-- Objective Questions --

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

Ans – **MYSQL QUERY**

select \* from album;

select distinct \* from album; -- no duplicate

select \* from artist;

select distinct \* from artist; -- no duplicate

select \* from customer;

select distinct \* from customer; -- no duplicate

select count(\*) from customer

where company is null; -- count of null in company 49

select count(\*) from customer

where state is null; -- count of null in state 29

select count(\*) from customer

where fax is null; -- count of null in fax 47

select \* from employee; -- there is 1 null value in report\_to for employee\_id = 1

select distinct \* from employee; -- no duplicate

select \* from genre;

select distinct \* from genre; -- no duplicate

select \* from invoice;

select distinct \* from invoice; -- no duplicate

select \* from invoice\_line;

select distinct \* from invoice\_line; -- no duplicate

select \* from media\_type;

select distinct \* from media\_type; -- no duplicate

select \* from playlist;

select distinct \* from playlist; -- no duplicate

select \* from playlist\_track;

select distinct \* from playlist\_track; -- no duplicate

select \* from track;

select distinct \* from track; -- no duplicate

select count(\*) from track

where composer is null; -- 978 null value in composer

There are no duplicate value in any table of whole dataset

some null/missing values following tables :

* 49 company, 29 state, 47 fax values are null in the customer

table.

* 1 reports\_to value is null in the employee table.
* 978 composer values are null in the track table.

I will use 'COALESCE' to handle null values

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

Ans – **MYSQL QUERY**

select Top\_selling\_tracks , Top\_artist , Most\_famous\_genres

from (

select t.name as Top\_selling\_tracks , ar.name as Top\_artist , g.name as Most\_famous\_genres ,

sum(t.unit\_price \* il.quantity) as total\_sales from track t

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

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

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

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

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

where billing\_country = "USA"

group by t.name , ar.name , g.name

order by total\_sales desc

limit 15

) Top\_Track\_Artist\_Genre;

-- For top selling genre

select Top\_Genre from (

select g.name as Top\_Genre

from track t

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

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

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

where i.billing\_country = 'USA'

group by g.name

order by sum(il.quantity) desc

limit 15 ) Most\_Famous\_genre;

Top selling tracks , top artist and top genre in USA



Top top-selling genre in USA :



**Insights**

* .Jim Hendrix is the artist in most of top tracks.
* "War Pigs" by Cake is the top track.
* Rock dominates US sales

**Recommendations**

* Prioritise Rock in inventory, marketing and promotions.
* Other than Rock, R&B should also be prioritised for genre-specific promotions.
* Jim Hendrix could be a great choice for brand collaborations.

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

Ans – **MYSQL QUERY**

select count(distinct country) from customer ; -- total country 24

select country , count(customer\_id) as customer\_count

from customer

group by country

order by customer\_count desc;

-- USA : 13 , Canada : 8 , Brazil : 5 , France : 5 , Germany : 4 , United Kingdom : 3 ,

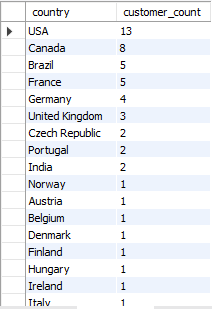
-- Czech Republic : 2 , Pourtgal : 2, India :2 , and from remaining all country 1 customer each

/\*

The customer base of Chinook music store is across 24 countries. According to our data USA is the country with most number of customers 13.

age and gender not available in given customer table columns to understand the customer breakdown.

\*/



**Insights**

* Across 24 countries we have customer spread.
* Most number of customers are present in the USA.

**Recommendations**

* Focus on the US market due to its large customer base.
* Make special focus on Canada, Brazil, France, and Germany to increase the count and make them reach a higher base size.

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

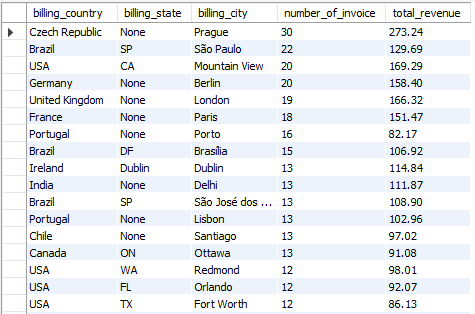
Ans – **MYSQL QUERY**

select billing\_country , billing\_state , billing\_city , count(\*) as number\_of\_invoice , sum(total) as total\_revenue

from invoice

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

order by number\_of\_invoice desc , total\_revenue desc;



**Insights**

* Prague (Czech Republic) generates the highest revenue .
* São Paulo (Brazil) has a high number of invoice but lower revenue compared to None.

**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.
* **Address Underperforming Regions :** find reasons behind low performance in cities like Copenhagen, Buenos Aires, and Edmonton.

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

Ans – **MYSQL QUERY**

with customer\_wise\_total\_revenue as (

select customer\_id , sum(total) as total\_revenue

from invoice

group by customer\_id ) ,

Customer\_ranking as (

select concat(c.first\_name , ' ' , c.last\_name) as Customer\_Name , c.country ,

dense\_rank() over (partition by c.country order by cr.total\_revenue desc) as ranking

from customer c

right join customer\_wise\_total\_revenue cr on cr.customer\_id = c.customer\_id)

select Customer\_name, country, ranking from Customer\_ranking where ranking <=5

order by country;

|  |  |  |
| --- | --- | --- |
| **Customer\_name** | **country** | **ranking** |
| Diego Gutiérrez | Argentina | 1 |
| Mark Taylor | Australia | 1 |
| Astrid Gruber | Austria | 1 |
| Daan Peeters | Belgium | 1 |
| Luís Gonçalves | Brazil | 1 |
| Fernanda Ramos | Brazil | 2 |
| Roberto Almeida | Brazil | 3 |
| Alexandre Rocha | Brazil | 4 |
| Eduardo Martins | Brazil | 5 |
| François Tremblay | Canada | 1 |
| Edward Francis | Canada | 2 |
| Ellie Sullivan | Canada | 3 |
| Aaron Mitchell | Canada | 4 |
| Jennifer Peterson | Canada | 5 |
| Luis Rojas | Chile | 1 |
| František Wichterlová | Czech Republic | 1 |
| Helena Holý | Czech Republic | 2 |
| Kara Nielsen | Denmark | 1 |
| Terhi Hämäläinen | Finland | 1 |
| Wyatt Girard | France | 1 |
| Camille Bernard | France | 2 |
| Isabelle Mercier | France | 3 |
| Dominique Lefebvre | France | 4 |
| Marc Dubois | France | 5 |
| Fynn Zimmermann | Germany | 1 |
| Hannah Schneider | Germany | 2 |
| Leonie Köhler | Germany | 3 |
| Niklas Schröder | Germany | 4 |
| Ladislav Kovács | Hungary | 1 |
| Manoj Pareek | India | 1 |
| Puja Srivastava | India | 2 |
| Hugh O'Reilly | Ireland | 1 |
| Lucas Mancini | Italy | 1 |
| Johannes Van der Berg | Netherlands | 1 |
| Bjørn Hansen | Norway | 1 |
| Stanisław Wójcik | Poland | 1 |
| João Fernandes | Portugal | 1 |
| Madalena Sampaio | Portugal | 2 |
| Enrique Muñoz | Spain | 1 |
| Joakim Johansson | Sweden | 1 |
| Phil Hughes | United Kingdom | 1 |
| Steve Murray | United Kingdom | 2 |
| Emma Jones | United Kingdom | 3 |
| Jack Smith | USA | 1 |
| Dan Miller | USA | 2 |
| Heather Leacock | USA | 3 |
| Kathy Chase | USA | 4 |
| Richard Cunningham | USA | 5 |

**Insights**

* Customers are number of consumers on same rank across countries.
* USA, Brazil, Canada and France got customers on all rankings below 5.

**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.

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

Ans – **MYSQL QUERY**

with Top\_selling\_Rank as (select c.customer\_id, concat(c.first\_name, ' ', c.last\_name) as customer\_name, t.name as track\_name, sum(il.quantity) as total\_sales,

row\_number() over (partition by c.customer\_id order by sum(il.quantity) desc) as ranking

from customer c

left join invoice i on i.customer\_id = c.customer\_id

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

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

group by c.customer\_id, concat(c.first\_name, ' ', c.last\_name), t.name)

select customer\_name, track\_name, total\_sales

from Top\_selling\_Rank

where ranking = 1;

