Chinook Project

Objective Questions

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

Ans:

Query:

```
use chinook;

-- Missing Values or not

select * from Employee;

select * from customer;

select * from invoice;

select * from invoice_line;

select * from track;

select * from playlist_track;

select * from playlist;

select * from album;

select * from artist;

select * from media_type;

select * from genre;
```

Missing Values Found:

Customers Table:

- Fax, State, Postal Code, Company columns had missing values.

Employee Table:

- Reports_to : one NULL value for employee_id = 1.

Track Table:

- Composer column has missing values.

-- Dealing with missing values

```
SET SQL_SAFE_UPDATES = 0;

UPDATE customer

SET

   company = 'N/A'

WHERE company IS NULL;
```

```
UPDATE customer
SET
   state = 'N/A'
WHERE state IS NULL;
UPDATE customer
SET
   postal_code = 'N/A'
WHERE postal_code IS NULL;
UPDATE customer
SET
   fax = 'N/A'
WHERE fax IS NULL;
UPDATE track
SET composer = 'Unknown'
WHERE composer IS NULL;
UPDATE Employee
SET
   reports_to = 0
WHERE reports_to IS NULL;
```

Updating Missing Values Found:

Customers Table:

- Fax, State, Postal Code, Company columns with N/A Status

Employee Table:

- Reports_to : with 0

Track Table:

- Composer column with Unknown Status

```
-- EMPLOYEE: Duplicates
SELECT employee_id, COUNT(*) AS dup_count
FROM employee
GROUP BY employee_id
HAVING COUNT(*) > 1;
```

```
-- CUSTOMER: Duplicates
SELECT customer_id, COUNT(*) AS dup_count
FROM customer
GROUP BY customer_id
HAVING COUNT(*) > 1;
-- INVOICE: Duplicates
SELECT invoice_id, COUNT(*) AS dup_count
FROM invoice
GROUP BY invoice_id
HAVING COUNT(*) > 1;
-- INVOICE_LINE: Duplicates
SELECT invoice_line_id, COUNT(*) AS dup_count
FROM invoice_line
GROUP BY invoice_line_id
HAVING COUNT(*) > 1;
-- TRACK: Duplicates
SELECT track_id, COUNT(*) AS dup_count
FROM track
GROUP BY track_id
HAVING COUNT(*) > 1;
-- PLAYLIST: Duplicates
SELECT playlist_id, COUNT(*) AS dup_count
FROM playlist
GROUP BY playlist_id
HAVING COUNT(*) > 1;
-- PLAYLIST_TRACK: Duplicates
SELECT playlist_id, track_id, COUNT(*) AS dup_count
FROM playlist_track
GROUP BY playlist_id, track_id
HAVING COUNT(*) > 1;
```

-- ALBUM: Duplicates

```
SELECT album_id, COUNT(*) AS dup_count
FROM album
GROUP BY album_id
HAVING COUNT(*) > 1;
```

-- ARTIST: Duplicates

```
SELECT artist_id, COUNT(*) AS dup_count
FROM artist
GROUP BY artist_id
HAVING COUNT(*) > 1;
```

-- GENRE: Duplicates

```
SELECT genre_id, COUNT(*) AS dup_count
FROM genre
GROUP BY genre_id
HAVING COUNT(*) > 1;
```

-- MEDIA_TYPE: Duplicates

```
SELECT media_type_id, COUNT(*) AS dup_count
FROM media_type
GROUP BY media_type_id
HAVING COUNT(*) > 1;
```

No Duplicates found

employee_id	last_name	first_name	title	reports_to	birthdate	hire_date	address
1	Adams	Andrew	General Manager	NULL	1962-02-18 00:00:00	2016-08-14 00:00:00	11120 Jasper Ave NW
2	Edwards	Nancy	Sales Manager	1	1958-12-08 00:00:00	2016-05-01 00:00:00	825 8 Ave SW
3	Peacock	Jane	Sales Support Agent	2	1973-08-29 00:00:00	2017-04-01 00:00:00	1111 6 Ave SW
4	Park	Margaret	Sales Support Agent	2	1947-09-19 00:00:00	2017-05-03 00:00:00	683 10 Street SW
5	Johnson	Steve	Sales Support Agent	2	1965-03-03 00:00:00	2017-10-17 00:00:00	7727B 41 Ave
6	Mitchell	Michael	IT Manager	1	1973-07-01 00:00:00	2016-10-17 00:00:00	5827 Bowness Road NW
7	King	Robert	IT Staff	6	1970-05-29 00:00:00	2017-01-02 00:00:00	590 Columbia Boulevard West
8	Callahan	Laura	IT Staff	6	1968-01-09 00:00:00	2017-03-04 00:00:00	923 7 ST NW

customer_id	first_name	last_name	company	address	city	state	country
1	Luís	Gonçalves	Embraer - Empresa Brasileira de Aeronáutica S.A.	Av. Brigadeiro Faria Lima, 2170	São José dos Campos	SP	Brazil
2	Leonie	Köhler	N/A	Theodor-Heuss-Straße 34	Stuttgart	N/A	Germany
3	François	Tremblay	N/A	1498 rue Bélanger	Montréal	QC	Canada
4	Bjørn	Hansen	N/A	Ullevålsveien 14	Oslo	N/A	Norway
5	František	Wichterlová	JetBrains s.r.o.	Klanova 9/506	Prague	N/A	Czech Republic
6	Helena	Holý	N/A	Rilská 3174/6	Prague	N/A	Czech Republic
7	Astrid	Gruber	N/A	Rotenturmstraße 4, 1010 Innere Stadt	Vienne	N/A	Austria
8	Daan	Peeters	N/A	Grétrystraat 63	Brussels	N/A	Belgium
9	Kara	Nielsen	N/A	Sønder Boulevard 51	Copenhagen	N/A	Denmark
10	Eduardo	Martins	Woodstock Discos	Rua Dr. Falcão Filho, 155	São Paulo	SP	Brazil

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

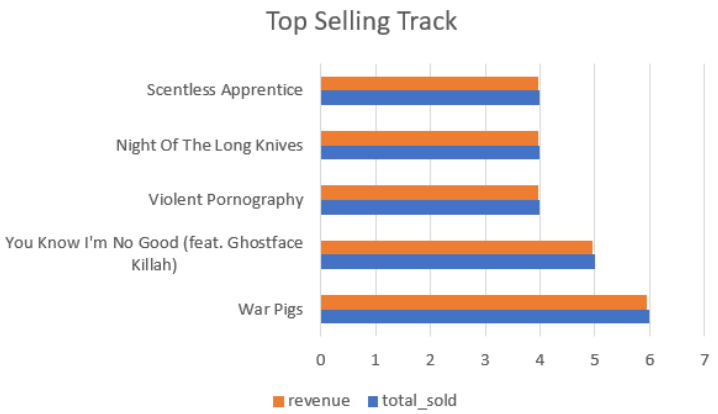
Ans:

Query:

```
-- Top selling tracks in USA
SELECT
    t.name AS track_name,
    a.title AS album_title,
    ar.name AS artist_name,
    g.name AS genre,
    SUM(il.quantity) AS total_sold,
    SUM(il.unit_price * il.quantity) AS revenue
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 a
ON t.album_id = a.album_id
JOIN artist ar
ON a.artist_id = ar.artist_id
JOIN genre g
ON t.genre_id = g.genre_id
WHERE i.billing_country = 'USA'
GROUP BY t.track_id, t.name, a.title, ar.name, g.name
ORDER BY revenue DESC
LIMIT 5;
```

Top selling tracks are:

track_name	album_title	artist_name	genre	total_sold	revenue
War Pigs	Cake: B-Sides and Rarities	Cake	Alternative	6	5.94
You Know I'm No Good (feat. Ghostface Killah)	Back to Black	Amy Winehouse	R&B/Soul	5	4.95
Violent Pornography	Mezmerize	System Of A Down	Metal	4	3.96
Night Of The Long Knives	For Those About To Rock We Salute You	AC/DC	Rock	4	3.96
Scentless Apprentice	From The Muddy Banks Of The Wishkah [live]	Nirvana	Rock	4	3.96



-- Top artist in USA with most famous genre

```
SELECT
   ar.name AS artist_name,
   g.name AS genre,
   SUM(il.unit_price * il.quantity) AS revenue
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 a ON t.album_id = a.album_id
JOIN artist ar ON a.artist_id = ar.artist_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE i.billing_country = 'USA'
GROUP BY ar.artist_id, ar.name, g.name
ORDER BY revenue DESC
LIMIT 1;
```

Top Artist with most famous genre:

artist_name	genre	revenue
Van Halen	Rock	42.57

Insights:

- Rock is leading in US sales
- War Pigs by Cake is best-selling track

Recommendations:

- Focus on Rock for inventory management
- Implement genre specific promotional strategies.

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

Ans:

Query:

-- Customers by Country

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

Customers by Country:

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

-- Customers by City (Top 10)

```
SELECT country, city, COUNT(*) AS customer_count
FROM customer
GROUP BY country, city
ORDER BY customer_count DESC
LIMIT 10;
```

Customers by City (Top 10):

country	city	customer_count
United Kingdom	London	2
France	Paris	2
Czech Republic	Prague	2
Brazil	São Paulo	2
Germany	Berlin	2
USA	Mountain View	2
Germany	Stuttgart	1
Norway	Oslo	1
Austria	Vienne	1
Belgium	Brussels	1

-- Customers by State

```
SELECT country, state, COUNT(*) AS customer_count
FROM customer
WHERE state IS NOT NULL
GROUP BY country, state
ORDER BY customer_count DESC;
```

Customers by State:

country	state	customer_count
France	N/A	5
Germany	N/A	4
Brazil	SP	3
USA	CA	3
United Kingdom	N/A	3
Czech Republic	N/A	2
Canada	ON	2
Portugal	N/A	2
India	N/A	2
Brazil	RJ	1
Brazil	DF	1
Canada	AB	1
Canada	BC	1
USA	WA	1
USA	NY	1
USA	NV	1
USA	FL	1
USA	MA	1
USA	IL	1
USA	WI	1

Insights:

- USA has the highest number of Customers

Recommendations:

- Prioritize US market due to its base.
- Identify growth opportunities in key countries like Canada, Brazil, France and Germany.

4. Calculate the total revenue and number of invoices for each country, state, and city?

Ans:

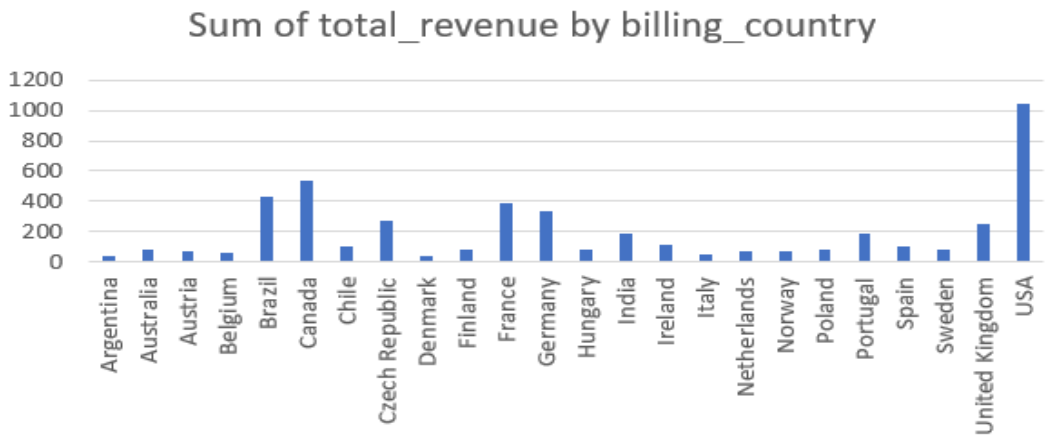
Query:

```
-- Total Revenue and Invoices based on Country, State and City

SELECT

    billing_country, billing_state, billing_city,

    SUM(total) AS total_revenue,

    COUNT(invoice_id) AS number_of_invoices

FROM invoice

GROUP BY

  billing_country, billing_state, billing_city

ORDER BY billing_country, billing_state, billing_city;
```

billing_country	billing_state	billing_city	total_revenue	number_of_invoices
Argentina	None	Buenos Aires	39.60	5
Australia	NSW	Sidney	81.18	10
Austria	None	Vienne	69.30	9
Belgium	None	Brussels	60.39	7
Brazil	DF	Brasîlia	106.92	15
Brazil	RJ	Rio de Janeiro	82.17	11
Brazil	SP	São José dos Campos	108.90	13
Brazil	SP	São Paulo	129.69	22
Canada	AB	Edmonton	29.70	10
Canada	BC	Vancouver	66.33	9
Canada	MB	Winnipeg	70.29	8
Canada	NS	Halifax	62.37	11
Canada	NT	Yellowknife	75.24	12
Canada	ON	Ottawa	91.08	13
Canada	ON	Toronto	40.59	4
Canada	QC	Montréal	99.99	9
Chile	None	Santiago	97.02	13
Czech Republic	None	Prague	273.24	30
Denmark	None	Copenhagen	37.62	10
Finland	None	Helsinki	79.20	11
France	None	Bordeaux	99.99	11
France	None	Dijon	73.26	12
France	None	Lyon	64.35	9

