Chinook Music Store Project

Objective Questions

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

ANS) There are three tables which has null values which are

1. Customer
    a. Company
    b. State
    c. Postal_code
    d. Phone
    e. Fax
2. Employee
    a. Reports_to
3. Track
    a. Composer

I will handle these columns that have null values using the COALESCE SQL function

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

ANS)

```sql
WITH TopTracksUSA AS (
    SELECT
        t.track_id,
        t.name AS track_name,
        g.name AS genre_name,
        ar.name AS artist_name,
        SUM(il.unit_price * il.quantity) AS sales
    FROM
        track t
    JOIN
        genre g ON t.genre_id = g.genre_id
    JOIN
        album al ON t.album_id = al.album_id
    JOIN
        artist ar ON al.artist_id = ar.artist_id
    JOIN
        invoice_line il ON il.track_id = t.track_id
    JOIN
        invoice i ON il.invoice_id = i.invoice_id
    WHERE
        i.billing_country = 'USA'
    GROUP BY
        t.track_id,t.name, g.name, ar.name
    ORDER BY
        sales DESC
    LIMIT 10
),
```

```
TopArtistUSA AS (
    SELECT
        ar.artist_id,
        ar.name as artist_name,
        SUM(il.unit_price * il.quantity) AS sales
    FROM
        artist ar
    JOIN
        album al ON al.artist_id = ar.artist_id
    JOIN
        track t ON t.album_id  = al.album_id
    JOIN
        invoice_line il ON il.track_id = t.track_id
    JOIN
        invoice i ON i.invoice_id = il.invoice_id
    WHERE
        i.billing_country = 'USA'
    GROUP BY
        ar.artist_id,ar.name
    ORDER BY
        sales DESC
    LIMIT 1
),
TopUSAArtistGenres AS (
    SELECT
        DISTINCT g.name as genre_name
    FROM
        genre g
    JOIN
        track t ON g.genre_id = t.genre_id
    JOIN
        album al ON al.album_id = t.album_id
    WHERE
        al.artist_id = (SELECT artist_id FROM TopArtistUSA)
)
```

I have created a Common Table Expression (CTE) named TopTracksUSA, which identifies the top ten tracks in the USA based on sales. Here are the top ten tracks in the USA.

```
SELECT * FROM TopTracksUSA;
```

track_id	track_name	genre_name	artist_name	sales
3336	War Pigs	Alternative	Cake	5.94
3465	You Know I'm No Good (feat. Ghostf…	R&B/Soul	Amy Winehouse	4.95
2560	Violent Pornography	Metal	System Of A Down	3.96
2647	End Of The Night	Rock	The Doors	3.96
13	Night Of The Long Knives	Rock	AC/DC	3.96
1995	Scentless Apprentice	Rock	Nirvana	3.96
153	Evil Woman	Metal	Black Sabbath	3.96
2646	I Looked At You	Rock	The Doors	3.96
1495	Highway Chile	Rock	Jimi Hendrix	3.96
55	I Can't Remember	Rock	Alice In Chains	2.97

I have created another CTE named TopArtistUSA, which identifies the top artist in the USA based on their track sales in the country. Accordingly, Van Halen is the top artist in the USA according to these sales.

```
SELECT * FROM TopArtistUSA;
```

artist_id	artist_name	sales
152	Van Halen	42.57

Finally, I compiled the names of the genres from the top ten rankings and the genres associated with the USA's top artist, Van Halen. I combined them using a UNION, resulting in a list of the most popular genres in the USA based on the sales of the top ten tracks and top artists.

```
SELECT DISTINCT genre_name FROM (
    SELECT genre_name FROM TopTracksUSA
    UNION ALL
    SELECT genre_name FROM TopUSAArtistGenres
) temp;
```

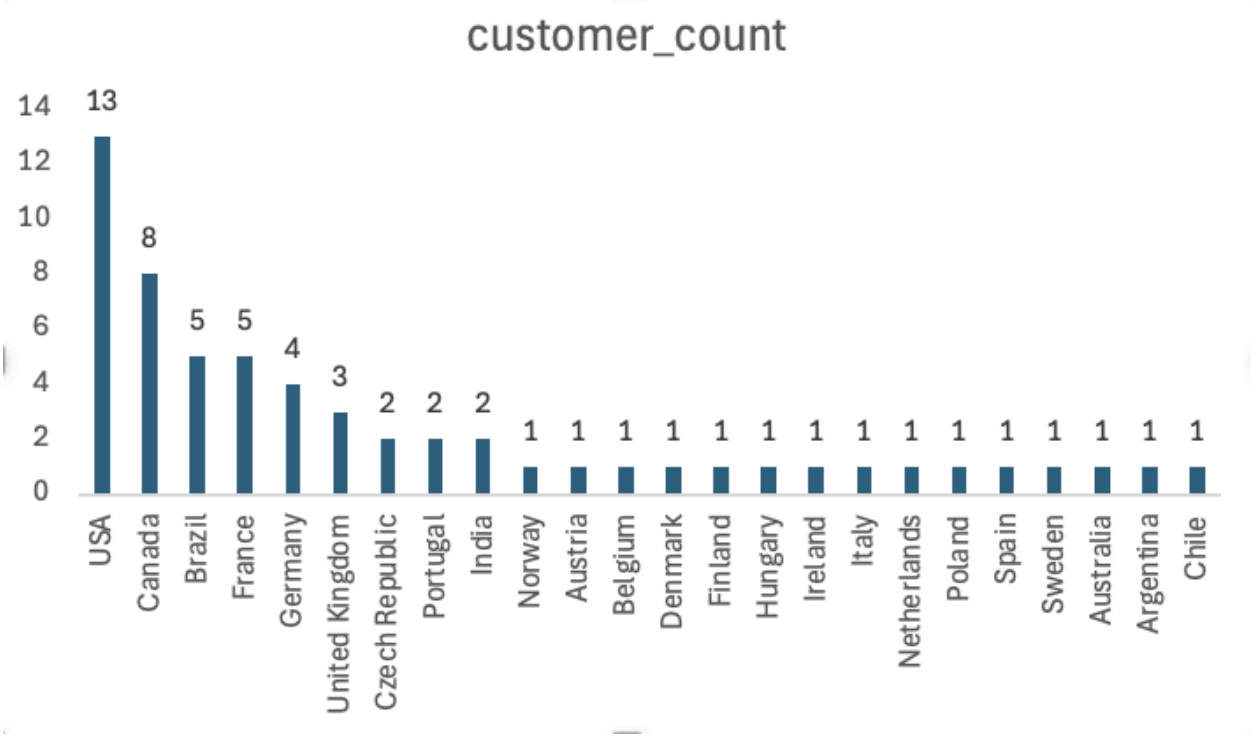
genre_name
Alternative
R&B/Soul
Metal
Rock

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

ANS)

```
SELECT
    c.country,
    COUNT(c.customer_id) AS customer_count
FROM
    customer c
GROUP BY
    c.country
ORDER BY
    customer_count DESC;
```

I have used the query above to determine the customer count in all the countries included in the database. Below, the bar chart provides a visual representation of the number of customers in each country.



From this bar chart, we can analyze that the United States and Canada are leading with the highest number of customers, with 13 and 8, respectively. Additionally, we can observe that in most other countries, the number of customers is the same, which is 1.

Q4) Calculate the total revenue and number of invoices for each country, state, and city

```
SELECT
    billing_country as country,
    billing_state as state,
    billing_city as city,
    COUNT(invoice_id) as Number_of_invoices,
    SUM(total) as Total_Revenue
FROM
    invoice
GROUP BY
    billing_country,billing_state,billing_city
ORDER BY
    Total_Revenue DESC, Number_of_invoices DESC;
```

country	state	city	Number_of_invoices	Total_Revenue
Czech Republic	None	Prague	30	273.24
USA	CA	Mountain View	20	169.29
United Kingdom	None	London	19	166.32
Germany	None	Berlin	20	158.40
France	None	Paris	18	151.47
Brazil	SP	São Paulo	22	129.69
Ireland	Dublin	Dublin	13	114.84
India	None	Delhi	13	111.87
Brazil	SP	São José dos Campos	13	108.90
Brazil	DF	Brasília	15	106.92
Portugal	None	Lisbon	13	102.96
France	None	Bordeaux	11	99.99
Canada	QC	Montréal	9	99.99
USA	WA	Redmond	12	98.01
Spain	None	Madrid	11	98.01
Chile	None	Santiago	13	97.02
Germany	None	Frankfurt	10	94.05
USA	FL	Orlando	12	92.07
Canada	ON	Ottawa	13	91.08
USA	NV	Reno	11	91.08
USA	TX	Fort Worth	12	86.13
USA	AZ	Tucson	9	84.15
Portugal	None	Porto	16	82.17
Germany	None	Stuttgart	11	82.17
Brazil	RJ	Rio de Janeiro	11	82.17
Australia	NSW	Sidney	10	81.18
Finland	None	Helsinki	11	79.20
United Kingdom	None	Edinburgh	9	79.20
USA	NY	New York	8	79.20
Hungary	None	Budapest	10	78.21
USA	WI	Madison	10	76.23
Poland	None	Warsaw	10	76.23
Canada	NT	Yellowknife	12	75.24
Sweden	None	Stockholm	10	75.24
France	None	Dijon	12	73.26
USA	UT	Salt Lake City	10	72.27
Norway	None	Oslo	9	72.27
USA	IL	Chicago	8	71.28
India	None	Bangalore	8	71.28
Canada	MB	Winnipeg	8	70.29
Austria	None	Vienne	9	69.30
USA	MA	Boston	10	66.33
Canada	BC	Vancouver	9	66.33

1. **United States is the highest earning country**: The US has a high number of cities listed in the table, contributing to a total revenue that is likely much higher than any other country.
2. **Brazil has a high number of invoices**: While the total revenue is not the highest, Brazil has many cities listed in the table. This suggests a high number of invoices generated, even if the individual invoice values are lower.
3. **European countries dominate the list:** There are a notable number of European countries represented in the table. This suggests strong business activity in Europe, particularly in countries like Germany, France, and the United Kingdom.