Top selling track for each customer grouping by customer\_id , customer\_name and track\_name

|  |  |  |
| --- | --- | --- |
| customer\_name | track\_name | total\_sales |
| Luís Gonçalves | Message in a Bottle | 1 |
| Leonie Köhler | When Evening Falls | 1 |
| François Tremblay | Sting Me | 2 |
| Bjørn Hansen | Animal | 1 |
| František Wichterlová | Bad Boy Boogie | 1 |
| Helena Holý | Summer Love | 1 |
| Astrid Gruber | Be Mine | 1 |
| Daan Peeters | Blow Your Mind | 1 |
| Kara Nielsen | Hypnotize | 1 |
| Eduardo Martins | Like A Bird | 2 |
| Alexandre Rocha | Inside Job | 1 |
| Roberto Almeida | Love And Marriage | 2 |
| Fernanda Ramos | 24 Caprices, Op. 1, No. 24, for Solo Violin, in A Minor | 2 |
| Mark Philips | Immigrant Song | 1 |
| Jennifer Peterson | Something In The Way | 1 |
| Frank Harris | Posso Perder Minha Mulher, Minha Mãe, Desde Que Eu Tenha O Rock And Roll | 1 |
| Jack Smith | Just Friends | 1 |
| Michelle Brooks | Suite for Solo Cello No. 1 in G Major, BWV 1007: I. Prélude | 1 |
| Tim Goyer | Turandot, Act III, Nessun dorma! | 1 |
| Dan Miller | I Don't Wanna Be Kissed (By Anyone But You) (Alternate Take) | 1 |
| Kathy Chase | The Worst | 1 |
| Heather Leacock | Amy Amy Amy (Outro) | 1 |
| John Gordon | My Time After Awhile | 1 |
| Frank Ralston | Who Wants To Live Forever | 1 |
| Victor Stevens | Untitled | 2 |
| Richard Cunningham | Hitchin' A Ride | 1 |
| Patrick Gray | War Pigs | 2 |
| Julia Barnett | Get What You Need | 2 |
| Robert Brown | Virginia Moon | 1 |
| Edward Francis | A Room At The Heartbreak Hotel | 1 |
| Martha Silk | She's A Rebel | 1 |
| Aaron Mitchell | You Know My Name | 1 |
| Ellie Sullivan | Indifference | 1 |
| João Fernandes | Train In Vain | 2 |
| Madalena Sampaio | Get In The Ring | 1 |
| Hannah Schneider | I Can't Explain | 2 |
| Fynn Zimmermann | Radio/Video | 2 |
| Niklas Schröder | Take the Box | 1 |
| Camille Bernard | Midnight | 1 |
| Dominique Lefebvre | Tomorrow's Dream | 1 |
| Marc Dubois | My Melancholy Blues | 1 |
| Wyatt Girard | Changes | 2 |
| Isabelle Mercier | Tease Me Please Me | 2 |
| Terhi Hämäläinen | Compulsion | 1 |
| Ladislav Kovács | Smoke On The Water | 1 |
| Hugh O'Reilly | Drain You | 2 |
| Lucas Mancini | Put The Finger On You | 1 |
| Johannes Van der Berg | Confusion | 2 |
| Stanisław Wójcik | Faceless | 2 |
| Enrique Muñoz | Highway Chile | 1 |
| Joakim Johansson | Run To The Hills | 1 |
| Emma Jones | It Ain't Like That | 1 |
| Phil Hughes | Message in a Bottle (new classic rock mix) | 1 |
| Steve Murray | Baby | 1 |
| Mark Taylor | Kayleigh | 1 |
| Diego Gutiérrez | My Way | 1 |
| Luis Rojas | Odara | 1 |
| Manoj Pareek | Who Wants To Live Forever | 1 |
| Puja Srivastava | Question! | 1 |

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

Ans – **MYSQL QUERY**

select customer\_id , avg(total) as average\_total\_value , count(invoice\_id) as number\_of\_ordder

from invoice

group by customer\_id

order by count(invoice\_id) ;

select count(invoice\_id) as monthly\_invoice\_count , date\_format(invoice\_date , '%m-%Y') as Month\_Year ,

round(avg(total),2) as Monthly\_average\_total , sum(total) as Monthly\_sum\_total

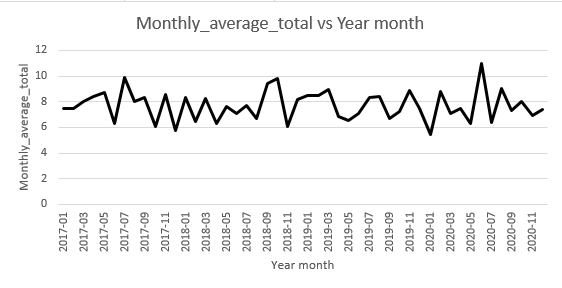
from invoice

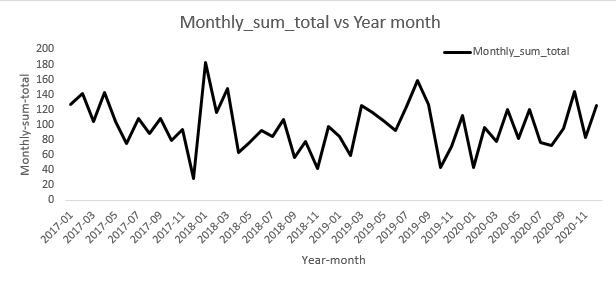
group by Month\_Year

order by Month\_Year;

|  |  |  |
| --- | --- | --- |
| customer\_id | average\_total\_value | number\_of\_ordder |
| 29 | 10.15 | 4 |
| 56 | 7.92 | 5 |
| 8 | 8.63 | 7 |
| 18 | 9.9 | 8 |
| 52 | 8.54 | 8 |
| 59 | 8.91 | 8 |
| 24 | 8.91 | 8 |
| 32 | 8.79 | 8 |
| 16 | 9.28 | 8 |
| 40 | 8.03 | 9 |
| 27 | 9.35 | 9 |
| 15 | 7.37 | 9 |
| 19 | 6.05 | 9 |
| 41 | 7.15 | 9 |
| 7 | 7.7 | 9 |
| 47 | 5.61 | 9 |
| 54 | 8.8 | 9 |
| 38 | 8.14 | 9 |
| 3 | 11.11 | 9 |
| 4 | 8.03 | 9 |
| 39 | 8.8 | 9 |
| 11 | 6.93 | 10 |
| 23 | 6.63 | 10 |
| 49 | 7.62 | 10 |
| 25 | 7.62 | 10 |
| 9 | 3.76 | 10 |
| 55 | 8.12 | 10 |
| 51 | 7.52 | 10 |
| 45 | 7.82 | 10 |
| 28 | 7.23 | 10 |
| 37 | 9.41 | 10 |
| 48 | 6.53 | 10 |
| 14 | 2.97 | 10 |
| 31 | 5.67 | 11 |
| 36 | 7.74 | 11 |
| 53 | 8.91 | 11 |
| 2 | 7.47 | 11 |
| 50 | 8.91 | 11 |
| 12 | 7.47 | 11 |
| 44 | 7.2 | 11 |
| 21 | 8.28 | 11 |
| 42 | 9.09 | 11 |
| 26 | 7.18 | 12 |
| 33 | 6.27 | 12 |
| 43 | 6.11 | 12 |
| 10 | 5.03 | 12 |
| 20 | 7.92 | 12 |
| 17 | 8.17 | 12 |
| 6 | 10.73 | 12 |
| 22 | 7.67 | 12 |
| 30 | 7.01 | 13 |
| 57 | 7.46 | 13 |
| 58 | 8.61 | 13 |
| 46 | 8.83 | 13 |
| 1 | 8.38 | 13 |
| 34 | 7.92 | 13 |
| 13 | 7.13 | 15 |
| 35 | 5.14 | 16 |
| 5 | 8.03 | 18 |

|  |  |  |  |
| --- | --- | --- | --- |
| monthly\_invoice\_count | Month\_Year | Monthly\_average\_total | Monthly\_sum\_total |
| 17 | 01-2017 | 7.45 | 126.72 |
| 22 | 01-2018 | 8.33 | 183.15 |
| 10 | 01-2019 | 8.51 | 85.14 |
| 8 | 01-2020 | 5.45 | 43.56 |
| 19 | 02-2017 | 7.45 | 141.57 |
| 18 | 02-2018 | 6.49 | 116.82 |
| 7 | 02-2019 | 8.49 | 59.4 |
| 11 | 02-2020 | 8.82 | 97.02 |
| 13 | 03-2017 | 8 | 103.95 |
| 18 | 03-2018 | 8.25 | 148.5 |
| 14 | 03-2019 | 8.98 | 125.73 |
| 11 | 03-2020 | 7.11 | 78.21 |
| 17 | 04-2017 | 8.39 | 142.56 |
| 10 | 04-2018 | 6.34 | 63.36 |
| 17 | 04-2019 | 6.87 | 116.82 |
| 16 | 04-2020 | 7.49 | 119.79 |
| 12 | 05-2017 | 8.75 | 104.94 |
| 10 | 05-2018 | 7.62 | 76.23 |
| 16 | 05-2019 | 6.56 | 104.94 |
| 13 | 05-2020 | 6.32 | 82.17 |
| 12 | 06-2017 | 6.27 | 75.24 |
| 13 | 06-2018 | 7.08 | 92.07 |
| 13 | 06-2019 | 7.08 | 92.07 |
| 11 | 06-2020 | 10.98 | 120.78 |
| 11 | 07-2017 | 9.9 | 108.9 |
| 11 | 07-2018 | 7.74 | 85.14 |
| 15 | 07-2019 | 8.32 | 124.74 |
| 12 | 07-2020 | 6.35 | 76.23 |
| 11 | 08-2017 | 8.01 | 88.11 |
| 16 | 08-2018 | 6.68 | 106.92 |
| 19 | 08-2019 | 8.39 | 159.39 |
| 8 | 08-2020 | 9.03 | 72.27 |
| 13 | 09-2017 | 8.3 | 107.91 |
| 6 | 09-2018 | 9.41 | 56.43 |
| 19 | 09-2019 | 6.67 | 126.72 |
| 13 | 09-2020 | 7.31 | 95.04 |
| 13 | 10-2017 | 6.09 | 79.2 |
| 8 | 10-2018 | 9.78 | 78.21 |
| 6 | 10-2019 | 7.26 | 43.56 |
| 18 | 10-2020 | 8.03 | 144.54 |
| 11 | 11-2017 | 8.55 | 94.05 |
| 7 | 11-2018 | 6.08 | 42.57 |
| 8 | 11-2019 | 8.91 | 71.28 |
| 12 | 11-2020 | 6.93 | 83.16 |
| 5 | 12-2017 | 5.74 | 28.71 |
| 12 | 12-2018 | 8.17 | 98.01 |
| 15 | 12-2019 | 7.46 | 111.87 |
| 17 | 12-2020 | 7.4 | 125.73 |





