Objective Questions:

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

➢ To ensure data integrity and completeness, it is crucial to identify and handle any missing values (NULLs) in the database tables. Below is the approach I followed to detect and address these null values across different tables.

Step 1: Checking for Null Values

I started by checking the customer table for any null values across all columns. The following SQL query was used to identify rows where any of the columns might contain a null value:

```sql
-- Checking for Null Values in customer table
SELECT *
FROM customer
WHERE first_name IS NULL
   OR last_name IS NULL
   OR company IS NULL
   OR address IS NULL
   OR city IS NULL
   OR state IS NULL
   OR country IS NULL
   OR postal_code IS NULL
   OR phone IS NULL
   OR fax IS NULL
   OR email IS NULL
   OR support_rep_id IS NULL;
```

Step 2: Handling Null Values

I observed that some columns contained NULL values. To manage this, I replaced NULL values with meaningful defaults based on their data type. Text fields like company, state, and fax were updated with placeholder values like 'UNKNOWN', 'NONE', and 'N/A' by using COALESCE function. Similarly, for numerical fields, appropriate default values were assigned. This ensures better data consistency and improves query results readability.

For the customer table, the following SQL query was executed:

```sql
-- ******** Handling Null Values ********
SELECT customer_id,
       COALESCE(company, 'UNKNOWN') AS company,
       COALESCE(state, 'NONE') AS state,
       COALESCE(postal_code, 'N/A') AS postal_code,
       COALESCE(phone, 'N/A') AS phone,
       COALESCE(fax, 'N/A') AS fax
FROM customer;
```

OUTPUT TABLE:

customer_id	company	state	postal_code	phone	fax
1	Embraer - Empresa Brasileira de Aeronáutica S.A.	SP	12227-000	+55 (12) 3923-5555	+55 (12) 3923-5566
2	UNKNOWN	NONE	70174	+49 0711 2842222	N/A
3	UNKNOWN	QC	H2G 1A7	+1 (514) 721-4711	N/A
4	UNKNOWN	NONE	0171	+47 22 44 22 22	N/A
5	JetBrains s.r.o.	NONE	14700	+420 2 4172 5555	+420 2 4172 5555
6	UNKNOWN	NONE	14300	+420 2 4177 0449	N/A
7	UNKNOWN	NONE	1010	+43 01 5134505	N/A
8	UNKNOWN	NONE	1000	+32 02 219 03 03	N/A
9	UNKNOWN	NONE	1720	+453 3331 9991	N/A
10	Woodstock Discos	SP	01007-010	+55 (11) 3033-5446	+55 (11) 3033-4564
11	Banco do Brasil S.A.	SP	01310-200	+55 (11) 3055-3278	+55 (11) 3055-8131
12	Riotur	RJ	20040-020	+55 (21) 2271-7000	+55 (21) 2271-7070
13	UNKNOWN	DF	71020-677	+55 (61) 3363-5547	+55 (61) 3363-7855
14	Telus	AB	T6G 2C7	+1 (780) 434-4554	+1 (780) 434-5565
15	Rogers Canada	BC	V6C 1G8	+1 (604) 688-2255	+1 (604) 688-8756
16	Google Inc.	CA	94043-1351	+1 (650) 253-0000	+1 (650) 253-0000
17	Microsoft Corporation	WA	98052-8300	+1 (425) 882-8080	+1 (425) 882-8081
18	UNKNOWN	NY	10012-2612	+1 (212) 221-3546	+1 (212) 221-4679
19	Apple Inc.	CA	95014	+1 (408) 996-1010	+1 (408) 996-1011
20	UNKNOWN	CA	94040-111	+1 (650) 644-3358	N/A
21	UNKNOWN	NV	89503	+1 (775) 223-7665	N/A
22	UNKNOWN	FL	32801	+1 (407) 999-7788	N/A
23	UNKNOWN	MA	2113	+1 (617) 522-1333	N/A
24	UNKNOWN	IL	60611	+1 (312) 332-3232	N/A
25	UNKNOWN	WI	53703	+1 (608) 257-0597	N/A
26	UNKNOWN	TX	76110	+1 (817) 924-7272	N/A

Step 3: Applying the Same Approach to Other Tables

I applied a similar strategy to other tables in the database.
For instance, to check for null values in the “employee” and “track” table and replace them, the following SQL queries were used:

```sql
-- Checking for Null Values in employee table
SELECT *
FROM employee
WHERE last_name IS NULL
   OR first_name IS NULL
   OR title IS NULL
   OR reports_to IS NULL
   OR birthdate IS NULL
   OR hire_date IS NULL
   OR address IS NULL
   OR city IS NULL
   OR state IS NULL
   OR country IS NULL
   OR postal_code IS NULL
   OR phone IS NULL
   OR fax IS NULL
   OR email IS NULL;
```

```sql
-- Checking for Null Values in track table
SELECT *
FROM track
WHERE name IS NULL
   OR album_id IS NULL
   OR media_type_id IS NULL
   OR genre_id IS NULL
   OR composer IS NULL
   OR milliseconds IS NULL
   OR bytes IS NULL
   OR unit_price IS NULL;
```

Handling Null Values:

```sql
-- Handling Null Values
SELECT
  employee_id,
  first_name,
  last_name,
  COALESCE(reports_to, 'N/A') AS reports_to
FROM employee;
```

```sql
-- Handling Null Values
SELECT
  track_id,
  name,
  COALESCE(composer, 'N/A') AS composer
FROM track;
```