Q5) Find the top 5 customers by total revenue in each country

```sql
WITH CustomerRevenueRanked AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_full_name,
        c.country,
        SUM(i.total) AS total_revenue,
        RANK() OVER(PARTITION BY c.country ORDER BY SUM(i.total) DESC) AS rank_num
    FROM
        customer c
    JOIN
        invoice i ON i.customer_id = c.customer_id
    GROUP BY
        c.customer_id, customer_full_name, c.country
)
SELECT
    customer_id,
    customer_full_name,
    country,
    total_revenue,
    rank_num
FROM
    CustomerRevenueRanked
WHERE
    rank_num <= 5
ORDER BY
    country, rank_num;
```

I have identified the top five customers in each country based on their total revenue using the query described above. Initially, I calculated the total amount each customer spent and ranked the customers within their respective countries according to this total. Finally, I wrote a query to select only those customers who are ranked in the top five. Below is an image of the data obtained from running this query; due to the large volume of data, this image displays only a portion of it.

customer_id	customer_full_name	country	total_revenue	rank_num
56	Diego Gutiérrez	Argentina	39.60	1
55	Mark Taylor	Australia	81.18	1
7	Astrid Gruber	Austria	69.30	1
8	Daan Peeters	Belgium	60.39	1
1	Luís Gonçalves	Brazil	108.90	1
13	Fernanda Ramos	Brazil	106.92	2
12	Roberto Almeida	Brazil	82.17	3
11	Alexandre Rocha	Brazil	69.30	4
10	Eduardo Martins	Brazil	60.39	5
3	François Tremblay	Canada	99.99	1
30	Edward Francis	Canada	91.08	2
33	Ellie Sullivan	Canada	75.24	3
32	Aaron Mitchell	Canada	70.29	4
15	Jennifer Peterson	Canada	66.33	5
57	Luis Rojas	Chile	97.02	1
5	František Wichterl…	Czech Republic	144.54	1
6	Helena Holý	Czech Republic	128.70	2
9	Kara Nielsen	Denmark	37.62	1
44	Terhi Hämäläinen	Finland	79.20	1
42	Wyatt Girard	France	99.99	1
39	Camille Bernard	France	79.20	2
43	Isabelle Mercier	France	73.26	3
40	Dominique Lefebvre	France	72.27	4
41	Marc Dubois	France	64.35	5
37	Fynn Zimmermann	Germany	94.05	1
36	Hannah Schneider	Germany	85.14	2
2	Leonie Köhler	Germany	82.17	3
38	Niklas Schröder	Germany	73.26	4
45	Ladislav Kovács	Hungary	78.21	1
58	Manoj Pareek	India	111.87	1
59	Puja Srivastava	India	71.28	2
46	Hugh O'Reilly	Ireland	114.84	1
47	Lucas Mancini	Italy	50.49	1
48	Johannes Van der…	Netherlands	65.34	1
4	Bjørn Hansen	Norway	72.27	1
49	Stanisław Wójcik	Poland	76.23	1
34	João Fernandes	Portugal	102.96	1
35	Madalena Sampaio	Portugal	82.17	2
50	Enrique Muñoz	Spain	98.01	1
51	Joakim Johansson	Sweden	75.24	1
53	Phil Hughes	United Kingdom	98.01	1
54	Steve Murray	United Kingdom	79.20	2
52	Emma Jones	United Kingdom	68.31	3

Q6) Identify the top-selling track for each customer

```sql
WITH CustomerTrackSales AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_full_name,
        t.name AS track_name,
        SUM(il.unit_price * il.quantity) AS total_revenue,
        ROW_NUMBER() OVER (PARTITION BY c.customer_id ORDER BY SUM(il.unit_price * il.quantity) DESC) AS rank_num
    FROM
        customer c
    JOIN
        invoice i ON c.customer_id = i.customer_id
    JOIN
        invoice_line il ON i.invoice_id = il.invoice_id
    JOIN
        track t ON il.track_id = t.track_id
    GROUP BY
        c.customer_id, customer_full_name, t.name
)
SELECT
    customer_id,
    customer_full_name,
    track_name,
    total_revenue
FROM
    CustomerTrackSales
WHERE
    rank_num = 1
ORDER BY
    customer_id;
```

I utilized the query mentioned above to identify the top tracks for each customer, and the image attached below displays a partial result from this query on the dataset.

My primary observation from the complete results is that each customer has a unique track that is their top-selling choice.

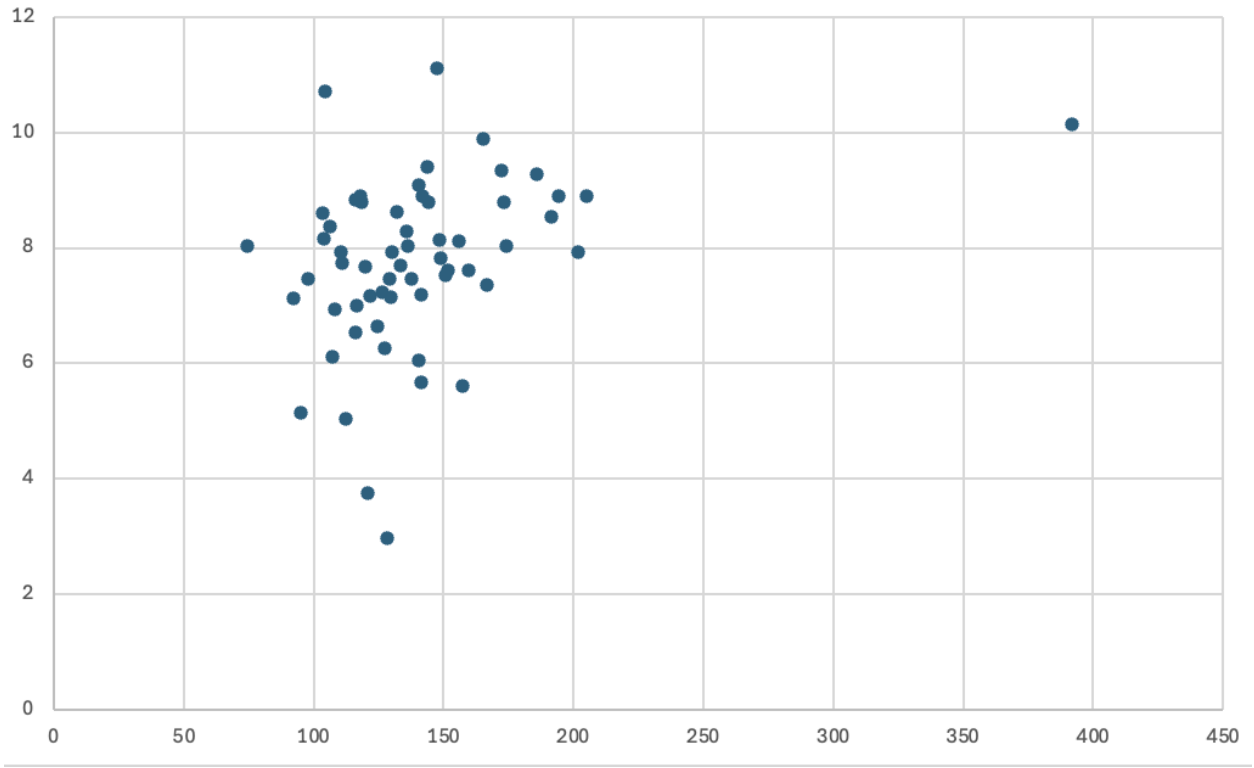
customer_id	customer_full_name	track_name	total_revenue
1	Luís Gonçalves	Put Your Lights On	0.99
2	Leonie Köhler	Cochise	0.99
3	François Tremblay	Sting Me	1.98
4	Bjørn Hansen	Welcome to the Jungle	0.99
5	František Wichterlová	Low Desert	0.99
6	Helena Holý	Foxy Lady	0.99
7	Astrid Gruber	Lady Double Dealer	0.99
8	Daan Peeters	I Wish It Would Rain	0.99
9	Kara Nielsen	The Crystal Ship	0.99
10	Eduardo Martins	Like A Bird	1.98
11	Alexandre Rocha	Battery	0.99
12	Roberto Almeida	Love And Marriage	1.98
13	Fernanda Ramos	24 Caprices, Op. 1, No. 24, for S…	1.98
14	Mark Philips	Us And Them	0.99
15	Jennifer Peterson	Hell Ain't A Bad Place To Be	0.99
16	Frank Harris	Too Young To Die	0.99
17	Jack Smith	Fly Away	0.99
18	Michelle Brooks	Right Next Door to Hell	0.99
19	Tim Goyer	Message in a Bottle (new classic…	0.99
20	Dan Miller	I Like Dirt	0.99

Q7) Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?

```
WITH InvoiceDates AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
        i.invoice_date,
        DATEDIFF(
            LEAD(i.invoice_date) OVER (PARTITION BY c.customer_id ORDER BY i.invoice_date),
            i.invoice_date
        ) AS days_between_purchases
    FROM
        customer c
    JOIN
        invoice i ON c.customer_id = i.customer_id
),
```

```sql
PurchaseFrequency AS (
    SELECT
        customer_id,
        customer_name,
        COUNT(*) AS total_purchases,
        AVG(days_between_purchases) AS avg_days_between_purchases
    FROM
        InvoiceDates
    GROUP BY
        customer_id, customer_name
),
CustomerSpending AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_full_name,
        COUNT(i.invoice_id) AS total_purchases,
        SUM(i.total) AS total_spent,
        AVG(i.total) AS avg_order_value
    FROM
        customer c
    JOIN
        invoice i ON c.customer_id = i.customer_id
    GROUP BY
        c.customer_id, customer_full_name
)
SELECT
    f.customer_id,
    f.customer_name,
    f.total_purchases,
    f.avg_days_between_purchases,
    s.total_spent,
    s.avg_order_value
FROM
    PurchaseFrequency f
JOIN
    CustomerSpending s ON f.customer_id = s.customer_id
ORDER BY
    f.total_purchases DESC;
```

customer_id	customer_name	total_purchases	avg_days_between_purchases	total_spent	avg_order_value
5	František Wichterlová	18	74.2941	144.54	8.030000
35	Madalena Sampaio	16	95.0000	82.17	5.135625
13	Fernanda Ramos	15	92.1429	106.92	7.128000
30	Edward Francis	13	116.4167	91.08	7.006154
46	Hugh O'Reilly	13	116.0833	114.84	8.833846
58	Manoj Pareek	13	103.3333	111.87	8.605385
57	Luis Rojas	13	97.6667	97.02	7.463077
1	Luís Gonçalves	13	106.2500	108.90	8.376923
34	João Fernandes	13	110.5000	102.96	7.920000
6	Helena Holý	12	104.4545	128.70	10.725000
10	Eduardo Martins	12	112.2727	60.39	5.032500
43	Isabelle Mercier	12	107.3636	73.26	6.105000
17	Jack Smith	12	104.1818	98.01	8.167500
20	Dan Miller	12	130.0909	95.04	7.920000
22	Heather Leacock	12	120.0909	92.07	7.672500
33	Ellie Sullivan	12	127.5455	75.24	6.270000
26	Richard Cunningham	12	121.5455	86.13	7.177500
50	Enrique Muñoz	11	118.1000	98.01	8.910000
44	Terhi Hämäläinen	11	141.3000	79.20	7.200000
12	Roberto Almeida	11	137.5000	82.17	7.470000



I have used the provided query to investigate potential patterns between the frequency of purchases and the average order value. Unfortunately, the dataset lacks data related to payment methods, limiting our analysis in that regard.

In the attached image, I've plotted a scatter graph based on partial data from the query results. This graph displays the relationship between the frequency of purchases (in days) and the average order value for each customer. Upon analyzing the scatter plot, it becomes evident that no strong correlation exists between these two variables. The data points are quite scattered, indicating that higher or lower purchase frequencies do not consistently correspond to higher or lower average order values.

Key Insights:

- **Minimum frequency of purchase:** Approximately 74 days.
- **Maximum frequency of purchase:** 392 days, which spans over a year.
- **Minimum average order cost:** $2.97.
- **Maximum average order cost:** $11.11.

The graph also suggests that most customers tend to cluster in a similar range of both average order value (around $6–$10) and frequency of purchases (around 100–200 days). However, there are a few outliers with much higher frequencies or order values. This lack of a clear trend might suggest that purchasing behavior varies significantly across the customer base, with no consistent link between how often customers buy and how much they spend per order.