**Insights:**

* **Seasonal Peaks**: Reflecting the seasonal trend, a high level of activity has been recorded in January, April and August.
* **Customer Frequency**:. The reduction in counts in some months presents a problematic feature in the analysis due to variations in the levels of purchase being made.
* **Average Order Value (AOV**): Total revenue (AOV) is relatively flat (~6-10), with small fluctuations to the up suggesting large transactions made within certain months of the year, like November 2020.
* **Yearly Growth**: Evidently, aggregations of totals over years showing a gradual increase means that customer spending is on the rise, or business operations are becoming more efficient.

**Recommendations:**

* **Target Peak Months**: This should be done to perfectly coincide with January, April, and August so that the marketed products can generate high season sale.
* **Boost Low-Performing Months**: You should therefore provide discounts or a loyalty program to encourage sales during the slow season, which includes December, February and other such months.
* **Increase AOV**: By adjusting the approach used in selling, one can introduce upper selling or packaging of products to help increase the average order value.
* **Customer Retention**: Review people’s purchasing habits: Many customers who used the company’s services several times, and those who only once, and make useful adjustments.
* **Data-Driven Actions**: Constantly summarize invoice patterns and customer’s demands for further adjustments of stock and prices.

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

Ans – **MYSQL QUERY**

with number\_of\_customer\_in\_1st\_3months as (

select count(customer\_id) as customer\_1st\_3months

from invoice

where invoice\_date between '2017-01-01' and '2017-03-31') ,

-- I have taken the assumption that total number of customers in the beginning is equal to the customers joining in the first 3 months.

-- 49 customer in 1st 3 months

number\_of\_customer\_in\_last\_2months as (select count(customer\_id) as customer\_in\_last\_2months

from invoice

where invoice\_date between '2020-11-01' and '2020-12-31' )

-- I have taken the assumption that churn rate will be calculated on the basis of the number of customers left in the last two months.

-- 29 customer in last 2 months

select round((((select customer\_1st\_3months from number\_of\_customer\_in\_1st\_3months) - (select customer\_in\_last\_2months

from number\_of\_customer\_in\_last\_2months))/(select customer\_1st\_3months

from number\_of\_customer\_in\_1st\_3months)\* 100),2) as churn\_rate;

/\*

As per given data the customer churn rate of the company is 40.82% based on the total number of customer

in first 3 months is 49 and the number of customer present in the last 2 months is 29

So, number of customers lost = 49-29 = 20 , churn rate = (49-29)/49 \* 100

\*/

**Insights from Churn Rate Analysis:**

* **Churn Rate**: The calculated churn rate of 40.82% shows much customer attrition occurred between the beginning of the business period (2017 Q1) and the ending two months (2020 Nov-Dec).
* **Customer Base Decline**: The initial customer base was 49 and out of them 20 customers chucked the service, an indication towards low customer retention.
* **Recommendations:**
* **Retention Strategies:** Loyalty programs or coupons or customer satisfaction surveys to determine and minimize customer attrition.
* **Engagement Initiatives**: Use emails, promotions, and targeted campaigns, mostly in the later stages of the lifecycle with customers.
* **Focus on At-Risk Customers:** To provide an understanding on the type of customers who may be at high risk of leaving their service providers and provide early intervention by implementing retention strategies.
* **Improve Onboarding:** One has to bear in mind that it is important having a good on boarding to enhance customer satisfaction and likely patronage.

**Analyze Exit Reasons:** To obtain specific reasons and also to solve some matter, it conducted survey to churned customers.

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

Ans – **MYSQL QUERY**

with Total\_USA\_Revenue as ( select sum(total) as total\_rev

from invoice

where billing\_country = 'USA' ),

Genre\_wise\_total\_revenue as ( select g.name as genre\_name, sum(t.unit\_price \* il.quantity) as total\_genre\_revenue

from genre g

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

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

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

where billing\_country = 'USA'

group by g.name

order by total\_genre\_revenue desc ),

Genre\_Ranking as ( select genre\_name, round(total\_genre\_revenue \*100/(select total\_rev from Total\_USA\_Revenue),2) percentage\_contribution,

dense\_rank() over(order by round(total\_genre\_revenue \*100/(select total\_rev from Total\_USA\_Revenue),2) desc) Ranking

from Genre\_wise\_total\_revenue )

select genre\_name, percentage\_contribution, Ranking from Genre\_Ranking ;

-- Best selling genre

/\* genre\_name, percentage\_contribution, Ranking

Rock 53.38 1

Alternative & Punk 12.37 2

Metal 11.80 3

R&B/Soul 5.04 4

Blues 3.43 5

Alternative 3.33 6

Latin 2.09 7

Pop 2.09 7

Hip Hop/Rap 1.90 8

Jazz 1.33 9

Easy Listening 1.24 10

Reggae 0.57 11

Electronica/Dance 0.48 12

Classical 0.38 13

Heavy Metal 0.29 14

TV Shows 0.19 15

Soundtrack 0.19 15

\*/

**Best selling Artist wise**

with Total\_USA\_Revenue as ( select sum(total) as total\_rev

from invoice

where billing\_country = 'USA' ),

Artist\_wise\_total\_revenue as ( select a.name as Artist\_name, sum(t.unit\_price \* il.quantity) as total\_revenue\_Artist\_wise

from artist a

left join album al on a.artist\_id = al.artist\_id

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

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

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

where billing\_country = 'USA'

group by a.name

order by total\_revenue\_Artist\_wise desc ),

Artist\_Ranking as ( select Artist\_name, round(total\_revenue\_Artist\_wise \*100/(select total\_rev from Total\_USA\_Revenue),2) percentage\_contribution,

dense\_rank() over(order by round(total\_revenue\_Artist\_wise \*100/(select total\_rev from Total\_USA\_Revenue),2) desc) Ranking

from Artist\_wise\_total\_revenue )

select Artist\_name, percentage\_contribution, Ranking

from Artist\_Ranking

limit 10 ;

-- Top 10 selling Artist wise

/\* Artist\_name, percentage\_contribution, Ranking

Van Halen 4.09 1

R.E.M. 3.62 2

The Rolling Stones 3.52 3

Nirvana 3.33 4

Foo Fighters 3.24 5

Eric Clapton 3.24 5

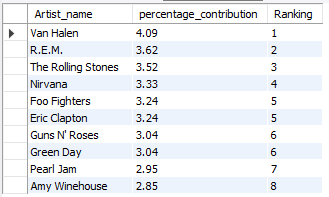
Guns N' Roses 3.04 6

Green Day 3.04 6

Pearl Jam 2.95 7

Amy Winehouse 2.85 8

\*/



|  |  |  |
| --- | --- | --- |
| genre\_name | percentage\_contribution | Ranking |
| Rock | 53.38 | 1 |
| Alternative & Punk | 12.37 | 2 |
| Metal | 11.8 | 3 |
| R&B/Soul | 5.04 | 4 |
| Blues | 3.43 | 5 |
| Alternative | 3.33 | 6 |
| Latin | 2.09 | 7 |
| Pop | 2.09 | 7 |
| Hip Hop/Rap | 1.9 | 8 |
| Jazz | 1.33 | 9 |
| Easy Listening | 1.24 | 10 |
| Reggae | 0.57 | 11 |
| Electronica/Dance | 0.48 | 12 |
| Classical | 0.38 | 13 |
| Heavy Metal | 0.29 | 14 |
| TV Shows | 0.19 | 15 |
| Soundtrack | 0.19 | 15 |

**Insights:**

* **Top Genres**: The genre most represented is R&B/Soul/Blues, followed by R&B/Soul/Blues, Blues, and Alternative, which respectively represent 5.04%, .
* **Underperforming Genres:** contributes least including Heavy Metal (0.29%), Classical (0.38%), and TV Shows/Soundtracks (0.19).
* **Tie in Contribution:** Pop and Latin share the same percentage (2.09%} which means that both genres are equally popular.
* **Niche Preferences:** Slightly in the middle of the list with Low, Medium and Low to Moderate contributions are the Genres Jazz, Easy Listening, Reggae.