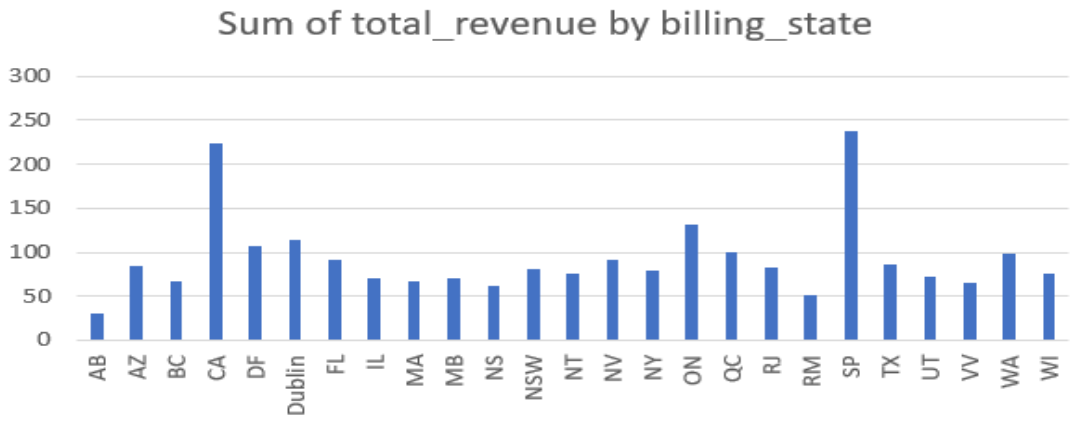
billing_country	billing_state	billing_city	total_revenue	number_of_invoices
France	None	Dijon	73.26	12
France	None	Lyon	64.35	9
France	None	Paris	151.47	18
Germany	None	Berlin	158.40	20
Germany	None	Frankfurt	94.05	10
Germany	None	Stuttgart	82.17	11
Hungary	None	Budapest	78.21	10
India	None	Bangalore	71.28	8
India	None	Delhi	111.87	13
Ireland	Dublin	Dublin	114.84	13
Italy	RM	Rome	50.49	9
Netherlands	VV	Amsterdam	65.34	10
Norway	None	Oslo	72.27	9
Poland	None	Warsaw	76.23	10
Portugal	None	Lisbon	102.96	13
Portugal	None	Porto	82.17	16
Spain	None	Madrid	98.01	11
Sweden	None	Stockholm	75.24	10
United Kingdom	None	Edinburgh	79.20	9
United Kingdom	None	London	166.32	19
USA	AZ	Tucson	84.15	9
USA	CA	Cupertino	54.45	9
USA	CA	Mountain View	169.29	20
USA	FL	Orlando	92.07	12
USA	IL	Chicago	71.28	8
USA	MA	Boston	66.33	10
USA	NV	Reno	91.08	11
USA	NY	New York	79.20	8
USA	TX	Fort Worth	86.13	12
USA	UT	Salt Lake City	72.27	10
USA	WA	Redmond	98.01	12
USA	WI	Madison	76.23	10

Sum of total_revenue



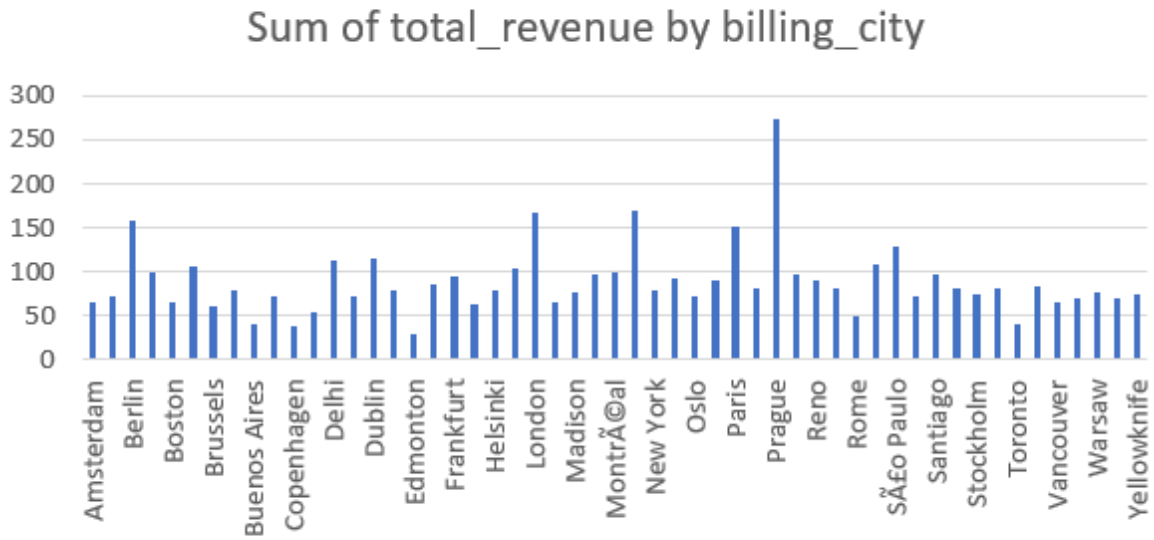
billing_country

Sum of total_revenue



billing_state

Sum of total_revenue



billing_city

Insights:

- Prague leads in revenue generation.

Recommendations:

- Analyse the factors behind Prague high revenue and adapt successful practices for other cities.

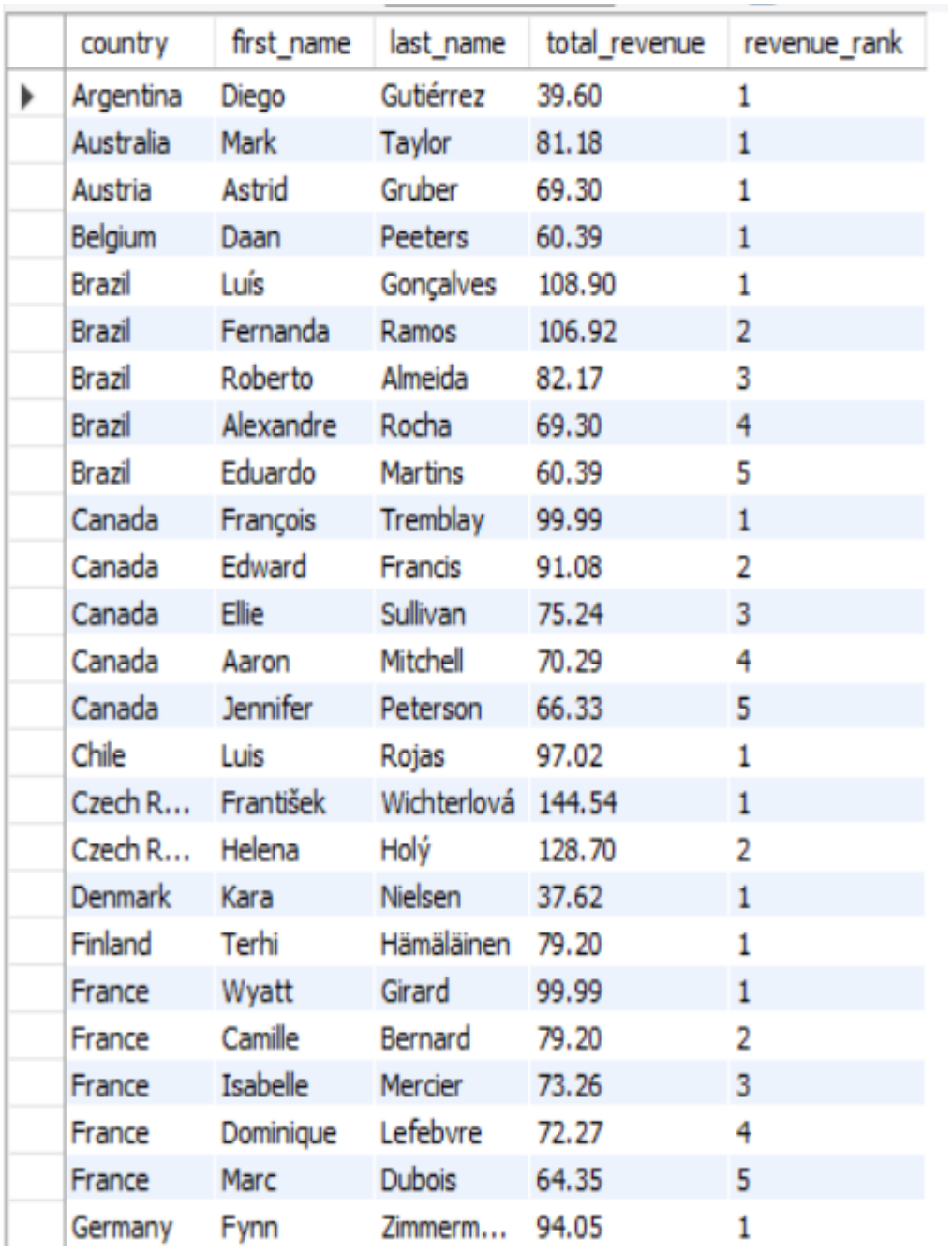
5. Find the top 5 customers by total revenue in each country?

Ans:

Query:

```
WITH CustomerRevenue AS (
  SELECT
    c.country,
    c.first_name,
    c.last_name,
    SUM(i.total) AS total_revenue,
    RANK() OVER (PARTITION BY c.country ORDER BY SUM(i.total) DESC) AS revenue_rank
  FROM
    customer AS c
  JOIN
    invoice AS i ON c.customer_id = i.customer_id
  GROUP BY
    c.country,
    c.first_name,
    c.last_name
)
SELECT
  country,
  first_name,
  last_name,
  total_revenue,
  revenue_rank
FROM
  CustomerRevenue
```

```
WHERE
    revenue_rank <= 5
ORDER BY
    country,
    revenue_rank;
```



country	first_name	last_name	total_revenue	revenue_rank
Argentina	Diego	Gutiérrez	39.60	1
Australia	Mark	Taylor	81.18	1
Austria	Astrid	Gruber	69.30	1
Belgium	Daan	Peeters	60.39	1
Brazil	Luís	Gonçalves	108.90	1
Brazil	Fernanda	Ramos	106.92	2
Brazil	Roberto	Almeida	82.17	3
Brazil	Alexandre	Rocha	69.30	4
Brazil	Eduardo	Martins	60.39	5
Canada	François	Tremblay	99.99	1
Canada	Edward	Francis	91.08	2
Canada	Ellie	Sullivan	75.24	3
Canada	Aaron	Mitchell	70.29	4
Canada	Jennifer	Peterson	66.33	5
Chile	Luis	Rojas	97.02	1
Czech R...	František	Wichterlová	144.54	1
Czech R...	Helena	Holý	128.70	2
Denmark	Kara	Nielsen	37.62	1
Finland	Terhi	Hämäläinen	79.20	1
France	Wyatt	Girard	99.99	1
France	Camille	Bernard	79.20	2
France	Isabelle	Mercier	73.26	3
France	Dominique	Lefebvre	72.27	4
France	Marc	Dubois	64.35	5
Germany	Fynn	Zimmerm...	94.05	1

country	first_name	last_name	total_revenue	revenue_rank
France	Marc	Dubois	64.35	5
Germany	Fynn	Zimmerm...	94.05	1
Germany	Hannah	Schneider	85.14	2
Germany	Leonie	Köhler	82.17	3
Germany	Niklas	Schröder	73.26	4
Hungary	Ladislav	Kovács	78.21	1
India	Manoj	Pareek	111.87	1
India	Puja	Srivastava	71.28	2
Ireland	Hugh	O'Reilly	114.84	1
Italy	Lucas	Mancini	50.49	1
Netherla...	Johannes	Van der B...	65.34	1
Norway	Bjørn	Hansen	72.27	1
Poland	Stanisław	Wójcik	76.23	1
Portugal	João	Fernandes	102.96	1
Portugal	Madalena	Sampaio	82.17	2
Spain	Enrique	Muñoz	98.01	1
Sweden	Joakim	Johansson	75.24	1
United Ki...	Phil	Hughes	98.01	1
United Ki...	Steve	Murray	79.20	2
United Ki...	Emma	Jones	68.31	3
USA	Jack	Smith	98.01	1
USA	Dan	Miller	95.04	2
USA	Heather	Leacock	92.07	3
USA	Kathy	Chase	91.08	4
USA	Richard	Cunningh...	86.13	5

Insights:

- **Top Performers by Country:**
  - i) František Wichterlová and Helena Holý are the top-performing clients in the Czech Republic, bringing in the most money (144.54 and 128.7, respectively).
  - ii) Brazil also has a high revenue, with Luís Gonçalves (108.9), Fernanda Ramos (106.92), and other prominent clients.
  - iii) Jack Smith (98.01), Dan Miller (95.04), and Heather Leacock (92.07) are among the top 5 in the USA, which displays a range of high-performing clients.
- **Geographic Trends:**
  - i) North America (USA, Canada) seems to have a lot of high-performing clients, with Edward Francis (91.08) and François Tremblay (99.99) from Canada holding the top spots.
  - ii) Notable performances include France, Germany, and India, demonstrating a wide range of global revenue.
- **Revenue Allocation:**
  - i) Argentina, Denmark, Italy, and Norway generate less revenue compared to other nations on the list.
  - ii) Countries such as Brazil, Czech Republic, and the USA display a more significant concentration of high revenue, indicating potential key markets.

