**Objective Questions**

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

Ans: The data provided possess 0 duplicates although there are missing values in 3 tables.

Customer table : 47 fax, 29 state and 49 company values are null in the customer table

Employee table: 1 reports\_to value is null for employee\_id = 1

Track table: 978 composers are not assigned any value/are null in the track table

These Null values could be handled easily by using the Coalesce function in SQL.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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

Answer:Below are the top 10 selling Tracks of the USA, War Pigs by Cake being on top.



Below are the top most preferred genre in USA by number of records sold:



Rock is the most preferred genre in USA.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

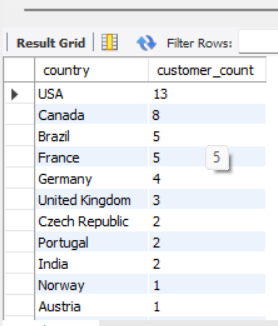
* Rock dominates US sales.
* "War Pigs" by Cake is the top track.

**Recommendations**

* Prioritise Rock in inventory and marketing.
* Use genre-specific promotions.
* Consider long-term genre diversification.

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

Ans. The customer base of Chinook music store is spread across 24 countries. USA is the country with most number of customers accounting to 13.



None of the provided tables have the data for age and gender analysis.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

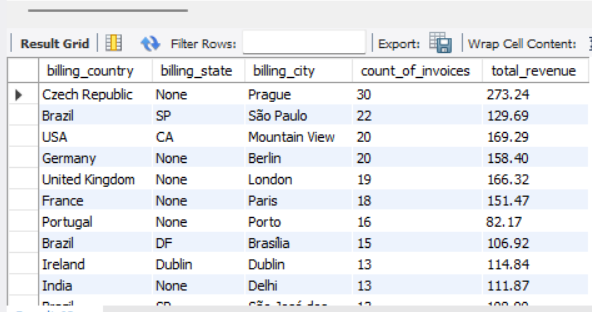
**Insights**

* Customers are spread across 24 countries.
* The USA has the most customers.

**Recommendations**

* Focus on the US market due to its large customer base.
* Explore growth opportunities in other countries, particularly Canada, Brazil, France, and Germany.
* Tailor marketing strategies by region.

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

Ans. 

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

* Prague (Czech Republic) generates the highest revenue.
* São Paulo (Brazil) has a high number of invoices but lower revenue.

**Recommendations**

* **Analyse Prague:** Investigate why Prague has high revenue per invoice; replicate this success in other cities.
* **Evaluate Revenue per Invoice:** Analyse this metric across all locations to identify areas for improvement and optimise pricing strategies.

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

Ans:



**Insights**

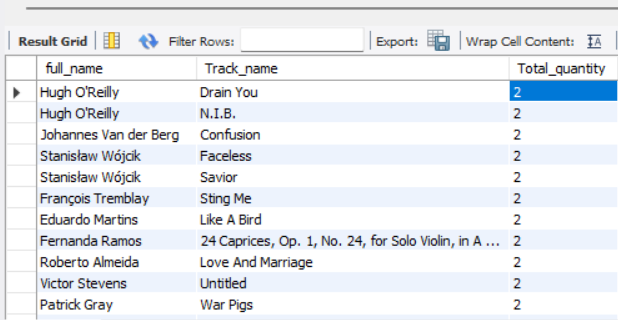
* Top customers vary by country.

**Recommendations**

* **Personalized Marketing:** Implement targeted campaigns for top customers in each country.
* **Loyalty Programs:** Consider tiered programs based on spending.
* **Customer Segmentation:** Group customers based on value and location for tailored communication.
* **Analyse Customer Behaviour:** Study the purchase history and preferences of top customers to understand what drives high value.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

