

# Objective Questions

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

- These are the columns in different tables with null values.

table_name	column_name	null_count
customer	company	49
customer	fax	47
customer	state	29
customer	postal_code	4
customer	phone	1
employee	reports_to	1
track	composer	978

- CUSTOMER Table:

Company: 49 null values.

Fax: 47 null values.

State: 29 null values.

Postal\_code: 4 null values.

- Employees Table:

Reports\_to: 1 null value.

- Track Table:

Composer: 978 null values.

- For duplicate values, the CHINOOK DB does not have any fully duplicate rows. Every table in the DB has primary key, it ensures each row in a table is unique. The DB system will automatically prevent any attempt to insert a row that would duplicate a primary key.
- The existing null value columns can be handled using COALESCE function in SQL.

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

#### A. Top Selling Tracks in USA

- **Approach:**

- Joined track, invoice, invoice\_line, album, artist, genre tables.
- Applied where clause on country as USA.
- The applied group by on track\_name, then calculated the total\_units\_sold and ordered the units sold in descending order and limited the results to 10.

## SQL Query:

```
▶ SELECT
    t.track_id,
    t.name as track_name,
    g.name as genre,
    a.name as artist_name,
    sum(il.quantity)total_units_sold
  FROM invoice i
  JOIN invoice_line il
  on i.invoice_id=il.invoice_id
  join track t
  on il.track_id=t.track_id
  join genre g
  on t.genre_id=g.genre_id
  join album al
  on t.album_id = al.album_id
  join artist a
  on al.artist_id = a.artist_id
  where i.billing_country="USA"
  group by t.track_id,t.name,g.name,a.name
  order by total_units_sold desc,t.track_id
  limit 10;
```

## Output:

track_id	track_name	genre	artist_name	total_units_sold
3336	War Pigs	Alternative	Cake	6
3465	You Know I'm No Good (feat. Ghostface Killah)	R&B/Soul	Amy Winehouse	5
13	Night Of The Long Knives	Rock	AC/DC	4
153	Evil Woman	Metal	Black Sabbath	4
1495	Highway Chile	Rock	Jimi Hendrix	4
1995	Scentless Apprentice	Rock	Nirvana	4
2560	Violent Pornography	Metal	System Of A Down	4
2646	I Looked At You	Rock	The Doors	4
2647	End Of The Night	Rock	The Doors	4
16	Dog Eat Dog	Rock	AC/DC	3

## **Insights:**

- Top selling tracks in USA are War Pigs, You Know Iam No Good, Night Of the long knives, EvilWomen, Highway Chile...etc. We can see majority of top tracks come in Rock genre.

## **B. Top Artist in USA by total tracks sold**

### **• Approach:**

- Joined the Invoice, invoice\_line, track, album, and artist tables.
- Applied where clause on country as USA.
- Applied Group by on ArtistName.
- Applied where clause on country as USA.
- Applied Order by on TotalTracksSold in desc order.
- Applied Limit of 10 rows.

## SQL Query:

```
SELECT
    ar.name AS ArtistName,
    SUM(il.Quantity) AS TotalTracksSold
FROM invoice_line AS il
JOIN Invoice AS I
    ON IL.Invoice_Id = I.Invoice_Id
JOIN Track AS T
    ON IL.Track_Id = T.Track_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'
GROUP BY
    ArtistName
ORDER BY
    TotalTracksSold DESC
LIMIT 10;
```

## **Output:**

ArtistName	TotalTracksSold
Van Halen	43
R.E.M.	38
The Rolling Stones	37
Nirvana	35
Foo Fighters	34
Eric Clapton	34
Guns N' Roses	32
Green Day	32
Pearl Jam	31
Amy Winehouse	30

## **Insights:**

- The top artists are Van Halen,REM,The Rolling Stones, Nirvana,Foo Fighters..etc.

## **C. Most Popular Genre in USA by Units sold.**

- **Approach:**
  - Joined invoice\_line,track,invoice and genre tables.
  - Applied group by on genere\_name.
  - Applied where clause on country as USA.
  - Applied order by on total\_revenue in desc order.
  - Applied limit of 10 rows.

## SQL Query:

```
SELECT
    g.genre_id,
    g.name as genre_name,
    COUNT(DISTINCT t.track_id) as unique_tracks_sold,
    SUM(il.quantity) as total_units_sold,
    SUM(il.quantity * il.unit_price) as total_revenue
FROM invoice_line il
JOIN track t ON il.track_id = t.track_id
JOIN invoice i ON il.invoice_id = i.invoice_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE i.billing_country = 'USA'
GROUP BY g.genre_id, g.name
ORDER BY total_revenue desc
LIMIT 10;
```

## Output:

genre_id	genre_name	unique_tracks_sold	total_units_sold	total_revenue
1	Rock	418	561	555.39
4	Alternative & Punk	95	130	128.70
3	Metal	87	124	122.76
14	R&B/Soul	36	53	52.47
6	Blues	28	36	35.64
23	Alternative	23	35	34.65
7	Latin	22	22	21.78
9	Pop	19	22	21.78
17	Hip Hop/Rap	19	20	19.80
2	Jazz	14	14	13.86

## Insights:

- The top genres are Rock with a huge revenue and units sold, followed by Alternative & Punk, Metal, R&B Soul ...etc.

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

- **Approach:**
  - Used Customers table, Applied group by on country.
  - Calculated count of customers in each country using COUNT() function.
  - Applied Order by on customer\_count in desc order.

