Objective Questions

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

**Approach:**

* To see if a column has null values, we need to run below query on every table.
* And to see if a column has duplicate value, we can run below query on every table.

**Query:**

SELECT \*from Track

WHERE track\_id is null or genre\_id is null or name is null or album\_id is null



For every table, the results would be similar, as there is no null values in the tables

**Query:**

SELECT Track\_id, count(Track\_id)

FROM Track

GROUP BY Track\_id

HAVING COUNT(Track\_id) > 1



**Conclusion:**

With the above query we can conclude that there are no null values and no duplicate values in the given tables.

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

**Approach**-

For Top selling Tracks, KPI’s to be considered are Track\_id and sum of quantity from Invoice\_line with the country filter as USA.

**Query and there reference image:**

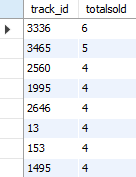
\*Top Selling Tracks (Below code gives tracks sold from most to the least)

select track\_id, sum(quantity) as totalsold from invoice\_line il

join invoice i on il.invoice\_id=i.invoice\_id

where billing\_country="USA"

group by track\_id



\*Top Artist in USA (Below code gives Artist’s name whose tracks were sold from most to the least)

**Approach:** For Top artist, KPI’s to be considered are name from artist table and sum of quantity from invoice\_line

select ar.name, sum(quantity) as totalsold from invoice\_line il

join track t on il.track\_id=t.track\_id

join album al on t.album\_id=al.album\_id

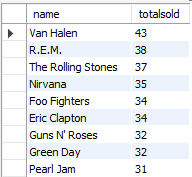
join artist ar on al.artist\_id=ar.artist\_id

join invoice i on il.invoice\_id=i.invoice\_id

where billing\_country="USA"

group by name

order by totalsold desc



\*Most famous Genre of the Artists (Below code gives Artist names and there most famous genre sold, from highest to lowest)

**Approach:** For most famous Genres of every artist, KPI’s to be considered are artist’s name, Genre ID and sold quantity

with cte as (select ar.name as ar\_name, g.name as genre\_name, sum(quantity) as totalsold, rank() over(partition by ar.name order by sum(quantity)) as rnk from invoice\_line il

join track t on il.track\_id=t.track\_id

join album al on t.album\_id=al.album\_id

join artist ar on al.artist\_id=ar.artist\_id

join genre g on t.genre\_id=g.genre\_id

group by ar.name, g.name)

select ar\_name, genre\_name from cte where rnk=1

order by totalsold desc



**Conclusion:**

From the above analysis we can conclude that most liked and purchased genre is Rock, and most purchased trackId is 3336, further most purchased artist is Van Helen.

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

**Approach:**

For this question, we need Customer’s percentage contribution per country in sales and location as well.

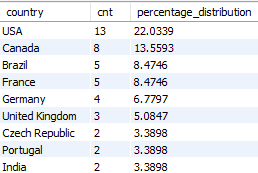
**Query (For Customer distribution in countries):**

with cte as (select country, count(customer\_id) as cnt from customer

group by country)

select\*, cnt\*100/(select sum(cnt) from cte) as percentage\_distribution from cte

order by cnt desc



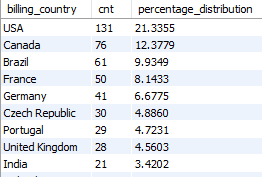
**Query (For Sales distribution of customers per country):**

with cte as (select billing\_country, count(customer\_id) as cnt from invoice

group by billing\_country)

select\*, cnt\*100/(select sum(cnt) from cte) as percentage\_distribution from cte

order by cnt desc



**Conclusion:**

From the above the query, we can see that most contribution of a country in its sales is USA and Canada.

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

**Approach:**

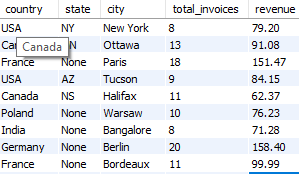
To calculate Total Revenue, and number of invoices with each country, state and city, we need address from invoice table, sum of Total and sum of Quantity sold.