Further investigation into other factors like customer segments, purchase history, or regional behaviors could provide deeper insights into purchasing behavior.

Q8) What is the customer churn rate?

```sql
WITH MostRecentInvoice AS (
    -- Find the most recent invoice date
    SELECT MAX(invoice_date) AS most_recent_invoice_date
    FROM invoice
),
CutoffDate AS (
    -- Calculate the cutoff date as 1 year before the most recent invoice date
    SELECT DATE_SUB(most_recent_invoice_date, INTERVAL 1 YEAR) AS cutoff_date
    FROM MostRecentInvoice
),
ChurnedCustomers AS (
    -- Find customers who haven't made a purchase in over a year or never made a purchase
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_full_name,
        MAX(i.invoice_date) AS last_purchase_date
    FROM
        customer c
    LEFT JOIN invoice i ON c.customer_id = i.customer_id
    GROUP BY
        c.customer_id, customer_full_name
    HAVING
        MAX(i.invoice_date) IS NULL OR MAX(i.invoice_date) < (SELECT cutoff_date FROM CutoffDate)
)
-- Calculate the churn rate
SELECT
    COUNT(*) AS churned_customers,
    (SELECT COUNT(*) FROM customer) AS total_customers,
    ROUND((COUNT(*) / (SELECT COUNT(*) FROM customer)) * 100, 2) AS churn_rate
FROM
    ChurnedCustomers;
```

churned_customers	total_customers	churn_rate
1	59	1.69

The query calculates the customer churn rate, which is an important metric for understanding customer retention and how many customers have stopped purchasing within a given period—in this case, one year. Here's an analysis of the query and the result:

Breakdown of the Query:

1. **MostRecentInvoice Subquery:**
    - **It retrieves the most recent purchase date across all**

customers (`MAX(invoice_date)`) to establish the current point in time for the analysis.

2. **CutoffDate Subquery:**
    - **This calculates the "cutoff date," which is exactly one year before the most recent purchase. Customers who have not made a purchase after this cutoff date are considered churned.**
3. **ChurnedCustomers Subquery:**
    - **This identifies customers who either:**
        - **Never made a purchase (i.e., `MAX(invoice_date) IS NULL`).**
        - **Haven't made a purchase in over a year (i.e., `MAX(invoice_date) < cutoff_date`).**
4. **Final Churn Rate Calculation:**
    - **The final query calculates the churn rate by dividing the number of churned customers by the total number of customers and multiplying the result by 100 to get a percentage.**

Result:

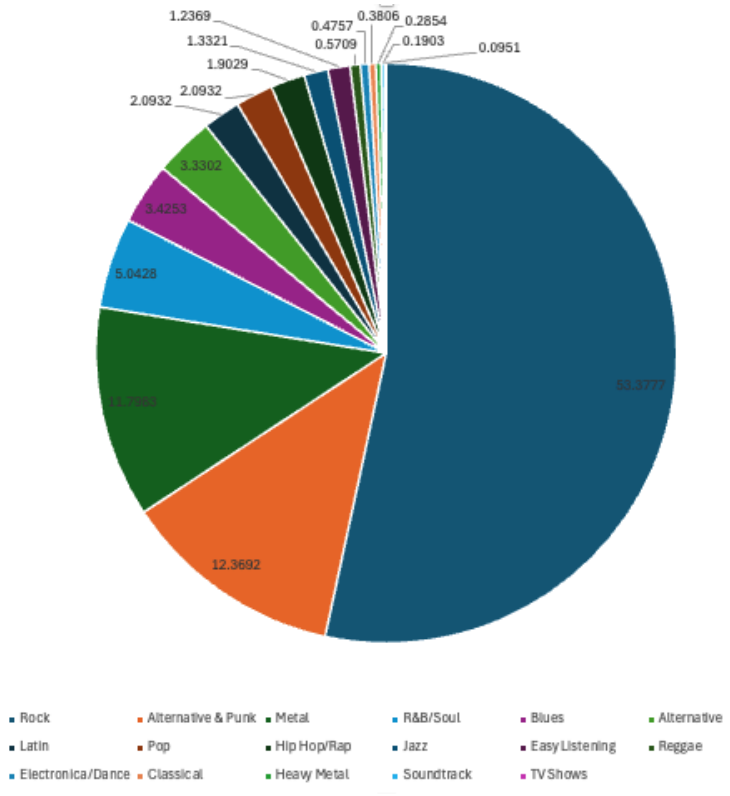
- **Churned Customers: 1.**
- **Total Customers: 59.**
- **Churn Rate: 1.69%.**

Analysis:

- **Churn Rate of 1.69%:** The churn rate is relatively low, which suggests that customer retention is quite high, with only one customer having stopped making purchases within the last year.
- **Implication**: A churn rate below 2% indicates that most customers are retained. However, it's still essential to investigate why this particular customer churned. Understanding the reasons behind customer churn, even for a small number of customers, can help in enhancing the customer experience and preventing future churn.

Q9) Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.

```sql
WITH TotalTrackSalesInUSA AS (
    SELECT
        SUM(quantity) as total_sales
    FROM
        invoice_line il
    JOIN
        invoice i ON i.invoice_id = il.invoice_id
    WHERE
        billing_country = 'USA'
)
SELECT
    g.name as genre_name,
    SUM(il.quantity) as total_genre_sales,
    (SUM(il.quantity)/(SELECT total_sales FROM TotalTrackSalesInUSA) * 100) as percentage_of_sales_contributed
FROM
    genre g
JOIN
    track t ON g.genre_id = t.genre_id
JOIN
    invoice_line il ON il.track_id = t.track_id
JOIN
    invoice i ON i.invoice_id  = il.invoice_id
WHERE
    i.billing_country = 'USA'
GROUP BY
    genre_name
ORDER BY
    percentage_of_sales_contributed DESC;
```




```sql
WITH TotalTrackSalesInUSA AS (
    SELECT
        SUM(il.quantity) AS total_sales
    FROM
        invoice_line il
    JOIN
        invoice i ON i.invoice_id = il.invoice_id
    WHERE
        i.billing_country = 'USA'
)
SELECT
    a.artist_id,
    a.name AS artist_name,
    SUM(il.quantity) AS total_artist_sales,
    ROUND((SUM(il.quantity) / (SELECT total_sales FROM TotalTrackSalesInUSA)) * 100, 2) AS percentage_of_sales_contributed
FROM
    artist a
JOIN
    album ab ON a.artist_id = ab.artist_id
JOIN
    track t ON ab.album_id = t.album_id
JOIN
    invoice_line il ON il.track_id = t.track_id
JOIN
    invoice i ON i.invoice_id = il.invoice_id
WHERE
    i.billing_country = 'USA'
GROUP BY
    a.artist_id,a.name
ORDER BY
    percentage_of_sales_contributed DESC;
```

artist_id	artist_name	total_artist_sales	percentage_of_sales_contributed
152	Van Halen	43	4.09
124	R.E.M.	38	3.62
142	The Rolling Stones	37	3.52
110	Nirvana	35	3.33
81	Eric Clapton	34	3.24
84	Foo Fighters	34	3.24
88	Guns N' Roses	32	3.04
54	Green Day	32	3.04
118	Pearl Jam	31	2.95
252	Amy Winehouse	30	2.85
1	AC/DC	29	2.76
94	Jimi Hendrix	28	2.66
140	The Doors	27	2.57
104	Marvin Gaye	26	2.47
12	Black Sabbath	24	2.28
150	U2	23	2.19
135	System Of A Down	22	2.09
179	Scorpions	21	2.00
5	Alice In Chains	21	2.00
180	House Of Pain	20	1.90

Genre Sales Contribution:

The pie chart highlights the distribution of total sales across various music genres in the USA. The leading genres are:

1. Rock – Dominating with over 53.37% of the total sales, Rock is the top-selling genre. This suggests that Rock has a significant influence and popularity in the music market, capturing more than half of the sales.
2. Alternative & Punk – this genre accounts for 12.37%, showing a moderate yet dedicated following.

Other genres like Metal (11.79%), R&B/Soul (5.42%), and Blues (3.42%) contribute smaller but significant portions to the total sales.

Top-Selling Artists:

The leading artists, based on total sales, are as follows:

- Van Halen – As the top-selling artist, Van Halen's contributions to the Rock genre likely play a major role in its dominant position in the pie chart.
- R.E.M. – Another major contributor, R.E.M., ranks second in sales, further solidifying Rock's overwhelming presence in the music market.

These artists' popularity and extensive catalog have cemented their place at the top, reinforcing Rock's dominance in the music industry.

Conclusion:

From this analysis, it is clear that Rock music is the dominant genre in the USA, contributing more than half of the total sales. Alternative & Punk music follow but is significantly behind in terms of sales percentage. The most successful artists, like Van Halen and R.E.M., are crucial to Rock’s stronghold in the market.

Q10)Find customers who have purchased tracks from at least 3 different+ genres

```sql
SELECT
    c.customer_id,
    CONCAT(c.first_name,' ',c.last_name) as customer_name,
    COUNT(DISTINCT t.genre_id) as total_different_genres
FROM
    customer c
JOIN
    invoice i ON c.customer_id = i.customer_id
JOIN
    invoice_line il ON i.invoice_id = il.invoice_id
JOIN
    track t ON t.track_id = il.track_id
GROUP BY
    c.customer_id, customer_name
HAVING
    total_different_genres >= 3
ORDER BY
    total_different_genres DESC;
```

customer_id	customer_name	total_different_genres
2	Leonie Köhler	14
30	Edward Francis	13
5	František Wichterlová	13
44	Terhi Hämäläinen	13
35	Madalena Sampaio	13
22	Heather Leacock	13
46	Hugh O'Reilly	12
13	Fernanda Ramos	12
42	Wyatt Girard	12
41	Marc Dubois	12
17	Jack Smith	12
18	Michelle Brooks	12
38	Niklas Schröder	12
23	John Gordon	12
28	Julia Barnett	12
57	Luis Rojas	11
55	Mark Taylor	11
6	Helena Holý	11
54	Steve Murray	11
50	Enrique Muñoz	11
48	Johannes Van der B...	11
45	Ladislav Kovács	11
39	Camille Bernard	11

Total Customers:

A total of 59 customers have purchased tracks from at least three different genres. This indicates that a significant portion of the customer base enjoys a diverse range of music, highlighting a trend of genre-crossing musical preferences among these listeners.

Customer with the Most Genres:

- Leonie Köhler stands out as the customer with the most varied taste, having purchased tracks from 14 different genres. This reflects a deep and broad interest in music across a wide spectrum of styles. Such a diverse musical preference suggests that Leonie Köhler may have an eclectic listening habit and likely enjoys exploring multiple music styles without sticking to a single genre.

Customer with the Least Genres:

- On the other hand, Robert Brown, with purchases spanning 5 genres, represents the lower end of the diversity spectrum among this group. While still purchasing from multiple genres, Robert Brown's choices are relatively more concentrated compared to other customers in this list. This suggests that although he enjoys a variety of music, his tastes are somewhat more focused compared to someone like Leonie Köhler.

