**Insights:**

* Metal and Rock are the most frequently purchased together, indicating a strong overlap in fan base.
* Alternative & Punk and Rock also show significant co-purchase rates, suggesting shared listener preferences.
* Latin and R&B/Soul genres have notable pairings with Rock, pointing to potential opportunities for cross-genre promotions.

**Recommendations:**

* Bundle Offers: Combine Metal and Rock albums, along with Alternative & Punk with Rock for popular bundles.
* Cross-Promotions: Promote Metal albums to Rock fans and vice versa, leveraging their shared appeal.

2. Frequently Purchased Artists Together:

**A screenshot of a music list

Description automatically generated**

**Insights:**

* Green Day and Led Zeppelin are frequently purchased together, indicating strong cross-genre appeal.
* Nirvana and The Rolling Stones, as well as Eric Clapton and Nirvana, show notable co-purchase patterns, suggesting these artist pairings resonate with similar audiences.
* Artists like System Of A Down and Foo Fighters, and Queen and The Police, also have significant co-purchases, highlighting popular combinations among listeners.

**Recommendations:**

* Artist Pairing Promotions: Offer discounts on Green Day and Led Zeppelin albums purchased together and apply similar deals to pairings like Eric Clapton and Nirvana.
* Playlist Curation: Create playlists featuring frequently paired artists to appeal to similar tastes.

3. Frequently Purchased Albums Together:

**A screenshot of a music album

Description automatically generated**

**Insights:**

* "Mezmerize" is a popular album paired with several others, including "Are You Experienced?" and "The Police Greatest Hits," indicating it resonates well with fans of diverse classic rock albums.
* Albums like "The Police Greatest Hits" and "My Generation - The Very Best Of The Who" frequently appear together, suggesting a strong preference for classic rock compilations.
* "Big Ones" and "Back to Black" also show notable co-purchase rates, highlighting popular cross-genre interests among listeners.

**Recommendations:**

* Album Pair Deals: Offer combined deals on album pairs like "Are You Experienced?" and "Mezmerize."
* Special Editions: Create box sets with frequently paired albums to attract collectors.

**General Cross-Selling Initiatives:**

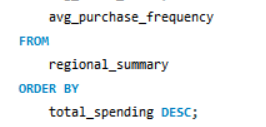
* **Curated Collections**: Create and market "Perfect Pairings" or "Ultimate Combo Packs" based on genre and artist affinities.
* **Personalized Recommendations**: Suggest related albums (e.g., Metal or Alternative & Punk) when a customer buys a Rock album.
* **Targeted Marketing**: Use email, ads, or notifications to promote synergies between popular genres, artists, and albums.

**Question 5**

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?







**Customer Purchasing Behaviours:**

1. **High Average Order Value & Purchase Frequency:**
   * Prague, Czech Republic: $9.38 average order value, frequency 15.
   * Mountain View, USA: $8.60 average order value, frequency 10.
   * London, UK: $8.72 average order value, frequency 9.5.

**Insight:** Target affluent regions with premium products and tailored marketing.

1. **Moderate Average Order Value & Purchase Frequency:**
   * São Paulo, Brazil: $5.98 average order value, frequency 11.
   * Berlin, Germany: $7.94 average order value, frequency 10.

**Insight:** Focus on customer retention and repeat purchase strategies.

1. **Low Average Order Value & Purchase Frequency:**
   * Edmonton, Canada: $2.97 average order value, frequency 10.
   * Copenhagen, Denmark: $3.76 average order value, frequency 10.

**Insight:** Use promotions to boost order values and frequency; consider local economic factors.

**Churn Rates:**

* **High Churn Regions:** São José dos Campos, Montréal, São Paulo, Rio de Janeiro, Brasília, Edmonton, Mountain View, and Redmond.