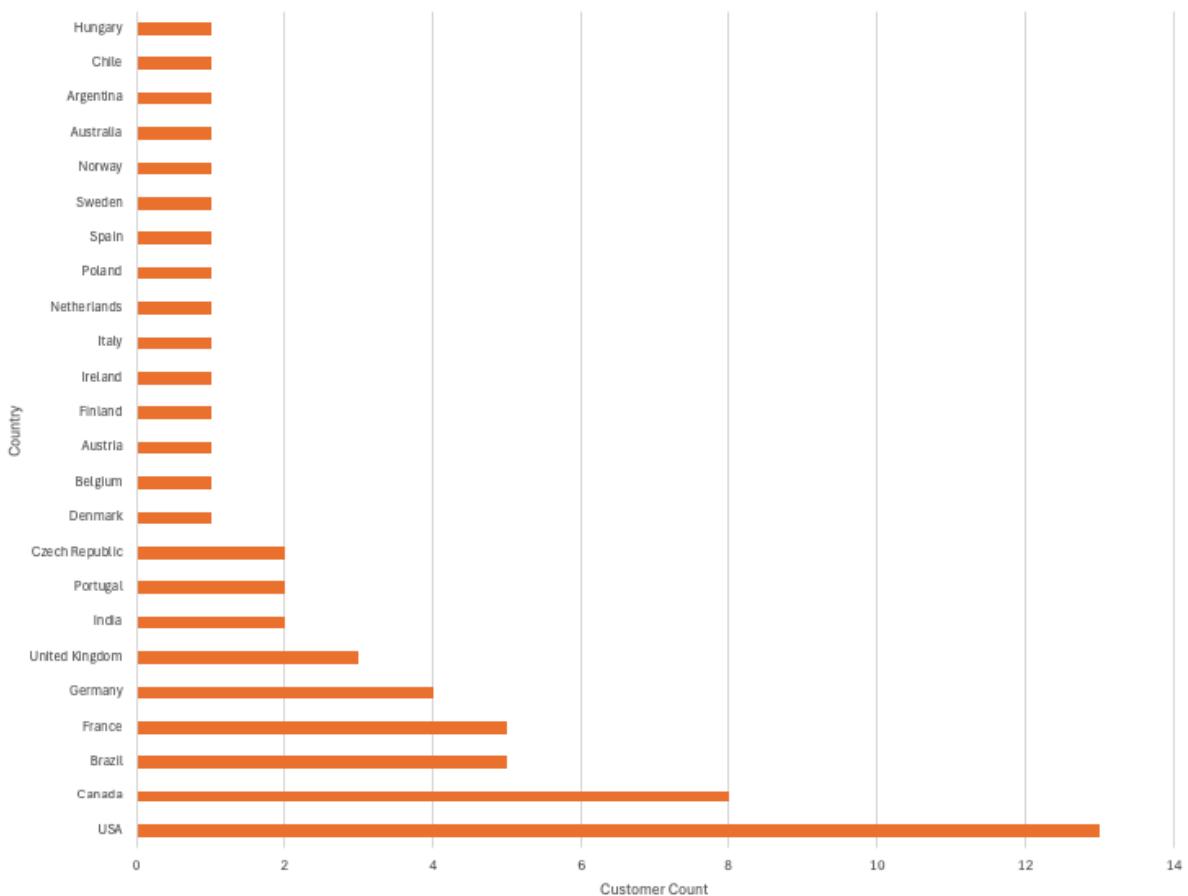
**SQL Query:**

```
SELECT country,COUNT(*) customer_count,
ROUND(COUNT(customer_id) * 100.0 / (SELECT COUNT(*) FROM customer), 2) as percentage
FROM CUSTOMER
GROUP BY country
ORDER BY customer_count desc;
```

**Output:**

country	customer_count	percentage
USA	13	22.03
Canada	8	13.56
Brazil	5	8.47
France	5	8.47
Germany	4	6.78
United Kingdom	3	5.08
India	2	3.39
Portugal	2	3.39
Czech Republic	2	3.39
Denmark	1	1.69
Belgium	1	1.69
Austria	1	1.69
Finland	1	1.69
Ireland	1	1.69
Italy	1	1.69
Netherlands	1	1.69
Poland	1	1.69
Spain	1	1.69
Sweden	1	1.69
Norway	1	1.69
Australia	1	1.69
Argentina	1	1.69
Chile	1	1.69
Hungary	1	1.69

Country Vs Customer Count



## Insights:

- Since we don't have age or gender of customers in our dataset. However, we have the location information of customers, so we will use country wise customer demography.
- There are 24 countries where chinook customers are from, In 53 different cities.
- USA has highest number of customers.

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

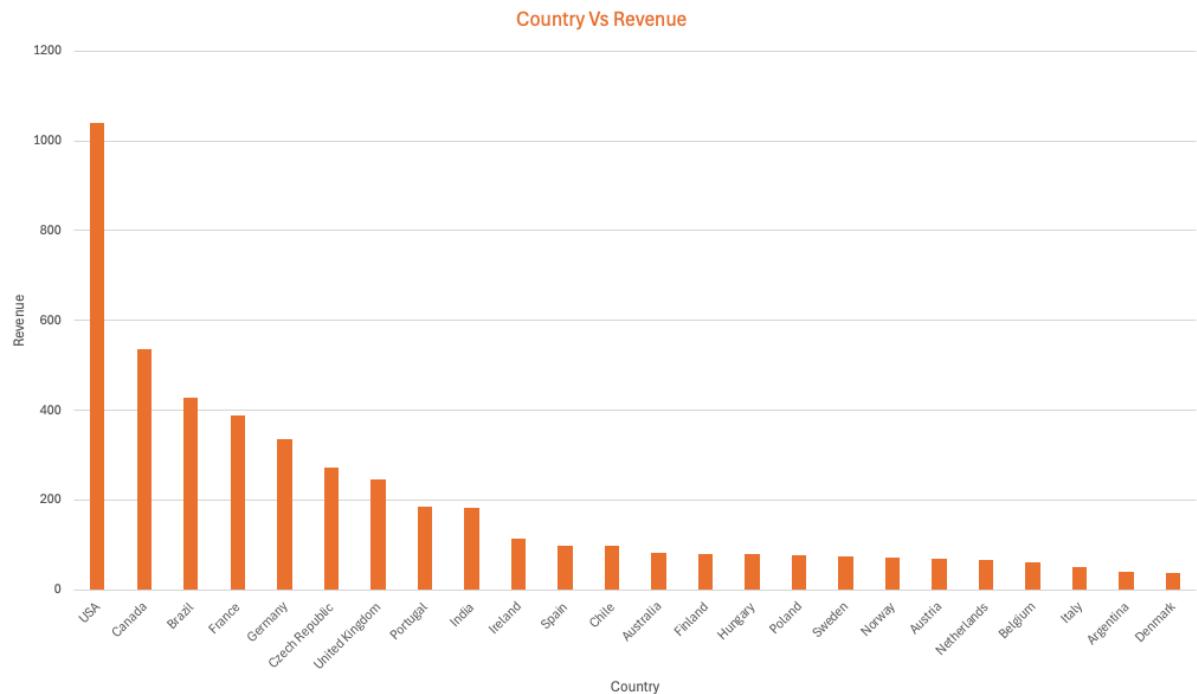
- **Approach:**
  - To find total revenue and number of invoices for each country, state and city. The query includes grouping the results with country, state and city, also using COALESCE to show “Not Specified” for null values in city and state columns. Also used aggregate functions like SUM and ROUND to find the total revenue.
  -