**Recommendations:**

* **Focus on Popular Genres:** Among the genres, concentrating efforts in marketing and content development, including playlists, focusing on R&B/Soul, Blues, and Alternative should generate the highest result.
* **Promote Emerging Genres:** Create ads for increasing recognition for such weak performers such as Classical and Heavy Metal, although they might have a small market share.
* **Leverage Tied Genres:** Market both Latin and Pop genres together because both types of music are an equal market.
* **Expand Niche Appeal:** Promote Jazz and Easy Listening of Music through the showcasing of more playlists or special collaborations to excite their large following.

**Data-Driven Investments:** Work on analyzing trends in order to bring focus on how much to invest into which genres in order to capitalize on their potential.

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

Ans – **MYSQL QUERY**

select Customer\_Name , Total

from ( select concat(first\_name, ' ', last\_name) as Customer\_Name , count(distinct g.name) as Total

from customer c

left join invoice i on i.customer\_id = c.customer\_id

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

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

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

group by 1

having count(distinct g.name) >= 3

order by count(distinct g.name) desc

) as Customer\_Total\_Purchase ;

|  |  |
| --- | --- |
| Customer\_Name | Total |
| Leonie Köhler | 14 |
| Terhi Hämäläinen | 13 |
| Edward Francis | 13 |
| Madalena Sampaio | 13 |
| František Wichterlová | 13 |
| Heather Leacock | 13 |
| Fernanda Ramos | 12 |
| John Gordon | 12 |
| Jack Smith | 12 |
| Niklas Schröder | 12 |
| Michelle Brooks | 12 |
| Wyatt Girard | 12 |
| Marc Dubois | 12 |
| Julia Barnett | 12 |
| Hugh O'Reilly | 12 |
| Steve Murray | 11 |
| Camille Bernard | 11 |
| Johannes Van der Berg | 11 |
| Ellie Sullivan | 11 |
| Mark Taylor | 11 |
| Enrique Muñoz | 11 |
| Luis Rojas | 11 |
| Ladislav Kovács | 11 |
| Helena Holý | 11 |
| Tim Goyer | 10 |
| Daan Peeters | 10 |
| Hannah Schneider | 10 |
| Isabelle Mercier | 10 |
| João Fernandes | 10 |
| Diego Gutiérrez | 10 |
| Kathy Chase | 10 |
| Lucas Mancini | 10 |
| Luís Gonçalves | 10 |
| Manoj Pareek | 10 |
| Richard Cunningham | 10 |
| Joakim Johansson | 9 |
| Aaron Mitchell | 9 |
| Mark Philips | 9 |
| Eduardo Martins | 9 |
| Dan Miller | 9 |
| Patrick Gray | 9 |
| Stanisław Wójcik | 9 |
| Bjørn Hansen | 9 |
| Alexandre Rocha | 8 |
| Frank Harris | 8 |
| François Tremblay | 8 |
| Dominique Lefebvre | 8 |
| Phil Hughes | 8 |
| Puja Srivastava | 8 |
| Martha Silk | 8 |
| Roberto Almeida | 8 |
| Astrid Gruber | 8 |
| Frank Ralston | 8 |
| Emma Jones | 7 |
| Jennifer Peterson | 6 |
| Kara Nielsen | 6 |
| Fynn Zimmermann | 6 |
| Victor Stevens | 6 |
| Robert Brown | 5 |

**Insights:**

* **Top Customers**: At the moment Terhi Hämäläinen and others comprise the list of thirteen transactions, while the leader is **Leonie Köhler** with **fourteen** transactions.
* **Low Engagement Customers**: For example, customers such as Robert Brown, Victor Stevens, Kara Nielsen have low levels of interaction: with 6 and fewer transactions.
* **Customer Distribution**: The values drop significantly after the initial ~15 customers, which means that there is a potential for other retention strategies.

**Recommendations**:

* **Reward Top Customers**: Provide incentives or special operational rate reductions for significant purchasers (e.g., Leonie Köhler, Terhi Hämäläinen) to continue utilising the service.
* **Re-engage Low Spenders**: Those customers who have made 6 or less transactions (such as Robert Brown) should be targeted with offer or campaign that will make them’re more active.
* **Analyze Patterns**: From an analysis of the customer’s buying patterns, determine the parameters that make the largest customers loyal and apply it to the rest of the customers.
* **Broaden Reach**: Concentrate to fasten the frequency of transactions with infrequent or rarely transacting customers while specifying and overcoming barriers such as price, product portfolio or contact issues.

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

Ans – **MYSQL QUERY**

with Sales\_Genre as (

select t.genre\_id , g.name , sum(t.unit\_price \* il.quantity) sales\_performance

from track t

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

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

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

where billing\_country = 'USA'

group by 1, 2 )

select name , sales\_performance ,

dense\_rank() over (order by sales\_performance desc) `Rank`

from Sales\_Genre;

|  |  |  |
| --- | --- | --- |
| name | sales\_performance | Rank |
| Rock | 555.39 | 1 |
| Alternative & Punk | 128.7 | 2 |
| Metal | 122.76 | 3 |
| R&B/Soul | 52.47 | 4 |
| Blues | 35.64 | 5 |
| Alternative | 34.65 | 6 |
| Latin | 21.78 | 7 |
| Pop | 21.78 | 7 |
| Hip Hop/Rap | 19.8 | 8 |
| Jazz | 13.86 | 9 |
| Easy Listening | 12.87 | 10 |
| Reggae | 5.94 | 11 |
| Electronica/Dance | 4.95 | 12 |
| Classical | 3.96 | 13 |
| Heavy Metal | 2.97 | 14 |
| TV Shows | 1.99 | 15 |
| Soundtrack | 1.98 | 16 |

**Insights:**

* Rock sells the most with 555.39 in sales while the other categories are far behind this figure.
* Alternative & Punk and Metal claim 2nd and 3rd positions with fairly decent sales of 128.7 and 122.76 however considerably lower that Rock.
* Sometimes a movie falls under certain genre and makes only a little contribution; this is what we see with Soundtrack (1.98) and TV Shows (1.99).
* Pop has same sale as Latin (21.78), hence, groping the same level of audience interest.
* These are Classical, Heavy Metal, and Electronica/Dance music progammes which registered low audience involvement even though they target a select audience.

**Recommendations:**

* Focus on Rock and spend more time on the particular content, Rock merchandise, and specially selected playlists to solidify its position.
* Promote More of Alternative & Punk and Metal with well designed marketing strategies to reach even more listeners and thus expanding their market.
* It is essential to assess the strength of weak-b Lyric and Under perform Evaluation of Soundtrack and TV Shows one must use niche marketing or bundle them with other successful genres.
* Marketing Latin and Pop music together is effective due to shared services customers and the chance to attract more consumers.
* Offer more options in targeted areas (like Classical, Jazz) providing special club events or encased material that will interest target audience most.
* Continuously review and assess the patterns of sales performance based on obtained data and several adjustments for increasing the income for each genre.

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

Ans – **MYSQL QUERY**

select c.customer\_id , concat(c.first\_name , ' ' , c.last\_name) as customer\_name

from customer c

where c.customer\_id not in ( select distinct i.customer\_id

from invoice i

where i.invoice\_date >= date\_sub('2020-12-31', interval 3 month ) );

|  |  |
| --- | --- |
| customer\_id | customer\_name |
| 1 | LuÃ­s GonÃ§alves |
| 3 | FranÃ§ois Tremblay |
| 4 | BjÃ¸rn Hansen |
| 7 | Astrid Gruber |
| 8 | Daan Peeters |
| 9 | Kara Nielsen |
| 10 | Eduardo Martins |
| 11 | Alexandre Rocha |
| 17 | Jack Smith |
| 18 | Michelle Brooks |
| 19 | Tim Goyer |
| 36 | Hannah Schneider |
| 37 | Fynn Zimmermann |
| 38 | Niklas SchrÃ¶der |
| 39 | Camille Bernard |
| 43 | Isabelle Mercier |
| 48 | Johannes Van der Berg |
| 50 | Enrique MuÃ±oz |
| 54 | Steve Murray |
| 56 | Diego GutiÃ©rrez |
| 57 | Luis Rojas |
| 58 | Manoj Pareek |

**Insights:**

* Customers can be grouped according to regions or even culture for product advertising.
* This should be complemented with the recent activity data to discover the customers to re-target those that have not made any purchase recently.

**Recommendations:**

* Use names to make customers feel appreciated; this way, there is improved customer loyalty.
* Make regional or cultural specific promotions.
* Any customer who has been inactive in the recent past should be considered for a ultimatum offer or a campaign.
* Identify those customers who are valuable to the company (e.g., Leonie Köhler, Luís Gonçalves) and then try to replicate their customer experience.