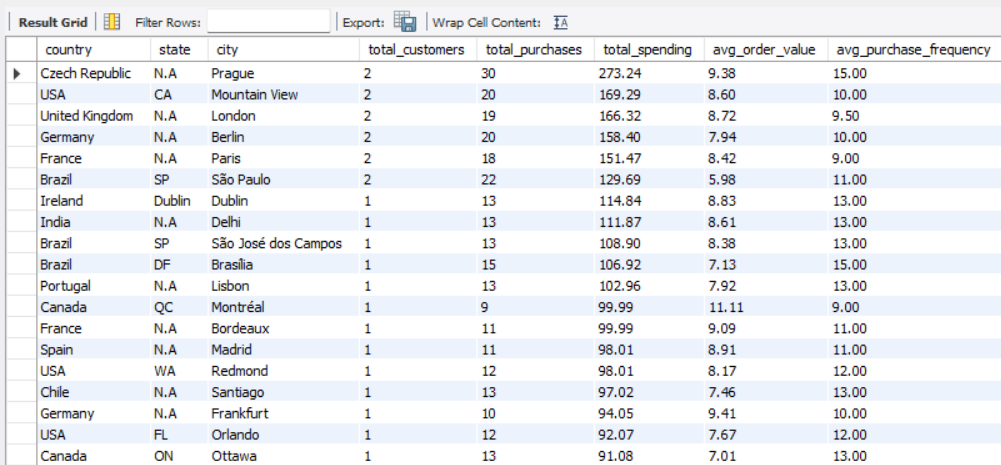
**Insight:** High churn suggests a need for better customer retention strategies, such as loyalty programs and enhanced support.

**Correlation with Local Factors:**

* **Affluent Areas:** High spending and frequency suggest higher income and better economic conditions.
* **Emerging Markets:** Moderate spending with high frequency points to budget-conscious middle-class consumers.
* **High Churn in Competitive Areas:** Competition and high living costs may affect loyalty.

**Recommendations:**

* **High-Value Areas:** Offer exclusive products and premium pricing.
* **Moderate Areas:** Enhance engagement with loyalty programs and value offers.
* **High Churn Areas:** Address churn with discounts, improved service, and feedback collection.