OUTPUT TABLE:

employee_id	first_name	last_name	reports_to
1	Andrew	Adams	N/A
2	Nancy	Edwards	1
3	Jane	Peacock	2
4	Margaret	Park	2
5	Steve	Johnson	2
6	Michael	Mitchell	1
7	Robert	King	6
8	Laura	Callahan	6

track_id	name	composer
58	Sunshine	Jerry Cantrell
59	Put You D...	Jerry Cantrell
60	Confusion	Jerry Cantrell, Michael Starr, Layne Staley
61	I Know So...	Jerry Cantrell
62	Real Thing	Jerry Cantrell, Layne Staley
63	Desafinado	N/A
64	Garota De ...	N/A
65	Samba De ...	N/A
66	Por Causa ...	N/A
67	Ligia	N/A
68	Fotografia	N/A
69	Dindi (Dindi)	N/A
70	Se Todos ...	N/A
71	Falando D...	N/A

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

➢

❖ **Top – selling tracks in USA**

```sql
-- Top Selling Track in USA
SELECT
  t.track_id,
  t.name AS track_name,
  SUM(il.quantity) AS total_sold,
  g.name AS genre,
  a.name AS artist
FROM
  invoice_line il
  INNER JOIN invoice i ON il.invoice_id = i.invoice_id
  INNER JOIN customer c ON i.customer_id = c.customer_id
  INNER JOIN track t ON il.track_id = t.track_id
  INNER JOIN album al ON t.album_id = al.album_id
  INNER JOIN artist a ON al.artist_id = a.artist_id
  INNER JOIN genre g ON t.genre_id = g.genre_id
WHERE c.country = 'USA'
GROUP BY t.track_id, t.name, g.name, a.name
ORDER BY total_sold DESC
LIMIT 10;
```

OUTPUT TABLE:

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

track_id	track_name	total_sold	genre	artist
3336	War Pigs	6	Alternative	Cake
3465	You Know I'm No Good (feat. Ghostface Killah)	5	R&B/Soul	Amy Winehouse
2560	Violent Pornography	4	Metal	System Of A Down
2647	End Of The Night	4	Rock	The Doors
13	Night Of The Long Knives	4	Rock	AC/DC
1995	Scentless Apprentice	4	Rock	Nirvana
153	Evil Woman	4	Metal	Black Sabbath
2646	I Looked At You	4	Rock	The Doors
1495	Highway Chile	4	Rock	Jimi Hendrix
55	I Can't Remember	3	Rock	Alice In Chains

INSIGHTS:

- **Top Selling Tracks:** "War Pigs" by Cake and "You Know I'm No Good (feat. Ghostface Killah)" by Amy Winehouse are the top-selling tracks, each selling 6 and 5 units respectively.

- **Genre Distribution:** Rock is the dominant genre among the top-selling tracks, with several entries like "End Of The Night" by The Doors, "Night Of The Long Knives" by AC/DC, and others.
- **Artist Popularity:** The Doors appear twice in the top-selling list with "End Of The Night" and "I Looked At You", highlighting their popularity. Other notable artists include Nirvana, Black Sabbath, and Jimi Hendrix.
- **Diverse Genres:** The top-selling tracks span various genres including Rock, Alternative, Metal, and R&B/Soul, indicating a diverse customer preference.
- **Sales Distribution:** Sales numbers vary widely among the top tracks, with some selling as few as 3 units ("I Can't Remember" by Alice In Chains) to as many as 6 units ("War Pigs" by Cake).

  These insights provide a snapshot of the sales performance, genre preferences, and artist popularity among the listed tracks.

❖ Top Artist in USA and Most Famous Genres of the Top Artist:

```sql
-- Top Artist in USA and Most Famous Genres of the Top Artist
SELECT
  a.artist_id,
  a.name AS artist_name,
  g.name AS genre_name,
  SUM(il.quantity) AS total_sold
FROM
  invoice_line il
  INNER JOIN invoice i ON il.invoice_id = i.invoice_id
  INNER JOIN customer c ON i.customer_id = c.customer_id
  INNER JOIN track t ON il.track_id = t.track_id
  INNER JOIN album al ON t.album_id = al.album_id
  INNER JOIN artist a ON al.artist_id = a.artist_id
  INNER JOIN genre g ON t.genre_id = g.genre_id
WHERE c.country = 'USA'
GROUP BY a.artist_id, a.name, g.name
ORDER BY total_sold DESC
LIMIT 1;
```

OUTPUT TABLE:

Result Grid | Filter Rows:

artist_id	artist_name	genre_name	total_sold
152	Van Halen	Rock	43

INSIGHTS:

The data reveals that **Van Halen** is a **top artist in the USA**, highlighting their widespread popularity and strong sales performance. The total sales of **43 units** indicate a well-established fan base and the enduring appeal of their music. Additionally, the most **famous genre of this top artist is Rock**, which signifies the genre's strong presence and influence in the music industry. This suggests that Rock continues to attract a dedicated audience, contributing to Van Halen's success.