**SQL Query:**

```
SELECT
    billing_country as country,
    COALESCE(billing_state, 'Not Specified') as state,
    COALESCE(billing_city, 'Not Specified') as city,
    COUNT(invoice_id) as number_of_invoices,
    ROUND(SUM(total), 2) as total_revenue
FROM invoice
GROUP BY billing_country, billing_state, billing_city
ORDER BY total_revenue DESC;
```

## Output:

country	state	city	number_of_invoices	total_revenue
Czech Republic	None	Prague	30	273.24
USA	CA	Mountain View	20	169.29
United Kingdom	None	London	19	166.32
Germany	None	Berlin	20	158.40
France	None	Paris	18	151.47
Brazil	SP	São Paulo	22	129.69
Ireland	Dublin	Dublin	13	114.84
India	None	Delhi	13	111.87
Brazil	SP	São José dos Campos	13	108.90
Brazil	DF	Brasília	15	106.92
Portugal	None	Lisbon	13	102.96
France	None	Bordeaux	11	99.99
Canada	QC	Montréal	9	99.99
Spain	None	Madrid	11	98.01
USA	WA	Redmond	12	98.01
Chile	None	Santiago	13	97.02
Germany	None	Frankfurt	10	94.05
USA	FL	Orlando	12	92.07
Canada	ON	Ottawa	13	91.08
USA	NV	Reno	11	91.08
USA	TX	Fort Worth	12	86.13
USA	AZ	Tucson	9	84.15
Germany	None	Stuttgart	11	82.17
Brazil	RJ	Rio de Janeiro	11	82.17



## Insights:

- Prague in Czech Republic has the highest revenue 273.24 with 30 invoices.Followed by Mt.View in USA, London in UK and Berlin in Germany.
- USA , Canada and Brazil are the highest revenue markets globally.

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

- **Approach:**
  - Joined invoice and customer tables.
  - Applied Group by on customer
  - Used CTE (common table expression).
  - Applied RANK() window function, partition by on country and order by sum of total amount from invoice.

## SQL Query:

```
) WITH customer_revenue_by_country AS (
    SELECT
        c.customer_id,
        concat(c.first_name, " ", c.last_name) AS customer_name,
        c.country,
        c.city,
        COUNT(i.invoice_id) AS total_invoices,
        ROUND(SUM(i.total), 2) AS total_revenue,
        RANK() OVER (PARTITION BY c.country ORDER BY SUM(i.total) DESC) AS revenue_rank
    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, c.city, c.company
)
```

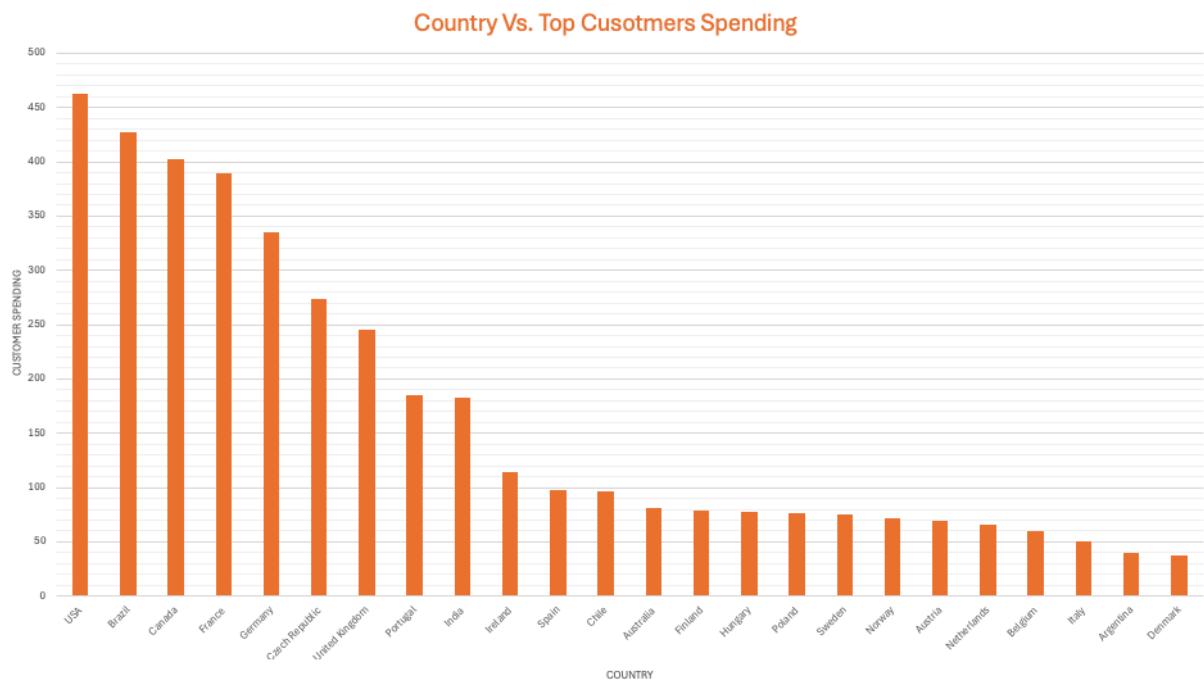
```

SELECT
    country,
    customer_name,
    city,
    total_invoices,
    total_revenue,
    revenue_rank
FROM customer_revenue_by_country
WHERE revenue_rank <= 5
ORDER BY country, revenue_rank;