Q11) Rank genres based on their sales performance in the USA

```sql
WITH genreSalesInUsa AS (
    SELECT
        g.genre_id,
        g.name as genre_name,
        SUM(il.unit_price * quantity) as total_sales
    FROM
        genre g
    JOIN
        track t ON g.genre_id = t.genre_id
    JOIN
        invoice_line il ON t.track_id = il.track_id
    JOIN
        invoice i ON i.invoice_id = il.invoice_id
    WHERE
        i.billing_country = 'USA'
    GROUP BY
        g.genre_id, genre_name
)
SELECT
    genre_id,
    genre_name,
    total_sales,
    RANK() OVER(ORDER BY total_sales DESC) as genre_rank
FROM
    genreSalesInUsa
ORDER BY
    genre_rank;
```

genre_id	genre_name	total_sales	genre_rank
1	Rock	555.39	1
4	Alternative & Punk	128.70	2
3	Metal	122.76	3
14	R&B/Soul	52.47	4
6	Blues	35.64	5
23	Alternative	34.65	6
7	Latin	21.78	7
9	Pop	21.78	7
17	Hip Hop/Rap	19.80	9
2	Jazz	13.86	10
12	Easy Listening	12.87	11
8	Reggae	5.94	12
15	Electronica/Dance	4.95	13
24	Classical	3.96	14
13	Heavy Metal	2.97	15
10	Soundtrack	1.98	16
19	TV Shows	0.99	17

Top Genres Based on Sales Performance:

1. Rock – With total sales of 555.39, Rock is the dominant genre, clearly leading the market.
2. Alternative & Punk – Following Rock, this genre has sales of 128.70, making it the second-best performing genre.
3. Metal – Taking the third spot, Metal has total sales of 122.76.
4. R&B/Soul – With sales amounting to 52.47, R&B/Soul ranks fourth in sales performance.
5. Blues – Ranking fifth, Blues has total sales of 35.64.
6. Alternative – Close behind, Alternative comes in sixth with sales of 34.65.
7. Latin and Pop – Both genres share the seventh position with identical sales of 21.78.
8. Hip Hop/Rap – Hip Hop/Rap ranks ninth with total sales of 19.80.
9. Jazz – With sales of 13.86, Jazz ranks tenth.

Conclusion:

Rock is the clear leader in the USA music market with significantly higher sales compared to other genres. Alternative & Punk and Metal also perform well, but the gap between Rock and the rest is substantial. On the lower end, genres like TV Shows, Soundtrack, and Heavy Metal contribute the least to total sales.

Q12) Identify customers who have not made a purchase in the last 3 months

```sql
SELECT
    c.customer_id,
    CONCAT(c.first_name,' ',c.last_name) as customer_name
FROM
    customer c
JOIN
    invoice i ON c.customer_id = i.customer_id
GROUP BY
    c.customer_id,customer_name
HAVING
    MAX(i.invoice_date) <= DATE_SUB(current_timestamp(), INTERVAL 3 MONTH)
ORDER BY
    customer_name, customer_id;
```

customer_id	customer_name
32	Aaron Mitchell
11	Alexandre Rocha
7	Astrid Gruber
4	Bjørn Hansen
39	Camille Bernard
8	Daan Peeters
20	Dan Miller
56	Diego Gutiérrez
40	Dominique Lefebvre
10	Eduardo Martins
30	Edward Francis
33	Ellie Sullivan
52	Emma Jones
50	Enrique Muñoz
13	Fernanda Ramos
3	François Tremblay
16	Frank Harris
24	Frank Ralston
5	František Wichterl…

All Customers Identified:

According to the data retrieved from the query, all customers in the dataset have not made any purchases in the last 3 months. The most recent purchases recorded in the invoice data are from a long time ago, meaning no transactions have occurred within the recent 3-month period.

Analysis:

This result could indicate that customer engagement has significantly dropped, or the data being used is outdated. In either case, the lack of recent purchases points to a possible gap in customer activity, which may require intervention, such as marketing campaigns, promotional offers, or customer outreach, to reignite interest and encourage future purchases.

SUBJECTIVE QUESTIONS

Q1) 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.

```sql
WITH TopGenreSalesInUsa AS (
    SELECT
        g.genre_id,
        g.name as genre_name,
        SUM(il.unit_price * il.quantity) as total_sales,
        RANK() OVER(ORDER BY SUM(il.unit_price * il.quantity) DESC) as genre_rank
    FROM
        genre g
    JOIN
        track t on t.genre_id = g.genre_id
    JOIN
        invoice_line il ON il.track_id = t.track_id
    JOIN
        invoice i ON i.invoice_id = il.invoice_id
    WHERE
        i.billing_country = 'USA'
    GROUP BY
        g.genre_id, g.name
    ORDER BY
        total_sales DESC
)
SELECT
    al.title as album_name,
    g.name as genre_name,
    SUM(il.unit_price * il.quantity) as total_sales
FROM
    album al
JOIN
    track t ON al.album_id = t.album_id
JOIN
    invoice_line il ON il.track_id = t.track_id
JOIN
    genre g ON t.genre_id = g.genre_id
WHERE
    t.genre_id IN (SELECT genre_id FROM TopGenreSalesInUsa WHERE genre_rank < 3)
GROUP BY
    al.title, g.name
ORDER BY
    total_sales DESC;
```

album_name	genre_name	total_sales
Are You Experienced?	Rock	185.13
Get Born	Alternative & Punk	89.10
The Doors	Rock	82.17
The Police Greatest Hits	Rock	79.20
Big Ones	Rock	79.20
Greatest Hits I	Rock	79.20
From The Muddy Banks Of The Wish…	Rock	77.22
My Generation - The Very Best Of Th…	Rock	75.24
Jagged Little Pill	Rock	74.25

Based on the analysis of genre sales in the USA, we can prioritize the following three albums for advertising and promotion:

1. **Are You Experienced?**
    - **Genre**: Rock
    - **Reason**: Rock continues to be one of the top-performing genres in the USA, consistently contributing a large share of total music sales. "Are You Experienced?" is a prominent album within this genre, making it a strong candidate for promotion. Targeting fans of the rock genre, which has a proven record of high sales, will likely yield significant returns.
2. **Get Born**
    - **Genre**: Alternative & Punk
    - **Reason**: Alternative & Punk is another high-selling genre in the USA. "Get Born" stands out in this genre due to its total sales performance. The album resonates with a diverse audience who are inclined toward more contemporary or alternative styles of music. Promoting this album will allow the label to capture interest in a genre that has demonstrated growth and sustained popularity.
3. **The Doors**
    - **Genre**: Rock
    - **Reason**: As mentioned earlier, Rock is a dominant genre in the market. "The Doors" album is another top seller in this genre, cementing its relevance in today's market. This album has

historical significance and continues to engage fans of classic rock. Promoting "The Doors" would not only leverage the popularity of the genre but also tap into the nostalgia and loyalty associated with classic rock bands.

Analysis:

The selection of these albums is grounded in a clear understanding of current market trends in the USA, particularly in terms of **genre performance**. Both **Rock** and **Alternative & Punk** genres consistently perform well, accounting for a large portion of sales. These three albums, being top sellers in their respective genres, provide a strong opportunity to target established and passionate fanbases.

- **Genre Popularity**: Rock continues to dominate, while Alternative & Punk represents a younger, vibrant audience. Promoting albums in these genres will help capture both older, more loyal fans (for Rock) and a younger audience (for Alternative & Punk).
- **Total Sales Performance**: The albums selected have shown excellent sales performance, indicating that they are already resonating with the audience. By amplifying their visibility through advertising and promotion, we can further boost sales and maximize revenue.
- **Advertising Focus**: Rock and Alternative & Punk should be the core focus of any promotion strategy in the USA due to their continued sales success. These genres also tend to have dedicated fanbases, making them ideal for targeted marketing campaigns.

Conclusion:

These three albums—**"Are You Experienced?"**, **"Get Born"**, and **"The Doors"**—are well-positioned to drive future sales. By prioritizing them for promotion, we can align our efforts with the genres and albums that are already performing strongly in the USA market. This strategy will help maximize the impact of our marketing efforts and further establish these albums in their respective fanbases.

Q2) Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

```sql
WITH GenreSalesByCountry AS (
    SELECT
        i.billing_country AS country,
        g.name AS genre_name,
        SUM(il.quantity) AS total_sales,
        RANK() OVER (PARTITION BY i.billing_country ORDER BY SUM(il.quantity) DESC) AS genre_rank
    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'  -- Exclude the USA
    GROUP BY
        i.billing_country, g.name
)
SELECT
    country,
    genre_name,
    total_sales
FROM
    GenreSalesByCountry
WHERE
    genre_rank < 3  -- Top genre for each country
ORDER BY
    total_sales DESC;
```

country	genre_name	total_sales
Canada	Rock	333
France	Rock	211
Brazil	Rock	205
Germany	Rock	194
United Kingdom	Rock	166
Czech Republic	Rock	143
Portugal	Rock	108
India	Rock	102
Brazil	Alternative & Punk	74
Canada	Metal	72
Ireland	Rock	72
Chile	Rock	61
Sweden	Rock	60
France	Metal	54
Spain	Rock	46
Finland	Rock	46
Germany	Metal	44
Hungary	Rock	44
Czech Republic	Alternative & Punk	42
Austria	Rock	40
Poland	Rock	40
Norway	Rock	40
Italy	Rock	35
Australia	Rock	34
Netherlands	Rock	33
United Kingdom	Metal	31
Portugal	Metal	28
Belgium	Rock	26
India	Alternative & Punk	25
Denmark	Rock	24
Australia	Alternative & Punk	22
Ireland	Latin	20
Hungary	Metal	19

Based on the analysis of top-selling genres in countries other than the USA, the following insights can be drawn:

Summary of Top 2 Genres Across Different Countries:

- Rock: 23 countries
- Metal: 12 countries
- Alternative & Punk: 6 countries

- Jazz: 2 countries
- Blues: 1 country
- Latin: 1 country
- R&B/Soul: 1 country

Key Observations and Trends:

1. Rock's Global Dominance:
    - Rock is the clear global leader, appearing in the top 2 genres in 23 countries. This genre has widespread appeal across multiple regions, suggesting that Rock has a broad international fan base. It is popular across diverse markets and remains a cornerstone of the global music industry.
2. Metal's Strong Global Presence:
    - Metal ranks in the top 2 genres in 12 countries, showcasing its popularity, especially in regions with strong rock or alternative music cultures. While not as universally dominant as Rock, Metal has a dedicated and passionate global audience.
3. Alternative & Punk's Mid-level Appeal:
    - Alternative & Punk ranks in the top 2 in 6 countries, indicating that it holds a significant presence but does not have the same global reach as Rock or Metal. This genre is more niche, with strong followings in specific markets where alternative or rebellious musical styles are embraced.
4. Niche Genres with Regional Popularity:
    - Genres like Blues, Jazz, Latin, and R&B/Soul show up in the top 2 genres for only 1 or 2 countries. This suggests that these genres have strong appeal in specific regions but lack widespread global dominance. For example, Latin music likely performs well in regions with a cultural affinity for Latin rhythms, while Blues and Jazz remain more niche and genre-specific in their appeal.