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

```sql
WITH customer_information_cte as (
  SELECT
    customer_id,
    first_name,
    last_name,
    city,
    COALESCE(state,'N.A') as state,
    country
  FROM customer
)
SELECT
  country,
  state,
  city,
  COUNT(customer_id) as total_customers
FROM customer_information_cte
GROUP BY country, state, city
ORDER BY country, state, city;
```

OUTPUT TABLE:

Result Grid | Filter Rows: | Export:

country	state	city	total_customers
Argentina	N.A	Buenos Aires	1
Australia	NSW	Sidney	1
Austria	N.A	Vienne	1
Belgium	N.A	Brussels	1
Brazil	DF	Brasília	1
Brazil	RJ	Rio de Janeiro	1
Brazil	SP	São José dos Campos	1
Brazil	SP	São Paulo	2
Canada	AB	Edmonton	1
Canada	BC	Vancouver	1
Canada	MB	Winnipeg	1
Canada	NS	Halifax	1
Canada	NT	Yellowknife	1
Canada	ON	Ottawa	1
Canada	ON	Toronto	1

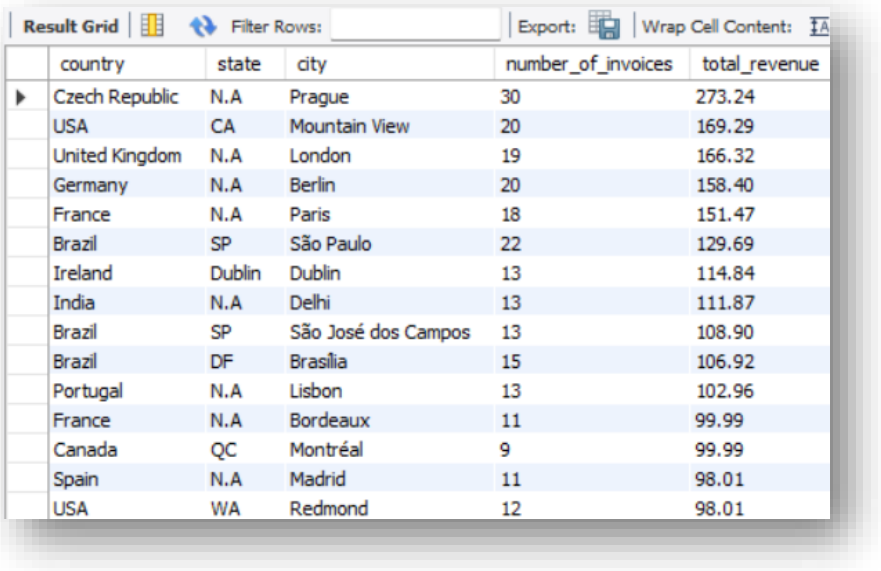
INSIGHTS:

- The data shows that **São Paulo** in Brazil, **Paris** in France, and **London** in the UK are the cities with the highest number of customers, each having 2 customers.
- **Brazil** has multiple cities listed, with **São Paulo** having the highest customer count among them, indicating a strong presence in this country.
- The **USA** also shows a diverse spread with multiple cities having customers, including **Mountain View and New York**, each with 2 customers.
- Many countries have only one city listed, suggesting a more limited customer base in those locations compared to cities with higher counts.
- There is a notable concentration of customers in major cities and capitals, such as **Paris, London, and New York**, reflecting a trend where larger, prominent cities are more likely to have higher customer counts.

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

```sql
SELECT
  c.country,
    COALESCE(c.state,'N.A') as state,
    c.city,
    SUM(i.total) as total_revenue,
    COUNT(i.invoice_id) as number_of_invoices
FROM customer c
INNER JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY c.country, c.state, c.city
ORDER BY total_revenue DESC, number_of_invoices DESC;
```

OUTPUT TABLE:



country	state	city	number_of_invoices	total_revenue
Czech Republic	N.A	Prague	30	273.24
USA	CA	Mountain View	20	169.29
United Kingdom	N.A	London	19	166.32
Germany	N.A	Berlin	20	158.40
France	N.A	Paris	18	151.47
Brazil	SP	São Paulo	22	129.69
Ireland	Dublin	Dublin	13	114.84
India	N.A	Delhi	13	111.87
Brazil	SP	São José dos Campos	13	108.90
Brazil	DF	Brasília	15	106.92
Portugal	N.A	Lisbon	13	102.96
France	N.A	Bordeaux	11	99.99
Canada	QC	Montréal	9	99.99
Spain	N.A	Madrid	11	98.01
USA	WA	Redmond	12	98.01

INSIGHTS:

- **São Paulo** in Brazil has the highest total revenue at **$129.69** despite having **22 invoices**, highlighting its significant market potential.
- **Paris** in France and **London** in the UK show strong performance with high total revenues of **$151.47 and $166.32 respectively**, reflecting their high volume of transactions and possibly higher pricing.
- The **Czech Republic's Prague** leads in the number of invoices with **30**, and also generates significant revenue at **$273.24**, indicating a robust market presence.
- The **USA** shows a diverse revenue distribution with cities like **Mountain View and New York** achieving notable revenues, **$169.29 and $79.20 respectively**, showcasing the strength of various US cities.
- Several countries like **Canada and Australia** have lower revenue figures despite a decent number of invoices, suggesting that while there is consistent business, the average revenue per invoice might be lower.