```

## **Output:**

country	customer_name	city	total_invoices	total_revenue	revenue_rank
Argentina	Diego Gutiérrez	Buenos Aires	5	39.60	1
Australia	Mark Taylor	Sidney	10	81.18	1
Austria	Astrid Gruber	Vienne	9	69.30	1
Belgium	Daan Peeters	Brussels	7	60.39	1
Brazil	Luís Gonçalves	São José dos Campos	13	108.90	1
Brazil	Fernanda Ramos	Brasília	15	106.92	2
Brazil	Roberto Almeida	Rio de Janeiro	11	82.17	3
Brazil	Alexandre Rocha	São Paulo	10	69.30	4
Brazil	Eduardo Martins	São Paulo	12	60.39	5
Canada	François Tremblay	Montréal	9	99.99	1
Canada	Edward Francis	Ottawa	13	91.08	2
Canada	Ellie Sullivan	Yellowknife	12	75.24	3
Canada	Aaron Mitchell	Winnipeg	8	70.29	4



### Insights:

- We can see the top 5 highest spending customers in each country, the chart shows how the spending habit of top customers from each country differs.
- Customers from USA, Brazil and Canada have the highest spending capacity.

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

- **Approach:**
  - Joined customer, invoice, invoice\_line, track, album and artist tables, to get more info.
  - Aggregated total quantity sold per customer and revenue.
  - Applied group by on customer and track name, along with RANK() to rank top selling tracks.

## SQL Query:

```
WITH customer_track_purchases AS (
    SELECT
        c.customer_id,
        concat(c.first_name, " ", c.last_name) AS customer_name,
        t.track_id,
        t.name AS track_name,
        ar.name AS artist_name,
        g.name AS genre,
        COUNT(il.invoice_line_id) AS times_purchased,
        SUM(il.quantity) AS total_units,
        ROUND(SUM(il.quantity * il.unit_price), 2) AS total_spent,
        ROW_NUMBER() OVER (PARTITION BY c.customer_id ORDER BY SUM(il.quantity*il.unit_price) DESC, COUNT(il.invoice_line_id) DESC) AS track_rank
    FROM customer c
    JOIN invoice i
    ON c.customer_id = i.customer_id
    JOIN invoice_line il
    ON i.invoice_id = il.invoice_id
    JOIN track t
    ON il.track_id = t.track_id
    JOIN album al
    ON t.album_id = al.album_id
    JOIN artist ar
    ON al.artist_id = ar.artist_id
    JOIN genre g
    ON t.genre_id = g.genre_id
    GROUP BY c.customer_id, c.first_name, c.last_name,
        t.track_id, t.name, ar.name, g.name
)
-----
```

```
SELECT
    customer_id,
    customer_name,
    track_name,
    artist_name,
    genre,
    times_purchased,
    total_units,
    total_spent
FROM customer_track_purchases
WHERE track_rank = 1
ORDER BY customer_name;
```

## Output:

customer_id	customer_name	track_name	artist_name	genre	times_purchased	total_units	total_spent
32	Aaron Mitchell	Please Please Please	James Brown	R&B/Soul	1	1	0.99
11	Alexandre Rocha	Come Back	Pearl Jam	Alternative & Punk	1	1	0.99
7	Astrid Gruber	Always On The Run	Lenny Kravitz	Rock	1	1	0.99
4	Bjørn Hansen	Hit The Lights	Metallica	Metal	1	1	0.99
39	Camille Bernard	Say Hello 2 Heaven	Temple of the Dog	Alternative	1	1	0.99
8	Daan Peeters	Give Peace a Chance	U2	Pop	1	1	0.99
20	Dan Miller	I Like Dirt	Red Hot Chili Peppers	Rock	1	1	0.99
56	Diego Gutiérrez	Make It Funky Pt.1	James Brown	R&B/Soul	1	1	0.99
40	Dominique Lefebvre	Black Hole Sun	Soundgarden	Rock	1	1	0.99
10	Eduardo Martins	Like A Bird	Black Label Society	Metal	2	2	1.98
30	Edward Francis	The Unforgiven	Apocalyptica	Metal	1	1	0.99

## Insights:

- Each customer has more than one track as top selling because most of the tracks are purchased one time by a single time.

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

- **Approach:**
  - The data set doesn't contain any data about preferred payment method. So we have utilized the aggregates of customer purchase data.
  - Joined customer and invoice table.
  - Applied group by on customer data.
  - Calculated the purchase metrics like totalorders, total\_spent, average order value, the tenure days and day since last purchase.

## SQL Query:

```
WITH customer_purchase_patterns AS (
    SELECT
        c.customer_id,
        CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
        c.country,
        -- Purchase metrics
        COUNT(i.invoice_id) AS total_orders,
        ROUND(SUM(i.total), 2) AS total_spent,
        ROUND(AVG(i.total), 2) AS avg_order_value,
        -- Time-based metrics
        DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS customer_tenure_days,
        -- Recency
        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_name,
    country,
    total_orders,
    total_spent,
    avg_order_value,
    customer_tenure_days,
    days_since_last_purchase
FROM customer_purchase_patterns
ORDER BY total_spent DESC;
```

## Output:

customer_name	country	total_orders	total_spent	avg_order_value	customer_tenure_days	days_since_last_purchase
František Wichterlová	Czech Republic	18	144.54	8.03	1263	1800
Helena Holy	Czech Republic	12	128.70	10.73	1149	1820
Hugh O'Reilly	Ireland	13	114.84	8.83	1393	1770
Manoj Pareek	India	13	111.87	8.61	1240	1920
Luis Gonçalves	Brazil	13	108.90	8.38	1275	1911
Fernanda Ramos	Brazil	15	106.92	7.13	1290	1784
João Fernandes	Portugal	13	102.96	7.92	1326	1830
François Tremblay	Canada	9	99.99	11.11	1180	1980
Wyatt Girard	France	11	99.99	9.09	1406	1785
Enrique Muñoz	Spain	11	98.01	8.91	1181	1955
Jack Smith	USA	12	98.01	8.17	1146	1862
Phil Hughes	United Kingdom	11	98.01	8.91	1420	1752

## **Insights:**

- We have frequent customers who have orders more than 10 times. The analysis also shows the down trend that recent purchase or activity of customers are so low. Recently there are no purchases, the most recent purchase on the store was 1740 days old.
- The tenure days of all customers are high which shows long-term customer loyalty, customer retention...etc.
- The average order value of top customers indicate how much the customers are willing to pay.