Suggestions:

- **Concentrate on High Revenue Markets:**
  Focus on expanding into countries like the Czech Republic, Brazil, and the USA for future growth and high-value initiatives. The strong performance in these areas is noteworthy, and strategies can be customized to further leverage this advantage.
- **Customer Segmentation:**
  Analyze customers from Argentina, Denmark, and Italy, as their reduced revenue indicates opportunities for growth. Explore the factors influencing their performance and develop tailored marketing strategies to boost engagement and sales.
- **Utilize High Achievers for Marketing:**
  Employ top performers in countries such as the Czech Republic, Germany, and the USA as case studies or brand ambassadors to enhance reach and establish credibility in comparable markets.

6. Identify the top-selling track for each customer

Ans:

Query:

```
WITH RankedTracks AS (
    SELECT
        c.customer_id,
        c.first_name,
        c.last_name,
        t.name AS track_name,
        SUM(il.quantity) AS total_quantity_sold,
        ROW_NUMBER() OVER (PARTITION BY c.customer_id ORDER BY SUM(il.quantity) DESC)
AS rank_num
    FROM
        customer AS c
    JOIN
        invoice AS i ON c.customer_id = i.customer_id
    JOIN
        invoice_line AS il ON i.invoice_id = il.invoice_id
    JOIN
        track AS t ON il.track_id = t.track_id
    GROUP BY
```

```
        c.customer_id,
        c.first_name,
        c.last_name,
        t.name
)
SELECT
    customer_id,
    first_name,
    last_name,
    track_name,
    total_quantity_sold
FROM
    RankedTracks
WHERE
    rank_num = 1
ORDER BY
    customer_id;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_id	first_name	last_name	track_name	total_quantity_sold
1	Luís	Gonçalves	Put Your Lights On	1
2	Leonie	Köhler	Cochise	1
3	François	Tremblay	Sting Me	2
4	Bjørn	Hansen	Welcome to the Jungle	1
5	František	Wichterlová	Razor	1
6	Helena	Holý	Foxy Lady	1
7	Astrid	Gruber	Lady Double Dealer	1
8	Daan	Peeters	I Wish It Would Rain	1
9	Kara	Nielsen	Put The Finger On You	1
10	Eduardo	Martins	Like A Bird	2
11	Alexandre	Rocha	Battery	1
12	Roberto	Almeida	Love And Marriage	2
13	Fernanda	Ramos	24 Caprices, Op. 1, N...	2
14	Mark	Philips	Us And Them	1
15	Jennifer	Peterson	Hell Ain't A Bad Place ...	1
16	Frank	Harris	Too Young To Die	1
17	Jack	Smith	Fly Away	1
18	Michelle	Brooks	Right Next Door to Hell	1
19	Tim	Goyer	Spank Thru	1
20	Dan	Miller	I Like Dirt	1
21	Kathy	Chase	Foxy Lady	1
22	Heather	Leacock	Alberta	1
23	John	Gordon	Whatsername	1
24	Frank	Ralston	See You	1

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_id	first_name	last_name	track_name	total_quantity_sold
24	Frank	Ralston	See You	1
25	Victor	Stevens	Untitled	2
26	Richard	Cunningham	Wherever I Lay My Hat	1
27	Patrick	Gray	War Pigs	2
28	Julia	Barnett	Get What You Need	2
29	Robert	Brown	Children Of The Grave	1
30	Edward	Francis	Help Yourself	1
31	Martha	Silk	Boris The Spider	1
32	Aaron	Mitchell	Please Please Please	1
33	Ellie	Sullivan	Dude (Looks Like A La...	1
34	João	Fernandes	Train In Vain	2
35	Madalena	Sampaio	Highway Chile	1
36	Hannah	Schneider	I Can't Explain	2
37	Fynn	Zimmermann	Radio/Video	2
38	Niklas	Schröder	English Civil War	1
39	Camille	Bernard	It's Electric	1
40	Dominique	Lefebvre	Black Hole Sun	1
41	Marc	Dubois	Runnin' With The Devil	1
42	Wyatt	Girard	Changes	2
43	Isabelle	Mercier	Tease Me Please Me	2
44	Terhi	Hämäläinen	Baba O'Riley	1
45	Ladislav	Kovács	Up In Arms	1
46	Hugh	O'Reilly	N.I.B.	2
47	Lucas	Mancini	Love Is Strong	1

47	Lucas	Mancini	Love Is Strong	1
48	Johannes	Van der Berg	Confusion	2
49	Stanisław	Wójcik	Faceless	2
50	Enrique	Muñoz	Até Que Enfim Encont...	1
51	Joakim	Johansson	Stairway To Heaven	1
52	Emma	Jones	Speak To Me/Breathe	1
53	Phil	Hughes	For Those About To R...	1
54	Steve	Murray	Sparks Will Fly	1
55	Mark	Taylor	Mother's Little Helper	1
56	Diego	Gutiérrez	My Way	1
57	Luis	Rojas	Shape of Things to Co...	1
58	Manoj	Pareek	Dirty Little Thing	1
59	Puja	Srivastava	Time	1

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

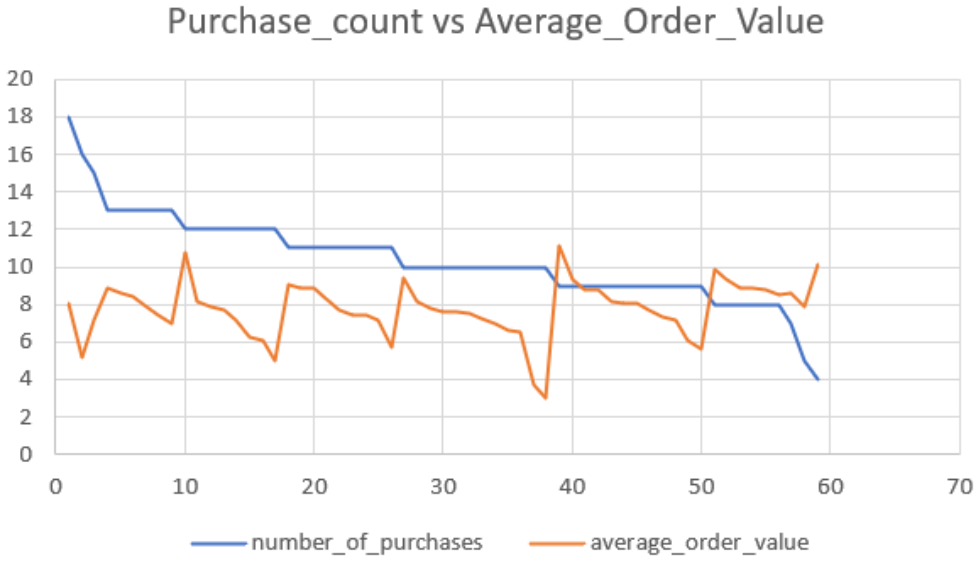
Ans:

Query:

```
SELECT
        c.first_name,
        c.last_name,
        COUNT(i.invoice_id) AS number_of_purchases,
        AVG(i.total) AS average_order_value
FROM
        customer AS c
JOIN
        invoice AS i ON c.customer_id = i.customer_id
GROUP BY
        c.customer_id
ORDER BY
        number_of_purchases DESC, average_order_value DESC;
```

first_name	last_name	number_of_purchases	average_order_value
František	Wichterlová	18	8.030000
Madalena	Sampaio	16	5.135625
Fernanda	Ramos	15	7.128000
Hugh	O'Reilly	13	8.833846
Manoj	Pareek	13	8.605385
Luís	Gonçalves	13	8.376923
João	Fernandes	13	7.920000
Luis	Rojas	13	7.463077
Edward	Francis	13	7.006154
Helena	Holý	12	10.725000
Jack	Smith	12	8.167500
Dan	Miller	12	7.920000
Heather	Leacock	12	7.672500
Richard	Cunningham	12	7.177500
Ellie	Sullivan	12	6.270000
Isabelle	Mercier	12	6.105000
Eduardo	Martins	12	5.032500
Wyatt	Girard	11	9.090000
Phil	Hughes	11	8.910000
Enrique	Muñoz	11	8.910000
Kathy	Chase	11	8.280000
Hannah	Schneider	11	7.740000
Leonie	Köhler	11	7.470000
Roberto	Almeida	11	7.470000
Terhi	Hämäläinen	11	7.200000

first_name	last_name	number_of_purchases	average_order_value
Terhi	Hämäläinen	11	7.200000
Martha	Silk	11	5.670000
Fynn	Zimmermann	10	9.405000
Mark	Taylor	10	8.118000
Ladislav	Kovács	10	7.821000
Stanisław	Wójcik	10	7.623000
Victor	Stevens	10	7.623000
Joakim	Johansson	10	7.524000
Julia	Barnett	10	7.227000
Alexandre	Rocha	10	6.930000
John	Gordon	10	6.633000
Johannes	Van der Berg	10	6.534000
Kara	Nielsen	10	3.762000
Mark	Philips	10	2.970000
François	Tremblay	9	11.110000
Patrick	Gray	9	9.350000
Camille	Bernard	9	8.800000
Steve	Murray	9	8.800000
Niklas	Schröder	9	8.140000
Bjørn	Hansen	9	8.030000
Dominique	Lefebvre	9	8.030000
Astrid	Gruber	9	7.700000
Jennifer	Peterson	9	7.370000
Marc	Dubois	9	7.150000
Tim	Goyer	9	6.050000
Lucas	Mancini	9	5.610000
Michelle	Brooks	8	9.900000
Frank	Harris	8	9.281250
Frank	Ralston	8	8.910000
Puja	Srivastava	8	8.910000
Aaron	Mitchell	8	8.786250
Emma	Jones	8	8.538750
Daan	Peeters	7	8.627143
Diego	Gutiérrez	5	7.920000
Robert	Brown	4	10.147500



These are some discernible patterns and trends in consumer purchase behaviour based on the data

Frequency of Purchases vs Average Order Value:

High-Frequency Customers: The average order value (about $8.03) is typically lower for customers who make the most purchases (e.g., František Wichterlová, with 18 orders). This implies that rather than purchasing entire albums, they are probably purchasing single tracks or small collections of songs.

High-Value Clients: Conversely, clients that make less purchases typically have larger average order values. Robert Brown, for instance, has an average order value of $10.15 despite only making four transactions, suggesting that he is more inclined to acquire full albums or more costly tracks in a single transaction.

This pattern is not absolute, as certain customers such as Helena Holý have a high purchase frequency (12 purchases) and as somewhat high average order value ($10.73). This indicates a valuable and loyal client who probably makes a combination of small and large purchases.

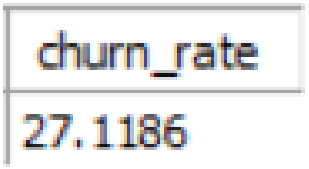
8. What is the customer churn rate?

Ans:

Query:

```
WITH last_purchase AS (
  SELECT
    customer_id,
    MAX(invoice_date) AS last_order_date
  FROM invoice
  GROUP BY customer_id
),
churned_customers AS (
  SELECT
    COUNT(customer_id) AS churned_count
  FROM last_purchase
  WHERE last_order_date < DATE_SUB(( SELECT max(invoice_date) from invoice),
INTERVAL 6 MONTH)
),
total_customers AS (
  SELECT COUNT(DISTINCT customer_id) AS total_count FROM customer
```

```
)
SELECT
   (c.churned_count / t.total_count) * 100 AS churn_rate
FROM churned_customers c, total_customers t;
```



churn_rate
27.1186

Explanation:

The churn rate of customers stands at 27.1186, based on the most recent invoice date over the last 6 months.

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

Ans:

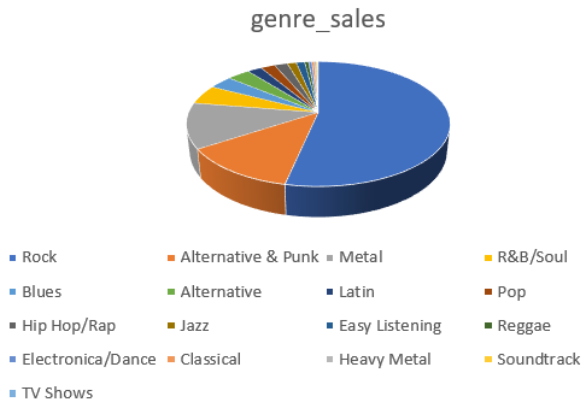
Query:

```
-- Genre Sales Analysis
SELECT
   G.name AS genre_name,
   SUM(IL.quantity * IL.unit_price) AS genre_sales,
   (
      SELECT SUM(IL2.quantity * IL2.unit_price)
      FROM invoice_line AS IL2
      JOIN invoice AS I2 ON IL2.invoice_id = I2.invoice_id
      JOIN customer AS C2 ON I2.customer_id = C2.customer_id
      WHERE C2.country = 'USA'
   ) AS total_usa_sales,
   (SUM(IL.quantity * IL.unit_price) / (
      SELECT SUM(IL3.quantity * IL3.unit_price)
      FROM invoice_line AS IL3
      JOIN invoice AS I3 ON IL3.invoice_id = I3.invoice_id
      JOIN customer AS C3 ON I3.customer_id = C3.customer_id
      WHERE C3.country = 'USA'
```

```
    )) * 100 AS percentage_of_total_sales
FROM
    invoice_line AS IL
JOIN
    track AS T ON IL.track_id = T.track_id
JOIN
    genre AS G ON T.genre_id = G.genre_id
JOIN
    invoice AS I ON IL.invoice_id = I.invoice_id
JOIN
    customer AS C ON I.customer_id = C.customer_id
WHERE
    C.country = 'USA'
GROUP BY
    G.name
ORDER BY
    percentage_of_total_sales DESC;
```

Genre Sales Analysis

genre_name	genre_sales	total_usa_sales	percentage_of_total_sales
Rock	555.39	1040.49	53.377735
Alternative & Punk	128.70	1040.49	12.369172
Metal	122.76	1040.49	11.798287
R&B/Soul	52.47	1040.49	5.042816
Blues	35.64	1040.49	3.425309
Alternative	34.65	1040.49	3.330162
Latin	21.78	1040.49	2.093245
Pop	21.78	1040.49	2.093245
Hip Hop/Rap	19.80	1040.49	1.902950
Jazz	13.86	1040.49	1.332065
Easy Listening	12.87	1040.49	1.236917
Reggae	5.94	1040.49	0.570885
Electronica/Dance	4.95	1040.49	0.475737
Classical	3.96	1040.49	0.380590
Heavy Metal	2.97	1040.49	0.285442
Soundtrack	1.98	1040.49	0.190295
TV Shows	0.99	1040.49	0.095147

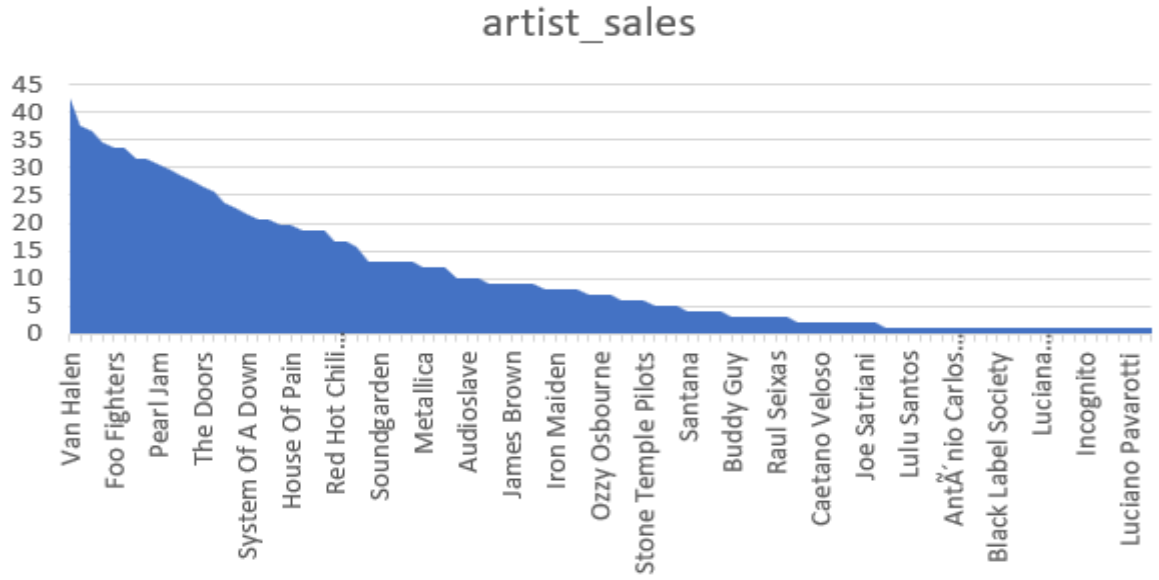


Query:

```
-- Best Selling Artists
SELECT
    A.name AS artist_name,
    SUM(IL.quantity * IL.unit_price) AS artist_sales
FROM
    invoice_line AS IL
JOIN
    track AS T ON IL.track_id = T.track_id
JOIN
    album AS AL ON T.album_id = AL.album_id
JOIN
    artist AS A ON AL.artist_id = A.artist_id
JOIN
    invoice AS I ON IL.invoice_id = I.invoice_id
JOIN
    customer AS C ON I.customer_id = C.customer_id
WHERE
    C.country = 'USA'
GROUP BY
    A.name
ORDER BY
    artist_sales DESC;
```

Best Selling Artists

artist_name	artist_sales	artist_name	artist_sale	artist_name	artist_sales
The Clash	12.87	UB40	5.94	Titãs	0.99
Soundgarden	12.87	Stone Temple Pilots	5.94	Jorge Ben	0.99
Jamiroquai	12.87	Motörhead	4.95	Lulu Santos	0.99
Skank	12.87	Mötley Crüe	4.95	The King's Singers	0.99
Frank Sinatra	12.87	Rush	4.95	Spyro Gyra	0.99
Metallica	11.88	Santana	3.96	Passengers	0.99
Deep Purple	11.88	Men At Work	3.96	Antônio Carlos Jo...	0.99
Creedence Clear...	11.88	Marillion	3.96	Funk Como Le Gu...	0.99
Def Leppard	9.90	Kiss	3.96	Gilberto Gil	0.99
Audioslave	9.90	Buddy Guy	2.97	Aisha Duo	0.99
Pink Floyd	9.90	Marisa Monte	2.97	Black Label Society	0.99
Velvet Revolver	8.91	Raimundos	2.97	Jota Quest	0.99
Miles Davis	8.91	Cássia Eller	2.97	Maurizio Pollini	0.99
James Brown	8.91	Raul Seixas	2.97	Body Count	0.99
The Police	8.91	Faith No More	2.97	Luciana Souza/Ro...	0.99
JET	8.91	Legião Urbana	1.98	Lost	0.99
Judas Priest	7.92	Billy Cobham	1.98	Karsh Kale	0.99
Iron Maiden	7.92	Caetano Veloso	1.98	Calexico	0.99
Alanis Morissette	7.92	Accept	1.98	Incognito	0.99
Apocalyptica	7.92	The Black Crowes	1.98	Bruce Dickinson	0.99
Lenny Kravitz	6.93	The Cult	1.98	Yo-Yo Ma	0.99
Ozzy Osbourne	6.93	Joe Satriani	1.98	Tim Maia	0.99
Aerosmith	6.93	Temple of the Dog	1.98	Luciano Pavarotti	0.99
Cake	5.94	Titãs	0.99	Vinícius De Moraes	0.99
UB40	5.94	Jorge Ben	0.99	Ed Motta	0.99



Insights:

- Rock genre was 53.38% of total, which is dominating.
- Rock, Alternative & Punk and Metal together contribute more than 77% of total with Rock being the standout genre
- A few like Classical, Heavy Metal and soundtrack and TV shows are Niche, they are collectively accounting for only 1.91% of total.

Recommendations:

- Focus on high popular genres for promoting or creating content to attract large audience.
- Try experimenting with Pop, Rap which have the potential but contribute less.

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

Ans:

Query:

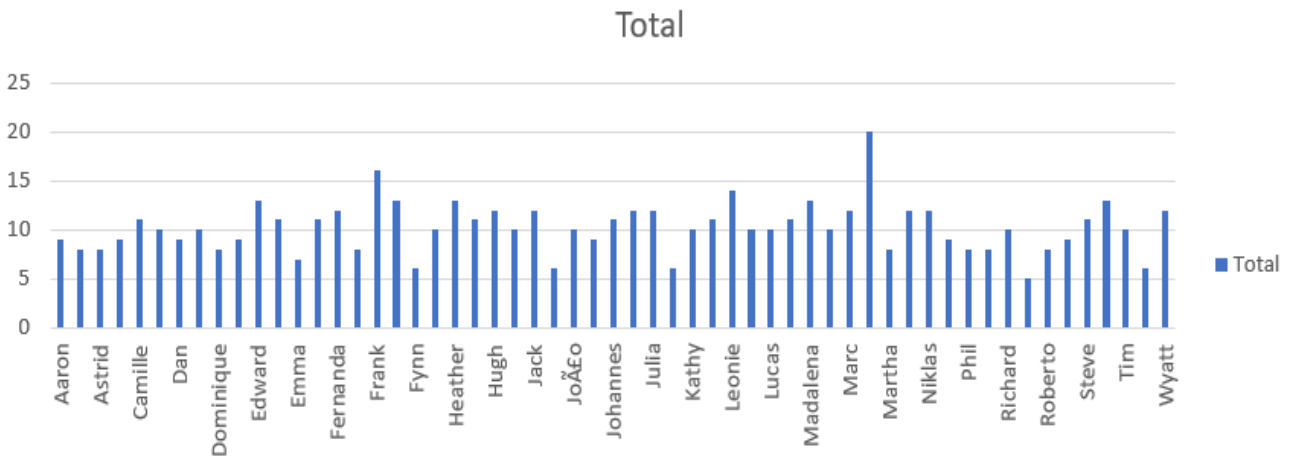
```
SELECT
    C.customer_id,
   c.first_name,
   c.last_name,
   COUNT(DISTINCT T.genre_id) AS number_of_genres
FROM
   invoice_line AS IL
JOIN
   invoice AS I ON IL.invoice_id = I.invoice_id
JOIN
   customer AS C ON I.customer_id = C.customer_id
JOIN
   track AS T ON IL.track_id = T.track_id
GROUP BY
   c.customer_id, c.first_name, c.last_name
HAVING
   COUNT(DISTINCT T.genre_id) >= 3;
```

first_name	last_name	number_of_genres
Luís	Gonçalves	10
Leonie	Köhler	14
François	Tremblay	8
Bjørn	Hansen	9
František	Wichterlová	13
Helena	Holý	11
Astrid	Gruber	8
Daan	Peeters	10
Kara	Nielsen	6
Eduardo	Martins	9
Alexandre	Rocha	8
Roberto	Almeida	8
Fernanda	Ramos	12
Mark	Philips	9
Jennifer	Peterson	6
Frank	Harris	8
Jack	Smith	12
Michelle	Brooks	12
Tim	Goyer	10
Dan	Miller	9
Kathy	Chase	10
Heather	Leacock	13
John	Gordon	12
Frank	Ralston	8
Victor	Stevens	6

first_name	last_name	number_of_genres
Victor	Stevens	6
Richard	Cunningham	10
Patrick	Gray	9
Julia	Barnett	12
Robert	Brown	5
Edward	Francis	13
Martha	Silk	8
Aaron	Mitchell	9
Ellie	Sullivan	11
João	Fernandes	10
Madalena	Sampaio	13
Hannah	Schneider	10
Fynn	Zimmermann	6
Niklas	Schröder	12
Camille	Bernard	11
Dominique	Lefebvre	8
Marc	Dubois	12
Wyatt	Girard	12
Isabelle	Mercier	10
Terhi	Hämäläinen	13
Ladislav	Kovács	11
Hugh	O'Reilly	12
Lucas	Mancini	10
Johannes	Van der Berg	11
Stanisław	Wójcik	9

Enrique	Muñoz	11
Joakim	Johansson	9
Emma	Jones	7
Phil	Hughes	8
Steve	Murray	11
Mark	Taylor	11
Diego	Gutiérrez	10
Luis	Rojas	11
Manoj	Pareek	10
Puja	Srivastava	8

Sum of number_of_genres



first_name

Insights:

- Individuals like Leonie, Edward, Madalena, Frantisek, heather and terhi, show strong engagement in genre.
- Consistent people like Fernanda, Niklas, Michele and others show high level of participation.
- Aaron, Eduardo, Dan and others had the lower engagement.

Recommendations:

- Recognize top contributors and make them to further improve their involvement.
- Investigate why few engagements are very less and analyse the genre preferences.

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

Ans:

Query:

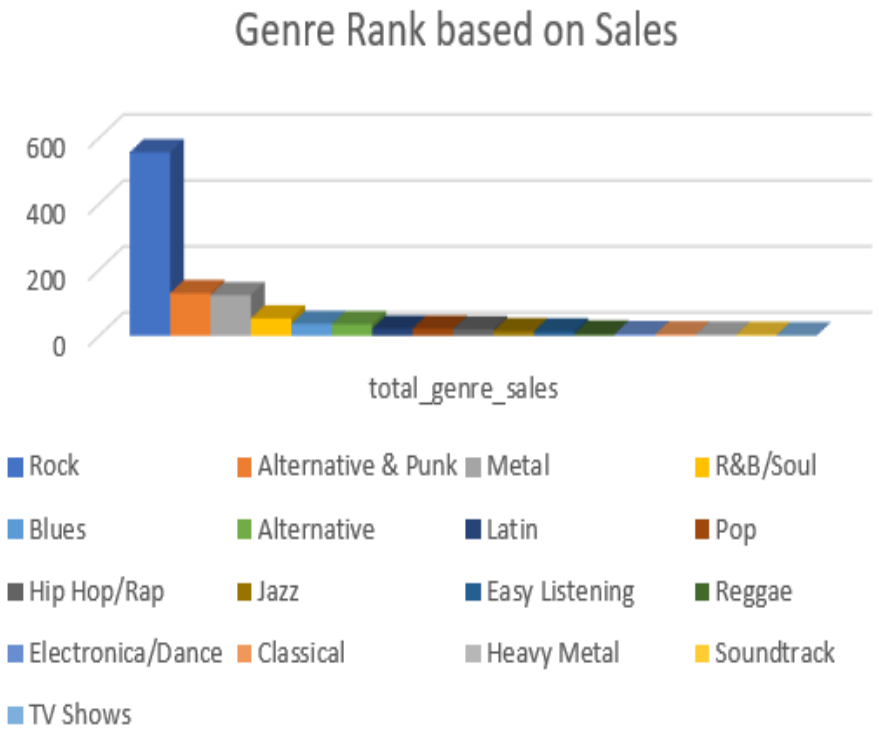
```
WITH GenreSales AS (
  SELECT
    G.name AS genre_name,
    SUM(IL.quantity * IL.unit_price) AS total_genre_sales
  FROM
    invoice_line AS IL
  JOIN
    track AS T ON IL.track_id = T.track_id
  JOIN
    genre AS G ON T.genre_id = G.genre_id
  JOIN
    invoice AS I ON IL.invoice_id = I.invoice_id
  WHERE
    I.billing_country = 'USA'
  GROUP BY
    G.name
)
SELECT
  genre_name,
  total_genre_sales,
  RANK() OVER (ORDER BY total_genre_sales DESC) AS genre_rank
```