1. Identify the top-selling track for each customer

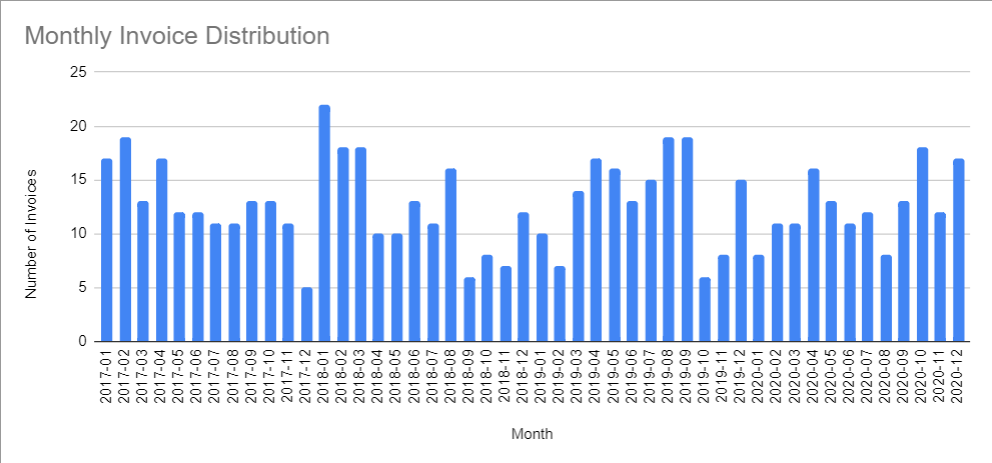
Ans: 

Identifying a top selling track for each customer is difficult as the distribution of tracks bought by customers doesn’t seem to be differentiating.

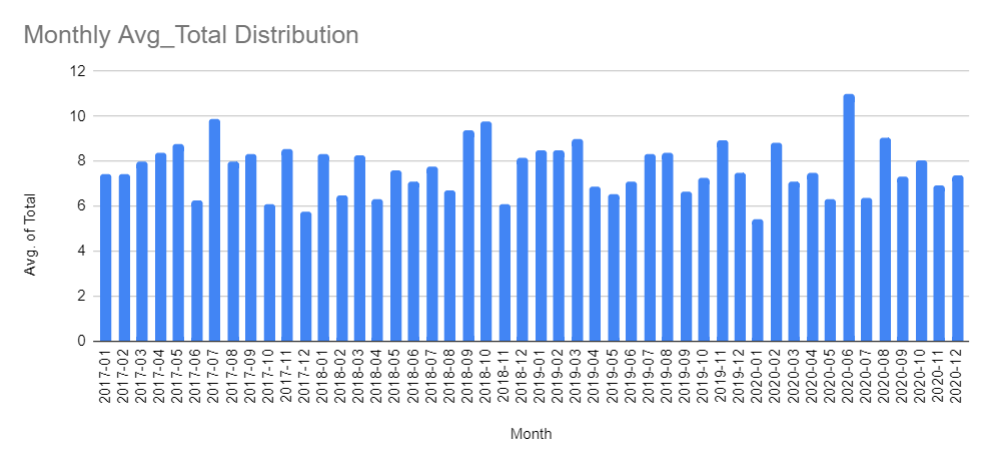
MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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

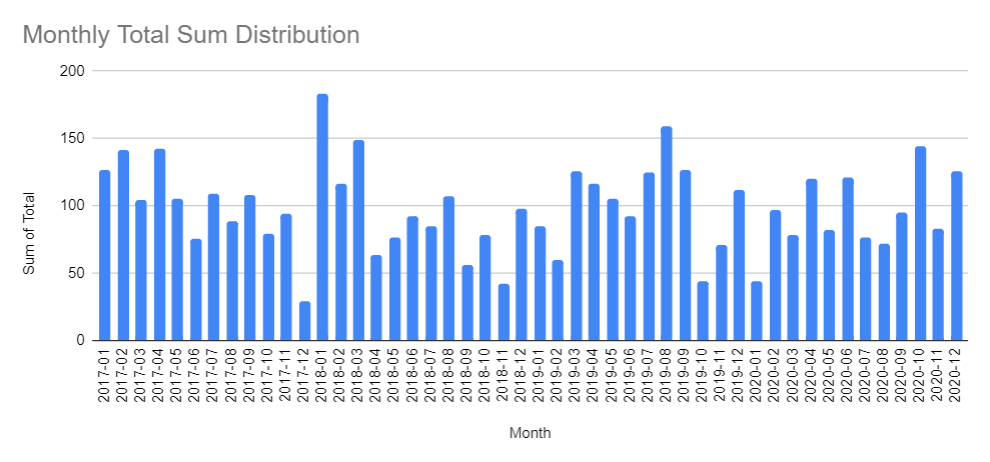
Ans:



The given chart summarises the monthly distribution of Invoices across the 4 year period.