## **Customer Purchase trends based on geographics data:**

- **Approach:**
  - Used invoice table data.
  - Applied group by on billing\_country.
  - Calculated aggregates such as total\_order, total\_revenue, average order value, max\_order\_value, min order value, unique customers, orders per customer and revenue per customer.

## **SQL Query:**

```
SELECT
    billing_country as country,
    COUNT(invoice_id) as total_orders,
    ROUND(SUM(total), 2) as total_revenue,
    ROUND(AVG(total), 2) as avg_order_value,
    ROUND(MAX(total), 2) as max_order_value,
    ROUND(MIN(total), 2) as min_order_value,
    COUNT(DISTINCT customer_id) as unique_customers,
    ROUND(COUNT(invoice_id) * 1.0 / COUNT(DISTINCT customer_id), 2) as orders_per_customer,
    ROUND(SUM(total) / COUNT(DISTINCT customer_id), 2) as revenue_per_customer
FROM invoice
GROUP BY billing_country
ORDER BY total_revenue DESC;
```

## Output:

country	total_orders	total_revenue	avg_order_value	max_order_value	min_order_value	unique_customers	orders_per_cus
USA	131	1040.49	7.94	18.81	0.99	13	10.08
Canada	76	535.59	7.05	19.80	0.99	8	9.50
Brazil	61	427.68	7.01	17.82	0.99	5	12.20
France	50	389.07	7.78	23.76	0.99	5	10.00
Germany	41	334.62	8.16	16.83	0.99	4	10.25
Czech Republic	30	273.24	9.11	19.80	0.99	2	15.00
United Kingdom	28	245.52	8.77	15.84	1.98	3	9.33
Portugal	29	185.13	6.38	16.83	0.99	2	14.50
India	21	183.15	8.72	16.83	2.97	2	10.50
Ireland	13	114.84	8.83	17.82	0.99	1	13.00
Spain	11	98.01	8.91	15.84	0.99	1	11.00
Chile	13	97.02	7.46	14.85	0.99	1	13.00

## Insights:

- The data based on customer behaviour geographically shows that customer from USA, CANADA and BRAZIL have high revenue purchases.
- Customers from Chez Republic, Ireland and Spain delivers high revenue per customers.
- USA, Canada, Brazil and France has high number of customers.
- Customers from France, Canada and Chez consistently generate the highest average order values, indicating strong spending capacity in this market.
- Customers from Belgium, India and Finland account for the highest volume of orders in the economy price tier, suggesting different purchasing patterns compared to other regions.

## 8. What is the customer churn rate?

- **Approach:**

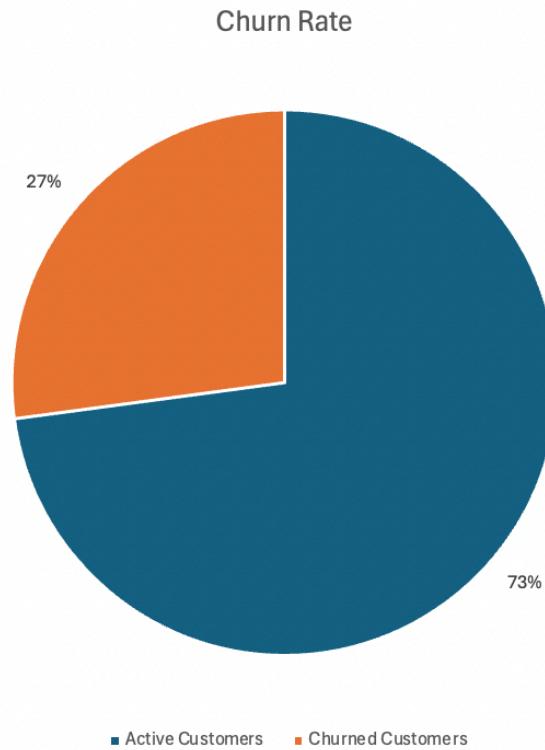
- We will take the latest date from invoice table.
- Joined customers and invoice table.(considering customer as churned if customer has not purchased anything within last 180 days from latest invoice date).
- Utilized case statement to categorize churned and active customers.

### SQL Query

```
WITH latest_date AS (
    SELECT MAX(invoice_date) as max_date FROM invoice
),
customer_status AS (
    SELECT
        c.customer_id,
        CASE
            WHEN MAX(i.invoice_date) >= DATE_SUB((SELECT max_date FROM latest_date) , INTERVAL 180 day )
            THEN 'Active'
            ELSE 'Churned'
        END as status
    FROM customer c
    LEFT JOIN invoice i ON c.customer_id = i.customer_id
    GROUP BY c.customer_id
)
SELECT
    COUNT(*) as total_customers,
    SUM(CASE WHEN status = 'Active' THEN 1 ELSE 0 END) as active_customers,
    SUM(CASE WHEN status = 'Churned' THEN 1 ELSE 0 END) as churned_customers,
    ROUND((SUM(CASE WHEN status = 'Churned' THEN 1 ELSE 0 END) * 100.0 / COUNT(*)), 2) as churn_rate_percentage
FROM customer_status;
```

### Output:

total_customers	active_customers	churned_customers	churn_rate_percentage
59	43	16	27.12



### Insights:

- I used each customers last purchase date, and considered customer whose last purchase was before 180 days from current date is considered as churned.
- The overall churn rate is 27%, affecting more than quarter of the customer base, this require targeted retention strategies.
- We have a total of 59 customer, in that 43 customers are active and remaining 16 are considered to be churned as they are inactive for past 180 days or 6 months.
- A churn rate this high is severe and signals significant dissatisfaction with the service or product, unsustainable pricing, or strong competitive pressure.
- It necessitates immediate and drastic intervention in areas such as customer success, product value, and retention strategies, as the business model may be at risk if this trend continues.