-- **Subjective Questions** –

**Q.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 – **MYSQL QUERY**

select \* from track

order by album\_id , genre\_id ;

select genre\_id , name from genre ;

with Sales as ( select g.name as genre\_name, sum(il.quantity \* il.unit\_price) as total\_sales

from invoice i

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

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

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

where i.billing\_country = 'usa'

group by g.name

order by total\_sales desc ),

Top\_Albums as ( select al.title as album\_title, ar.name as artist\_name, g.name as genre\_name, sum(il.quantity \* il.unit\_price) as album\_sales

from invoice i

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

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

where i.billing\_country = 'USA'

group by al.title, ar.name, g.name

order by album\_sales desc )

select album\_title, artist\_name, genre\_name, album\_sales

from Top\_Albums

limit 3;

/\*

Following albums should be prioritised for advertisement and promotion:

album\_title artist\_name genre\_name

From The Muddy Banks Of The Wishkah [live] Nirvana Rock

Are You Experienced? Jimi Hendrix Rock

The Doors The Doors Rock

\*/



### **Approach**:

* **Data Collection**: Gather sales data from invoices, invoice lines, tracks, albums, artists, and genres. This data will be used to analyze sales performance.
* **Sales Analysis by Genre**:
  + Create a Common Table Expression (CTE) named Sales to calculate total sales for each genre in the USA. This will help identify which genres are performing well in terms of sales.
* **Top Albums Identification**:
  + Create another CTE named Top\_Albums to calculate total sales for each album, along with the associated artist and genre. This will allow us to see which albums are the top sellers in the USA.
* **Final Selection**:
  + Select the top three albums based on total sales from the Top\_Albums CTE. This will provide the albums that should be prioritized for advertising and promotion.

**Insights:**

* From the Live album, From The Muddy Banks Of The Wishkah [live] is the number one Rock album in the USA.
* Are You Experienced? and The Best Of Van Halen, Vol. Next are which I closely follow.

**Recommendations**:

* Considering this a marketing manager should ensure that there are more promotion for the three most popular albums in order to increase sales even further.
* Utilise the appeal of nostalgia and that these albums will create a bundle selling point.

**Q.2) Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.**

Ans – **MYSQL QUERY**

select g.genre\_id , g.name , sum(t.unit\_price \* il.quantity) as Total\_Revenue\_Each\_Genre

from track t

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

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

left join invoice i on i.invoice\_id = t.track\_id

where billing\_country != 'USA'

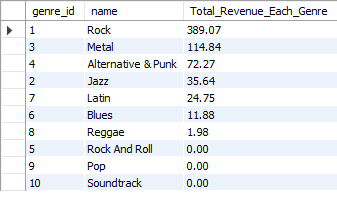
group by 1,2

order by Total\_Revenue\_Each\_Genre desc ;

/\* Rock genre is on Top ,

Metal is 2nd & Alternative and Punk on 3rd.

\*/



### **Approach** :

* **Data Collection**: Gather sales data from tracks, genres, invoice lines, and invoices. This data will be used to analyze genre performance in countries excluding the USA.
* **Sales Analysis by Genre**:
* Use a SQL query to calculate the total revenue for each genre by summing the product of the unit price and quantity sold.
  + Filter out records where the billing country is the USA.
* **Grouping and Ordering**:
  + Group the results by genre to aggregate the total revenue for each genre.
  + Order the results by total revenue in descending order to identify the top-selling genres.

### **Insights:**

* It is Interesting to see Rock genre is commonly on Top.
* METAL is 2nd & ALTERNATIVE and Punk on 3rd.

### **Recommendations:**

* Focus global campaigns on Rock, Metal, and Alternative & Punk to maximize sales.
* Target regional markets with localized promotions for Latin, Jazz, and other mid-tier genres.

**Q.3) 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 – **MYSQL QUERY**

with Purchase\_freq as ( select i.customer\_id , max(invoice\_date) , min(invoice\_date) ,

abs(timestampdiff(month , max(invoice\_date) , min(invoice\_date))) as each\_cust\_time ,

sum(total) as sales , sum(quantity) as item\_count , count(invoice\_date) as frequency

from invoice i

left join customer c on c.customer\_id = i.customer\_id

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

group by i.customer\_id

order by each\_cust\_time desc ) ,

average\_time as ( select round(avg(each\_cust\_time),2) as average

from Purchase\_freq ) , -- 40.36 Months

Category\_Define as (select \* , case

when each\_cust\_time > ( select average from average\_time)

then "Long-term Customer" else "Short-term Customer"

end as category

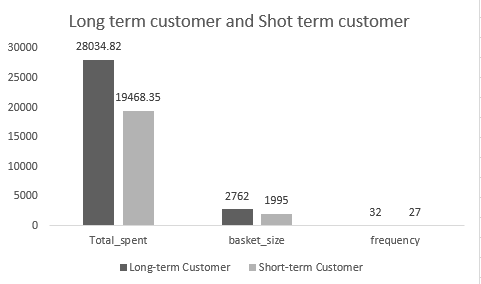
from Purchase\_freq )

select category , sum(sales) as Total\_spent , sum(item\_count) as basket\_size , count(frequency) as frequency

from Category\_Define

group by category ;





### **Approach:**

* **Data Collection**: Gather data from invoices, customers, and invoice lines to analyze customer purchasing behavior.
* **Purchase Frequency Calculation**:
  + Create a Common Table Expression (CTE) named Purchase\_freq to calculate metrics for each customer, including the maximum and minimum invoice dates, the time span between these dates, total sales, item count, and purchase frequency.
* **Average Time Calculation**:
  + Create another CTE named average\_time to calculate the average time span of customer purchases across all customers.
* **Customer Categorization**:
  + Create a CTE named Category\_Define to categorize customers as "Long-term" or "Short-term" based on whether their time span exceeds the average time calculated in the previous step.
* **Final Aggregation**:
  + Select and aggregate the total spending, basket size, and purchase frequency for each customer category.

**Insights:**

Long-term customers tend to spend more, have larger basket sizes, and shop more frequently compared to short-term customers.

**Recommendations:**

The company should focus on offering a wider variety of genres tailored to the preferences of long-term customers.

These customers contribute significantly more revenue, as loyalty plays a crucial role in driving sales.

Building and nurturing relationships with long-term customers is key, as they tend to make more purchases

over time compared to short-term customers.

**Q.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?**

Ans – **MYSQL QUERY**

-- ( select \* from invoice\_line ; )

-- Purchase by customer over an invoice but different genre

select il.invoice\_id , g.name

from invoice\_line il

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

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

group by 1 , 2 ;

-- Different Artists prefered in single invoice

select il.invoice\_id, ar.name

from invoice\_line il

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

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

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

group by 1 , 2 ;

-- Different Albums purchased over an invoice

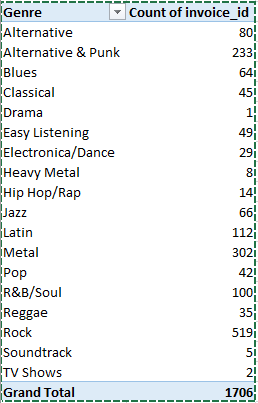
select il.invoice\_id, al.title

from invoice\_line il

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

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

group by 1 , 2 ;



**Genre Preferences:**

* Top Genres: Rock, Metal, and Alternative & Punk are the most popular among users.
* Cross-Selling Opportunity: If a user buys one, sales staff can recommend trying the other two.

**Top-Selling Artists:**

* Key Performers: Green Day, U2, Foo Fighters, Nirvana, The Rolling Stones, Queens, and System of a Down dominate sales.
* Recommendation Strategy: Use this data to suggest tracks or albums from these artists to users.

**Top Albums:**

* Top Performers: *Mesmerize*, *Are You Experienced?*, and *The Doors* are the leading albums.
* Sales Opportunity: Highlight these albums in promotions or bundles to drive sales.

**Q.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?**

Ans – **MYSQL QUERY**

with first\_six\_months as (select billing\_country, count(customer\_id) count\_of\_cust from invoice

where invoice\_date between '2017-01-01' and '2017-06-30'

group by billing\_country ),

last\_six\_months as ( select billing\_country, count(customer\_id) count\_of\_cust from invoice

where invoice\_date between '2020-07-01' and '2020-12-31'

group by billing\_country )

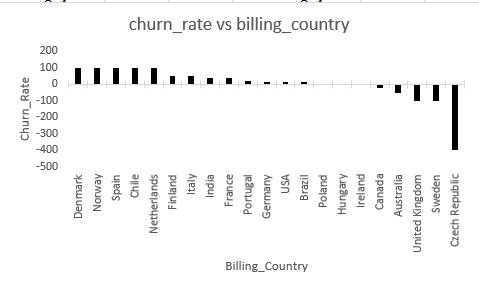
select f6.billing\_country, round((f6.count\_of\_cust - coalesce(l6.count\_of\_cust,0))/f6.count\_of\_cust \* 100 , 2) churn\_rate

from first\_six\_months f6

left join last\_six\_months l6 on f6.billing\_country = l6.billing\_country

order by churn\_rate desc ;

|  |  |
| --- | --- |
| billing\_country | churn\_rate |
| Denmark | 100 |
| Norway | 100 |
| Spain | 100 |
| Chile | 100 |
| Netherlands | 100 |
| Finland | 50 |
| Italy | 50 |
| India | 40 |
| France | 37.5 |
| Portugal | 20 |
| Germany | 16.67 |
| USA | 15 |
| Brazil | 14.29 |
| Poland | 0 |
| Hungary | 0 |
| Ireland | 0 |
| Canada | -21.43 |
| Australia | -50 |
| United Kingdom | -100 |
| Sweden | -100 |
| Czech Republic | -400 |



### **Approach:**

* **Data Collection**: Gather data from invoices to analyze customer counts in different billing countries over specified time periods.
* **Customer Count for the First Six Months**:
  + Create a Common Table Expression (CTE) named first\_six\_months to count the number of unique customers in each billing country for the first half of 2017.
* **Customer Count for the Last Six Months**:
  + Create another CTE named last\_six\_months to count the number of unique customers in each billing country for the last half of 2020.
* **Churn Rate Calculation**:
  + Calculate the churn rate for each country by comparing the customer counts from the two periods. The churn rate is defined as the percentage change in the number of customers from the first period to the second.
* **Final Selection**:
  + Select the billing country and the calculated churn rate, ordering the results by churn rate in descending order.