The given chart summarises the monthly distribution of Average Total billing of invoices across the 4 year period.



The given chart summarises the monthly distribution of Total billing of invoices across the 4 year period.

**Insights:**

* Seasonality in Total Sum: The Monthly Total Sum Distribution has pronounced tops for December 2018 and January 2019, presuming higher sales activity during the holiday season and year end.
* Invoice Stability: Monthly Invoice Distribution is more or less consistent, even though there are lows in mid year segments (around June to July) and bouts of increase as we approach the year end with a bit of seasonality that is comparable to the grand total.
* Average Total Fluctuations: There is steadiness around the average of 7 to 8 units per month, though the Monthly Average Total Distribution every now and again breaks over 9 (like mid-2018 and around early 2020), possibly suggesting high value transactions might occur during that time.

**Recommendations:**

* Leverage High Sales Periods: Targeted campaigns and promotions during peak times (such as December and early January) can capitalize on times when naturally customer spending tends to peak.
* Address Mid-Year Lulls: Ride out the late April/mid May high and put in place promotions or campaigns around mid year (June/July) when both the invoice numbers and total sales drop for consistent revenue.
* Investigate High-Value Transactions: Success of strategies or customer behaviors in these top few months might inform you how to do it in different months with higher-than average totals (e.g., mid 2018, early 2020).

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

1. What is the customer churn rate?

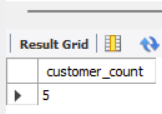
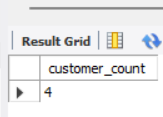
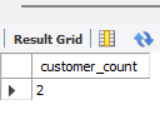
Ans:

**Overall Churn Rate :**

Methodology used:

Churn Rate = Total Churned Customers\*100​/Average Customer Base

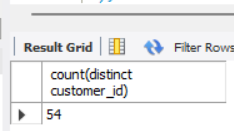
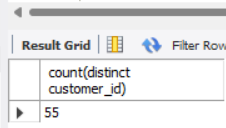
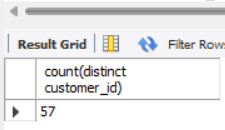
Total Churned Customers = Churned in Year 2018 + Churned in Year 2019 + Churned in Year 2020 = 2 + 4 + 5 = 11



Average Customer Base =

(Customers at Start of Year 2018 + Customers at Start of Year 2019 + Customers at Start of Year 2020)/3

= (57+55+54)/3 = 55.33



Churn Rate = 11 \*100/55.33 = 19.88%

**Insights:**

* Churn Rate (19.88%): A high churn rate means that almost 20% leave annually.
* Stable Customer Base: On the one hand, customer numbers are stable, while retention is weakening.

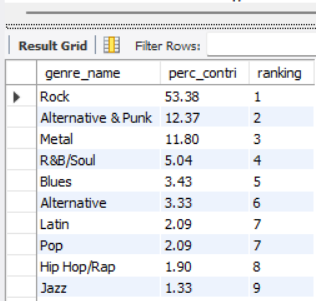
**Recommendations:**

* Boost Retention Programs: Loyalty program and personalised offer.
* Investigate Churn Causes: Let customers give you feedback to know where you are losing them.
* Improve Onboarding: Raise new customer engagement to strengthen retention.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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

Ans.



Best selling genre for the USA is Rock with a 53.38% contribution of the total revenue.

However, bestselling artist is Cake(refer Question 2)

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

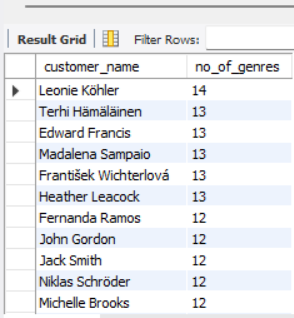
* Rock Dominates: Rock music alone accounts for more than 50% of the sales in the US.
* Cake is a Key Artist: Cake preferences have more people opting for this artist’s music which is why they are named the best selling artist.

**Recommendations**

* Focus on Rock: EmphasizeRock music in the actual stock and promotional campaigns.
* Promote Cake: Use Feature Cake in promotions and recommendations.
* Explore Related Genres: It also to consider advertising Alternative & Punk and Metal since its sales are quite high.
* Diversify Long-Term: Thus,izing Rock’s popularity, consider the ways to present and advertise other genres for extending the customer audience.