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

### Total Sales by genre

- **Approach:**

- Joined the invoiceline, invoice, track and genre.
- Applied where clause for USA as country.
- Applied group by on genre.
- Calculated the total\_tracks\_sold and total\_revenue.

### SQL Query:

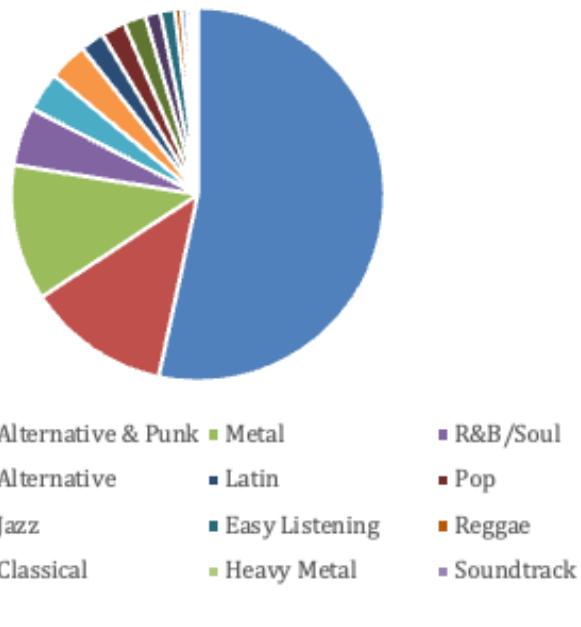
```
④ WITH usa_sales AS (
    SELECT
        g.genre_id,
        g.name AS genre_name,
        COUNT(il.invoice_line_id) AS total_tracks_sold,
        ROUND(SUM(il.quantity * il.unit_price), 2) AS total_revenue
    FROM invoice_line il
    JOIN invoice i ON il.invoice_id = i.invoice_id
    JOIN track t ON il.track_id = t.track_id
    JOIN genre g ON t.genre_id = g.genre_id
    WHERE i.billing_country = 'USA'
    GROUP BY g.genre_id, g.name
)

SELECT
    genre_name,
    total_tracks_sold,
    total_revenue,
    ROUND((total_revenue / (SELECT sum(total_revenue) FROM usa_sales )) * 100, 2) AS percentage_of_revenue
FROM usa_sales
ORDER BY total_revenue DESC;
```

## Output:

genre_name	total_tracks_sold	total_revenue	percentage_of_revenue
Rock	561	555.39	53.38
Alternative & Punk	130	128.70	12.37
Metal	124	122.76	11.80
R&B/Soul	53	52.47	5.04
Blues	36	35.64	3.43
Alternative	35	34.65	3.33
Latin	22	21.78	2.09
Pop	22	21.78	2.09
Hip Hop/Rap	20	19.80	1.90
Jazz	14	13.86	1.33
Easy Listening	13	12.87	1.24
Reggae	6	5.94	0.57
Electronica/Dance	5	4.95	0.48
Classical	4	3.96	0.38
Heavy Metal	3	2.97	0.29
Soundtrack	2	1.98	0.19
TV Shows	1	0.99	0.10

Genre / Revenue



## Insights:

- The analysis shows that Rock genre leads with 53.38% in revenue, followed by Alternative & Punk with 12.37% and Metal with 11.80% and so on.

## Best Selling Artists

- Approach:

- In each genre we can rank the artists with their total revenue.
- Joined invoice\_line, track, genre, album and artist tables.
- Used ROW\_NUMBER() function rank and then filter out top 5 artists in a particular genre.
- Calculated tracks\_sold, total\_units sold, total\_revenue and genre market share for an artist in that genre.

## SQL Query:

```
WITH genre_artist_sales AS (
    SELECT
        g.genre_id,
        g.name AS genre_name,
        ar.artist_id,
        ar.name AS artist_name,
        COUNT(il.invoice_line_id) AS tracks_sold,
        SUM(il.quantity) AS total_units,
        ROUND(SUM(il.quantity * il.unit_price), 2) AS total_revenue,
        ROW_NUMBER() OVER (PARTITION BY g.genre_id ORDER BY SUM(il.quantity * il.unit_price) DESC) AS artist_rank
    FROM invoice_line il
    JOIN invoice i ON il.invoice_id = i.invoice_id
    JOIN track t ON il.track_id = t.track_id
    JOIN genre g ON t.genre_id = g.genre_id
    JOIN album al ON t.album_id = al.album_id
    JOIN artist ar ON al.artist_id = ar.artist_id
    WHERE i.billing_country = 'USA'
    GROUP BY g.genre_id, g.name, ar.artist_id, ar.name
)
SELECT
    genre_name,
    artist_name,
    tracks_sold,
    total_units,
    total_revenue,
    ROUND((total_revenue / SUM(total_revenue) OVER (PARTITION BY genre_name)) * 100, 2) AS genre_market_share
FROM genre_artist_sales
WHERE artist_rank <= 5
ORDER BY genre_name, total_revenue DESC;
```

## **Output:**

genre_name	artist_name	tracks_sold	total_units	total_revenue	genre_market_share
Alternative	Chris Cornell	19	19	18.81	54.29
Alternative	Audioslave	7	7	6.93	20.00
Alternative	Cake	6	6	5.94	17.14
Alternative	Temple of the Dog	2	2	1.98	5.71
Alternative	Calexico	1	1	0.99	2.86
Alternative & Punk	R.E.M.	32	32	31.68	29.36
Alternative & Punk	Green Day	32	32	31.68	29.36
Alternative & Punk	Os Mutantes	19	19	18.81	17.43
Alternative & Punk	The Clash	13	13	12.87	11.93
Alternative & Punk	Foo Fighters	13	13	12.87	11.93
Blues	Eric Clapton	31	31	30.69	86.11
Blues	Buddy Guy	3	3	2.97	8.33
Blues	The Black Crowes	2	2	1.98	5.56
Classical	Luciano Pavarotti	1	1	0.99	25.00

## **Insights:**

- We can see many of the customers are dominating the genre market share, that show the customer interests towards this particular genre.