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

➤

```sql
SELECT * FROM customer;
SELECT * FROM invoice;
WITH customer_wise_revenue_cte1 as(
  SELECT
    c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) as customers,
        c.country,
        SUM(i.total) as total_revenue
  FROM customer c
  INNER JOIN invoice i ON c.customer_id = i.customer_id
    GROUP BY c.customer_id, customers, c.country
  ORDER BY c.country, total_revenue
),
ranked_customers_cte2 as (
  SELECT
    customer_id,
        customers,
        country,
        total_revenue,
        RANK() OVER (PARTITION BY country ORDER BY total_revenue desc) as customer_rank
  FROM customer_wise_revenue_cte1
)
SELECT
  customer_id,
  customers,
  country,
  total_revenue,
    customer_rank
FROM ranked_customers_cte2
WHERE customer_rank <= 5
ORDER BY country, customer_rank;
```

OUTPUT TABLE:

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

customer_id	customers	country	total_revenue	customer_rank
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

INSIGHTS:

- **Top Customers by Revenue:** František Wichterlová from the Czech Republic is the highest revenue-generating customer with a total of $144.54. This suggests strong individual customer value in this region.
- **Diverse High Revenue Contributions:** Several countries, including Brazil, India, Ireland, and Portugal, have customers with revenues exceeding $100, indicating strong sales from diverse regions.
- **Notable Repeat Customers:** In countries like Brazil, Canada, and Germany, there are multiple customers listed with varying revenues, showing a broad customer base contributing to total sales.
- **Consistent Revenue Across Regions:** There is a notable consistency in customer revenues across different countries, with many customers contributing between $60 and $100, indicating steady performance across global markets.
- **Unique High Performers:** Certain countries have standout high-revenue customers like Manoj Pareek in India ($111.87) and Hugh O'Reilly in Ireland ($114.84), indicating exceptional individual performances in these regions.

6. Identify the top-selling track for each customer.

➢

```sql
WITH Customer_track as (
  SELECT
    c.customer_id,
    CONCAT(c.first_name, ' ', c.last_name) as customers,
    SUM(il.quantity) as total_quantity
  FROM customer c
  INNER JOIN invoice i ON c.customer_id = i.customer_id
  INNER JOIN invoice_line il ON i.invoice_id = il.invoice_id
  INNER JOIN track t ON t.track_id = il.track_id
  GROUP BY c.customer_id, customers
),
ranked_track as(
  SELECT
    Customer_track.customer_id,
        Customer_track.customers,
        Customer_track.total_quantity,
        t.track_id,
        t.name as track_name,
        ROW_NUMBER() OVER (PARTITION BY Customer_track.customer_id ORDER BY Customer_track.total_quantity DESC) as track_rank
  FROM Customer_track
  INNER JOIN invoice i ON Customer_track.customer_id = i.customer_id
  INNER JOIN invoice_line il ON i.invoice_id = il.invoice_id
  INNER JOIN track t ON t.track_id = il.track_id
)
SELECT
  customer_id,
    customers,
    track_id,
    track_name,
    total_quantity
FROM ranked_track
WHERE track_rank = 1
ORDER BY total_quantity DESC;
```

OUTPUT TABLE:

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

customer_id	customers	track_id	track_name	total_quantity
1	Luís Gonçalves	572	Put Your Lights On	110
2	Leonie Köhler	85	Cochise	83
3	François Tremblay	33	The Other Side	101
4	Bjørn Hansen	1146	Welcome to the Jungle	73
5	František Wichterlová	3106	Big Machine	146
6	Helena Holý	1479	Foxy Lady	130
7	Astrid Gruber	546	Lady Double Dealer	70
8	Daan Peeters	1784	I Wish It Would Rain	61
9	Kara Nielsen	6	Put The Finger On You	38
10	Eduardo Martins	105	The Worm	61
11	Alexandre Rocha	1853	Battery	70
12	Roberto Almeida	36	Angel	83
13	Fernanda Ramos	711	Molina	108
14	Mark Philips	2234	Us And Them	30
15	Jennifer Peterson	21	Hell Ain't A Bad Place ...	67

INSIGHTS:

- "**Big Machine**", "**Foxy Lady**", and "**House Of Pain Anthem**" are the **most popular tracks**, with over 100 units sold each.
- Several customers show high engagement, purchasing large quantities of their preferred tracks.
- There's a wide variety of tracks purchased, indicating diverse musical tastes among customers.
- Some tracks, like "**Foxy Lady**" are popular with multiple customers, showing broad appeal.
- High purchase quantities are seen across different countries, reflecting strong global demand.

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

➢

❖ **Frequency of Purchases:**

```sql
-- Frequency of Purchases
SELECT
  c.customer_id,
  CONCAT(c.first_name, ' ', c.last_name) as customers,
  YEAR(i.invoice_date) AS year,
  COUNT(i.invoice_id) AS purchase_count
FROM customer c
INNER JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY c.customer_id, customers, YEAR(i.invoice_date)
ORDER BY c.customer_id, customers, YEAR(i.invoice_date);
```

OUTPUT TABLE:

Result Grid | Filter Rows: | Export:

customer_id	customers	year	purchase_count
1	Luís Gonçalves	2017	4
1	Luís Gonçalves	2018	5
1	Luís Gonçalves	2019	2
1	Luís Gonçalves	2020	2
2	Leonie Köhler	2017	3
2	Leonie Köhler	2018	3
2	Leonie Köhler	2019	1
2	Leonie Köhler	2020	4
3	François Tremblay	2017	3
3	François Tremblay	2018	2
3	François Tremblay	2019	3
3	François Tremblay	2020	1
4	Bjørn Hansen	2017	4

INSIGHTS:

- **František Wichterlová** had a significant increase in purchases, from **2 in 2017 to 8 in 2020**, showing growing engagement.
- Many customers, like **LuÃs GonÃ§alves and Fernanda Ramos**, showed varied purchase behavior across years, indicating inconsistent buying patterns.
- Some customers, such as **Hugh O'Reilly and Edward Francis**, had more stable or slightly fluctuating purchase counts over the years.
- **2020** saw a mix of increased and decreased purchases among different customers, reflecting diverse trends possibly influenced by external factors.
- Several customers had notable high purchase years, such as Martha Silk in **2017 with 6 purchases** and Edward Francis in **2020 with 6 purchases**.

❖ **Calculate the average order value for each customer:**

```sql
-- Calculate the average order value for each customer
SELECT
  c.customer_id,
  CONCAT(c.first_name, ' ', c.last_name) as customers,
    ROUND(AVG(i.total), 2) AS avg_order_value
FROM customer c
INNER JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY c.customer_id, customers
ORDER BY avg_order_value desc;
```

OUTPUT TABLE:

Result Grid | Filter Rows:

customer_id	customers	avg_order_value
3	François Tremblay	11.11
6	Helena Holý	10.73
29	Robert Brown	10.15
18	Michelle Brooks	9.90
37	Fynn Zimmermann	9.41
27	Patrick Gray	9.35
16	Frank Harris	9.28
42	Wyatt Girard	9.09
59	Puja Srivastava	8.91
24	Frank Ralston	8.91
53	Phil Hughes	8.91
50	Enrique Muñoz	8.91
46	Hugh O'Reilly	8.83
54	Steve Murray	8.80
39	Camille Bernard	8.80

INSIGHTS:

- **François Tremblay** has the highest average order value **at $11.11**, followed closely by **Helena Holý and Robert Brown**.
- Most customers have an average order value **between $7 and $9**, indicating a **moderate spending range**.
- **Kara Nielsen and Mark Philips** have the **lowest average order values**, suggesting they make smaller or less frequent purchases.
- A few customers, like **Michelle Brooks and Fynn Zimmermann**, have relatively **high average order values** around **$9**, highlighting their higher spending per transaction.

❖ **Calculate the total revenue generated by each customer:**

```sql
-- Calculate the total revenue generated by each customer
SELECT
  c.customer_id,
  CONCAT(c.first_name, ' ', c.last_name) as customers,
    SUM(i.total) AS total_revenue
FROM customer c
INNER JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY c.customer_id, customers
ORDER BY total_revenue desc;
```

OUTPUT TABLE:

Result Grid | Filter Rows:

customer_id	customers	total_revenue
5	František Wichterlová	144.54
6	Helena Holý	128.70
46	Hugh O'Reilly	114.84
58	Manoj Pareek	111.87
1	Luís Gonçalves	108.90
13	Fernanda Ramos	106.92
34	João Fernandes	102.96
3	François Tremblay	99.99
42	Wyatt Girard	99.99
53	Phil Hughes	98.01
17	Jack Smith	98.01
50	Enrique Muñoz	98.01
57	Luis Rojas	97.02
20	Dan Miller	95.04
37	Fynn Zimmermann	94.05

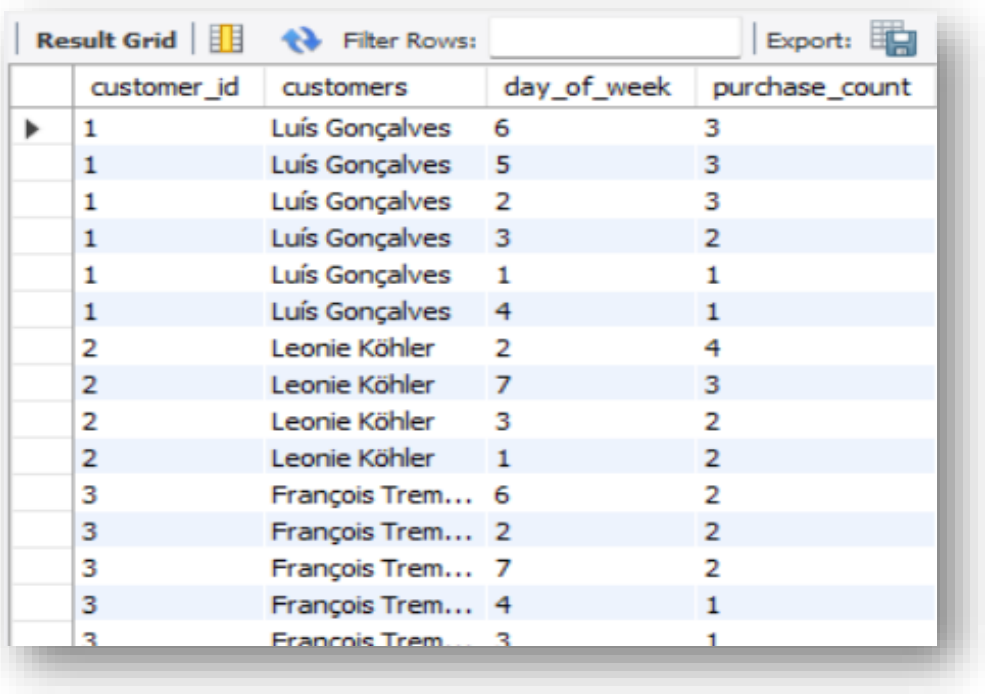
INSIGHTS:

- **František Wichterlová** leads with the **highest total revenue of $144.54**, followed by Helena Holý and Hugh O'Reilly.
- The top 10 customers all have **total revenues above $98**, indicating strong purchasing behaviour.
- Most customers have total revenues **between $60 and $100**, showing consistent spending across the customer base.
- Lower revenue customers, such as **Diego Gutiérrez and Kara Nielsen**, have total revenues **below $40**, suggesting lower engagement or fewer purchases.

❖ **Identify the preferred purchase periods:**

```sql
-- Identify the preferred purchase periods
SELECT
    c.customer_id,
  CONCAT(c.first_name, ' ', c.last_name) as customers,
    DAYOFWEEK(i.invoice_date) AS day_of_week,
    COUNT(i.invoice_id) AS purchase_count
FROM customer c
INNER JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY c.customer_id, c.first_name, c.last_name, DAYOFWEEK(i.invoice_date)
ORDER BY c.customer_id, customers, purchase_count DESC;
```

OUTPUT TABLE:



Result Grid | Filter Rows: | Export:

customer_id	customers	day_of_week	purchase_count
1	Luís Gonçalves	6	3
1	Luís Gonçalves	5	3
1	Luís Gonçalves	2	3
1	Luís Gonçalves	3	2
1	Luís Gonçalves	1	1
1	Luís Gonçalves	4	1
2	Leonie Köhler	2	4
2	Leonie Köhler	7	3
2	Leonie Köhler	3	2
2	Leonie Köhler	1	2
3	François Trem...	6	2
3	François Trem...	2	2
3	François Trem...	7	2
3	François Trem...	4	1
3	François Trem...	3	1

INSIGHTS:

- Most customers have their highest purchase counts spread across multiple days of the week, indicating varied shopping habits.
- **František Wichterlová and Dan Miller** are among the **top** with significant purchases on specific days, showing concentrated buying behavior.
- Several customers, like **Luís Gonçalves and Hugh O'Reilly**, make purchases consistently across different days, indicating steady engagement.
- **Weekends (day 7)** show notable purchases for many customers, suggesting increased activity during weekends.

8. What is the customer churn rate?

➢

```sql
WITH MostRecentInvoice AS (
    SELECT MAX(invoice_date) AS most_recent_invoice_date
    FROM invoice
),
CutoffDate AS (
    SELECT DATE_SUB(most_recent_invoice_date, INTERVAL 1 YEAR) AS cutoff_date
    FROM MostRecentInvoice
),
ChurnedCustomers AS (
    SELECT
        c.customer_id,
        COALESCE(c.first_name, ' ',c.last_name) as customers,
        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, customers
    HAVING MAX(i.invoice_date) IS NULL OR MAX(i.invoice_date) < (SELECT cutoff_date FROM CutoffDate)
)
-- ******** Calculate the churn rate ********
SELECT (SELECT COUNT(*) FROM ChurnedCustomers) / (SELECT COUNT(*) FROM customer) * 100 AS churn_rate;
```

OUTPUT TABLE:

Result Grid

	churn_rate
▶	1.6949

INSIGHTS:

- The customer **churn rate is 1.6949**, indicating that approximately **1.7%** of customers are leaving over a specific period.
- **A churn rate below 2%** suggests relatively low customer turnover, indicating stable customer retention.
- Maintaining a low churn rate is crucial for long-term business growth and customer loyalty.
- Strategies to further reduce churn could involve improving customer engagement, enhancing service quality, and addressing customer feedback.

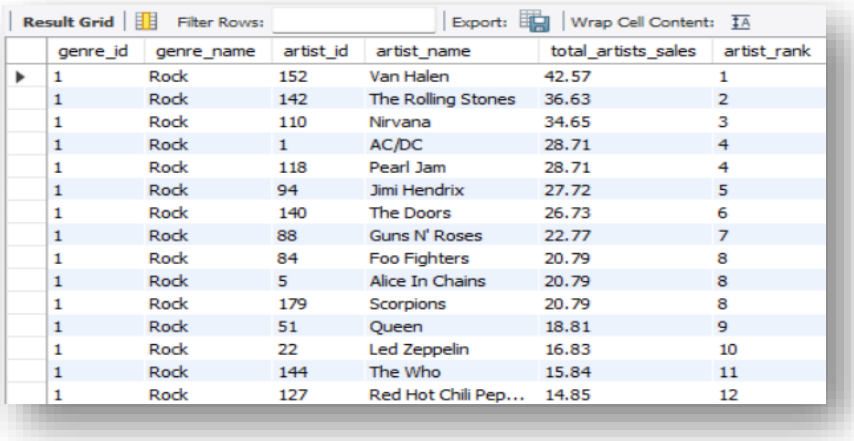
9. Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.