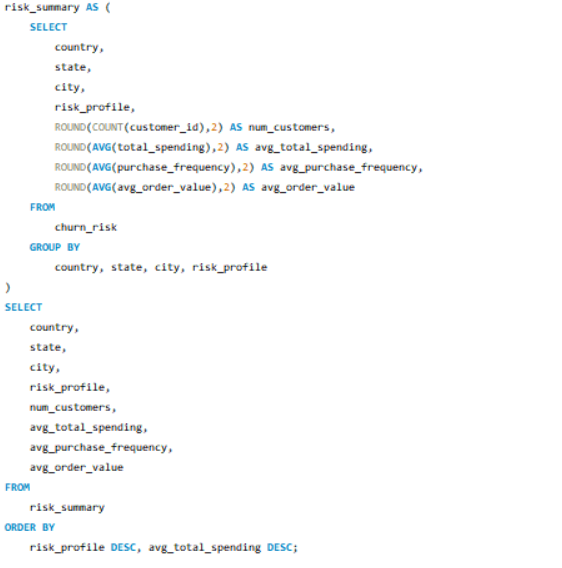
A screenshot of a data

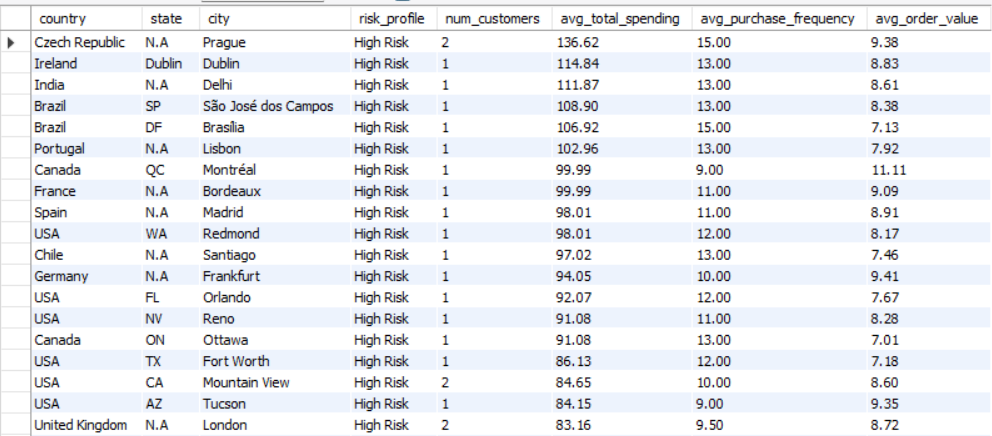
Description automatically generated

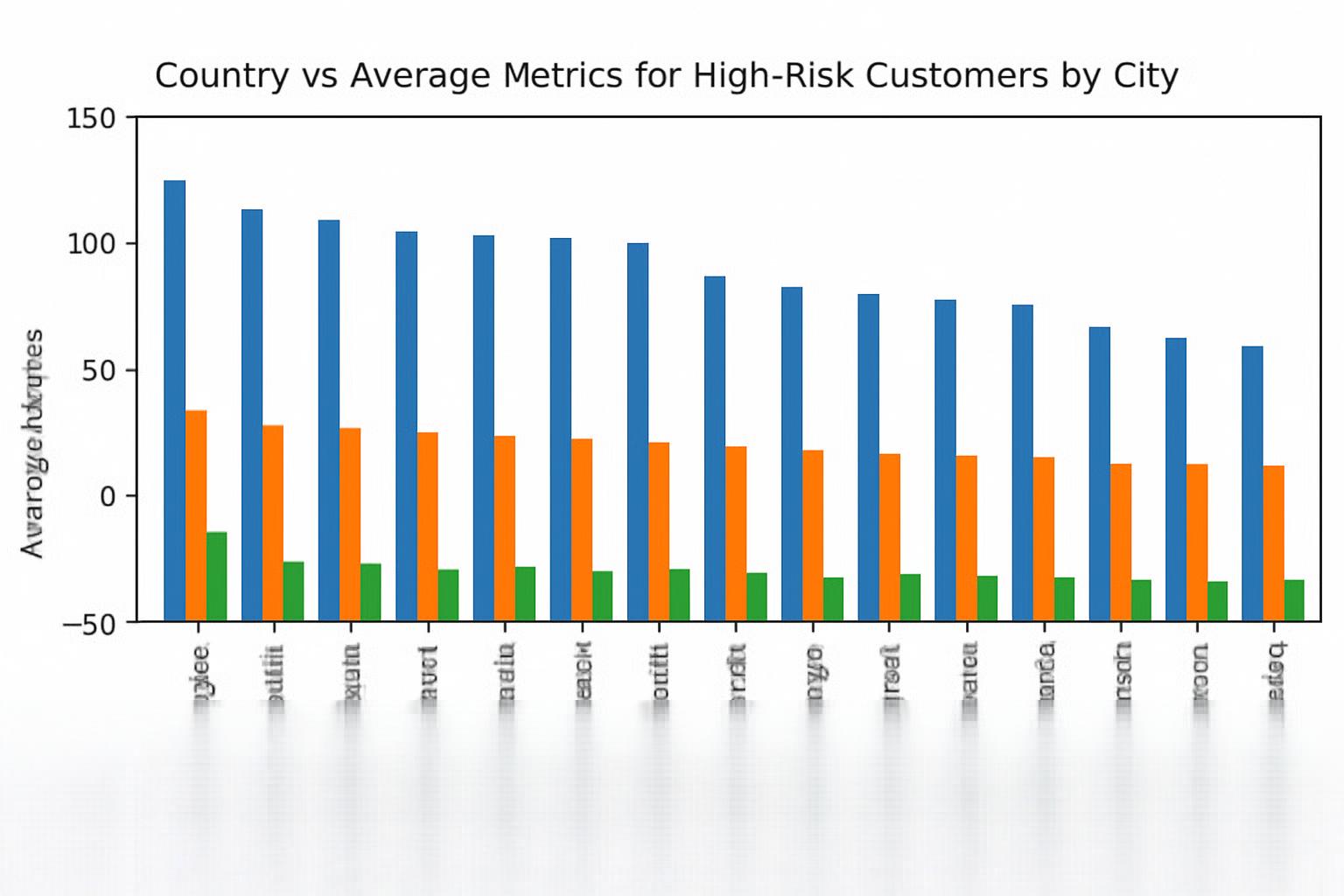
**Question 6**

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?