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

- **Approach:**
  - Joined customer, invoice, invoice\_line, track and genre tables.
  - Applied grouping on customer\_id and name.
  - Applied having on condition of customers with distinct genre count greater than 3.
  - Applied order by genre\_count in desc order.

**SQL Query:**

```
SELECT c.customer_id,
concat(c.first_name, " ", c.last_name)as customer_name ,
count(distinct g.genre_id) genre_count
FROM customer c
join invoice i
on c.customer_id = i.customer_id
join invoice_line il
on i.invoice_id=il.invoice_id
join track t
on t.track_id=il.track_id
join genre g
on g.genre_id=t.genre_id
group by c.customer_id,first_name
having genre_count>=3
order by genre_count desc;
```

**Output:**

customer_id	customer_name	genre_count
2	Leonie Köhler	14
44	Terhi Hämäläinen	13
22	Heather Leacock	13
35	Madalena Sampaio	13
5	František Wichterlová	13
30	Edward Francis	13
13	Fernanda Ramos	12
46	Hugh O'Reilly	12
42	Wyatt Girard	12

**Insights:**

- Many customers are interested in multiple variety of genres.
- Leonie Kohler is the customer who have interest in multiple different genres, purchased tracks from 14 different genres.

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

- **Approach:**
  - Joined invoice\_line, invoice, track and genre tables.
  - Applied group by on genre\_id and name.
  - Applied where clause with condition of USA as country.
  - Calculated total\_sales\_revenue and ranked on revenue.

## SQL Query:

```
SELECT
    g.genre_id,
    g.name as genre_name,
    ROUND(SUM(il.quantity * il.unit_price), 2) as total_sales_revenue,
    rank() over(order by sum(il.quantity*il.unit_price) desc) genre_rank
FROM invoice_line il
JOIN invoice i ON il.invoice_id = i.invoice_id
JOIN track t ON il.track_id = t.track_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE i.billing_country = 'USA'
GROUP BY g.genre_id, g.name
ORDER BY total_sales_revenue DESC;
```

## Output:

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

## Insights:

- Rock table has highest revenue in USA market.Followed by Alternative & Punk and Metal.

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

- **Approach:**
  - Joined customers and invoice table.
  - Applied group by on customer\_id.
  - Took latest date from invoice and considered that as current date.
  - Utilized date\_sub() function calculate the past 3 month date and filtered the customers.

### SQL Query:

```
with not_in as(
SELECT c.customer_id
FROM customer c
join invoice i
on c.customer_id=i.customer_id
group by c.customer_id
having max(invoice_date)>=(select date_sub(max(invoice_date),interval 3 month) from invoice)
)
SELECT customer_id,
CONCAT(first_name," ",last_name) full_name
FROM customer
WHERE customer_id not in(select * from not_in);
```

## Output:

customer_id	full_name
1	Luís Gonçalves
3	François Tremblay
4	Bjørn Hansen
7	Astrid Gruber
8	Daan Peeters
9	Kara Nielsen
10	Eduardo Martins
11	Alexandre Rocha
17	Jack Smith
18	Michelle Brooks
19	Tim Goyer
36	Hannah Schneider
37	Fynn Zimmermann
38	Niklas Schröder
39	Camille Bernard
43	Isabelle Mercier
48	Johannes Van der Berg
50	Enrique Muñoz
54	Steve Murray
56	Diego Gutiérrez
57	Luis Rojas
58	Manoj Pareek

- **Insights:**

- There are a total of 22 customers who have not made any purchase for past 3 months.

# 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.

## A. Approach:

- o First I analysed the top genres in USA. Then I took top performing genre the first 3 i.e Rock, Alternative & Punk and Metal have high number of tracks sold than rest of the genre in USA.
- o Now to find the new label albums I will take the latest 100 albums in the top performing genre.

## SQL Query:

```
with top_genre as(
SELECT
g.genre_id,
g.name AS genre_name,
SUM(il.quantity) AS tracks_sold
FROM invoice i
JOIN invoice_line il ON i.invoice_id = il.invoice_id
JOIN track t ON il.track_id = t.track_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE i.billing_country = 'USA'
GROUP BY g.genre_id
ORDER BY tracks_sold DESC
limit 3)
```

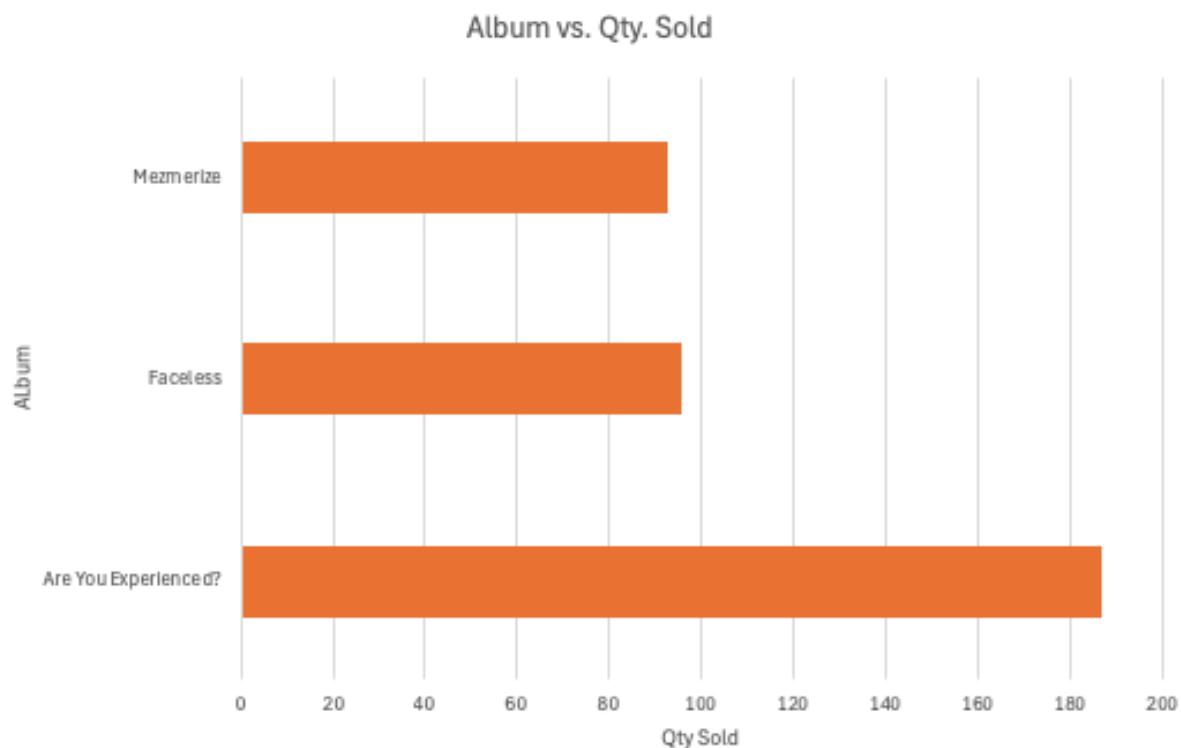
```

select al.title album_title,g.name Genre,ar.name artist_name,
sum(il.quantity) quantity_sold
from album al
join track t
on al.album_id =t.album_id
join genre g
on g.genre_id=t.genre_id
join artist ar
on al.artist_id=ar.artist_id
join invoice_line il
on il.track_id=t.track_id
where g.genre_id in (select genre_id from top_genre) and al.album_id >=(select max(al.album_id)-100 from album)
group by g.name,al.title,ar.name
order by quantity_sold desc
limit 3;