Commonalities and Differences:

- Commonalities:
    - Rock and Metal are the most common high-performing genres globally (outside the USA). They appear frequently across multiple regions, demonstrating their universal appeal. Both genres resonate with international audiences, transcending cultural and linguistic barriers.
- Differences:
    - Some genres, such as Jazz, Blues, Latin, and R&B/Soul, have more limited geographic appeal. Their popularity is often confined to specific countries or regions, indicating cultural or regional preferences. For instance, Latin music's popularity may be concentrated in Spanish-speaking countries, while Jazz may have stronger followings in regions with a rich musical heritage tied to jazz traditions.

Conclusion:

- Rock and Metal are the most globally dominant genres outside the USA, with significant appeal in a wide variety of markets.
- Alternative & Punk has a moderate but notable presence in specific countries.
- Blues, Jazz, Latin, and R&B/Soul show strong regional preferences but lack broad international presence

Q3) 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?

```
WITH CustomerClassification AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
        MIN(i.invoice_date) AS first_purchase_date,
        CASE
            WHEN MIN(i.invoice_date) < DATE_SUB(CURDATE(), INTERVAL 1 YEAR) THEN 'Long-term'
            ELSE 'New'
        END AS customer_type
    FROM
        customer c
    JOIN
        invoice i ON c.customer_id = i.customer_id
    GROUP BY
        c.customer_id, customer_name
),
CustomerBehavior AS (
    SELECT
        cc.customer_id,
        cc.customer_name,
        cc.customer_type,
        COUNT(i.invoice_id) AS total_purchases,
        AVG(il.quantity) AS avg_basket_size,
        AVG(i.total) AS avg_spending
    FROM
        CustomerClassification cc
    JOIN
        invoice i ON cc.customer_id = i.customer_id
    JOIN
        invoice_line il ON i.invoice_id = il.invoice_id
    GROUP BY
        cc.customer_id, cc.customer_name, cc.customer_type
)
```

```sql
-- Compare purchasing habits between long-term and new customers
SELECT
    customer_type,
    AVG(total_purchases) AS avg_total_purchases,
    AVG(avg_basket_size) AS avg_basket_size,
    AVG(avg_spending) AS avg_spending
FROM
    CustomerBehavior
GROUP BY
    customer_type;
```

customer_type	avg_total_purchases	avg_basket_size	avg_spending
Long-term	80.6271	1.00000000	9.7689326610

Based on the results of the query analyzing customer purchasing behavior, we can derive some meaningful insights about the purchasing habits of long-term customers compared to new customers:

Analysis of Customer Purchasing Behavior:

The query provided the following results for **long-term customers**:

- **Average Total Purchases**: 80.62
- **Average Basket Size** (number of items per purchase): 1.00
- **Average Spending per Order**: $9.77

However, there was no data generated for **new customers**, which suggests that all of the customers in the dataset may fall under the "long-term" category. This indicates a potential data limitation where either there is no recent influx of new customers, or the classification cutoff for new customers (within the last year) may not be capturing enough customers to provide comparable results.

Insights:

1. **High Engagement Among Long-Term Customers**:
    - On average, long-term customers have made **80.62 total purchases**, which indicates a high level of engagement and loyalty. This is a significant number and shows that long-term

customers have consistently returned to make repeat purchases.

2. **Consistent Basket Size**:
    - The **average basket size** (1.00) indicates that, on average, customers are purchasing a single item per transaction. While this is consistent, it may highlight an opportunity for increasing the basket size through cross-selling or bundling offers to encourage larger purchases per transaction.
3. **Steady Spending Patterns**:
    - The **average spending per order** is around **$9.77**. This suggests that while customers are making frequent purchases, each transaction remains relatively modest in value. Understanding the types of products that are typically purchased can help refine marketing strategies aimed at increasing the average order value.

Retention and Loyalty Insights:

- **High Purchase Frequency Reflects Loyalty**:
    - The high number of total purchases (80.62) among long-term customers suggests strong brand loyalty. These customers have likely developed a habit of purchasing from the store, making them a valuable segment for retention efforts.
- **Retention Strategies for Long-Term Customers**:
    - To further enhance loyalty, offering personalized recommendations based on previous purchase history, loyalty rewards, or exclusive discounts could incentivize long-term customers to continue shopping and potentially increase basket size.
- **Opportunity for Cross-Selling**:
    - Given that the average basket size is 1.00, cross-selling initiatives could be introduced to encourage customers to purchase additional items during each transaction. This can increase the overall value of each purchase without requiring customers to increase their purchase frequency.

Missing Data for New Customers:

- The absence of data for **new customers** suggests that there may be a need to refine the classification criteria or address the lack of recent customer acquisition. By understanding why fewer new customers are being captured, the company could adjust its marketing strategies to attract and retain newer customers.

Conclusion:

The purchasing behavior analysis reveals that long-term customers exhibit high frequency in purchases but tend to purchase only one item per transaction with an average spending of $9.77. There is an opportunity to drive growth by increasing the basket size through cross-selling and personalized recommendations. Additionally, it is important to address the lack of data for new customers to understand and improve customer acquisition and retention strategies.

Q4)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?

```sql
WITH GenreCoPurchase AS (
    -- Find all combinations of genres purchased in the same invoice
    SELECT
        il1.invoice_id,
        g1.name AS genre_1,
        g2.name AS genre_2,
        COUNT(*) AS co_purchase_count
    FROM
        invoice_line il1
    JOIN
        track t1 ON il1.track_id = t1.track_id
    JOIN
        genre g1 ON t1.genre_id = g1.genre_id
    JOIN
        invoice_line il2 ON il1.invoice_id = il2.invoice_id
    JOIN
        track t2 ON il2.track_id = t2.track_id
    JOIN
        genre g2 ON t2.genre_id = g2.genre_id
    WHERE
        il1.track_id <> il2.track_id
    GROUP BY
        g1.name, g2.name, il1.invoice_id
)
SELECT
    genre_1,
    genre_2,
    SUM(co_purchase_count) AS total_co_purchases
FROM
    GenreCoPurchase
GROUP BY
    genre_1, genre_2
ORDER BY
    total_co_purchases DESC;
```

genre_1	genre_2	total_co_purchases
Rock	Metal	1622
Metal	Rock	1622
Rock	Alternative & Punk	1056
Alternative & Punk	Rock	1056
Latin	Rock	427
Rock	Latin	427
R&B/Soul	Rock	407
Rock	R&B/Soul	407
Metal	Alternative & Punk	315
Alternative & Punk	Metal	315
Rock	Alternative	309
Alternative	Rock	309
Blues	Rock	281
Rock	Blues	281
Rock	Jazz	237
Jazz	Rock	237
Rock	Easy Listening	184
Easy Listening	Rock	184
Rock	Classical	164
Classical	Rock	164
Rock	Pop	143
Pop	Rock	143
Latin	Metal	123
Metal	Latin	123

Product Affinity Analysis:

Based on the query results, we have identified patterns of co-purchased music **genres**, **artists**, and **albums**. This data provides valuable insights for guiding product recommendations and cross-selling initiatives in the store.

1. Top Co-Purchased Music Genres:

- **Metal and Rock**: Co-purchased 1,622 times
- **Rock and Alternative & Punk**: Co-purchased 1,056 times
- **Latin and Rock**: Co-purchased 426 times

Analysis:

- **Rock** appears frequently in combination with other genres such as Metal, Alternative & Punk, and Latin. This suggests that **Rock** is a versatile genre that resonates with customers who enjoy a variety of music styles.

- Customers who purchase **Metal** often also purchase **Rock**, which highlights a potential opportunity for cross-promoting these genres together. Additionally, **Alternative & Punk** and **Rock** show a strong affinity, suggesting that customers who enjoy **Alternative & Punk** are also inclined towards **Rock**.
- **Latin** music, when paired with **Rock**, also shows significant co-purchasing. This indicates an opportunity to recommend Latin music to customers already interested in **Rock** and vice versa, tapping into customers' varied tastes

```sql
WITH ArtistCoPurchase AS (
    -- Find all combinations of artists purchased in the same invoice
    SELECT
        il1.invoice_id,
        a1.name AS artist_1,
        a2.name AS artist_2,
        COUNT(*) AS co_purchase_count
    FROM
        invoice_line il1
    JOIN
        track t1 ON il1.track_id = t1.track_id
    JOIN
        album al1 ON t1.album_id = al1.album_id
    JOIN
        artist a1 ON al1.artist_id = a1.artist_id
    JOIN
        invoice_line il2 ON il1.invoice_id = il2.invoice_id
    JOIN
        track t2 ON il2.track_id = t2.track_id
    JOIN
        album al2 ON t2.album_id = al2.album_id
    JOIN
        artist a2 ON al2.artist_id = a2.artist_id
    WHERE
        il1.track_id <> il2.track_id AND a1.artist_id <> a2.artist_id
    GROUP BY
        a1.name, a2.name, il1.invoice_id
)
SELECT
    artist_1,
    artist_2,
    SUM(co_purchase_count) AS total_co_purchases
FROM
    ArtistCoPurchase
GROUP BY
    artist_1, artist_2
ORDER BY
    total_co_purchases DESC;
```

artist_1	artist_2	total_co_purchases
Led Zeppelin	Green Day	24
Green Day	Led Zeppelin	24
Green Day	Foo Fighters	20
Foo Fighters	Green Day	20
Nirvana	Eric Clapton	19
Eric Clapton	Nirvana	19
Nirvana	The Rolling Stones	19
The Rolling Stones	Nirvana	19
Metallica	Green Day	18
Green Day	Metallica	18
Queen	Green Day	17
Green Day	Queen	17
The Doors	Amy Winehouse	16
Amy Winehouse	The Doors	16
Queen	U2	16
U2	Queen	16
Aerosmith	Guns N' Roses	16
Guns N' Roses	Aerosmith	16
Green Day	Guns N' Roses	16
Guns N' Roses	Green Day	16

2. **Top Co-Purchased Artists:**

- **Led Zeppelin and Green Day**: Co-purchased 24 times
- **Green Day and Foo Fighters**: Co-purchased 20 times
- **Nirvana and Eric Clapton**: Co-purchased 19 times

Analysis:

- The combination of **Led Zeppelin** and **Green Day**, as well as **Green Day** and **Foo Fighters**, shows that customers who enjoy **classic rock** (Led Zeppelin) are also likely to appreciate **punk rock** (Green Day) and **modern rock** (Foo Fighters).
- **Nirvana** (grunge) and **Eric Clapton** (blues/rock) also show a noteworthy co-purchase pattern, suggesting that customers who are drawn to **rock sub-genres** are open to exploring different styles within rock and blues.
- These insights can guide **cross-selling** by recommending albums from **Green Day** and **Foo Fighters** to customers purchasing from **Led Zeppelin** or vice versa, as well as suggesting both modern and classic artists together to appeal to customers with a broad taste in rock music.

```sql
WITH AlbumCoPurchase AS (
    -- Find all combinations of artists purchased in the same invoice
    SELECT
        il1.invoice_id,
        al1.title AS album_1,
        al2.title AS album_2,
        COUNT(*) AS co_purchase_count
    FROM
        invoice_line il1
    JOIN
        track t1 ON il1.track_id = t1.track_id
    JOIN
        album al1 ON t1.album_id = al1.album_id
    JOIN
        invoice_line il2 ON il1.invoice_id = il2.invoice_id
    JOIN
        track t2 ON il2.track_id = t2.track_id
    JOIN
        album al2 ON t2.album_id = al2.album_id
    WHERE
        il1.track_id <> il2.track_id AND al1.album_id <> al2.album_id
    GROUP BY
        al1.title, al2.title, il1.invoice_id
)
SELECT
    album_1,
    album_2,
    SUM(co_purchase_count) AS total_co_purchases
FROM
    AlbumCoPurchase
GROUP BY
    album_1, album_2
ORDER BY
    total_co_purchases DESC;
```

album_1	album_2	total_co_purchases
Are You Experienced?	Mezmerize	16
Mezmerize	Are You Experienced?	16
Vault: Def Leppard's Greatest Hits	Mezmerize	12
Mezmerize	Vault: Def Leppard's Greatest Hits	12
My Generation - The Very Best Of Th...	Mezmerize	12
Mezmerize	My Generation - The Very Best Of Th...	12
Mezmerize	The Police Greatest Hits	11
The Police Greatest Hits	Mezmerize	11
Dark Side Of The Moon	The Singles	11
The Singles	Dark Side Of The Moon	11
The Singles	My Generation - The Very Best Of Th...	11
My Generation - The Very Best Of Th...	The Singles	11
My Way: The Best Of Frank Sinatra [...	The Police Greatest Hits	10
The Police Greatest Hits	My Way: The Best Of Frank Sinatra [...	10
The Singles	From The Muddy Banks Of The Wish...	10
From The Muddy Banks Of The Wish...	The Singles	10
Mezmerize	Facelift	10
Facelift	Mezmerize	10
The Doors	Back to Black	10
Back to Black	The Doors	10
Get Born	Mezmerize	10
Mezmerize	Get Born	10
Big Ones	Jagged Little Pill	10
Jagged Little Pill	Big Ones	10
Faceless	My Generation - The Very Best Of Th...	10
My Generation - The Very Best Of Th...	Faceless	10
Big Ones	Back to Black	9

3. Top Co-Purchased Albums:

- **Are You Experienced?** (Jimi Hendrix) and **Mezmerize** (System of a Down): Co-purchased 16 times
- **Vault** (Def Leppard) and **Mezmerize**: Co-purchased 12 times
- **My Generation** (The Who) and **Mezmerize**: Co-purchased 12 times

Analysis:

- **Mezmerize** (a metal album by System of a Down) consistently appears with other classic rock albums such as **Are You Experienced?**, **Vault**, and **My Generation**. This indicates that customers who enjoy **heavy metal** are also interested in **classic rock**, which spans multiple decades of music.
- The consistent pairing of **Mezmerize** with different rock albums suggests an opportunity to cross-promote metal albums with rock albums, particularly targeting fans of **classic rock** who may also enjoy exploring heavier, modern genres like metal.

Recommendations for Product Strategy:

1. **Cross-Selling Based on Genre Affinity**:
    - **Metal and Rock** should be heavily promoted together. If a customer is purchasing from one genre, recommending products from the other could increase basket size.
    - Additionally, pairing **Rock** with **Alternative & Punk** or **Latin** genres could be effective, especially for customers with diverse music tastes.
2. **Artist-Based Recommendations**:
    - For customers buying from artists like **Led Zeppelin**, recommending other iconic rock bands like **Green Day** or **Foo Fighters** could be a natural fit, driving additional sales.
    - Customers who are interested in **Nirvana** or **Eric Clapton** should be targeted with promotions that highlight both **grunge** and **blues/rock** albums.
3. **Album Cross-Promotions**:
    - Given the strong co-purchase of **Mezmerize** with several classic rock albums, marketing campaigns could bundle metal albums with classic rock collections, offering special deals or curated playlists to appeal to these crossover interests.

- Albums like **Are You Experienced?**, **Vault**, and **My Generation** should be strategically promoted alongside **Mezmerize** to tap into this cross-genre interest.

Q5)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?

```sql
-- customer purchasing behaviour per billing country
WITH RegionCustomerBehavior AS (
    SELECT
        i.billing_country AS region,
        COUNT(DISTINCT i.customer_id) AS total_customers,
        COUNT(i.invoice_id) AS total_purchases,
        AVG(il.quantity) AS avg_basket_size,
        AVG(i.total) AS avg_spending
    FROM
        invoice i
    JOIN
        invoice_line il ON i.invoice_id = il.invoice_id
    GROUP BY
        i.billing_country
)
SELECT
    region,
    total_customers,
    total_purchases,
    avg_basket_size,
    avg_spending
FROM
    RegionCustomerBehavior
ORDER BY
    avg_spending DESC;
```

region	total_customers	total_purchases	avg_basket_size	avg_spending
Ireland	1	116	1.0000	12.357931
Czech Republic	2	276	1.0000	11.535652
Australia	1	82	1.0000	11.469512
Spain	1	99	1.0000	10.870000
Hungary	1	79	1.0000	10.514051
India	2	185	1.0000	10.205027
Germany	4	338	1.0000	10.181183
Canada	8	541	1.0000	10.147043
France	5	393	1.0000	10.109084
United Kingdom	3	248	1.0000	10.075645
Sweden	1	76	1.0000	9.926053
USA	13	1051	1.0000	9.900942
Argentina	1	40	1.0000	9.900000
Brazil	5	432	1.0000	9.395833
Portugal	2	187	1.0000	9.386471
Chile	1	98	1.0000	9.314082
Belgium	1	61	1.0000	9.299508
Austria	1	70	1.0000	9.277714
Italy	1	51	1.0000	9.181765
Norway	1	73	1.0000	9.099863
Poland	1	77	1.0000	8.961429
Finland	1	80	1.0000	8.563500
Netherlands	1	66	1.0000	8.250000
Denmark	1	38	1.0000	5.158421

```sql
WITH ChurnedCustomersByRegion AS (
    SELECT
        c.customer_id,
        i.billing_country AS region,
        MAX(i.invoice_date) AS last_purchase_date,
        CASE
            WHEN MAX(i.invoice_date) < DATE_SUB(CURDATE(), INTERVAL 1 YEAR) THEN 1
            ELSE 0
        END AS is_churned
    FROM
        customer c
    LEFT JOIN
        invoice i ON c.customer_id = i.customer_id
    GROUP BY
        c.customer_id, region
),
ChurnRateByRegion AS (
    SELECT
        region,
        COUNT(*) AS total_customers,
        SUM(is_churned) AS churned_customers,
        (SUM(is_churned) / COUNT(*)) * 100 AS churn_rate
    FROM
        ChurnedCustomersByRegion
    GROUP BY
        region
)
SELECT
    region,
    total_customers,
    churned_customers,
    churn_rate
FROM
    ChurnRateByRegion
ORDER BY
    churn_rate DESC;
```

region	total_customers	churned_customers	churn_rate
Brazil	5	5	100.0000
Germany	4	4	100.0000
Canada	8	8	100.0000
Norway	1	1	100.0000
Czech Republic	2	2	100.0000
Austria	1	1	100.0000
Belgium	1	1	100.0000
Denmark	1	1	100.0000
USA	13	13	100.0000
Portugal	2	2	100.0000
France	5	5	100.0000
Finland	1	1	100.0000
Hungary	1	1	100.0000
Ireland	1	1	100.0000
Italy	1	1	100.0000
Netherlands	1	1	100.0000
Poland	1	1	100.0000
Spain	1	1	100.0000
Sweden	1	1	100.0000
United Kingdom	3	3	100.0000
Australia	1	1	100.0000
Argentina	1	1	100.0000
Chile	1	1	100.0000
India	2	2	100.0000

1. **Customer Purchasing Behaviors Across Regions:**

- **Highest Number of Customers**:
  - **USA** has the highest number of customers (13), followed by **Canada** (8). This indicates that the store has the strongest presence in these two regions, with more active customers making purchases.
- **Total Purchases**:
  - The **USA** also leads in terms of total purchases (1,051), which suggests a high level of customer engagement in this region.
  - Interestingly, **Brazil** ranks second in total purchases (432), despite having fewer customers than **Canada**, indicating that Brazilian customers may have a higher purchase frequency, making it a potentially lucrative market.
- **Average Basket Size**:
  - All regions exhibit the same average basket size of **1**, which indicates that customers typically buy one item per transaction. This suggests that the store could explore opportunities to increase basket size by introducing **cross-selling** or **bundling** strategies across regions.
- **Average Spending**:
  - **Ireland** has the highest average spending, followed by **Czech Republic**, with **Denmark** having the lowest. This suggests that

customers in **Ireland** and **Czech Republic** are willing to spend more per transaction, possibly reflecting **stronger economic conditions** or **higher customer purchasing power** in these regions.
    - On the other hand, **Denmark's** low average spending indicates either a more price-sensitive market or fewer premium purchases.

2. Churn Rate Across Regions:

- **Churn Rate**:
    - The data reveals that all regions exhibit a **churn rate of 100%**, meaning that no customers have made a purchase in the last year or more.
    - This uniformly high churn rate suggests there may be broader issues affecting customer retention, such as lack of engagement, inadequate marketing, or failure to incentivize repeat purchases. The high churn could be attributed to various factors, including lack of **loyalty programs**, insufficient **regional marketing efforts**, or changing customer preferences.

3. Correlations and Insights:

- **USA** and **Canada**, with the highest number of customers, present key markets for customer retention strategies. Despite their strong customer base and purchase activity, both regions have a 100% churn rate, which indicates a **lack of customer loyalty** or **failure to bring back customers for repeat purchases**. This presents a clear opportunity to introduce **retention programs** such as **personalized offers**, **email marketing campaigns**, or **loyalty rewards**.
- **Brazil**, despite having fewer customers than Canada, exhibits a high total purchase count, suggesting that Brazilian customers are more engaged with frequent purchases. However, the 100% churn rate highlights the need to retain these highly active customers through **targeted regional promotions** or **exclusive offers**.
- **Ireland** and **Czech Republic** show the highest average spending, indicating that customers in these regions are willing to make higher-value purchases. Focusing on **premium offerings** and

exclusive deals in these markets could capitalize on the higher spending habits of customers in these regions.
- **Denmark**, with the lowest average spending, presents an opportunity to explore **budget-friendly options** or **price-sensitive promotions** that cater to the purchasing behavior of customers in this region.

4. Strategic Recommendations:

- **Retention Focus**: Given the high churn rates across all regions, the store should prioritize **retention strategies** such as:
    - Implementing **loyalty programs** that reward frequent purchases.
    - Offering **personalized discounts** to re-engage churned customers.
    - Sending **targeted email campaigns** reminding customers of promotions or new releases.
- **Regional Targeting**:
    - In **Ireland** and **Czech Republic**, leverage higher spending by promoting **premium products** and offering **exclusive packages**.
    - In **Brazil**, focus on keeping the existing customers engaged with **frequent promotions**, as their purchase frequency indicates a strong potential for continued engagement.
    - For **Denmark** and similar markets with lower spending, the store could experiment with **affordable bundles** or **seasonal discounts** to attract budget-conscious customers.
- **Increase Basket Size**: Since all regions report an average basket size of 1, there is significant potential to introduce **cross-selling** strategies. For example, recommend additional items during checkout or bundle complementary products to increase the basket size.

Conclusion:

The analysis shows that while some regions like the USA and Brazil lead in customer volume and purchase frequency, the universally high churn rate suggests a critical need for **improved retention strategies** across all regions. By tailoring strategies to individual markets based on spending behavior and customer engagement, the store can improve retention, increase basket sizes, and ultimately drive growth across diverse

geographic regions.