```
FROM
    GenreSales
ORDER BY
    genre_rank;
```

genre_name	total_genre_sales	genre_rank
Rock	555.39	1
Alternative & Punk	128.70	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	19.80	9
Jazz	13.86	10
Easy Listening	12.87	11
Reggae	5.94	12
Electronica/Dance	4.95	13
Classical	3.96	14
Heavy Metal	2.97	15
Soundtrack	1.98	16
TV Shows	0.99	17

Insights:

- Rock leads the significant margin and Alternative & Punk and Metal follow closely, ranking 2nd and 3rd respectively.
- TV Shows has lowest genre with other less popular genres like classical and soundtrack also falling at the lower end of the ranking.



Recommendations:

- Focus on High performing genres for further development and engagement.
- For Niche genres, make some efforts to grow their customer base.

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

Ans:

Query:

```
WITH latest_purchase as
(
SELECT max(invoice_date) as latest_purchase
FROM invoice
),
customer_latest_order as
(
SELECT customer_id, max(invoice_date) as customer_latest_purchase
FROM invoice
GROUP BY customer_id
),
3_months_difference as
(
SELECT DATE_SUB(latest_purchase, INTERVAL 3 MONTH) as last_three_months
FROM latest_purchase
)
SELECT c.customer_id, c.first_name, c.last_name, c.country
FROM customer c
JOIN customer_latest_order clo
ON c.customer_id = clo.customer_id
CROSS JOIN 3_months_difference tmd
WHERE clo.customer_latest_purchase < tmd.last_three_months;
```

customer_id	first_name	last_name	country
18	Michelle	Brooks	USA
38	Niklas	Schröder	Germany
1	Luís	Gonçalves	Brazil
10	Eduardo	Martins	Brazil
43	Isabelle	Mercier	France
9	Kara	Nielsen	Denmark
4	Bjørn	Hansen	Norway
39	Camille	Bernard	France
3	François	Tremblay	Canada
58	Manoj	Pareek	India
37	Fynn	Zimmerm...	Germany
50	Enrique	Muñoz	Spain
19	Tim	Goyer	USA
36	Hannah	Schneider	Germany
57	Luis	Rojas	Chile
48	Johannes	Van der B...	Netherl...
8	Daan	Peeters	Belgium
17	Jack	Smith	USA
54	Steve	Murray	United K...
7	Astrid	Gruber	Austria
11	Alexandre	Rocha	Brazil
56	Diego	Gutiérrez	Argentina

People who not purchased in past 3 months

Insights:

- A total of 22 customers has not made any purchase in the last 3 months.
- Inactivity is spread across regions, not limited to a single country - showing a global decline in recent purchasing activity.
- European customers (Germany, France, Norway, Austria, Netherlands, Belgium, Spain, UK, Denmark) make up 68% (approx.) of the inactive list.

Recommendations:

- Send personalized reactivation mails to inactive customers.
- Offer discounts or loyalty points on their next purchase within a limited time.
- Use past data to recommend related artists or genre.

Subjective Questions

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

Ans:

Approach:

i) Find the best-selling genres in USA
ii) Identify best-selling albums within top genres.
iii) Recommend the top 3 albums.

Query:

```
-- Best_Selling Genres in the USA

SELECT

   g.name AS genre_name,

   SUM(il.quantity * il.unit_price) AS total_sales

FROM

   invoice_line AS il

JOIN

   track AS t ON il.track_id = t.track_id

JOIN

   genre AS g ON t.genre_id = g.genre_id

JOIN

   invoice AS i ON il.invoice_id = i.invoice_id

WHERE

   i.billing_country = 'USA'

GROUP BY

   g.name

ORDER BY

   total_sales DESC

LIMIT 3;

-- Best_Selling Albums within Top Genres
```

```
SELECT
    a.title AS album_title,
    ar.name AS artist_name,
    SUM(il.quantity * il.unit_price) AS total_sales
FROM
    invoice_line AS il
JOIN
    track AS t ON il.track_id = t.track_id
JOIN
    genre AS g ON t.genre_id = g.genre_id
JOIN
    invoice AS i ON il.invoice_id = i.invoice_id
JOIN
    album AS a ON t.album_id = a.album_id
JOIN
    artist AS ar ON a.artist_id = ar.artist_id
WHERE
    i.billing_country = 'USA' AND g.name IN ('Rock', 'Alternative & Punk', 'Metal')
GROUP BY
    a.title, ar.name
ORDER BY
    total_sales DESC
LIMIT 3;
```

Step 1: Best-Selling Genres in the USA

First, the top-selling genres in the United States were identified by calculating the total sales for each genre.

genre_name	total_sales
Rock	555.39
Alternative & Punk	128.70
Metal	122.76

These three genres represent the strongest market for music sales.

Step 2: Best-Selling Albums within Top Genres

Next, we need to find the top-selling albums specifically within those high-performing genres. This ensures that the recommended albums align with the most popular and profitable music categories.

Query uses the same joins but adds a WHERE clause to focus only on the top genres. It then groups the sales data by album_title and artist_name to find the total sales for each album. The results are ordered to reveal the best-selling albums, which are the ones recommended for promotion.

album_title	artist_name	total_sales
Are You Experienced?	Jimi Hendrix	27.72
From The Muddy Banks Of The Wishkah [live]	Nirvana	27.72
The Doors	The Doors	26.73

Insights:

- Rock dominates the strong presence.
- Albums like “Are you Experienced?” and “From the Muddy Banks of the WishKah [Live]” are contributing significantly to sales, Each with a total_sale of 27.72.
- “The Doors” is also having the strong presence with their notable sales.

Recommendations:

- “Are you Experienced?” and “From the Muddy Banks of the WishKah [Live]” are top performers and should be heavily promoted based on its strong sales and popularity.
- Alternative & Punk, should be considered for targeted campaigns to reach fans, as these albums show strong performance.

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

Ans:

Approach:

i) Find sales by genre and country excluding USA.
ii) Identify top selling genre within each country.
iii) Identify patterns across each countries.

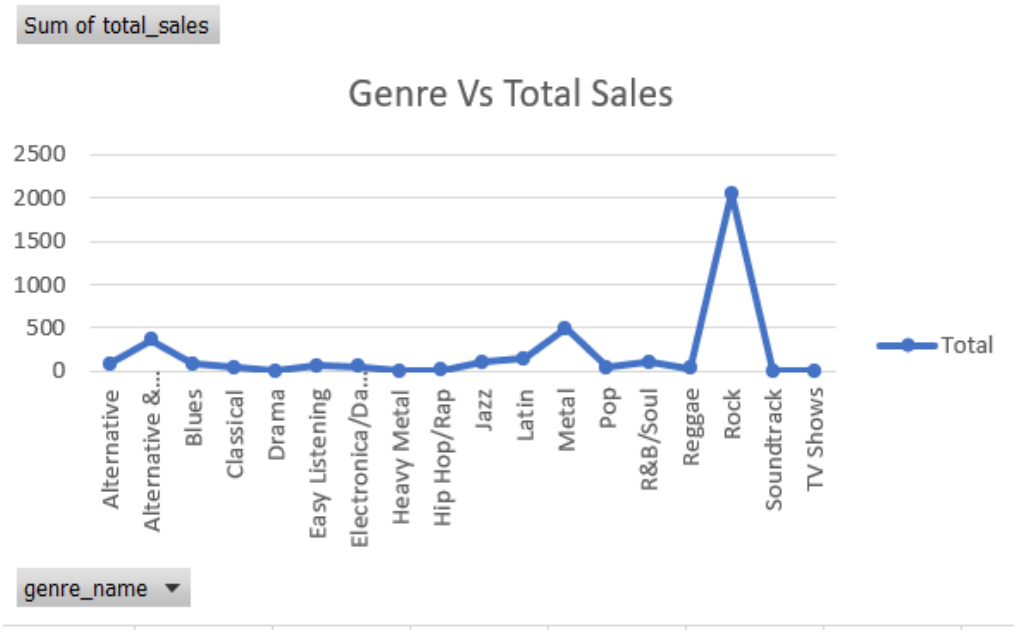
Query:

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

billing_country	genre_name	total_sales
Canada	Rock	329.67
France	Rock	208.89
Brazil	Rock	202.95
Germany	Rock	192.06
United Kingdom	Rock	164.34
Czech Republic	Rock	141.57
Portugal	Rock	106.92
India	Rock	100.98
Brazil	Alternative & Punk	73.26
Brazil	Metal	72.27
Canada	Metal	71.28
Ireland	Rock	71.28
Chile	Rock	60.39
Sweden	Rock	59.40
France	Metal	53.46
Finland	Rock	45.54
Spain	Rock	45.54
Germany	Metal	43.56
Hungary	Rock	43.56
Czech Republic	Alternative & Punk	41.58
Austria	Rock	39.60
Norway	Rock	39.60
Poland	Rock	39.60
Italy	Rock	34.65
Australia	Rock	33.66

billing_country	genre_name	total_sales
Australia	Rock	33.66
Netherlands	Rock	32.67
Canada	Alternative & Punk	30.69
United Kingdom	Metal	30.69
Canada	R&B/Soul	28.71
France	Easy Listening	28.71
Portugal	Metal	27.72
France	Latin	26.73
Belgium	Rock	25.74
India	Alternative & Punk	24.75
Denmark	Rock	23.76
France	Alternative & Punk	23.76
United Kingdom	Alternative & Punk	23.76
Czech Republic	Metal	22.77
Australia	Alternative & Punk	21.78
Canada	Alternative	21.78
India	Metal	21.78
Czech Republic	Latin	20.79
Brazil	Jazz	19.80
Ireland	Latin	19.80
Belgium	Metal	18.81
Hungary	Metal	18.81
India	Electronica/Dance	18.81
Czech Republic	Blues	17.82
Germany	Jazz	17.82

Only a few Shown in the above result, Entire output is in MySQL workbench.



After running the query, I found the following trends in genre popularity:

Insights:

Commonalities

- **Dominance of Rock:** Rock is the most popular genre in the USA, and it is also likely to be the top-selling genre in many other countries. This is because it is a broad category and a fundamental genre in the music industry.

- **Emergence of Mainstream Jazz:** Latin and Jazz genres are expected to show strong sales in many countries, reflecting their global appeal.

Differences

- **Regional Preferences:** R&B/Soul and Blues are sold more in the USA, while Latin and Jazz are sold most outside of USA

- **Market Size Discrepancies:** The total sales for a country like Canada or Brazil will likely be higher than those for smaller markets like the Netherlands or Finland. While the top-selling genre might be the same, the sheer volume of sales can vary significantly.

Recommendations:

- Focus on Rock, maintain a strong global strategy to maintain it.
- Adapt Inventory Management and promotions to cater specific genre for different regions.
- Promote Latin, Jazz in markets outside USA where these genres are gaining popularity.

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

Ans:

Approach:

i) Classify customers into long-term vs short-term based on active days.
ii) Compare their average spend, basket size, and frequency.
iii) Interpret differences to guide loyalty and retention strategies.

Query:

```
WITH CTE AS (

        SELECT

          i.customer_id,

          MAX(invoice_date) AS last_purchase_date,

          MIN(invoice_date) AS first_purchase_date,
```

```
    SUM(total) AS total_spent,
    SUM(quantity) AS items_bought,
    COUNT(i.customer_id) AS frequency,
    ABS(TIMESTAMPDIFF(DAY, MAX(invoice_date), MIN(invoice_date))) AS customer_since_days
  FROM invoice i
  LEFT JOIN invoice_line il ON il.invoice_id = i.invoice_id
  LEFT JOIN customer c ON c.customer_id = i.customer_id
  GROUP BY i.customer_id
),
long_short_term AS (
  SELECT
    total_spent,
    items_bought,
    frequency,
    CASE
      WHEN customer_since_days > (SELECT AVG(customer_since_days) AS average_days FROM cte)
        THEN 'Long Term'
        ELSE 'Short Term'
    END AS term
  FROM cte
)
SELECT
  term,
  AVG(total_spent) AS aveerage_spending,
  AVG(items_bought) AS average_basket_size,
  COUNT(frequency) AS number_of_customers
```

```
FROM long_short_term

GROUP BY term;
```

term	average_spending	average_basket_size	number_of_customers
Short Term	721.050000	73.8889	27
Long Term	876.088125	86.3125	32

Insights:

- Long term customers have spent ($28034.82) significantly more compared to short term customers ($19468.35) suggests long term customers are more likely to make larger purchases over time.
- Long term customers bought more items (2762) compared to (1995) by short term customers, indicates high frequency of purchases for long term customers.

Recommendations:

- Create loyalty programs for long term customers, offer exclusive discounts based on continuous purchasing behaviour.
- Implement personalized offers and promotions aimed at increasing their basket size.
- Deliver a seamless and personalized shopping experience to keep customers engaged, tailored recommendations could help in boosting retention of both short-term and long-term customers.

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:

Approach:

i) Group purchases by invoice to identify items frequently bought together (genres, artists, albums).
ii) Check pairings — different genres (cross-taste), same/different artists (loyalty vs cross-selling), different albums (variety).
iii) Use insights to design personalized recommendations and targeted cross-selling strategies.

Query:

-- Finding Co-purchased Genres (Different)

```
SELECT

  g1.name AS genre_1,

  g2.name AS genre_2,

  COUNT(*) AS times_bought_together
```

```
FROM invoice_line il1
JOIN invoice_line il2
    ON il1.invoice_id = il2.invoice_id
   AND il1.track_id < il2.track_id
JOIN track t1 ON il1.track_id = t1.track_id
JOIN track t2 ON il2.track_id = t2.track_id
JOIN genre g1 ON t1.genre_id = g1.genre_id
JOIN genre g2 ON t2.genre_id = g2.genre_id
WHERE g1.genre_id <> g2.genre_id
GROUP BY g1.name, g2.name
ORDER BY times_bought_together DESC
LIMIT 10;
```

genre_1	genre_2	times_bought_together
Metal	Rock	1002
Alternative & Punk	Rock	638
Rock	Metal	620
Rock	Alternative & Punk	418
Rock	Alternative	309
Latin	Rock	251
Rock	R&B/Soul	232
Blues	Rock	181
Metal	Alternative & Punk	177
Rock	Latin	176

-- Finding Co-purchased Artist (Same or different):

```
SELECT
    ar1.name AS artist_1,
    ar2.name AS artist_2,
    COUNT(*) AS times_bought_together
FROM invoice_line il1
JOIN invoice_line il2
    ON il1.invoice_id = il2.invoice_id
```

```
    AND il1.track_id < il2.track_id
JOIN track t1 ON il1.track_id = t1.track_id
JOIN track t2 ON il2.track_id = t2.track_id
JOIN album al1 ON t1.album_id = al1.album_id
JOIN album al2 ON t2.album_id = al2.album_id
JOIN artist ar1 ON al1.artist_id = ar1.artist_id
JOIN artist ar2 ON al2.artist_id = ar2.artist_id
GROUP BY ar1.name, ar2.name
ORDER BY times_bought_together DESC
LIMIT 10;
```

artist_1	artist_2	times_bought_together
Jimi Hendrix	Jimi Hendrix	1059
Queen	Queen	799
Red Hot Chili Peppers	Red Hot Chili Peppers	454
Pearl Jam	Pearl Jam	416
Nirvana	Nirvana	378
The Cult	The Cult	372
Guns N' Roses	Guns N' Roses	348
JET	JET	313
Marisa Monte	Marisa Monte	306
Kiss	Kiss	296

-- Finding Co-purchased Albums

```
SELECT
    al1.title AS album_1,
    al2.title AS album_2,
    COUNT(*) AS times_bought_together
FROM invoice_line il1
JOIN invoice_line il2
    ON il1.invoice_id = il2.invoice_id
   AND il1.track_id < il2.track_id
JOIN track t1 ON il1.track_id = t1.track_id
```

```
JOIN track t2 ON il2.track_id = t2.track_id

JOIN album al1 ON t1.album_id = al1.album_id

JOIN album al2 ON t2.album_id = al2.album_id

WHERE al1.album_id <> al2.album_id

GROUP BY al1.title, al2.title

ORDER BY times_bought_together DESC

LIMIT 10;
```

album_1	album_2	times_bought_together
Are You Experienced?	Mezmerize	16
Vault: Def Leppard's Greatest Hits	Mezmerize	12
Mezmerize	My Generation - The Very Best Of The Who	12
Dark Side Of The Moon	The Singles	11
The Singles	My Generation - The Very Best Of The Who	11
Mezmerize	The Police Greatest Hits	11
Get Born	Mezmerize	10
Facelift	Mezmerize	10
From The Muddy Banks Of The Wishkah [live]	The Singles	10
Big Ones	Jagged Little Pill	10

Insights:

- Most Commonly purchased genre combinations include:
    i) Rock and Metal
    ii) Alternative & Punk and Rock
    iii) Rock and R&B/Soul
    iv) Blues & Rock
- Most Commonly preferred Cross-Genre Preferences:
    i) Latin and Rock
    ii) Metal and Alternative & Punk
    iii) Jazz and Rock
    iv) Pop and Rock

Recommendations:

- Create bundle offers that include Rock and its frequent pairings, such as:
    i) Rock and Metal
    ii) Rock and Alternative & Punk
    iii) Rock and R&B/Soul
    iv) Blues and Rock
- Recommend similar genres that are frequently bought together for the customers who bought one, such as:
    i) Bought Rock, suggest Metal or R&B/Soul.
    ii) Bought Latin, suggest Rock or Blues based on purchasing patterns.
- Some genres appear less often but still show some level of affinity, consider cross-genre promotions to introduce customers to new music styles.

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

Ans:

Approach:

i) Measuring purchasing behaviour by region – track customers, revenue, average order value, and frequency.
ii) Identify churn patterns by region – calculate churned vs active customers and churn rates.
iii) Correlate results with local context – link differences in spending and churn to regional demographics or economic factors.

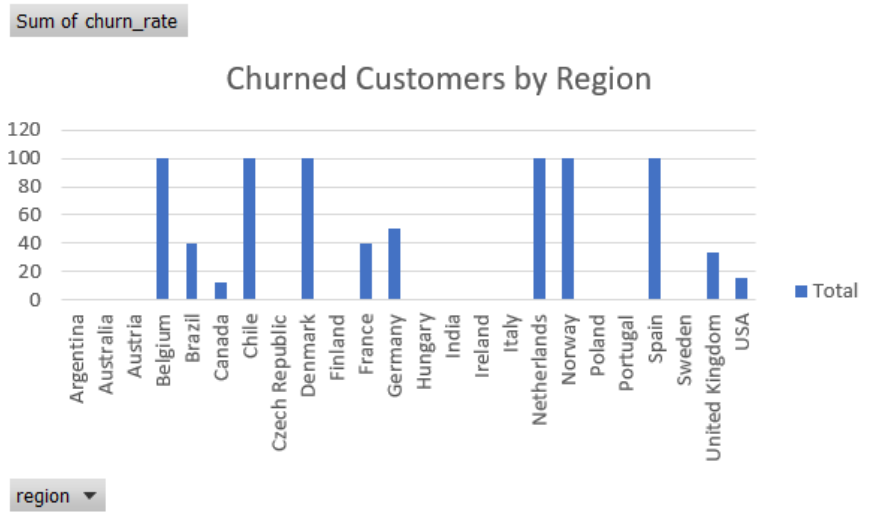
Query:

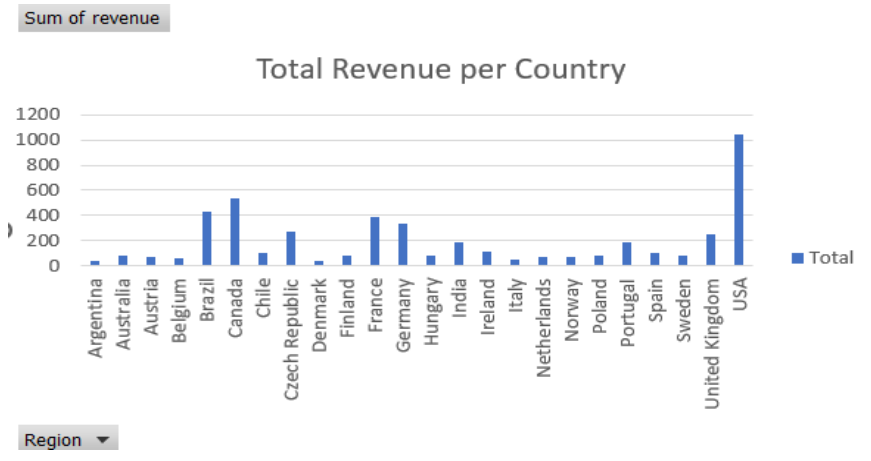
```
-- Purchasing Behavior by Region