**Insights:**

* Larger churn in Finland at 50%, India at 40% and France 37.5%; zero churn means total attrition in Spain, Denmark, and Norway.
* Only two of them have reported growth: the Canadian (-21.43%) and the Australian (-50%); this contextual evidence points to successful retention or acquisition.
* Stable churn in Poland Hungary and Ireland with zero churn which demonstrates customer loyalty.

**Recommendations:**

* Some of the high churn regions have been discovered to be Finland, India, and France and hence the need to develop loyalty programs and localized campaigns to regain customers’ loyalty.
* Build on growth in Canada and Australia by improving the ways that people engage with Priority’s marketing and service offerings.
* Sustain a level of penetration in Poland, Hungary and Ireland through continuous customer tempo.

**Q.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?**

Ans – **MYSQL QUERY**

First trying to find customer Days\_Since\_Last\_Purchase

select customer\_id , datediff('2020-12-31' , max(invoice\_date)) as Days\_Since\_Last\_Purchase

from invoice

group by customer\_id;

Assume a customer is at risk if they haven't purchased in the last 90 days.

with Risk\_Status as ( select customer\_id ,

case

when datediff('2020-12-31' , max(invoice\_date)) > 90 then 'At Risk'

else 'Active'

end as Customer\_Risk\_Status

from invoice

group by customer\_id )

select Customer\_Risk\_Status , count(\*) as Total\_Count

from Risk\_Status

group by Customer\_Risk\_Status;

Calculate Total Spending and Average Spending per Customer

select customer\_id , count(invoice\_id) as Total\_Invoice ,

sum(total) as Total\_Spent , round(avg(total),2) as Avg\_Spent\_Per\_invoice

from invoice

group by customer\_id ;

Compare spending across year by year

select customer\_id , year(invoice\_date) as Year , sum(total) as Yearly\_Spending

from invoice

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

order by customer\_id, Year;

Analyze Spending by Location (City , State and Country)

select billing\_city , billing\_state , billing\_country ,

count(distinct customer\_id) as Customer\_Count , sum(total) as Total\_Spent , round(avg(total),2) as Avg\_Spent

from invoice

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

order by total\_spent desc ;

**Factors Contributing to Risk**

High Recency (Days Since Last Purchase): Customers who haven’t purchased in a long time.

Low Frequency (Total Invoices): Customers with infrequent transactions.

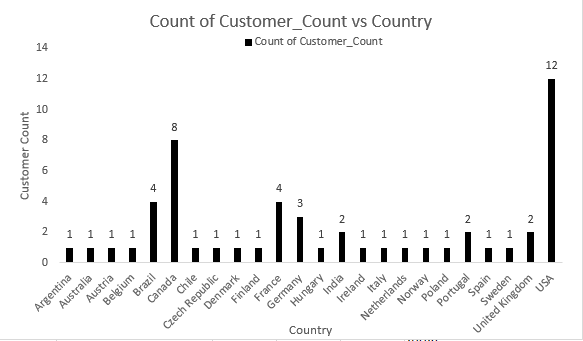
Declining Spending Trends: Customers whose spending decreases over time.

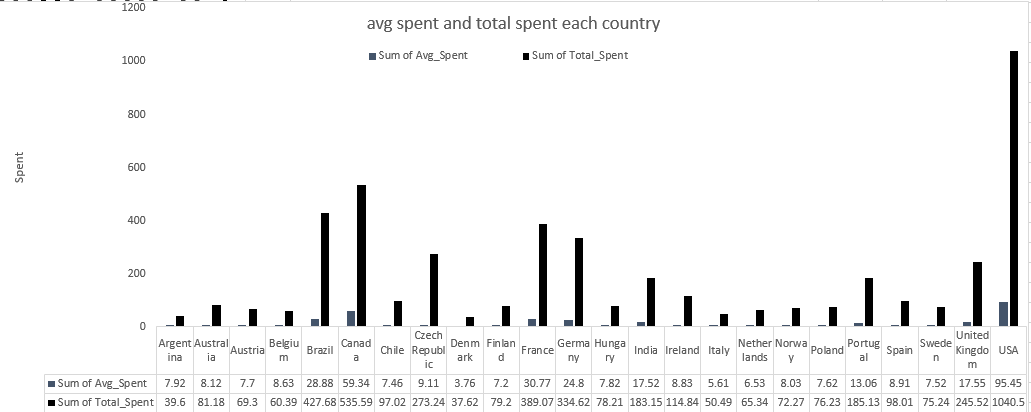
Demographic Location: Customers in regions with lower average spending or higher churn.



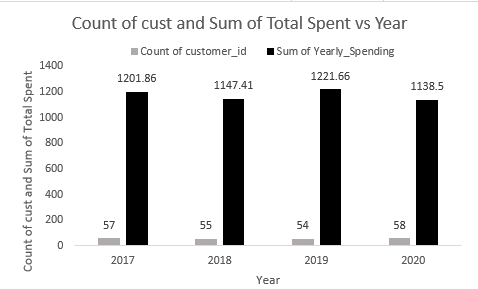
|  |  |  |  |
| --- | --- | --- | --- |
| customer\_id | Total\_Invoice | Total\_Spent | Avg\_Spent\_Per\_invoice |
| 18 | 8 | 79.2 | 9.9 |
| 30 | 13 | 91.08 | 7.01 |
| 40 | 9 | 72.27 | 8.03 |
| 27 | 9 | 84.15 | 9.35 |
| 31 | 11 | 62.37 | 5.67 |
| 49 | 10 | 76.23 | 7.62 |
| 59 | 8 | 71.28 | 8.91 |
| 38 | 9 | 73.26 | 8.14 |
| 42 | 11 | 99.99 | 9.09 |
| 35 | 16 | 82.17 | 5.14 |
| 25 | 10 | 76.23 | 7.62 |
| 44 | 11 | 79.2 | 7.2 |
| 1 | 13 | 108.9 | 8.38 |
| 20 | 12 | 95.04 | 7.92 |
| 24 | 8 | 71.28 | 8.91 |
| 10 | 12 | 60.39 | 5.03 |
| 43 | 12 | 73.26 | 6.11 |
| 9 | 10 | 37.62 | 3.76 |
| 53 | 11 | 98.01 | 8.91 |
| 4 | 9 | 72.27 | 8.03 |
| 39 | 9 | 79.2 | 8.8 |
| 55 | 10 | 81.18 | 8.12 |
| 12 | 11 | 82.17 | 7.47 |
| 45 | 10 | 78.21 | 7.82 |
| 46 | 13 | 114.84 | 8.83 |
| 3 | 9 | 99.99 | 11.11 |
| 58 | 13 | 111.87 | 8.61 |
| 33 | 12 | 75.24 | 6.27 |
| 34 | 13 | 102.96 | 7.92 |
| 51 | 10 | 75.24 | 7.52 |
| 37 | 10 | 94.05 | 9.41 |
| 50 | 11 | 98.01 | 8.91 |
| 19 | 9 | 54.45 | 6.05 |
| 22 | 12 | 92.07 | 7.67 |
| 21 | 11 | 91.08 | 8.28 |
| 36 | 11 | 85.14 | 7.74 |
| 57 | 13 | 97.02 | 7.46 |
| 26 | 12 | 86.13 | 7.18 |
| 15 | 9 | 66.33 | 7.37 |
| 52 | 8 | 68.31 | 8.54 |
| 47 | 9 | 50.49 | 5.61 |
| 16 | 8 | 74.25 | 9.28 |
| 2 | 11 | 82.17 | 7.47 |
| 13 | 15 | 106.92 | 7.13 |
| 5 | 18 | 144.54 | 8.03 |
| 48 | 10 | 65.34 | 6.53 |
| 32 | 8 | 70.29 | 8.79 |
| 8 | 7 | 60.39 | 8.63 |
| 17 | 12 | 98.01 | 8.17 |
| 54 | 9 | 79.2 | 8.8 |
| 6 | 12 | 128.7 | 10.73 |
| 28 | 10 | 72.27 | 7.23 |
| 29 | 4 | 40.59 | 10.15 |
| 7 | 9 | 69.3 | 7.7 |
| 14 | 10 | 29.7 | 2.97 |
| 11 | 10 | 69.3 | 6.93 |
| 23 | 10 | 66.33 | 6.63 |
| 41 | 9 | 64.35 | 7.15 |
| 56 | 5 | 39.6 | 7.92 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Billing\_county** | **Count of Customer\_Count** | **Sum of Avg\_Spent** | **Sum of Total\_Spent** |
| Argentina | 1 | 7.92 | 39.6 |
| Australia | 1 | 8.12 | 81.18 |
| Austria | 1 | 7.7 | 69.3 |
| Belgium | 1 | 8.63 | 60.39 |
| Brazil | 4 | 28.88 | 427.68 |
| Canada | 8 | 59.34 | 535.59 |
| Chile | 1 | 7.46 | 97.02 |
| Czech Republic | 1 | 9.11 | 273.24 |
| Denmark | 1 | 3.76 | 37.62 |
| Finland | 1 | 7.2 | 79.2 |
| France | 4 | 30.77 | 389.07 |
| Germany | 3 | 24.8 | 334.62 |
| Hungary | 1 | 7.82 | 78.21 |
| India | 2 | 17.52 | 183.15 |
| Ireland | 1 | 8.83 | 114.84 |
| Italy | 1 | 5.61 | 50.49 |
| Netherlands | 1 | 6.53 | 65.34 |
| Norway | 1 | 8.03 | 72.27 |
| Poland | 1 | 7.62 | 76.23 |
| Portugal | 2 | 13.06 | 185.13 |
| Spain | 1 | 8.91 | 98.01 |
| Sweden | 1 | 7.52 | 75.24 |
| United Kingdom | 2 | 17.55 | 245.52 |
| USA | 12 | 95.45 | 1040.49 |
| **Grand Total** | **53** | **408.14** | **4709.43** |