**Query:**

\*Below code gives number of invoices and total revenue per country, state and city.

select billing\_country as country, billing\_state as state, billing\_city as city, count(invoice\_id) as total\_invoices, sum(total) as revenue from invoice

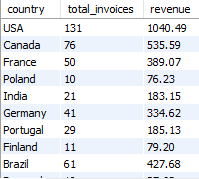
group by billing\_country, billing\_state, billing\_city



\*Below code gives number of invoices and total revenue per country

select billing\_country as country, count(invoice\_id) as total\_invoices, sum(total) as revenue from invoice

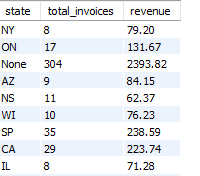
group by billing\_country



\*Below code gives number of invoices and total revenue per State

select billing\_state as state, count(invoice\_id) as total\_invoices, sum(total) as revenue from invoice

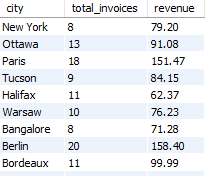
group by billing\_state



\*Below code gives number of invoices and total revenue per City

select billing\_city as city, count(invoice\_id) as total\_invoices, sum(total) as revenue from invoice

group by billing\_city



**Conclusion**:

With the results, we can conclude that USA has most contribution in the sales of Chinook’s whereas Berlin has most sales by any state.

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

**Approach:**

For this we need to take Customer\_id from Customers table and Group there count on the basis of Country.

**Query:**

\*Below query gives the name of customers with highest revenue in a country

with cte as (select customer\_id, billing\_country, sum(total) as revenue from invoice

group by customer\_id, billing\_country),

cte1 as (SELECT \*, rank() over(partition by billing\_country order by revenue desc) as rnk FROM cte)

select c.customer\_id,concat(first\_name," ",last\_name) as customer\_name, billing\_country as country, revenue from cte1

join customer c on cte1.customer\_id=c.customer\_id where rnk<=5



**Conclusion:**

Fernando Ramos, has the highest revenue contribution by any customer in Chinook’s platform.

1. Identify the top-selling track for each customer

**Approach:**

For this we need to take customer\_id and track\_id from invoice table in consideration.

**Query:**

\*Below query arranges the data with count of tracks purchased per customer per track

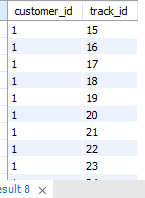
with cte as (select customer\_id, track\_id, count(track\_id) as cnt from invoice\_line il

join invoice i on il.invoice\_id=i.invoice\_id

group by customer\_id, track\_id),

cte1 as (select\*, rank() over(partition by customer\_id, track\_id order by cnt desc) as rnk from cte)

select customer\_id, track\_id from cte1 where rnk=1



**Conclusion:**

By above query we can get a customer’s most selling track.

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

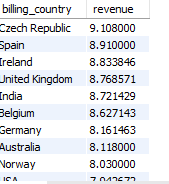
**Approach:** We need average order value for every country, and we also need average order value for every genre, with this we can evaluate highest purchasing country with value and most sold genre.

**Query:**

select billing\_country, avg(total) as revenue from invoice

group by billing\_country

order by revenue desc



**Query:**

select g.name, avg(total) as totals from track t

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

join genre g on t.genre\_id=g.genre\_id

group by g.name

order by totals desc



**Conclusion:**

With the above results we can deduce that Genre of “Hip Hop/Rap” category has highest revenue, whereas C.Republic and Spain contributes in higher average revenue.

1. What is the customer churn rate?

**Approach:**

* Customer churn rate refers to the no of unique customers who purchased next year.
* Here for Customer churn rate, we need to extract year from invoice\_date and count of unique Customer’s.

**Query:**

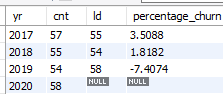
with cte as (select year(invoice\_date) as yr, count(distinct(customer\_id)) as cnt from invoice

group by yr),

cte1 as (select\*, lag(cnt) over(order by yr desc) as ld from cte

order by yr)

select \*, (cnt-ld)\*100/cnt as percentage\_churn from cte1



**Conclusion:**

We can see from the above result that there was a discernible level of customer churn in both 2017 and 2018, the rate of churn experienced a significant increase in 2019.

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

**Approach:**

To tackle this, we need Genre\_name from genre table and sum of total from Invoice table, later converting them in percentage.