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

Ans:



All 59 customers have bought records from at least 3 different genres

Leonie Kohler being the top buyer amongst all.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

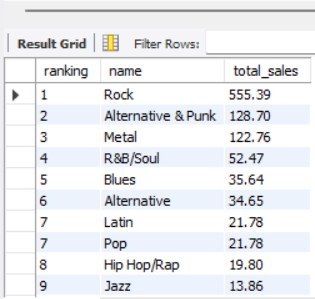
* Diverse Tastes: All customers have purchased from at least 3 different genres.
* Top Customer: Leonie Köhler purchased from the most genres (14) .

**Recommendations**

* Genre Recommendations: Draw information about favourite customer purchases and use it to recommend tracks in the newly considered genre.
* Targeted Promotions: Encourage promotions for types of products connected to a particular customer’s purchase.
* Reward Diverse Tastes: For example, think about a loyalty program for going to a concert of another genre.
* Analyze Genre Preferences: Examine how different genres overlap in order to find how they might benefit from synergizing.

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

Ans:



MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

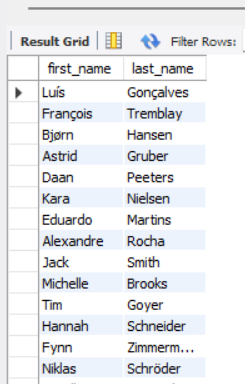
* Rock Leads: The top genre by a large margin is rock.
* Guitar-Driven Music: The top 3 genres are Rock, Alternative & Punk and Metal, so they have a lot of preference for this style of music.

**Recommendations**

* Inventory: Keep balanced inventory of Rock, Alternative & Punk,
* Markets concentrate on Rock music and related genres.
* Look into genre-specific promotions (like they do at the major labels: "We'll run a few thousand dollars per day in a 'Rocktober' sale.")
* Explore Subgenres: More targeted marketing: Investigate specific rock subgenres (e.g. classic rock, hard rock).
* Diversification (Long-term): Identify strategic ways to introduce and market less talked about genres in order to grow the customer base.

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

Ans:



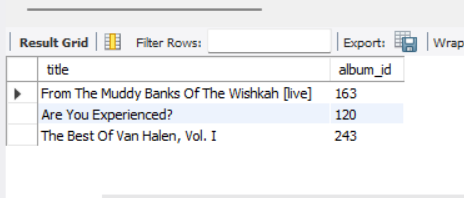
MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**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.

Ans: Based on previous analysis in question 11, Rock is the most preferred genre in the USA as per the sales. Therefore, top 3 albums from the Rock genre should be prioritised for advertisement and promotion in the USA.

Below are the top three albums based on quantity of tracks sold:



MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

* **Rock Rules:** Rock is the most preferred genre in the USA.
* **Top Albums:** The top 3 Rock albums are "From The Muddy Banks Of The Wishkah [live]", "Are You Experienced?", and "The Best Of Van Halen, Vol. I".

**Recommendations**

* **Prioritise Promotion:** Focus advertising and promotion efforts on above three Rock albums in the US market.
* **Consider Bundles:** Offer album bundles or special promotions featuring thealbums.
* **Targeted Marketing:** Use data to identify US customers who purchased Rock music and target them with ads for the albums.

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

Ans.

Excluding USA: USA:

Based on the data it is prominent that liking for Rock is the commonality between USA and the rest of the world as it has the top spot for both the cases. Apart from this there seems to be no commonality between the two data.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

**Insights**

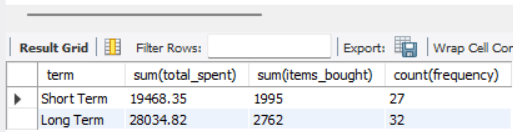
* Rock's Global Appeal: In the USA and all over the world the genre still is rock.
* US vs. International: In addition, the US differs quite a bit from other countries in terms of genre preference, other than Rock.
* Emerging Markets: Outside the USA Metal, Latin, and Jazz show stronger performance.