Q6)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?

```sql
WITH CustomerBehavior AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
        c.country AS customer_location,
        COUNT(i.invoice_id) AS total_purchases,
        SUM(i.total) AS total_spending,
        AVG(i.total) AS avg_spending_per_order,
        MAX(i.invoice_date) AS last_purchase_date,
        DATEDIFF(CURDATE(), MAX(i.invoice_date)) AS days_since_last_purchase
    FROM
        customer c
    JOIN
        invoice i ON c.customer_id = i.customer_id
    GROUP BY
        c.customer_id, customer_name, customer_location
),
```

```sql
CustomerRiskProfile AS (
    SELECT
        customer_id,
        customer_name,
        customer_location,
        total_purchases,
        total_spending,
        avg_spending_per_order,
        last_purchase_date,
        days_since_last_purchase,
        CASE
            WHEN days_since_last_purchase > 365 THEN 'High Risk'  -- Customer hasn't purchased in over 12 months
            WHEN total_purchases < 3 THEN 'Medium Risk'  -- Customer has made fewer than 3 purchases
            ELSE 'Low Risk'  -- Active and frequent customers
        END AS risk_category
    FROM
        CustomerBehavior
)
SELECT
    customer_id,
    customer_name,
    customer_location,
    total_purchases,
    total_spending,
    avg_spending_per_order,
    last_purchase_date,
    days_since_last_purchase,
    risk_category
FROM
    CustomerRiskProfile
ORDER BY
    risk_category DESC, days_since_last_purchase DESC;
```

```sql
WITH CustomerRiskProfileByDemographics AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
        c.country AS customer_location,
        SUM(i.total) AS total_spending,
        MAX(i.invoice_date) AS last_purchase_date,
        DATEDIFF(CURDATE(), MAX(i.invoice_date)) AS days_since_last_purchase,
        COUNT(i.invoice_id) AS total_purchases,
        CASE
            WHEN DATEDIFF(CURDATE(), MAX(i.invoice_date)) > 365 THEN 'High Risk'
            WHEN COUNT(i.invoice_id) < 3 THEN 'Medium Risk'
            ELSE 'Low Risk'
        END AS risk_category
    FROM
        customer c
    JOIN
        invoice i ON c.customer_id = i.customer_id
    GROUP BY
        c.customer_id, customer_name, customer_location
)
SELECT
    customer_location,
    risk_category,
    COUNT(customer_id) AS customer_count,
    AVG(total_spending) AS avg_spending_per_customer
FROM
    CustomerRiskProfileByDemographics
GROUP BY
    customer_location, risk_category
ORDER BY
    customer_count DESC;
```

Based on the query results, the customer base shows a universally high-risk profile, with all customers categorized as high-risk due to their inactivity. Here is a breakdown of the analysis and insights:

1. High Risk Across All Customers:

- All customers fall into the "High Risk" category, indicating that no purchases have been made in over a year. The minimum number of days since the last purchase is 1,392 days, and the maximum is 1,858 days, meaning customers have been inactive for a significant amount of time (almost 4-5 years).
- This suggests a critical problem with customer retention, as the entire customer base has disengaged and is at risk of complete churn.

2. Demographic Risk Breakdown:

- The second query, which groups customers by location and risk category, also shows that all countries are in the high-risk category. This suggests that the issue is not localized to one region but is prevalent across all geographic segments.
- Average spending per customer could indicate how valuable certain regions were in the past, but this information is overshadowed by the fact that all customers are now inactive.

3. Factors Contributing to Risk:

- Lack of Engagement: Customers have not made purchases in several years, suggesting the store has either failed to engage customers effectively through marketing, loyalty programs, or promotional efforts.
- No Recent Promotions: The absence of any recent purchase activity suggests that customers may not have been incentivized to return, possibly due to outdated product offerings, lack of promotions, or better alternatives elsewhere.
- Failure to Retain New Customers: The lack of any “Medium Risk” or “Low Risk” customers shows that the store may have also struggled to retain new customers. New customers typically should fall into a lower-risk category, but the absence of this tier points to a potential customer acquisition or onboarding issue.

4. Strategic Recommendations:

- Re-engagement Campaign: Since all customers are in the high-risk category, it is essential to launch an immediate re-engagement campaign to win back customers. This could include:
    - Personalized email campaigns offering discounts or special deals.
    - Introducing a loyalty program to incentivize return purchases.
    - Targeted ads on social media to remind customers of your brand and encourage them to come back.
- Revise Marketing Strategy: Explore why customers have become disengaged. This could include:
    - Conducting customer surveys to understand why they stopped purchasing.

    - Reviewing competitors to see if better deals or products are available elsewhere.
- Expand Product Offerings: Customers may have stopped purchasing due to outdated or irrelevant product offerings. Introducing new product lines or genres could help attract both new and returning customers.
- Segment Analysis: Although the risk profile shows all customers in the high-risk category, you could still perform a segmented analysis to prioritize re-engagement efforts based on:
    - Past spending patterns: Customers who spent more historically could be prioritized.
    - Geographic regions: Some countries might have greater potential for recovery if more targeted campaigns are deployed.
- Customer Retention Strategy: Long-term strategies should focus on preventing future churn through:
    - Regular follow-ups after a customer's first purchase to maintain engagement.
    - Incentives for returning customers, such as offering discounts after every few purchases to increase loyalty.

Conclusion:

The overall analysis reveals a critical need for an urgent focus on customer re-engagement. All customer segments are currently at high risk due to long periods of inactivity. Immediate corrective measures, including targeted campaigns, product refreshes, and loyalty programs, will be necessary to retain any remaining customer value and prevent complete churn across the customer base.

Q7)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?

```sql
WITH CustomerPurchaseHistory AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name,' ',c.last_name) as customer_name,
        c.country,
        COALESCE(c.state,'N.A') AS state,
        c.city,
        MIN(i.invoice_date) AS first_purchase_date,
        MAX(i.invoice_date) AS last_purchase_date,
        DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS tenure_days,
        COUNT(i.invoice_date) AS total_purchases,
        SUM(i.total) AS total_spent,
        AVG(i.total) AS avg_order_value,
        DATEDIFF(curdate(), MAX(i.invoice_date)) AS days_since_last_purchase
    FROM
        customer c
    JOIN
        invoice i ON i.customer_id = c.customer_id
    GROUP BY
        c.customer_id, customer_name
),
```

```sql
customer_life_time_analysis AS (
    SELECT
        customer_id,
        customer_name,
        country,
        state,
        city,
        tenure_days,
        total_purchases,
        total_spent,
        avg_order_value,
        CASE
            WHEN tenure_days >= 365 THEN 'Long-Term'
            ELSE 'Short-Term'
        END AS customer_segment,
        CASE
            WHEN last_purchase_date < DATE_SUB(CURDATE(), INTERVAL 1 YEAR) THEN 'Churned'
            ELSE 'Active'
        END AS customer_status,
        (total_spent / GREATEST(tenure_days, 1)) * 365 AS predicted_annual_value,
        total_spent AS lifetime_value
    FROM
        CustomerPurchaseHistory
),
```

```sql
segment_analysis AS (
    SELECT
        customer_segment,
        customer_status,
        COUNT(customer_id) AS num_customers,
        AVG(tenure_days) AS avg_tenure_days,
        AVG(total_spent) AS avg_lifetime_value,
        AVG(predicted_annual_value) AS avg_predicted_annual_value
    FROM
        customer_life_time_analysis
    GROUP BY
        customer_segment, customer_status
),
churn_analysis AS (
    SELECT
        country,
        state,
        city,
        customer_segment,
        COUNT(customer_id) AS churned_customers,
        AVG(total_spent) AS avg_lifetime_value
    FROM
        customer_life_time_analysis
    WHERE
        customer_status = 'Churned'
    GROUP BY
        country, state, city, customer_segment
)
```

```sql
SELECT
    *
FROM
    customer_life_time_analysis
ORDER BY
    lifetime_value DESC;

-- Segment Analysis
SELECT
    *
FROM
    segment_analysis
ORDER BY
    avg_lifetime_value DESC;

-- Churn Analysis
SELECT
    *
FROM
    churn_analysis
ORDER BY
    churned_customers DESC;
```

Based on the results of the query, here is an analysis addressing the question of leveraging customer data to predict lifetime value and understanding patterns among churned customers:

1. All Customers in the Long-Term Segment:

- According to your first query, all customers fall into the long-term segment (tenure ≥ 365 days). This indicates that the store has a historically loyal customer base, with no short-term customers in the dataset. However, this loyalty has not translated into active, repeat purchasing behavior, as all customers have eventually churned.

2. Churned Customers:

- From the segment analysis, it is clear that all customers have churned, which means that none of the long-term customers have made a purchase in over a year. Despite their tenure, they have disengaged from purchasing. This could indicate a lack of effective retention efforts or a disconnect between customer needs and product offerings.
- The segment analysis shows that for the long-term, churned customers, the average tenure was 1,244 days (about 3.4 years), indicating that they remained customers for a significant period before they stopped purchasing.
- The average lifetime value of these churned customers was 79.82, and the predicted annual value was 23.5, suggesting that their yearly contribution was relatively modest, but they accumulated a decent total over their tenure.

3. Churn Analysis by Geographic Location:

- In the churn analysis, all customers are reported as churned across different locations (countries, states, and cities). This widespread churn suggests that geographic factors such as regional economic conditions or cultural preferences may not have played a significant role in the customer churn. The lack of differentiation across locations makes it difficult to pinpoint geographic-specific strategies that could reduce churn.
- However, it would still be valuable to explore customer engagement strategies in regions where the customer base was previously larger

or generated more revenue, as they may represent higher recovery potential.

4. Common Characteristics or Patterns Among Churned Customers:

- Although no significant geographic or behavioral differences were observed in the churned customer data, a few patterns could be inferred:
    - Length of Tenure: Despite being in the long-term segment, all customers eventually churned, which may suggest that tenure alone is not a reliable predictor of ongoing engagement. This could imply that customer retention efforts need to be sustained even for long-term customers.
    - Average Spending: The average lifetime value of churned customers was relatively low (79.82), indicating that while customers stayed for a long time, their overall contribution was limited. The low predicted annual value (23.5) further suggests that spending was spread out over time without significant spikes.

5. Insights for Targeted Marketing and Loyalty Programs:

- Customer Segmentation: Since no customers are classified as short-term, it's clear that the focus should be on retaining long-term customers and re-engaging churned customers. Targeted loyalty programs, such as special offers for long-term customers who have not purchased recently, could help in recovering churned customers.
- Re-engagement Strategies: Given the widespread churn, re-engagement campaigns focused on high-tenure customers could offer incentives to make a return purchase. This could include personalized offers, reminders of past purchases, or loyalty rewards for those who return after a long hiatus.
- Predictive Retention Models: Since tenure alone doesn't predict churn or retention effectively, you might want to look at other indicators like basket size trends, spending spikes, or specific product engagement to identify at-risk customers earlier.

Q8)If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?