**Query:**

\*Below query gives total sales contribution by each genre in USA

with cte as (select g.name, sum(total) as totals from track t

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

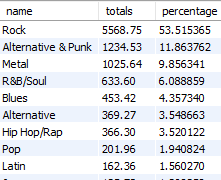
join genre g on t.genre\_id=g.genre\_id

where billing\_country="USA"

group by g.name)

select \*, totals\*100/(select sum(totals) from cte) as percentage from cte

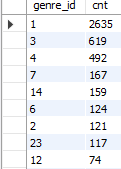
order by totals desc



**Approach:**

To take top bestselling Genres, we need Genre\_id from Tracks table and count of invoice\_id.

**Query:**



**Approach:**

To need to connect Top selling artists, we need the artist name from artist table and Invoice\_id from invoice table.

**Query:**

select ar.name, count(i.invoice\_id) as cnt from track t

join album a on t.album\_id=a.album\_id

join artist ar on a.artist\_id=ar.artist\_id

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

group by ar.name

order by cnt desc



**Conclusion**:

* Above query gives result that highest genre which gives highest revenue as well is Rock, sharing 57% of overall sales.
* Whereas top selling artist is Queen and J. Hendrix contributing above 180+ invoices.

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

**Approach:**

To have the answer for above question, we need to connect Customer\_id from invoice table and distinct count of Genre\_id from Genre Table.

\*Below query gives customer details who have purchased at least 3 different genres tracks

**Query:**

select customer\_id, count(distinct(genre\_id)) as cnt from track t

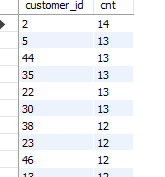
join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

group by customer\_id

having cnt>=3

order by cnt desc



**Conclusion:**

Highest buyer of different genres is the guy with customer\_id 2, who purchased 14 genres.

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

**Approach:** We need to connect Genre\_id from Genre table and quantity sold from invoice\_line table, and rank them accordingly.

**Query:**

\*Below query ranks genres based on their quantity sold

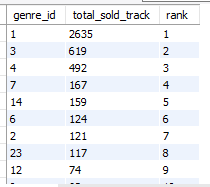
with cte as (select genre\_id, sum(quantity) as total\_sold\_track from track t

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

group by genre\_id)

select\*, rank() over(order by total\_sold\_track desc) as `rank` from cte



**Conclusion**:

Highest sold genre is Rock, having a genre\_id of 1, contributing more that 55% of overall sales.

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

**Approach:**

Here we need to take Customer\_id and there invoice\_date from invoice table and filter them with the 3 months interval.

**Query:**

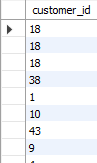
\*Below query gives customer\_id who didn’t purchased in last three months

with cte as (select distinct(customer\_id) from invoice

where invoice\_date>((select max(invoice\_date) from invoice)-interval 3 month))

select distinct(customer\_id) from invoice

where customer\_id not in(select customer\_id from cte)

****

**Conclusion:**

Above query can give the customer\_ids of customers who had not purchased in the last three months.

**Subjective Questions**

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

**Approach:**

Here we need to group data by Genre name and album name with the count of invoices sold, later adding the filter of country “USA**”**

**Query:**

with cte as (select g.name, a.title, count(i.invoice\_id) as cnt from track t

join album a on t.album\_id=a.album\_id

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

join genre g on t.genre\_id=g.genre\_id

where billing\_country="USA"

group by g.name, a.title

order by g.name, cnt)

select \*,rank() over(partition by name order by cnt desc) as rnk from cte

order by cnt desc



**Conclusion:**

* + With the above result we can conclude that the most sold genre in USA is “Rock” and “Alternatives”.
  + Narrowing down, which album would be the best to advertise, it should be from “Rock” Genre and based on the count of sold invoices, we can further narrow down on top 3 albums to promote or advertise.
  + In above case, the top albums to promote are:
    - Are You Experienced?
    - From The Muddy Banks Of The Wishkah [live]
    - The Doors

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

**Approach:**

For this question we need to take Genre\_id and countries, grouping them to count the number of invoices generated and later filtering them where country is not USA.