|  |  |  |
| --- | --- | --- |
| **Year** | **Count of customer\_id** | **Sum of Yearly\_Spending** |
| 2017 | 57 | 1201.86 |
| 2018 | 55 | 1147.41 |
| 2019 | 54 | 1221.66 |
| 2020 | 58 | 1138.5 |
| **Grand Total** | **224** | **4709.43** |



### **Approach:**

* **Days Since Last Purchase**:
  + Calculate the number of days since the last purchase for each customer to identify those who may be at risk of churning.
* **Risk Status Classification**:
  + Classify customers as "At Risk" if they haven't made a purchase in the last 90 days, and "Active" otherwise.
* **Total and Average Spending Calculation**:
  + Calculate total spending and average spending per customer to understand their purchasing behavior.
* **Yearly Spending Comparison**:
  + Analyze spending patterns year over year to identify trends in customer spending.
* **Spending Analysis by Location**:
  + Analyze spending by geographic location (city, state, country) to identify regional differences in customer behavior.

**Insights:**

* Inactivity: Prolonged periods between purchases boost the risk of churn, the study reveals.
* Low Purchase Frequency & Spending: Less often purchases along with the overall spendings indicate lesser interaction with the brands and the products.
* Demographics: Churn rates and spending may differ based on age and gender but beyond that, the data constraints make it difficult to go deeper on preferences and behaviour.

**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.

**Q.7) 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?**

Ans – **MYSQL QUERY**

with purchase\_history as ( select i.customer\_id , i.billing\_country , i.invoice\_date ,

concat(c.first\_name , ' ' , c.last\_name) as customer\_name , i.total

from invoice i

left join customer c on i.customer\_id = c.customer\_id

group by i.customer\_id , i.billing\_country , i.invoice\_date , i.total

order by customer\_name ) ,

Lifetime\_purchase as (select customer\_id , sum(total) as Lifetime\_value

from invoice

group by customer\_id )

select ph.customer\_id , ph.billing\_country , ph.invoice\_date ,

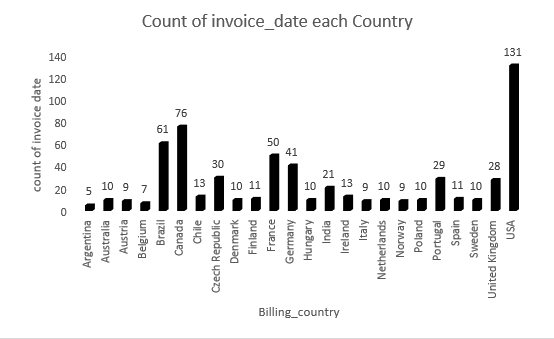
ph.customer\_name , ph.total , lp.Lifetime\_value

from purchase\_history ph

left join Lifetime\_purchase lp on lp.customer\_id = ph.customer\_id

order by lp.Lifetime\_value desc , ph.customer\_name , ph.invoice\_date ;

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Sum of Lifetime\_value** | **Count of invoice\_date** | **Sum of total** |
| Argentina | 198 | 5 | 39.6 |
| Australia | 811.8 | 10 | 81.18 |
| Austria | 623.7 | 9 | 69.3 |
| Belgium | 422.73 | 7 | 60.39 |
| Brazil | 5341.05 | 61 | 427.68 |
| Canada | 5291.55 | 76 | 535.59 |
| Chile | 1261.26 | 13 | 97.02 |
| Czech Republic | 4146.12 | 30 | 273.24 |
| Denmark | 376.2 | 10 | 37.62 |
| Finland | 871.2 | 11 | 79.2 |
| France | 3921.39 | 50 | 389.07 |
| Germany | 3440.25 | 41 | 334.62 |
| Hungary | 782.1 | 10 | 78.21 |
| India | 2024.55 | 21 | 183.15 |
| Ireland | 1492.92 | 13 | 114.84 |
| Italy | 454.41 | 9 | 50.49 |
| Netherlands | 653.4 | 10 | 65.34 |
| Norway | 650.43 | 9 | 72.27 |
| Poland | 762.3 | 10 | 76.23 |
| Portugal | 2653.2 | 29 | 185.13 |
| Spain | 1078.11 | 11 | 98.01 |
| Sweden | 752.4 | 10 | 75.24 |
| United Kingdom | 2337.39 | 28 | 245.52 |
| USA | 10650.42 | 131 | 1040.49 |



### **Approach:**

* **Purchase History Collection**:
  + Create a Common Table Expression (CTE) named purchase\_history to gather relevant customer data, including customer ID, billing country, invoice date, customer name, and total purchase amount.
* **Lifetime Value Calculation**:
  + Create another CTE named Lifetime\_purchase to calculate the total lifetime value for each customer by summing their total purchases.
* **Final Selection**:
  + Join the two CTEs to combine purchase history with lifetime value, allowing for a comprehensive view of each customer's purchasing behavior and their overall lifetime value.
* **Ordering Results**:
  + Order the results by lifetime value in descending order, followed by customer name and invoice date for better readability.

### **Insights from Customer Lifetime Value Modeling**

* **Identifying High-Value Customers**:
  + By calculating the lifetime value of customers, businesses can identify which segments contribute the most to revenue. This allows for targeted marketing efforts aimed at retaining these high-value customers.
* **Segmentation for Targeted Marketing**:
  + The analysis can reveal distinct customer segments based on their purchasing behavior, such as frequent buyers, high spenders, and occasional purchasers. Tailoring marketing strategies to these segments can improve engagement and conversion rates.
* **Churn Prediction**:
  + Customers with low or declining lifetime values may be at risk of churning. By identifying these customers early, businesses can implement retention strategies, such as personalized offers or loyalty programs, to encourage continued spending.
* **Understanding Purchase Patterns**:
  + Analyzing the purchase history can uncover trends in customer behavior, such as seasonal buying patterns or preferences for specific product categories. This information can inform inventory management and promotional strategies.
* **Impact of Engagement**:
  + Customers who engage with the brand through loyalty programs, newsletters, or social media may show higher lifetime values. Understanding the correlation between engagement and spending can help businesses design more effective engagement strategies.
* **Geographic Insights**:
  + The analysis of customer data by location can reveal regional differences in purchasing behavior. This can inform localized marketing strategies and help businesses allocate resources more effectively.

**Recommendations:**

**Customer Data Utilization:**

* Purchase History, Tenure, Engagement: These points of data can be used to calculate aspects about customers which include loyalty, value – high or low, and frequency of purchases.
* Recent High-Value Customers: Such customers may well become loyal in future and, therefore, represent potential for marketing or promotion.

**Segmentation & Targeted Marketing:**

Customer Segmentation: Due to the data processing techniques, customers can be divided into groups depending on their buying behavior, and the company can send targeted marketing messages to those groups.

**Churn in Less Developed Countries:**

* Higher Churn Rates: Mobile customers in LDCs have a higher churn rate possibly because of aspects such as product-market alignment or cost.

Addressing Issues: There are topics such as Product Market Fit and Pricing that need to be resolved to hack growth and minimize churn.

**Q.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?**

Ans – **MYSQL QUERY**

select count(\*) from track ; -- total available tracks according to our data

select distinct t.name

from track t

where t.track\_id not in ( select il.track\_id

from invoice\_line il

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

where i.invoice\_date between '2020-07-01' and '2020-12-31') ;

Identifying songs that were not Purchased by any customer in previous 6 months.

So promotions campaigns can applied over them to promote their sales.

select c.email , concat(c.first\_name, ' ', c.last\_name) as Customer\_Full\_Name

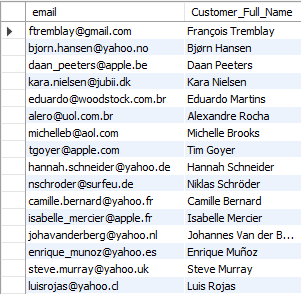
from customer c

where c.customer\_id not in (select distinct customer\_id

from invoice

where invoice\_date between '2020-07-01' and '2020-12-31' ) ;

Identifying those customer they have not made any purchase in previous 6 months.