**Recommendations**

* Global Strategy: Continue to focus strong on Rock music worldwide.
* Regional Adaptation: Manage inventory and promotions based on local preferences.
* Outside the US, promote more Metal, Latin and Jazz.
* Targeted Marketing: Create region-specific marketing campaigns talking to local tastes.
* Genre Diversification: Take an interest in how to brainstorm ways to introduce and promote other less popular genres to the US and international markets.

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

Ans.



Methodology used:

* Calculated the average of customer period of association with chinook
* Divided the customers based on their period of association:If any customer is associated term > avg of all customers, then he/she is a long term customer. Otherwise, short term.
* Calculated the other required metrics for analysis between short term and long term customers like, total expenditure, items bought, frequency of purchase.

Key findings:

Revenue from Long term customers is greater than those of short term therefore, they are crucial for the business.

**Recommendations**

Customer loyalty plays a very important role in any business and the same is prominent in the case of chinook. Therefore, the company should work towards keeping the customers happy which will bring in the revenue for a longer term. This could be achieved by offering deals and discounts to the customers. Also the user experience is also important and should be dealt with utmost care.

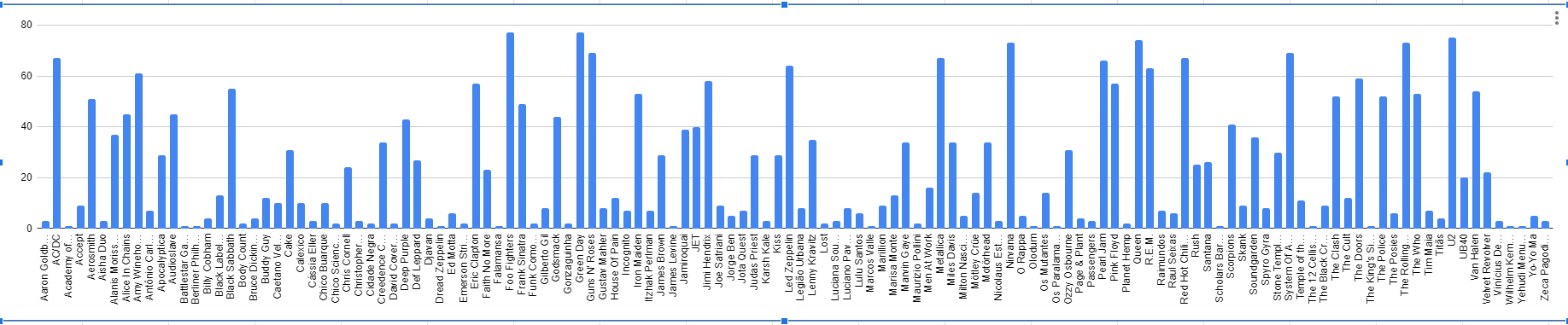
MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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?

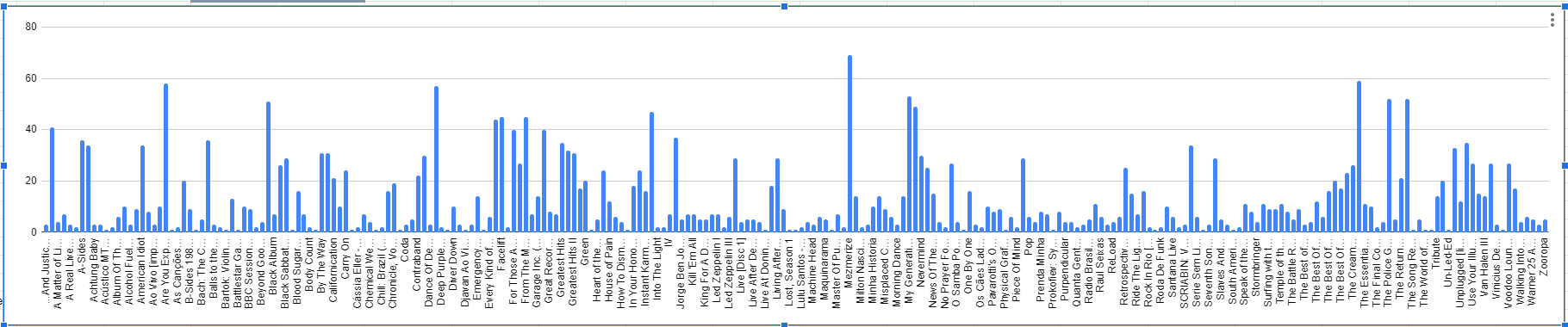
Ans:



The data obtained from the first query is used to plot this pivot table and it shows that Rock, Metal and Alternative & Punk are the most preferred Genres by the users. So, if a user buys one of these, the salesperson can recommend the buyer to try something from the other two.