**Query:**

with cte as (select billing\_country, g.name, count(i.invoice\_id) as total\_sold\_track from track t

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

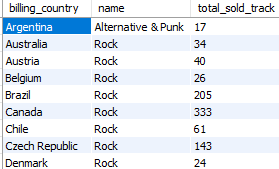
join Genre g on t.genre\_id=g.genre\_id

where billing\_country <> 'USA'

group by billing\_country, g.name),

cte1 as (select \*, rank() over(partition by billing\_country order by total\_sold\_track desc) as rnk from cte)

select\*from cte1 where rnk=1



**Conclusion:**

* The data reveals that Genre of Rock music, enjoys widespread popularity beyond USA’s border.
* In fact, the data shows that many other countries also have a substantial number of Rock music purchases.
* This suggests that Rock is a globally appreciated genre, resonating with audiences in various regions and not limited to a single market.

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?

**Approach:**

* For this question we need to take no of days a customer has been active, number of purchases customer did and the amount spent.
* Later grouping them together and finding an insight.

**Query:**

with cte as (select customer\_id, datediff(max(invoice\_date),min(invoice\_date)) as days\_active, count(invoice\_id) as cnt, sum(total) as total\_spent from invoice

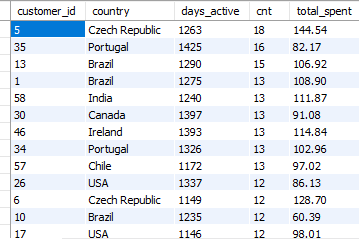
group by customer\_id

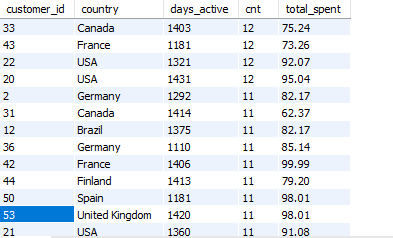
order by cnt, total\_spent)

select cte.customer\_id, country, days\_active, cnt, total\_spent from cte

join customer c on cte.customer\_id=c.customer\_Id

order by cnt desc





**Conclusion:**

* With the above Query, we can see that more the numbers days active a customer is, more is the quantity of sales that customer did and further increasing the total amount spent.
* Also, by seeing the above tables, we can conclude that by geography most of the sales came from USA, followed by Canada and Brazil.

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

**Approach:**

* This problem can be approached by dividing Customer\_id on Genre and Artist.
* Later ranking there invoice count and top ranking can be suggested to the customer to guide them in buying.

**Query:**

with cte as (select customer\_id, g.name, count(i.invoice\_id) as G\_pur\_cnt, rank() over(partition by customer\_id order by count(i.invoice\_id) desc) as rnk from track t

join invoice\_line il on t.track\_id=il.track\_id

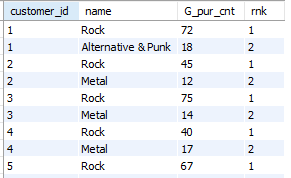
join invoice i on il.invoice\_id=i.invoice\_id

join Genre g on t.Genre\_id=g.Genre\_id

group by customer\_id, g.genre\_id

order by customer\_id, G\_pur\_cnt desc)

select\*from cte where rnk<3



with cte as (select customer\_id, ar.name, count(i.invoice\_id) as G\_pur\_cnt, rank() over(partition by customer\_id order by count(i.invoice\_id) desc) as rnk from track t

join invoice\_line il on t.track\_id=il.track\_id

join invoice i on il.invoice\_id=i.invoice\_id

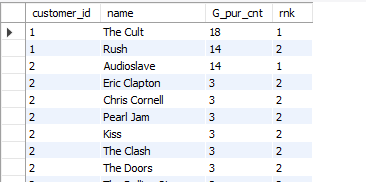
join album al on t.album\_id=al.album\_id

join artist ar on al.artist\_id=ar.artist\_id

group by customer\_id, ar.name

order by customer\_id, G\_pur\_cnt desc)

select\*from cte where rnk<3

****

**Conclusion:**

* The query above identifies the top two genres and artist’s name that a customer prefers based on their purchasing history.
* Using this information, we can make personalized recommendations to the customer, suggesting additional music or products that align with their preferred genres.
* This approach enhances their overall shopping experience and increasing the likelihood of making a purchase.

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?