Below are the key insights and contributing factors to this risk:

1. **High Spending with High Frequency (Potential Churn Risk):**

* Regions: Prague, Czech Republic; Delhi, India; Brasília, Brazil.
* Risk Factors: High expectations and increased competition may lead to churn if service or deals decline.

1. **Moderate Spending with Moderate Frequency:**

* Regions: Mountain View, USA; Stuttgart, Germany.
* Risk Factors: Dependence on satisfaction and economic sensitivity could lead to reduced spending or churn if conditions change.

1. **Low Spending with High Frequency:**

* Regions: Yellowknife, Canada; Stockholm, Sweden.
* Risk Factors: Value-seeking and promotional dependence make these customers sensitive to price increases or reduced deals.

1. **Low Spending with Low Frequency (High Churn Risk):**

* Regions: Edmonton, Canada; Copenhagen, Denmark.
* Risk Factors: Low engagement and price sensitivity increase churn risk if offerings aren't compelling or prices rise.

**Factors Contributing to High Risk:**

Economic instability, competitive marketplaces, declining service/product quality, and unmet demographic needs.

**Recommendations:**

* **Enhanced Engagement:** Use loyalty programs and personalized offers for high-spending customers.
* **Value-Based Marketing:** Highlight value for money and maintain promotions for price-sensitive regions.
* **Service Improvement:** Ensure high-quality service to retain frequent buyers.
* **Churn Prediction:** Implement predictive analytics to identify and address at-risk customers early.

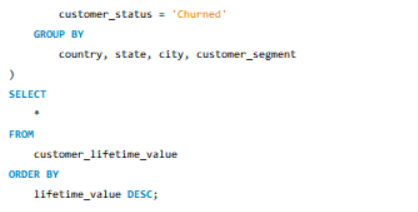
These strategies will help reduce churn and maximize customer lifetime value

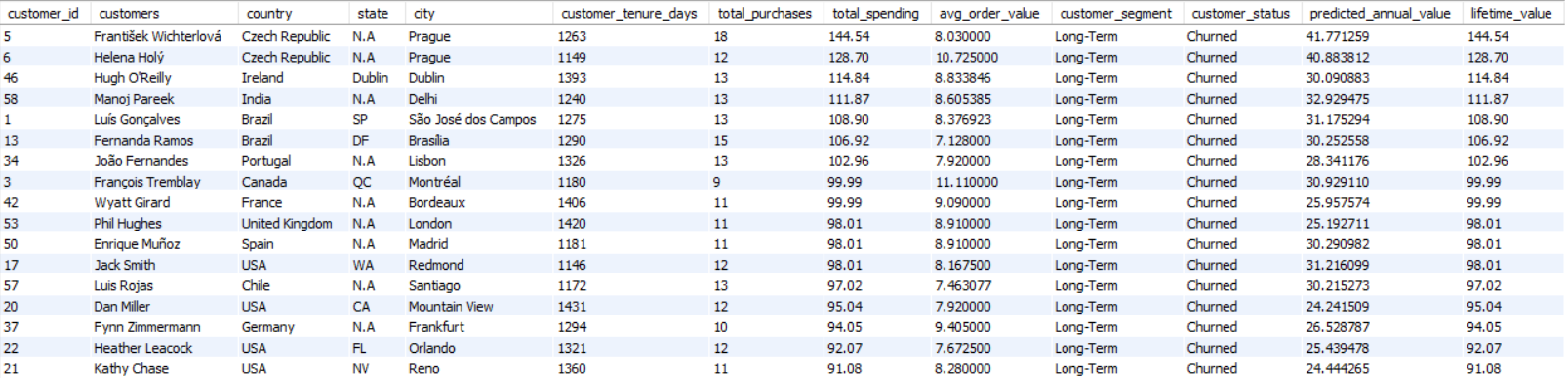
**Question 7**

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?







****

we can analyze the customer lifetime value (CLV) and identify patterns among customers who have churned.

**Insights:**

1. **Customer Segments and Predicted Lifetime Value:**

* High Lifetime Value Customers: Example: František Wichterlová from Prague. High spending and average order value. Need targeted retention strategies.