Similarly, analyzing the sales of Artists through chart we get, Green Day, U2, Foo Fighters, Nirvana, The Rolling Stones, Queens and System of a Down as the top sellers. This information could help in cross selling of records/tracks.



Doing the same analysis on albums we get, Mesmerize, Are you Experienced and The Doors on the top.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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

Ans:



**Insights**:

* From the above data we can see that countries like the USA, Canada, Germany, UK, Australia, Brazil have either very low churn rate or negative churn rate. As these nations come under developed economies with high GDP, this could be a factor supporting the trend in the data.
* On the other hand, developing nations or less developed nations like Norway, Netherlands, Chile, Denmark, India,etc. Have comparatively higher churn rate depicting decreasing scope of business.

**Recommendation**:

* More advertisements need to be aired in the countries with high economy while cheaper tracks should be sold in the countries with low economy.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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?

Ans:

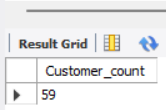
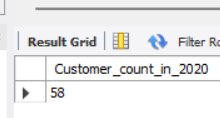
* Inactivity: Customers with a higher gap of purchase period between two purchases are more likely to churn.
* Low Purchase Frequency: Customers with fewer total purchases might be less engaged.
* Low Total Spending: Customers with lower total spending could be at risk of reducing spending further.
* Demographics: Certain age groups or gender segments in specific countries might display higher churn or lower spending compared to others. The effect of the economy on churn rate has been analysed in the previous question.
* The affect of age and gender on business could have been analysed if the data would have been available. This could have answered many questions like:

What’s the sales pattern over various ages?

What age/gender group prefer which music?

Which age/gender is more likely to churn?

Spending behaviour of different ages and genders.

Picture on the left showcases the overall customer count

Whereas, the picture on the right shows the customer count in 2020( latest year for the given data)

**Insights:**

* Customer base of Chinook seems to be loyal with very less attrition (only one customer didn’t do any purchase in the latest year).

**Recommendations:**

* As customer loyalty is crucial in any business and word of mouth play a significant role in assessing the business’ health therefore, such customers should be dealt with utmost care and specialised offers or campaigns should be launched to keep them loyal.

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?

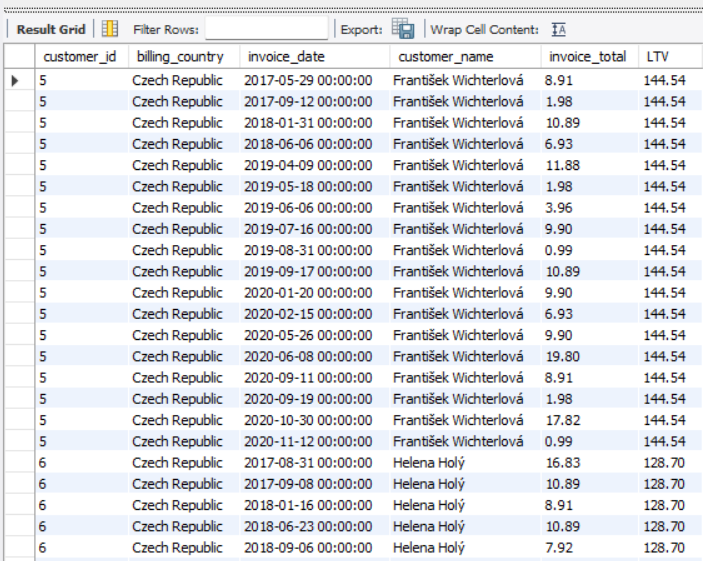
Ans:

Customer’s data such as purchase history, tenure, engagement,etc can be utilised to analyse traits such as the loyalty, high value/low value, frequent buyer or not ,etc

Recent customers with high purchase value can be future loyal customers and therefore, targeted marketing or promotional campaigns can be deployed for such potential buyers.