**Approach:**

To figure out customer churn on geographical are, we need to group our data with two factors, i.e., year and billing\_country, and later counting unique customers with their following year will give us an insight for analysis.

**Query:**

with cte as (select year(invoice\_date) as yr, billing\_country, count(distinct(customer\_id)) as cnt from invoice

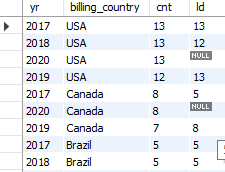
group by yr,billing\_country),

cte1 as (select\*, lag(cnt) over(partition by billing\_country order by yr desc) as ld from cte

order by yr)

select \*from cte1

order by cnt desc, billing\_country, yr



**Conclusion:**

* With the above query we can conclude that, churn is not specifically associated with high purchasing countries like USA, Canada and Brazil.
* Whereas with low purchasing countries they have declining customers like Australia, Argentina, Chile and other countries.

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?

**Approach:**

Here we need to categorize count of Invoice\_id and group them by year and country, which gives an insight which country has increasing churn rate and helps to focus there to improve in some way.

**Query:**

with cte as (select year(invoice\_date) as yr,billing\_country, count(distinct(customer\_id)) as cnt from invoice

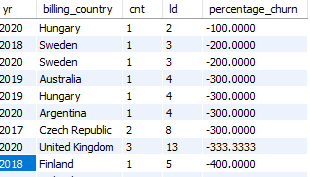
group by yr, billing\_country),

cte1 as (select\*, lag(cnt) over(order by yr desc) as ld from cte

order by yr)

select \*, (cnt-ld)\*100/cnt as percentage\_churn from cte1

order by percentage\_churn desc



**Conclusion:**

* With the above query we can conclude that UK and Canada is the country having stagnant increase in churn rate, which affects total sales and decreases overall revenue.
* With this we can further improve advertising and promotion in these countries to attract more customers and to increase customer retention.

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

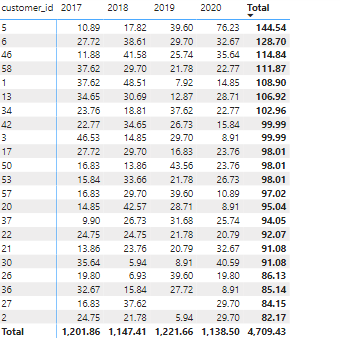
**Approach:**

This question can be analyzed using pivot table in powerbi, where we can group customer\_id’s, year purchased and total amount spent by them in particular years, later summing them to find overall a customer’s lifetime value.

**Query**

select customer\_id, year(invoice\_date) as yr, count(invoice\_id) as total\_purchase, sum(total) as total from invoice

group by customer\_id,year(invoice\_date)

****

**Conclusion:**

* With the above table we can conclude that, most of the customers have done there highest amount of purchase in the year 2018 and 2019.
* We can make them purchase again by reviewing further what are the most liked genres in both year and further expanding about the artists they have most purchased.

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?

**Approach:**

* To see the impact of customer acquisition, retention, and overall sales by advertising and promotion, we need Campaign Data, Customer Data and Sales Data.
* Later we can use this data to provide insights on how the campaign increased the customer base and further increase in sales.

**Conclusion:**

* The insights we can aim with the above data are about campaigns impact on customer retention, customer increment and sales overall sales.
* We can further add factors, if needed, about discounts, events or email marketing that helped in customer retention.

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

**Approach:**

If the questions were absent, first approach would have been by preparing the data, which includes handling duplicates and handling null values.

* After that, i would have taken geographical context to analyze data, that what geography is most active with music labels.
* Later, I would have analyzed the most active Genres, which are most sold and famous around the countries.
* Next factor to be considered is at what month people are more likely to make a sale, and which artist are famous among which country.
* Later, I would have taken Customer in consideration, which customer is more active on purchasing what genre or which artist it favours, and would try to make a sale to that customer.

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?

**Approach:**

To add column in a table, we use “Alter” with table\_name and add column with its datatype.

Its syntax is as follows:

ALTER TABLE table\_name

ADD column\_name datatype;

**Query:**

alter table Album

add Release\_year int



**Conclusion:**

* Above Query adds a column named “Release\_year” in Album table with the datatype Integer.
* Further we can add values using insert function and adding values by giving column name.

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.