➤

```sql
WITH genre_sales_in_usa AS (
  SELECT
    g.genre_id,
    g.name AS genre_name,
    SUM(il.unit_price * il.quantity) AS total_genre_sales
  FROM genre g
  INNER JOIN track t ON g.genre_id = t.genre_id
  INNER JOIN invoice_line il ON t.track_id = il.track_id
  INNER JOIN invoice i ON il.invoice_id = i.invoice_id
  INNER JOIN customer c ON i.customer_id = c.customer_id
  WHERE c.country = 'USA'
  GROUP BY g.genre_id, g.name
),
total_sales as(
  SELECT
    SUM(total_genre_sales) as total_usa_sales
  FROM genre_sales_in_usa
),

genre_sales_percentage AS(
  SELECT
    gs.genre_id,
        gs.genre_name,
        gs.total_genre_sales,
        ts.total_usa_sales,
        (gs.total_genre_sales/ts.total_usa_sales) * 100 AS percentage_contribution
  FROM genre_sales_in_usa gs
  CROSS JOIN total_sales ts
),
best_selling_artist AS (
  SELECT
    g.genre_id,
        g.name AS genre_name,
        a.artist_id,
        a.name AS artist_name,
        SUM(il.unit_price * il.quantity) AS total_artists_sales
  FROM genre g
  INNER JOIN track t ON g.genre_id = t.genre_id
    INNER JOIN album al ON al.album_id = t.album_id
    INNER JOIN artist a ON a.artist_id = al.artist_id
    INNER JOIN invoice_line il ON il.track_id = t.track_id
    INNER JOIN invoice i ON i.invoice_id = il.invoice_id
    INNER JOIN customer c ON c.customer_id = i.customer_id
  WHERE c.country = 'USA'
  GROUP BY g.genre_id, g.name, a.artist_id, a.name
)
SELECT
  genre_id,
  genre_name,
  artist_id,
  artist_name,
  total_artists_sales,
  DENSE_RANK() OVER (PARTITION BY genre_id ORDER BY total_artists_sales DESC) AS artist_rank
FROM best_selling_artist;
```

OUTPUT TABLE:



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

genre_id	genre_name	artist_id	artist_name	total_artists_sales	artist_rank
1	Rock	152	Van Halen	42.57	1
1	Rock	142	The Rolling Stones	36.63	2
1	Rock	110	Nirvana	34.65	3
1	Rock	1	AC/DC	28.71	4
1	Rock	118	Pearl Jam	28.71	4
1	Rock	94	Jimi Hendrix	27.72	5
1	Rock	140	The Doors	26.73	6
1	Rock	88	Guns N' Roses	22.77	7
1	Rock	84	Foo Fighters	20.79	8
1	Rock	5	Alice In Chains	20.79	8
1	Rock	179	Scorpions	20.79	8
1	Rock	51	Queen	18.81	9
1	Rock	22	Led Zeppelin	16.83	10
1	Rock	144	The Who	15.84	11
1	Rock	127	Red Hot Chili Pep...	14.85	12

INSIGHTS:

- **Van Halen leads Rock genre sales** with **42.57**, followed by T**he Rolling Stones and Nirvana**, showing their popularity.
- **Black Sabbath and System of a Down top Metal sales**, highlighting their dominance in the genre.
- **Eric Clapton is the highest-selling Blues artist** with significant sales compared to other Blues artists.
- **Amy Winehouse** is the **top-selling artist** in the **Pop and R&B/Soul genres**, indicating broad appeal across genres.
- **The Jazz and Classical genres** have relatively **low sales** compared to other genres, suggesting niche markets.

10. 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 customers,
    COUNT(DISTINCT g.genre_id) AS genre_count
FROM customer c
INNER JOIN invoice i ON c.customer_id = i.customer_id
INNER JOIN invoice_line il ON i.invoice_id = il.invoice_id
INNER JOIN track t ON il.track_id = t.track_id
INNER JOIN genre g ON t.genre_id = g.genre_id
GROUP BY c.customer_id, customers
HAVING COUNT(DISTINCT g.genre_id) >= 3
ORDER BY genre_count DESC;
```

OUTPUT TABLE:

Result Grid | Filter Rows:

customer_id	customers	genre_count
2	Leonie Köhler	14
5	František Wichterlová	13
44	Terhi Hämäläinen	13
35	Madalena Sampaio	13
22	Heather Leacock	13
30	Edward Francis	13
38	Niklas Schröder	12
23	John Gordon	12
46	Hugh O'Reilly	12
13	Fernanda Ramos	12
42	Wyatt Girard	12
17	Jack Smith	12
18	Michelle Brooks	12
41	Marc Dubois	12
28	Julia Barnett	12

INSIGHTS:

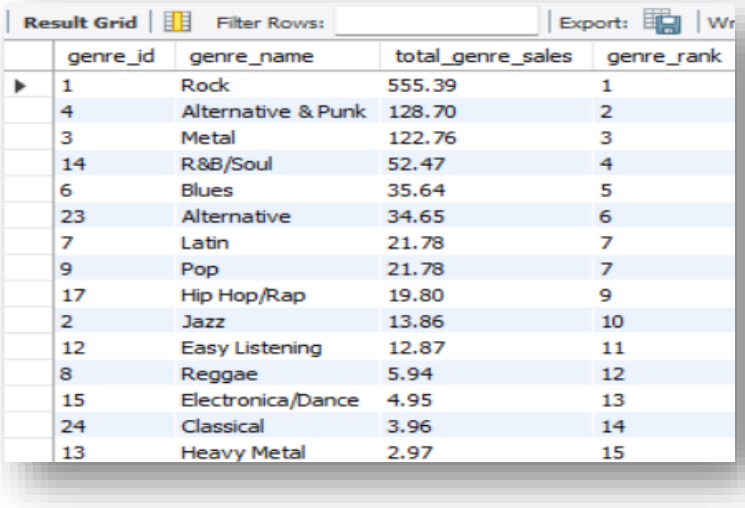
- **Leonie Köhler** explores the **most diverse range of genres**, engaging with **14 different types.**
- Several customers, including **František Wichterlová and Terhi Hämäläinen**, engage with **13 genres**, indicating a broad musical taste.
- A significant number of customers interact with **10 to 12 genres**, showing a preference for diverse music.
- Fewer customers, like **Robert Brown**, engage with only a **handful of genres**, suggesting more specialized tastes.

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

➢