To measure the impact of promotional campaigns (such as discounts, events, or email marketing) on **customer acquisition, retention, and overall sales**, several key metrics and analytical approaches can be employed. Here's a breakdown of how this analysis could be done:

1. Customer Acquisition Impact:

- **New Customer Count**: Track the number of **new customers acquired** during and immediately after the promotional campaign compared to a baseline period (before the campaign). If the campaign is successful, you should see a noticeable increase in new customers.
- **Customer Acquisition Cost (CAC)**: Calculate the cost of acquiring each new customer by dividing the total promotional spend by the number of new customers. A decrease in CAC during the promotional period would suggest that the campaign is efficient in attracting new customers.
- **Channel Analysis**: For campaigns like email marketing, track which **channels** (e.g., social media ads, email, website) bring the most new customers. This helps in identifying the most effective campaign channels for acquisition.

Key Metrics:

- Number of new customers
- Customer acquisition cost (CAC)
- Conversion rate from campaign to purchase
- Traffic source breakdown (website, social, email)

2. Customer Retention Impact:

- **Repeat Purchase Rate**: Measure the **percentage of customers** who make a repeat purchase after the campaign. Comparing this to the repeat purchase rate before the campaign will indicate whether the promotion has increased retention.
- **Churn Rate**: Analyze whether customers who engaged with the campaign (used a discount or attended an event) have a **lower churn rate** compared to those who didn't. This helps determine if promotions encourage long-term loyalty or if customers only make one-time purchases due to discounts.
- **Customer Lifetime Value (CLV)**: Track the **CLV** of customers who participated in the promotional campaigns to see if their value increases compared to non-participants. This will show whether the promotion is driving sustainable engagement and profitability.

Key Metrics:

- Repeat purchase rate
- Churn rate (before and after the campaign)
- Customer lifetime value (CLV)
- Percentage of reactivated customers (customers who returned after the campaign)

3. Overall Sales Impact:

- **Sales Growth**: Compare total **sales revenue** during the promotional campaign period against the same period in previous years or months (before the campaign) to evaluate the direct impact on sales.
- **Average Order Value (AOV)**: Analyze how **AOV** changes during the promotion. Sometimes, discounts can increase the total basket size, but if AOV decreases too much due to heavy discounting, this can hurt profitability.
- **Promotion-Driven Sales**: Identify the **percentage of sales directly attributed** to the promotional campaign (e.g., customers using a discount code). This helps in assessing whether the sales increase is primarily due to the promotion or other factors.
- **Profit Margins**: Calculate the impact of the campaign on **profit margins**. Although sales might increase during a campaign, aggressive discounting may reduce overall profit. It's crucial to assess if the campaign balances revenue growth with profitability.

Key Metrics:

- Total sales revenue (campaign vs. non-campaign period)
- Sales uplift percentage
- Average order value (AOV)
- Promotion-attributed sales percentage
- Profit margin analysis

4. Engagement Metrics for Email and Events:

- **Email Open and Click-Through Rates (CTR)**: For email marketing campaigns, track the **open rate** and **click-through rate** to measure customer interest and engagement. A high CTR suggests that the campaign message is resonating well with recipients.
- **Event Attendance and Conversion**: For promotional events, measure the number of attendees and how many converted into paying customers. Track if those who attended events make larger purchases or return for additional purchases after the event.

Key Metrics:

- Email open rate
- Email click-through rate (CTR)
- Event attendance rate
- Post-event sales conversion rate

5. Customer Segmentation Analysis:

- **Segment Performance**: Break down the impact of promotions by **customer segments** (e.g., new vs. existing customers, high-value vs. low-value customers). This helps in understanding which segments respond better to promotions and can guide future targeting efforts.
- **Behavioral Patterns**: Analyze **purchase behavior** for different segments. For example, do customers who regularly respond to discounts tend to churn after the promotion ends, or do they stay engaged?

Key Metrics:

- Segment-based purchase rates
- Segment-based CLV and AOV
- Response rates for high-value vs. low-value customers

6. Time Series and A/B Testing:

- **Time Series Analysis**: Perform a **pre- and post-promotion** time series analysis to measure the impact of campaigns on sales, customer acquisition, and retention over time. Identify trends, spikes, or drops in sales and customer activity during and after the promotion.
- **A/B Testing**: Run **A/B tests** by offering the promotion to a subset of customers and comparing their behavior to those who didn't receive the promotion. This helps isolate the impact of the promotion and provides more robust conclusions about its effectiveness.

Key Metrics:

- Time series comparison (sales, churn, retention before vs. after)
- A/B test performance metrics (conversion rates, sales lift)

Conclusion:

By employing these various metrics and analytical techniques, you can gain a comprehensive view of how promotional campaigns impact customer acquisition, retention, and overall sales. This analysis can also guide future campaign strategies to optimize marketing spend, enhance customer loyalty, and drive sustainable revenue growth.

Q9)How would you approach this problem, if the objective and subjective questions weren't given?

To analyze music record sales data and provide actionable insights for the company's strategy in the physical music market, I would approach the problem systematically by following these key steps:

1. Understand the Business Objectives:

- **What is the company trying to achieve?**
    - Is the goal to increase physical record sales, reduce churn, target new customer segments, or optimize inventory?
    - Understanding the broader business goals helps tailor the analysis to provide the most relevant insights and recommendations.
- **What is the current state of the physical music market?**
    - Researching industry trends (e.g., whether physical formats like vinyl are making a comeback) would provide context for the data analysis.

2. Data Exploration and Cleaning:

- **Inspect the available data**: Understand the data structure and contents (e.g., sales data, customer demographics, artist/album details, store locations).
    - Common data types might include sales transactions, customer information, product details, and geographic data.
- **Check for missing, inconsistent, or outlier data**: Cleaning the data is crucial for accurate analysis. I would remove duplicates, handle missing values, and standardize fields (e.g., genres, artist names).

3. Key Metrics to Analyze:

- **Sales Trends**:
    - **Monthly/Yearly Sales**: Track sales trends over time to understand if there are any seasonal patterns or year-over-year

growth/decline.
    - **Format Preferences**: Analyze the breakdown of sales by physical formats (e.g., vinyl, CDs, cassettes) to see which formats are gaining or losing popularity.
- **Genre, Artist, and Album Popularity**:
    - **Best-Selling Genres/Artists/Albums**: Identify which genres, artists, and albums contribute most to sales. This can help in understanding customer preferences and making inventory or promotional decisions.
    - **Co-Purchase Behavior**: Analyze which albums or artists are frequently purchased together, which can inform cross-selling strategies (e.g., bundling or store promotions).
- **Geographic Insights**:
    - **Regional Sales Performance**: Break down sales by region or store location to identify areas where physical music is more popular. Are there cities or regions where certain formats, genres, or artists are performing better?
    - **Demographic Trends**: If customer demographic data is available, I would analyze how different age groups, income brackets, or genders influence physical record sales.
- **Customer Behavior**:
    - **Customer Segmentation**: Group customers based on their purchasing behavior (e.g., high spenders, genre-specific buyers, repeat customers, etc.). This helps in tailoring marketing strategies.
    - **Churn Analysis**: Analyze which customer segments are most likely to stop purchasing physical music and which are loyal. This can guide strategies for customer retention (e.g., loyalty programs, exclusive offers).

6. Generate Insights & Recommendations:

Based on the analysis, I would generate the following insights and recommendations:

- **Target High-Growth Segments**: Focus marketing efforts on the fastest-growing genres, artists, or formats (e.g., vinyl if that's trending

upward).

- **Geographical Targeting**: Invest in regions or stores where physical music sales are still strong. Consider opening pop-up stores or hosting events in cities where demand is high.
- **Optimize Inventory**: Ensure that inventory is balanced to meet demand, particularly for popular genres and artists. Use demand forecasts to avoid stockouts or overstock situations.
- **Cross-Selling Opportunities**: Leverage co-purchase data to create bundles or recommendations. For example, if customers frequently buy rock and metal albums together, cross-sell these at the point of sale.
- **Revive Churned Customers**: Identify customers who have stopped purchasing physical music and re-engage them through targeted promotions, discounts, or exclusive offerings.
- **Experiment with New Formats**: If the analysis shows growth in niche formats like vinyl or cassette, consider expanding the product line or offering limited editions to cater to collectors.

7. Visualization & Reporting:

- **Dashboards**: Create visual dashboards to track key metrics (e.g., sales trends, format breakdown, geographic performance, customer segmentation) for quick insights.
- **Reports**: Summarize the key findings and recommendations in a detailed report. This can include visualizations like sales heat maps, genre performance over time, and customer purchase patterns.

Conclusion:

By leveraging sales data, customer behavior,the company can gain deep insights into its physical music market. These insights will allow it to:

- **Optimize inventory and marketing efforts**,
- **Engage high-value customers**, and
- **Strategically invest in growth areas** (e.g., genres, artists, or regions).

This data-driven approach will help the company stay competitive and maximize revenue in the physical music space.

Q10)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?

```
ALTER TABLE album
ADD COLUMN release_year INT NULL;
```

```
ALTER TABLE album ADD COLUMN release_year INT NULL
```

```
0 row(s) affected Records: 0  Duplicates: 0  Warnings: 0
```

```
0.018 sec
```

Q11)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 a SQL query to provide this information.

```sql
WITH CustomerTotalPurchaseBehaviour AS (
    SELECT
        c.customer_id,
        c.country,
        SUM(i.total) as total_spent,
        COUNT(DISTINCT il.track_id) AS number_of_tracks_purchased
    FROM
        customer c
    JOIN
        invoice i ON i.customer_id = c.customer_id
    JOIN
        invoice_line il ON il.invoice_id = i.invoice_id
    GROUP BY
        c.customer_id, c.country
)
SELECT
    country,
    COUNT(DISTINCT customer_id) AS number_of_customers,
    ROUND(AVG(total_spent),2) AS average_total_amount_spent_by_customer,
    ROUND(AVG(number_of_tracks_purchased),2) AS average_number_of_tracks_purchased
FROM
    CustomerTotalPurchaseBehaviour
GROUP BY
    country
ORDER BY
    number_of_customers DESC;
```

country	number_of_customers	average_total_amount_spent_by_customer	average_number_of_tracks_purchased
USA	13	800.45	80.77
Canada	8	686.19	67.50
Brazil	5	811.80	85.80
France	5	794.57	78.40
Germany	4	860.31	84.25
United Kingdom	3	832.92	82.67
Czech Republic	2	1591.92	138.00
India	2	943.97	92.50
Portugal	2	877.64	93.00
Argentina	1	396.00	40.00
Australia	1	940.50	82.00
Austria	1	649.44	70.00
Belgium	1	567.27	61.00
Chile	1	912.78	98.00
Denmark	1	196.02	38.00
Finland	1	685.08	80.00
Hungary	1	830.61	79.00
Ireland	1	1433.52	115.00
Italy	1	468.27	51.00
Netherlands	1	544.50	65.00
Norway	1	664.29	73.00
Poland	1	690.03	75.00
Spain	1	1076.13	99.00
Sweden	1	754.38	76.00

- Average Spending: The average amount spent by customers varies across countries, with Czech Republic having the highest average spending per customer.

- Track Purchases: The average number of tracks purchased also shows significant variation, with countries like Czech Republic and Ireland showing higher averages, indicating potentially more engaged customer bases.
- Customer Count: The number of customers varies, with the USA having the highest count, which may reflect its larger market size.