SELECT billing_country as Region,

    COUNT(DISTINCT customer_id) AS num_customers,

    SUM(total) AS revenue,

    AVG(total) AS avg_order_value,

    COUNT(invoice_id) / COUNT(DISTINCT customer_id) AS avg_frequency

FROM invoice

GROUP BY billing_country

ORDER BY revenue DESC;
```

Region	num_customers	revenue	avg_order_value	avg_frequen
USA	13	1040.49	7.942672	10.0769
Canada	8	535.59	7.047237	9.5000
Brazil	5	427.68	7.011148	12.2000
France	5	389.07	7.781400	10.0000
Germany	4	334.62	8.161463	10.2500
Czech Republic	2	273.24	9.108000	15.0000
United Kingdom	3	245.52	8.768571	9.3333
Portugal	2	185.13	6.383793	14.5000
India	2	183.15	8.721429	10.5000
Ireland	1	114.84	8.833846	13.0000
Spain	1	98.01	8.910000	11.0000
Chile	1	97.02	7.463077	13.0000
Australia	1	81.18	8.118000	10.0000
Finland	1	79.20	7.200000	11.0000
Hungary	1	78.21	7.821000	10.0000
Poland	1	76.23	7.623000	10.0000
Sweden	1	75.24	7.524000	10.0000
Norway	1	72.27	8.030000	9.0000
Austria	1	69.30	7.700000	9.0000
Netherlands	1	65.34	6.534000	10.0000
Belgium	1	60.39	8.627143	7.0000

```
-- Churn Rate by Region
WITH last_purchase AS (
   SELECT
      customer_id,
      billing_country AS region,
      MAX(invoice_date) AS last_order_date
   FROM invoice
   GROUP BY customer_id, billing_country
),
churned_customers AS (
   SELECT
      region,
      COUNT(customer_id) AS churned_count
   FROM last_purchase
   WHERE last_order_date < DATE_SUB('2020-12-30', INTERVAL 6 MONTH)
   GROUP BY region
)
SELECT
   l.region,
   COUNT(DISTINCT l.customer_id) AS total_customers,
   COALESCE((SELECT churned_count FROM churned_customers c WHERE c.region =
l.region),0) AS churned_customers,
   ROUND((COALESCE((SELECT churned_count FROM churned_customers c WHERE c.region =
l.region),0)/COUNT(DISTINCT l.customer_id))*100,2) AS churn_rate
FROM last_purchase l
GROUP BY l.region
ORDER BY churned_customers DESC;
```

region	total_customers	churned_customers	churn_rate
Brazil	5	2	40.00
France	5	2	40.00
Germany	4	2	50.00
USA	13	2	15.38
Belgium	1	1	100.00
Canada	8	1	12.50
Chile	1	1	100.00
Denmark	1	1	100.00
Netherlands	1	1	100.00
Norway	1	1	100.00
Spain	1	1	100.00
United King...	3	1	33.33
Argentina	1	0	0.00
Australia	1	0	0.00
Austria	1	0	0.00
Czech Rep...	2	0	0.00
Finland	1	0	0.00
Hungary	1	0	0.00
India	2	0	0.00
Ireland	1	0	0.00
Italy	1	0	0.00





Insights:

- USA had more customers, as well as highest total orders and total generated revenue, due to Rock music popularity.
- Canada has second highest number of customers, orders and revenue due to language and developed economies.
- Some Countries like Norway, Spain have 100% churn rate, but doesn't mean anything because they have only 1 customer each.
- Germany has 50% churn rate and France, Brazil have 40% churn rate, they may be less interest in English music.

Recommendations:

- Countries with one customer, initiatives should be taken to increase the number, give extra points to users for giving referrals.
- Countries with higher churned rates should be investigated properly.

Correlation with Local Demographics and Economic Factors:

i) **Economic Prosperity:** Countries with higher GDP per capita and disposable income, like Canada, Ireland, and France, show greater spending on music and entertainment, including premium content and subscriptions.
ii) **Cultural Influence:** Nations with rich musical traditions, such as Brazil, France, and India, have higher engagement with both local and international music, boosting sales across genres.
iii) **Digital Infrastructure:** Well-developed digital economies, such as in Ireland, Czech Republic, and Portugal, facilitate seamless online music purchases and streaming, encouraging frequent digital consumption.
iv) **Affluence and Leisure Spending:** Wealthier populations tend to allocate more income to entertainment, supporting both local and global music markets through diverse purchases
v) Finally Higher income, strong music cultures, and robust digital infrastructure correlate with increased music spending and more active customer engagement.

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:

Approach:

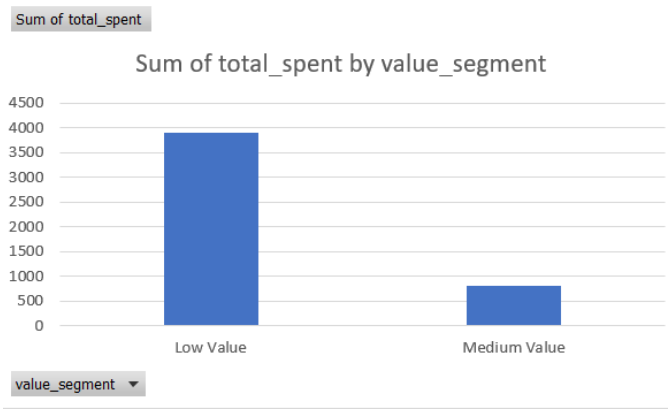
i) Summarize customer activity – calculate orders, spending, and recency (days since last purchase) for each customer.
ii) Assign risk levels – classify customers as High, Medium, or Low Risk based on inactivity duration.
iii) Segment by value – categorize spending into Low, Medium, or High Value to identify which risky groups also generate the most or least revenue.

Query:

```
-- Query for High-Risk Customer Segments
WITH customer_summary AS (
    SELECT
        c.customer_id,
        c.country,
        COUNT(i.invoice_id) AS total_orders,
        SUM(il.unit_price * il.quantity) AS total_spent,
        MAX(i.invoice_date) AS last_purchase_date,
        DATEDIFF(CURRENT_DATE, MAX(i.invoice_date)) AS days_since_last_purchase
    FROM customer c
    LEFT JOIN invoice i ON c.customer_id = i.customer_id
    LEFT JOIN invoice_line il ON i.invoice_id = il.invoice_id
    GROUP BY c.customer_id, c.country
),
customer_risk AS (
    SELECT
        customer_id,
        country,
        total_orders,
        total_spent,
        last_purchase_date,
        days_since_last_purchase,
        CASE
            WHEN days_since_last_purchase > 180 THEN 'High Risk'
            WHEN days_since_last_purchase BETWEEN 90 AND 180 THEN 'Medium Risk'
            ELSE 'Low Risk'
        END AS churn_risk,
```

```
        CASE
            WHEN total_spent < 100 THEN 'Low Value'
            WHEN total_spent BETWEEN 100 AND 500 THEN 'Medium Value'
            ELSE 'High Value'
        END AS value_segment
    FROM customer_summary
)
SELECT *
FROM customer_risk
ORDER BY churn_risk DESC, total_spent ASC;
```

customer_id	country	total_orders	total_spent	last_purchase_date	days_since_last_purchase	churn_risk	value_segment
14	Canada	30	29.70	2020-12-20 00:00:00	1747	High Risk	Low Value
9	Denmark	38	37.62	2020-01-29 00:00:00	2073	High Risk	Low Value
56	Argentina	40	39.60	2020-07-05 00:00:00	1915	High Risk	Low Value
29	Canada	41	40.59	2020-12-08 00:00:00	1759	High Risk	Low Value
47	Italy	51	50.49	2020-10-06 00:00:00	1822	High Risk	Low Value
19	USA	55	54.45	2020-04-14 00:00:00	1997	High Risk	Low Value
8	Belgium	61	60.39	2019-09-21 00:00:00	2203	High Risk	Low Value
10	Brazil	61	60.39	2020-06-25 00:00:00	1925	High Risk	Low Value
31	Canada	63	62.37	2020-11-24 00:00:00	1773	High Risk	Low Value
41	France	65	64.35	2020-11-12 00:00:00	1785	High Risk	Low Value
48	Netherla...	66	65.34	2020-04-27 00:00:00	1984	High Risk	Low Value
15	Canada	67	66.33	2020-12-18 00:00:00	1749	High Risk	Low Value
23	USA	67	66.33	2020-11-27 00:00:00	1770	High Risk	Low Value
52	United Ki...	69	68.31	2020-12-27 00:00:00	1740	High Risk	Low Value
7	Austria	70	69.30	2020-08-26 00:00:00	1863	High Risk	Low Value
11	Brazil	70	69.30	2020-06-24 00:00:00	1926	High Risk	Low Value
32	Canada	71	70.29	2020-10-26 00:00:00	1802	High Risk	Low Value
59	India	72	71.28	2020-12-19 00:00:00	1748	High Risk	Low Value
24	USA	72	71.28	2020-10-25 00:00:00	1803	High Risk	Low Value
40	France	73	72.27	2020-10-31 00:00:00	1797	High Risk	Low Value
4	Norway	73	72.27	2020-02-04 00:00:00	2067	High Risk	Low Value
28	USA	73	72.27	2020-10-20 00:00:00	1808	High Risk	Low Value
38	Germany	74	73.26	2020-04-22 00:00:00	1989	High Risk	Low Value
43	France	74	73.26	2020-05-02 00:00:00	1979	High Risk	Low Value
16	USA	75	74.25	2020-11-20 00:00:00	1777	High Risk	Low Value



Insights:

- Most customers fall into the High-Risk category due to very long inactivity periods (many with last purchases over 1700 days ago).
- Frequent buyers often belong to the Low-Value segment, suggesting small spending per transaction.
- Regions like Canada, Denmark, Argentina, Italy, USA, and Belgium have a concentration of high-risk, low-value customers.

Recommendations:

- Retention Campaigns: Re-engage high-risk, low-value customers through targeted offers, discounts, or email outreach.
- Loyalty Programs: Strengthen loyalty among high-value customers with rewards, points, or exclusive deals.
- Deeper Analysis: Explore why frequent customers spend little (e.g., pricing or product mix) and, where possible, enhance profiling with demographic data.

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:

Approach:

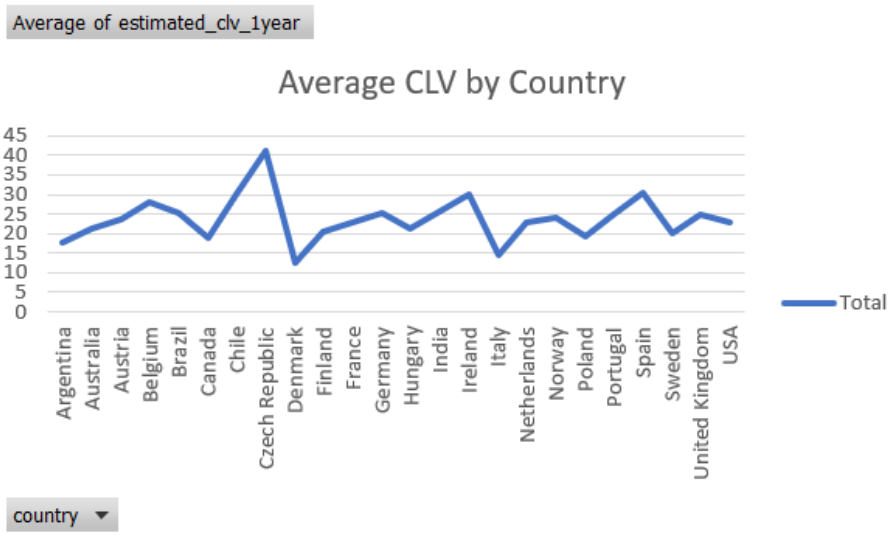
i) Build customer metrics – calculate tenure, frequency, total spend, average order value, and recency for each customer.
ii) Segment customers – classify them as Active, At-Risk, or Churned based on inactivity duration.
iii) Estimate CLV – project yearly value using revenue per day, and compare patterns (low frequency, long inactivity) among churned customers to identify common risk factors.

Query:

```
WITH CustomerMetrics AS (

SELECT

c.customer_id,

CONCAT(c.first_name, ' ', c.last_name) AS customer_name,

c.country,

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)) + 1 AS tenure_days,
```

```
        COUNT(i.invoice_id) AS purchase_frequency,
        SUM(i.total) AS total_spent,
        ROUND(SUM(i.total) / COUNT(i.invoice_id), 2) AS avg_order_value,
        DATEDIFF(CURRENT_DATE, 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, c.first_name, c.last_name, c.country
)
SELECT
    customer_id,
    customer_name,
    country,
    tenure_days,
    purchase_frequency,
    total_spent,
    avg_order_value,
    days_since_last_purchase,
    CASE
        WHEN days_since_last_purchase > 180 THEN 'Churned'
        WHEN days_since_last_purchase > 30 THEN 'At-Risk'
        ELSE 'Active'
    END AS customer_status,
ROUND((total_spent / tenure_days) * 365, 2) AS estimated_clv_1year,
    ROUND(total_spent / tenure_days, 2) AS revenue_per_day
FROM CustomerMetrics
ORDER BY days_since_last_purchase DESC;
```

customer_id	customer_name	country	tenure_days	purchase_frequency	total_spent	avg_order_value	days_since_last_purchase
18	Michelle Brooks	USA	1158	8	79.20	9.90	2037
54	Steve Murray	United Kingdom	950	9	79.20	8.80	2017
36	Hannah Schneider	Germany	1111	11	85.14	7.74	2004
39	Camille Bernard	France	1156	9	79.20	8.80	2000
19	Tim Goyer	USA	1124	9	54.45	6.05	1997
38	Niklas Schröder	Germany	1189	9	73.26	8.14	1989
48	Johannes Van der Berg	Netherlands	1048	10	65.34	6.53	1984
43	Isabelle Mercier	France	1182	12	73.26	6.11	1979
3	François Tremblay	Canada	1181	9	99.99	11.11	1965
57	Luis Rojas	Chile	1173	13	97.02	7.46	1941
50	Enrique Muñoz	Spain	1182	11	98.01	8.91	1940
11	Alexandre Rocha	Brazil	974	10	69.30	6.93	1926
10	Eduardo Martins	Brazil	1236	12	60.39	5.03	1925
56	Diego Gutiérrez	Argentina	809	5	39.60	7.92	1915
58	Manoj Pareek	India	1241	13	111.87	8.61	1905
1	Luís Gonçalves	Brazil	1276	13	108.90	8.38	1896
7	Astrid Gruber	Austria	1069	9	69.30	7.70	1863
17	Jack Smith	USA	1147	12	98.01	8.17	1847
37	Fynn Zimmermann	Germany	1295	10	94.05	9.41	1831
47	Lucas Mancini	Italy	1260	9	50.49	5.61	1822
34	João Fernandes	Portugal	1327	13	102.96	7.92	1815
27	Patrick Gray	USA	1379	9	84.15	9.35	1812
45	Ladislav Kovács	Hungary	1341	10	78.21	7.82	1809
28	Julia Barnett	USA	1137	10	72.27	7.23	1808



Insights:

- High-value customers are those with strong spending and frequent purchases, reflected in high CLV.
- Churned customers typically show long inactivity, low frequency, short tenure, and often cluster in certain countries.
- Short-tenure but high-spend customers present growth potential if nurtured through engagement.

Recommendations:

- Reward high-value customers with loyalty perks, VIP benefits, or early access to sustain engagement.
- Re-engage churned/at-risk segments using targeted offers, discounts, and region-specific campaigns.

- Retain promising new customers by recognizing high activity early and fostering loyalty.
- Leverage purchase patterns for cross-selling and personalized recommendations.

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: **Measuring Campaign Impact**

The methodology relies on comparing a **Targeted** group (customers who were exposed to the campaign) with a **Control** group (similar customers who were deliberately *not* exposed).

1. Impact on Overall Sales (Sales Lift)

Sales Lift measures the increase in revenue directly attributable to the promotion.

- **Methodology:** Compare the **average spending per customer** in the post-campaign period for the Targeted group versus the Control group.

- **Success Indicator:** If the **Targeted** group's average spending is significantly **higher** than the **Control** group, the campaign successfully drove purchases and generated a positive sales lift.

- **Required Data:** You need a **Campaign Log** (tracking start/end dates) and an **Exposure Log** (listing all customer_ids who received the promotion).

2. Impact on Customer Retention

Retention measures how effective the campaign was at keeping customers engaged and active after the promotion itself ended.

- **Methodology:** Track the **retention rate** (percentage of customers who made *any* purchase) in the **post-campaign period**.

- **Success Indicator:** A campaign is successful at retention if the **Targeted** group's retention rate is **higher** than the **Control** group. This is particularly valuable when targeting high-risk customers (like those with low purchase frequency).

- **Required Data:** The analysis must use a defined **cutoff date** (e.g., 90 days after the campaign ends) to check for repeat purchases.

3. Impact on Customer Acquisition

Acquisition measures how many **new** customers signed up or made their first purchase because of the campaign.

- **Methodology: Attribute** new customer sign-ups to the campaign's specific time frame and tracking ID.

- **Success Indicator:** Count the number of customers whose **first purchase date** (or account creation date) falls between the campaign's start and end dates and whose source can be linked to the promotion (e.g., a specific URL or discount code).

- **Required Data:** The customer table would need a **Join_Date** field and a **Source_Tracking_ID** field to accurately attribute new accounts.

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

Ans:

My approach to analyzing Chinook's music record sales data—without relying on pre-defined questions—follows a three-phase, top-down strategy designed to transform raw transactional data into strategic, revenue-driving recommendations.

The overarching goal is to answer one core question:

"How can Chinook optimize its product mix, customer engagement, and marketing strategy to maximize revenue in the physical music market?"

Phase One: Market and Product Performance Analysis

The first phase establishes a performance baseline for what Chinook is selling, where it's selling best, and which product lines are driving growth.

This foundation informs all subsequent marketing and inventory decisions.

Geographic Market Prioritization:

I would rank each country by three core metrics—Total Revenue, Number of Unique Customers, and Average Spending per Customer.

This triage identifies large, mature markets (e.g., USA, Canada) and smaller, high-value emerging markets (e.g., Czech Republic, Ireland).

Insight: This enables data-driven allocation of marketing budgets and logistics focus to regions with the highest profit potential.

Genre and Album Performance:

I would calculate total revenue by Genre and Album across all countries and within the top-performing regions.

Insight: This exposes the most commercially dominant genres (e.g., Rock, Metal, Classical) and their flagship albums, guiding promotional prioritization and future label partnerships.

Sales Representative Effectiveness:

Using customer.support_rep_id, I would evaluate sales performance across representatives.

Insight: Identifying high-performing reps helps Chinook replicate best practices across the global sales team.

Phase Two: Customer Value and Risk Profiling

The second phase focuses on understanding customer profitability, loyalty, and churn risk, allowing Chinook to develop more personalized and retention-focused marketing.

Customer Lifetime Value (CLV) Segmentation:

I would calculate each customer's Recency, Frequency, and Monetary metrics:

Monetary: Total historical spending per customer.

Frequency: Number of distinct purchase events (invoices).

Recency: Days since the last purchase, measuring engagement and churn risk.

These metrics collectively identify which customers drive long-term profitability.

High-Risk Churn Profiling:

By comparing behavioral metrics of inactive customers (no purchases in the last 3 months) to active ones, I would define the early-warning profile of a churning customer—typically low purchase frequency and low order value in early tenure.

Insight: Chinook can implement automated retention triggers for customers approaching this pattern.

High-Value Segment Identification:

I would classify customers into actionable segments such as:

High-Value & High-Frequency: Loyal power users deserving loyalty rewards.

High-Value but Low-Frequency: Profitable but disengaged customers needing reactivation campaigns.

Low-Value but Frequent: Price-sensitive users suitable for bundled or discount-based strategies.

This segmentation enables targeted marketing and more efficient resource allocation.

Phase Three: Product Affinity and Cross-Selling Strategy

The final phase focuses on leveraging transaction-level insights to design better marketing and sales experiences that drive incremental revenue.

Product Affinity (Market Basket Analysis):

I would analyze purchasing combinations using invoice-level co-occurrence of genres, artists, or albums.

Insight: This uncovers hidden product relationships (e.g., customers who buy "Jazz" also buy "Classical")—the foundation for AI-driven product recommendations and cross-sell campaigns.

Bundling and Upselling Opportunities:

Based on affinity results, I would propose curated product bundles (e.g., pairing a top-selling Rock album with a complementary Alternative one).

Insight: This improves average order value and strengthens customer engagement through relevant offers.

Inventory and Pricing Optimization:

Cross-referencing high-selling tracks with album-level sales helps detect profit leaks (e.g., frequent single-track purchases from high-performing albums).

Insight: Chinook can adjust pricing or bundle strategy to encourage full-album sales and improve margin efficiency.

Conclusion:

This three-phase analytical framework converts raw transactional data into strategic business intelligence.

By integrating market performance, customer segmentation, and product affinity modeling, I can help Chinook:

- Focus marketing on high-yield regions and genres,
- Retain valuable customers through targeted engagement, and
- Boost profitability via intelligent bundling and cross-selling.

Ultimately, this approach empowers Chinook to move from descriptive analytics to data-driven strategy, strengthening its position in the competitive physical music market.

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: To add a new column named ReleaseYear of type INTEGER to the Album table, we will use SQL ALTER TABLE statement with the ADD COLUMN command.

SQL Command:

ALTER TABLE album

ADD COLUMN ReleaseYear INT;

Updated Table:

album_id	title	artist_id	ReleaseYear
1	For Those About To Rock We Salute You	1	NULL
2	Balls to the Wall	2	NULL
3	Restless and Wild	2	NULL
4	Let There Be Rock	1	NULL
5	Big Ones	3	NULL
6	Jagged Little Pill	4	NULL
7	Facelift	5	NULL
8	Warner 25 Anos	6	NULL
9	Plays Metallica By Four Cellos	7	NULL
10	Audioslave	8	NULL
11	Out Of Exile	8	NULL
12	BackBeat Soundtrack	9	NULL
13	The Best Of Billy Cobham	10	NULL

11. Chinook is interested in understanding the purchasing behaviour of customers based on their geographical location. They want to know the average total amount spent by customers from each country, along with the number of customers and the average number of tracks purchased per customer. Write an SQL query to provide this information.

Ans:

Approach:

i) Measure how much customers in each country spend and how many tracks they purchase on average.
ii) Compare countries to identify differences in spending and purchasing behaviour.
iii) Use these differences to understand regional trends and guide market strategies.

Query:

```
WITH CustomerSummary AS (
  -- 1. Calculate Total Spent and Total Tracks Purchased for EACH CUSTOMER
  SELECT
    i.customer_id,
    SUM(i.total) AS total_spent,
    SUM(il.quantity) AS tracks_purchased
  FROM
    invoice AS i
  JOIN
    invoice_line AS il ON i.invoice_id = il.invoice_id
  GROUP BY
    i.customer_id
)
SELECT
  c.country,
  COUNT(c.customer_id) AS num_customers,
  -- Average total amount spent by customers from this country
  ROUND(AVG(cs.total_spent), 2) AS avg_spent_per_customer,
  -- Average number of tracks purchased by customers from this country
  ROUND(AVG(cs.tracks_purchased), 2) AS avg_tracks_per_customer
```

```
FROM
    customer AS c
JOIN
    CustomerSummary AS cs ON c.customer_id = cs.customer_id
GROUP BY
    c.country
ORDER BY
    num_customers DESC,
    avg_spent_per_customer DESC;
```

country	num_customers	avg_spent_per_customer	avg_tracks_per_customer
USA	13	800.45	80.85
Canada	8	686.19	67.63
Brazil	5	811.80	86.40
France	5	794.57	78.60
Germany	4	860.31	84.50
United Kingdom	3	832.92	82.67
Czech Republic	2	1591.92	138.00
India	2	943.97	92.50
Portugal	2	877.64	93.50
Ireland	1	1433.52	116.00
Spain	1	1076.13	99.00
Australia	1	940.50	82.00
Chile	1	912.78	98.00
Hungary	1	830.61	79.00
Sweden	1	754.38	76.00
Poland	1	690.03	77.00
Finland	1	685.08	80.00
Norway	1	664.29	73.00
Austria	1	649.44	70.00
Belgium	1	567.27	61.00
Netherlands	1	544.50	66.00
Italy	1	468.27	51.00
Argentina	1	396.00	40.00
Denmark	1	196.02	38.00

Insights:

- Highest spenders are in Czech Republic, Ireland, and Spain.
- USA contributes the most revenue overall.
- India and Brazil show rising customer activity.
- Lowest average spend is seen in Argentina, Denmark, and Italy.

Recommendations:

- Reward top spenders with premium offers or early access.
- Sustain engagement in the USA through loyalty programs.
- Promote growth markets like India and Brazil with targeted campaigns.
- Boost low-spend regions (Argentina, Denmark, Italy) using special offers or customer feedback surveys.