```sql
WITH genre_sales_in_usa AS (
  SELECT
    g.genre_id,
    g.name AS genre_name,
    SUM(il.unit_price * il.quantity) AS total_genre_sales
  FROM genre g
  INNER JOIN track t ON g.genre_id = t.genre_id
  INNER JOIN invoice_line il ON t.track_id = il.track_id
  INNER JOIN invoice i ON il.invoice_id = i.invoice_id
  INNER JOIN customer c ON i.customer_id = c.customer_id
  WHERE c.country = 'USA'
  GROUP BY g.genre_id, g.name
)
SELECT
  genre_id,
    genre_name,
    total_genre_sales,
    RANK() OVER (ORDER BY total_genre_sales DESC) AS genre_rank
FROM genre_sales_in_usa
ORDER BY genre_rank;
```

OUTPUT TABLE:



Result Grid | Filter Rows: | Export: | Wr

genre_id	genre_name	total_genre_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

INSIGHTS:

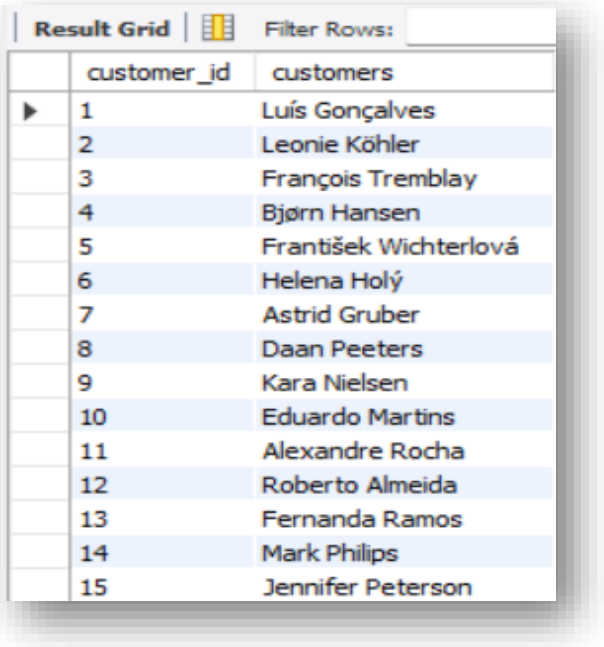
- **Rock is the top genre** by a significant margin, with **total sales of 555.39**, indicating its widespread popularity.
- **Alternative & Punk and Metal** follow as the **second and third most popular genres**, with **sales of 128.7 and 122.76, respectively**.
- **R&B/Soul and Blues** are **mid-ranked genres**, showing moderate popularity with **total sales of 52.47 and 35.64**.
- Genres like **Classical, Heavy Metal, Soundtrack, and TV Shows** have the **lowest sales**, reflecting niche or limited audience appeal.

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

➢

```sql
WITH recent_purchases AS (
  SELECT c.customer_id
  FROM customer c
  INNER JOIN invoice i ON c.customer_id = i.customer_id
  WHERE i.invoice_date >= CURDATE() - INTERVAL 3 MONTH
)
SELECT
  c.customer_id,
    CONCAT(c.first_name, ' ', c.last_name) as customers
FROM customer c
LEFT JOIN recent_purchases rp ON c.customer_id = rp.customer_id
WHERE rp.customer_id IS NULL
ORDER BY c.customer_id;
```

OUTPUT TABLE:



Result Grid | Filter Rows:

customer_id	customers
1	Luís Gonçalves
2	Leonie Köhler
3	François Tremblay
4	Bjørn Hansen
5	František Wichterlová
6	Helena Holý
7	Astrid Gruber
8	Daan Peeters
9	Kara Nielsen
10	Eduardo Martins
11	Alexandre Rocha
12	Roberto Almeida
13	Fernanda Ramos
14	Mark Philips
15	Jennifer Peterson

INSIGHTS:

- To identify customers who have not made a purchase in the last 3 months, we need data on their recent purchase dates.
- The list provides **59 customer names and IDs** but lacks specific transaction dates.
- Without recent transaction data, it's impossible to determine which customers are inactive.
- Additional data, such as a timestamp of the last purchase, is required to accurately identify inactive customers.

========== ***END OF OBJECTIVE QUESTIONS*** ==========