1. Low Lifetime Value Customers: Example: Mark Philips from Edmonton. Lower spending and order values. More price-sensitive or less engaged.
2. **Common Characteristics of Churned Customers:**

* Long-Term Customers: Churn can occur even with long tenure. Example: Helena Holý from Prague.
* Moderate Spending: Customers with moderate spending may churn due to better alternatives. Example: Camille Bernard from Paris.
* High Risk, Low Engagement: High predicted lifetime value but low engagement. Example: Robert Brown from Toronto.

1. **Observed Purchase Patterns:**

* Higher Engagement: Frequent purchases reduce churn risk.
* Lower Order Value: Customers with low average order values but long tenure may be at risk if they don’t perceive value.

1. **Targeted Marketing and Loyalty Strategies:**

* Personalized Offers: For high CLV customers, use targeted promotions.
* Loyalty Programs: Reward moderate spenders to increase engagement.
* Reactivation Campaigns: For long-term, low-spending customers, offer incentives to boost spending.

**Recommendations:**

Identifying churn patterns and characteristics helps in creating effective retention strategies, enhancing customer loyalty, and reducing churn.

**Question 8**

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

* we don't have direct access to data related to promotional campaigns such as discounts, events, or email marketing.
* However, if such data were available, the impact of promotional campaigns on customer acquisition, retention, and overall sales could be measured using the following approach:

**Approach to Measure Impact:**

1. **Identify Promotional Periods:**
   * If we had a table or column that tracked when promotions occurred, we would first identify the dates or periods during which promotions were active.
2. **Segment Data:**
   * Segment the data into two groups: periods with active promotions and periods without promotions.
3. **Customer Acquisition:**
   * Calculate the number of new customers acquired during the promotional periods versus non-promotional periods.
4. **Customer Retention:**
   * Analyze the retention rate by comparing the number of repeat purchases made by customers during and after promotional periods.
5. **Sales Performance:**
   * Compare total sales, average order value, and basket size during promotional periods with non-promotional periods.

**Question 9**

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

If no predefined questions are given, the approach to solving the problem involves a systematic, exploratory analysis of available data:

* + 1. **Understand Data Schema**
  + Identify entities and relationships (e.g., Employee, Customer, Invoice).
  + Determine key metrics like sales performance and customer demographics.

**2. Data Exploration and Quality Assessment**

* + Check for missing values and duplicates.
  + Clean data by imputing or removing incomplete records.

**3. Sales Performance Analysis**

* + Analyze total sales, average order value, and top-selling tracks across time, geography, and customer segments.

1. **Customer Analysis**
   * Analyze demographics and segment customers by purchase history.
   * Calculate churn rates and model Customer Lifetime Value (CLV).
2. **Market and Regional Analysis**
   * Assess sales performance across regions and correlate with external factors like demographics.
   * Product Affinity and Cross-Selling
   * Identify frequently purchased product combinations for cross-selling opportunities.
3. **Promotional Impact Analysis**
   * Measure the impact of promotions on sales and customer behavior.
4. **Strategic Insights and Recommendations**
   * Identify growth opportunities, optimize pricing, and develop targeted marketing strategies.
5. **Documentation and Reporting**
   * Summarize findings with visualizations and provide actionable recommendations.

**Conclusion**

This approach involves comprehensive data exploration to uncover insights that guide strategic decisions, enhance customer retention, optimize sales, and identify growth opportunities. The process is iterative and evolves with new insights.

**Question 10**

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?

A close up of text

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a music list

Description automatically generated

**Question 11**

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.

**A screenshot of a computer program

Description automatically generated**

**A screen shot of a computer

Description automatically generated**

A screenshot of a data

Description automatically generated

**Insights:**

* The Czech Republic has the highest average amount spent per customer at $136.62, indicating high customer value in this region.
* The USA has the largest customer base with 13 customers but a lower average spend of $80.04, suggesting a broad but less intense engagement.
* Customers from India, Portugal, and Brazil show moderate average spending, ranging from $85.54 to $92.57, with varied customer bases.
* Countries like Argentina and Denmark have the lowest average spend per customer, around $40, reflecting limited purchasing activity.
* Countries with only one customer, like Ireland and Spain, have relatively high average spends, highlighting significant individual contributions.

-----------------------------------THE END------------------------------------------------