Customers can be segmented based on their purchase history and therefore, targeted marketing campaign can be deployed for such customers.

Less developed countries have higher customer churn rates. There could be several reasons for this like product market fit, less appropriate pricing, etc. Such issues need to be addressed for the business to grow.



The above picture shows the result of customer’s purchase history as well as their Life Time Value to the business.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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?

Ans:

* Identifying the tracks that have not been sold in last 6 months(i.e. Between 01 July 2020 and 31 Dec 2020)



3006 such songs/tracks were Identified out of 3503 total songs/tracks

* Identifying customers that haven’t made any purchase in last 6 months.



16 such customers out of 59 were identified that haven’t made any purchase in last 6 months.

Recommendations:

* Targeted promotions could be released on such tracks in order to promote their sales.
* Similarly, lucrative offers can be launched for customers that have been inactive for too long. This would boost the revenue of the business.

If data on promotional campaigns (discounts, events, email marketing) was provided we could analyse:

Calculate Customer Acquisition Cost (CAC) for each campaign: Solve for the total amount spent on a particular campaign (for example a mail campaign or an offer) by the number of new customers who are result of that campaign. It is helpful to know which of your campaigns are providing the greatest value by driving the greatest numbers of new customers at the lowest possible cost.

Attribute new customers to specific campaigns: Ensure each campaign gets a unique link, coupon code, or landing page so that you can tell where your new customers are coming from. This enables you to expressly attribute the acquisition to particular campaign.

2. Customer Retention

Measure changes in customer churn rate:

It is possible to track churn rates of your subscribers before some campaigns, during their running, and after their completion. Franchising drive can cause a product to last longer or alternatively can it cause customer churn to go down? It can also show you trends that will help you understand the long term effects of your promotions.

Track repeat purchase rates:

Analyse how frequently a consumer who engages in a promotion purchase a product again as opposed to the consumers who didn’t. It brings out whether or not promotions help to foster another one and develop long-term customer relationships.

3. Overall Sales

Analyse sales data during the promotional period:

Relate sales figures obtained during the promotion to that obtained when there was no such promotion. This will indicate the degree to which the campaign has affected the sales revenue in one way or the other.

In the same manner, it should be possible to track the AOV each time there is promotion being carried out. Are customers able to buy more per visit with the use of Discounts or during Events?

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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

Ans:

Here's how I'd approach the Chinook music sales data analysis if I wasn't given specific objectives:

1. Data Understanding

Schema Familiarisation: Evaluating the schema to determine tables, column, data type and relationship among tables.

Data Exploration: Use SQL queries to:

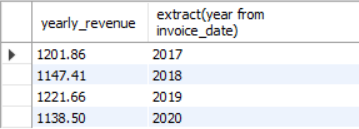
Detect for missing data and redundance

2. Ask questions pertaining to the data

Based on the initial exploration, formulate broad questions about the data, such as:

Sales Performance:

What is the general direction of sales?

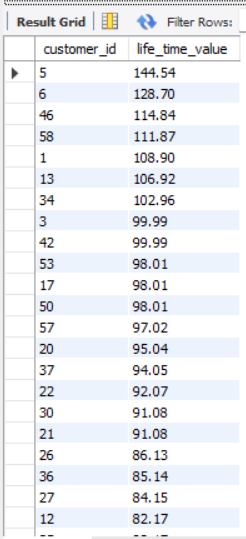


Are there fluctuations in sales over the year?

(Sales seem to be fluctuating as per the above query result.)

Customer Behaviour:

Which customers must be considered valuable?



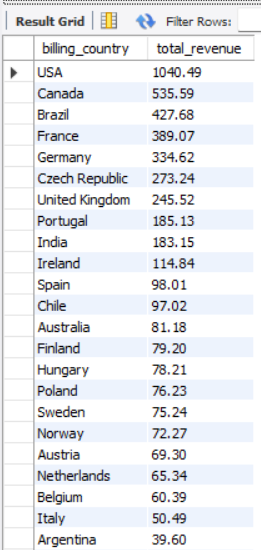
What should we know about our loyal customers?

(Define who are considered to be loyal first)

Do some customers make purchases differently from other customers?

Regional Analysis:

What is the trend in sales performance per geographic region?



Is there preference to regional genres or artists?