* 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.

### **Approach:**

* **Total Tracks Available**:
  + First, determine the total number of tracks available in the database. This provides a baseline for understanding the inventory.
* **Identifying Unpurchased Tracks**:
  + Identify tracks that have not been purchased by any customer in the last six months. This information can help target promotional campaigns to boost sales of these tracks.
* **Identifying Inactive Customers**:
  + Find customers who have not made any purchases in the last six months. This information can be used to target these customers with re-engagement campaigns, such as email marketing or special offers.

### **Insights from Promotional Campaign Analysis**

* **Targeted Marketing Effectiveness**:
  + By identifying tracks that have not been purchased recently, businesses can tailor promotional campaigns specifically to these products. This targeted approach can lead to increased sales for underperforming items, demonstrating the effectiveness of focused marketing efforts.
* **Re-engagement of Inactive Customers**:
  + Promotional campaigns aimed at inactive customers can significantly improve retention rates. If a notable percentage of these customers return to make purchases after receiving targeted promotions, it indicates that the campaigns are effective in re-engaging lapsed customers.
* **Customer Segmentation**:
  + Analyzing the response to promotional campaigns can reveal distinct customer segments that are more responsive to certain types of promotions (e.g., discounts, exclusive events). This insight allows businesses to refine their marketing strategies and tailor future campaigns to specific segments for better results.
* **Sales Growth from Promotions**:
  + By comparing sales data before and after promotional campaigns, businesses can quantify the impact of these initiatives. A significant increase in sales, particularly for previously unpurchased tracks, indicates that the promotions successfully stimulated demand.
* **Return on Investment (ROI)**:
  + Calculating the ROI of promotional campaigns provides insights into their financial effectiveness. If the revenue generated from increased sales exceeds the costs of the campaigns, it validates the investment in marketing efforts and can guide future budget allocations.
* **Customer Lifetime Value (CLV) Enhancement**:
  + Successful promotional campaigns that lead to repeat purchases from previously inactive customers can enhance their lifetime value. Understanding how promotions affect CLV can help businesses prioritize customer retention strategies that yield long-term benefits.
* **Feedback and Improvement**:
  + Gathering customer feedback on promotional campaigns can provide insights into customer preferences and perceptions. This qualitative data can inform future campaigns, helping businesses understand what types of promotions resonate most with their audience.
* **Seasonal Trends and Timing**:
  + Analyzing the timing of promotional campaigns in relation to sales spikes or declines can reveal seasonal trends. This insight can help businesses plan future promotions to align with peak purchasing periods, maximizing their impact.

**Recommendations:**

* Promote attractive with high profit margins to customers, who are long-term inactive, in order to reactivate them.
* Use promotions – coupons or bundles – on tracks that did not sell in the previous months.
* Monitor Customer Acquisition Cost (CAC) to know which campaigns are inexpensive and assign new customers to a particular campaign by codes or URLs.
* Establish the extent of how churn rate has changed over a period of time, and how repeat purchases have taken place and compare this with the same situation during or after a promotion in a bid to establishing how effective they are in growing long-term customer relations.
* Measure the performance of a campaign by comparing the number of items sold or revenue generated and the AOV during the promotion period against those of a normal period.

**Q.9) How would you approach this problem, if the objective and subjective questions weren't given?**

Ans – **MYSQL QUERY**

If objective and subjective questions weren't given . then first I'll try to approach for finding

Duplicate and null value in each table , also year and country wise revenue

select \* from album ;

select distinct \* from album ; -- No any Duplicate

select \* from artist ;

select distinct \* from artist ; -- No any Duplicate

select \* from customer ;

select distinct \* from customer ; -- No any Duplicate

select count(\*) from customer

where company is null ; -- 49 Null value

select count(\*) from customer

where state is null ; -- 29 Null value

select count(\*) from customer

where fax is null ; -- 47 Null value

select \* from employee ; -- employee\_id 1 report\_to is null

select distinct \* from employee ; -- No any Duplicate

select \* from genre ;

select distinct \* from genre ; -- No any Duplicate

select \* from invoice ;

select distinct \* from invoice ; -- No any Duplicate

select \* from invoice\_line ;

select distinct \* from invoice\_line ; -- No any Duplicate

select \* from media\_type ;

select distinct \* from media\_type ; -- No any Duplicate

select \* from playlist ;

select distinct \* from playlist ; -- No any Duplicate

select \* from playlist\_track ;

select distinct \* from playlist\_track ; -- No any Duplicate

select \* from track ;

select distinct \* from track ; -- No any Duplicate

select sum(total) as Yearly\_Revenue , extract(year from invoice\_date) as Year

from invoice

group by 2;

/\* Yearly revenue

1201.86 2017

1147.41 2018

1221.66 2019

1138.50 2020

\*/

select billing\_country , sum(total) as Total\_Revenue

from invoice

group by billing\_country

order by Total\_Revenue desc ;

select customer\_id , sum(total) as Lifetime\_Purchase

from invoice

group by customer\_id

order by sum(total) desc ;

**Understand the Problem Context:**

**Identify Chinook's primary business goals such as :**

· Increasing revenue.

· Improving customer retention.

· Enhancing customer acquisition.

**Define the Core Business Questions:**

· "Which geographical regions generate the most revenue?"

· "What is the average purchase behaviour of customers across countries?"

· "What drives customer retention, and where are we losing customers?"

· "Which products or services contribute the most to revenue?"

**Explore and Understand the Data**

· Review the available tables in the Chinook database (e.g., Customer, Invoice, Invoice Line, Track, Genre, etc.).

· Understand the relationships between these tables (e.g., how customers link to invoices, how tracks contribute to purchases).

**Identify Key Metrics**

**Customer Insights:**

· Total customers by country.

· Average spending per customer.

· Churn rate and retention rate.

· Average number of tracks purchased.

**Sales Insights:**

· Total revenue by country.

· Most purchased genres or tracks.

**Geographical Trends:**

· Which countries contribute most to revenue?

· Identify high-value and low-value regions.

**Conclusion**

Without predefined questions, the approach revolves around exploring the data, getting insights and aligning analysis with business goals.

**Q.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?**

Ans – **MYSQL QUERY**

alter table album

add column ReleaseYear int;

select \*

from album;

**Q.11) 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.**

Ans – **MYSQL QUERY**

with average\_spent as( select round(avg(total),2) as avg\_total\_amount\_spent,

count(distinct customer\_id) as num\_of\_customer , billing\_country

from invoice i

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

group by billing\_country ),

Purchase\_qty as (select i.customer\_id, sum(quantity) as quantity\_purchased from invoice i

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

group by i.customer\_id ),

average\_by\_country as ( select billing\_country, round(avg(quantity\_purchased),2) as avg\_tracks\_per\_country

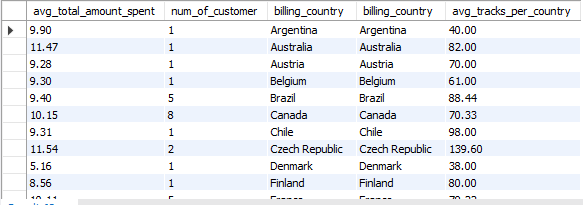
from invoice i

left join Purchase\_qty pq on pq.customer\_id = i.customer\_id

group by billing\_country )

select \* from average\_spent

left join average\_by\_country ac on ac.billing\_country = average\_spent.billing\_country ;



### **Approach**

* **Define Objectives**:
  + Determine the key metrics to analyze:
    - Average total amount spent by customers from each country.
    - Number of distinct customers in each country.
    - Average number of tracks purchased per customer in each country.
* **Data Sources**:
  + Identify the relevant tables in the database:
    - **Invoice Table**: Contains information about customer purchases, including total amounts and billing country.
    - **Invoice Line Table**: Contains details about individual items purchased in each invoice, including quantities and track IDs.
* **Calculate Average Total Amount Spent**:
  + Create a Common Table Expression (CTE) to calculate the average total amount spent by customers in each country:
    - Use the AVG() function to compute the average total from the invoice table.
    - Count distinct customers using COUNT(DISTINCT customer\_id) to get the number of customers per country.
    - Group the results by billing\_country.
* **Calculate Total Tracks Purchased**:
  + Create another CTE to calculate the total number of tracks purchased by each customer:
    - Sum the quantities from the invoice\_line table, grouping by customer\_id.
* **Calculate Average Tracks Purchased per Country**:
  + Create a third CTE to calculate the average number of tracks purchased per customer for each country:
    - Join the invoice table with the previous CTE (total tracks purchased) to compute the average number of tracks per country.
    - Use the AVG() function on the total quantities purchased, grouping by billing\_country.
* **Combine Results**:
  + Use a final SELECT statement to combine the results from the CTEs:
    - Perform a LEFT JOIN between the CTEs that calculate average spending and average tracks purchased, matching on billing\_country.
    - Select the relevant columns: average total amount spent, number of customers, billing country, and average tracks purchased.

**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.