(Data insufficient for analysis)

3. Analyze and Illustration

Deep Dive with SQL: Generate more particular SELECT statements to the questions created in step 2. This may include aggregation, denormalisation, sub-query, join, window functions and many others.

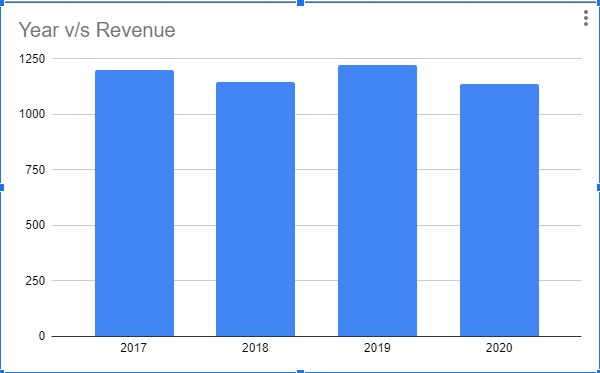
Visualization: Go ahead and analyze the raw data using analytical tools such as Tableau, Power BI or even excel to see the outcome of the analyzed patterns of the population. Create charts like:

The type of product graphs of sales dynamics for the identified period

Line graphs used in analyzing comparisons between sales, by genre, artist or by regions.

To describe relations between the two components, scatter plots are used.

Geographical maps in order to capture regional differences



**Insights and Recommendations**

Actionable Recommendations: offer recommendation to Chinook using the collected data, for instance:

* Methods of stores management
* Targeted marketing campaigns
* Fundamental and interconnected strategies which include customer segmentation and personalization.
* Recommend breaking the list into two parts, one for new products or a new genre to the viewer.
* Under this category, you will find the best strategies on how to set the right price and promote the product to the right market.

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?

Ans:

Query that can help in achieving this would be:

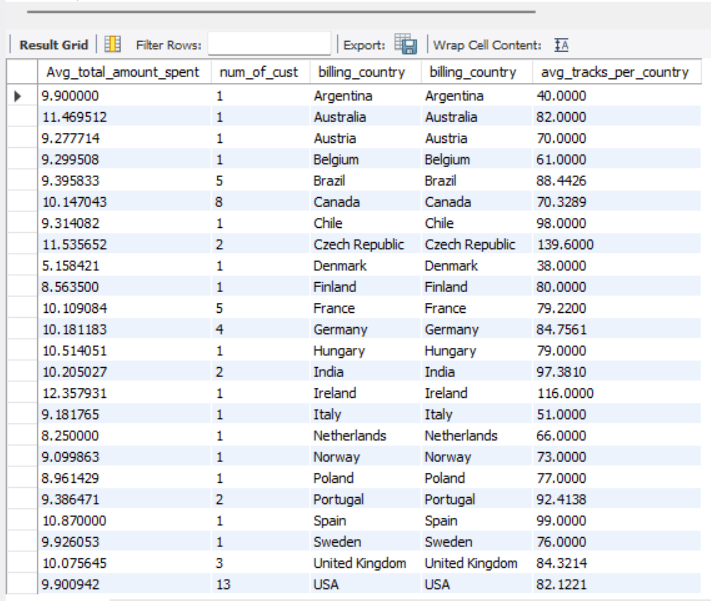
Alter table Albums add ReleaseYear int;



MySQL query:[sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)

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.

Ans:



**Insights:**

* The average total amount spent per customer seems to be in the range of 8 to 12 approximately with an outlier of 5.158 in Denmark.
* The number of customers are very few (single digits) in most of the countries.
* USA is the country with highest number of customers.

**Recommendations:**

* Focus on the USA: The US market has a good number of customers and therefore should be a top priority.
* Increase Penetration: Find ways to make your customer base bigger in other countries.
* Tailor Marketing: Looking at the types of genres, develop region specific marketing campaigns for each of them based on the genre preferences and spendings average.
* Address Low Spending: Check out the average spending of certain regions and countries and experiment with targeted promotions or product adjustment.
* Encourage Higher Spending: Offer customers with loyalty programs or personalised recommendations to encourage them to purchase more tracks.

MySQL query: [sql\_scripts.sql](https://drive.google.com/file/d/1LHc77KW40aXADIvpNC07hdaA-0iHpNIL/view?usp=sharing)