```

## Output:

album_title	Genre	artist_name	quantity_sold
Are You Experienced?	Rock	Jimi Hendrix	187
Faceless	Metal	Godsmack	96
Mezmerize	Metal	System Of A Down	93



## B. Insights:

- Based on the analysis on genre from USA Rock, Alternative & Punk and Metal are the top performing genres.
- Top artists are Jimi Hendrix, Godsmack and System of A Down.

## C. Recommendations:

- Based on both analysis , the following albums should be prioritized for advertising in the USA:
  - Are You Experienced? – Jimi Hendrix – Rock
  - Faceless – Godsmack – Metal
  - Mezmerize – System Of A Down - Metal
- 2. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

## A. Approach:

- To analyse the performance of genres in other countries than USA, I would check the sales of genres in other countries. This helps in determining where the global market is similar to US or have difference.

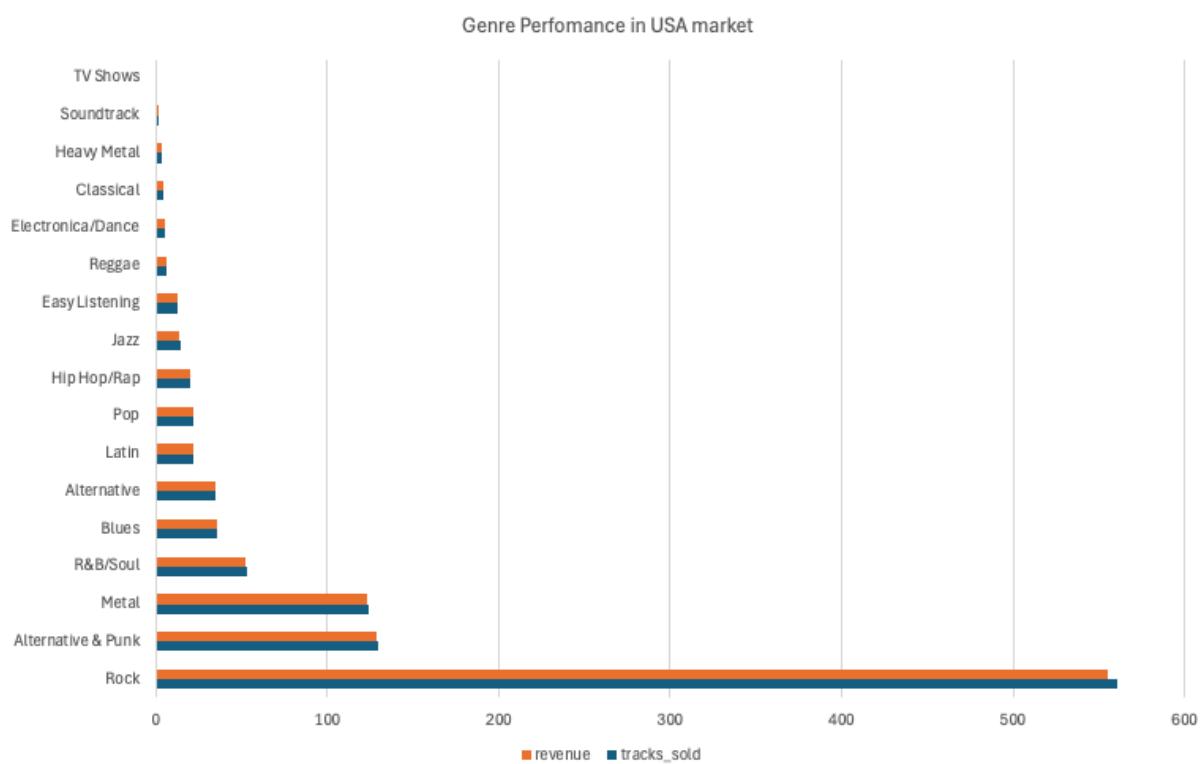
## Genre Performance in USA:

### SQL Query:

```
SELECT
    g.name AS genre_name,
    SUM(il.quantity) AS tracks_sold,
    ROUND(SUM(il.quantity * il.unit_price), 2) AS revenue,
    rank() over(order by sum(il.quantity* il.unit_price) desc) genre_rank
FROM invoice i
JOIN invoice_line il ON i.invoice_id = il.invoice_id
JOIN track t ON il.track_id = t.track_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE i.billing_country = 'USA'
GROUP BY g.genre_id
ORDER BY tracks_sold DESC;
```

## Output:

genre_name	tracks_sold	revenue	genre_rank
Rock	561	555.39	1
Alternative & Punk	130	128.70	2
Metal	124	122.76	3
R&B/Soul	53	52.47	4
Blues	36	35.64	5
Alternative	35	34.65	6
Latin	22	21.78	7



## Genre Performance in Other countries:

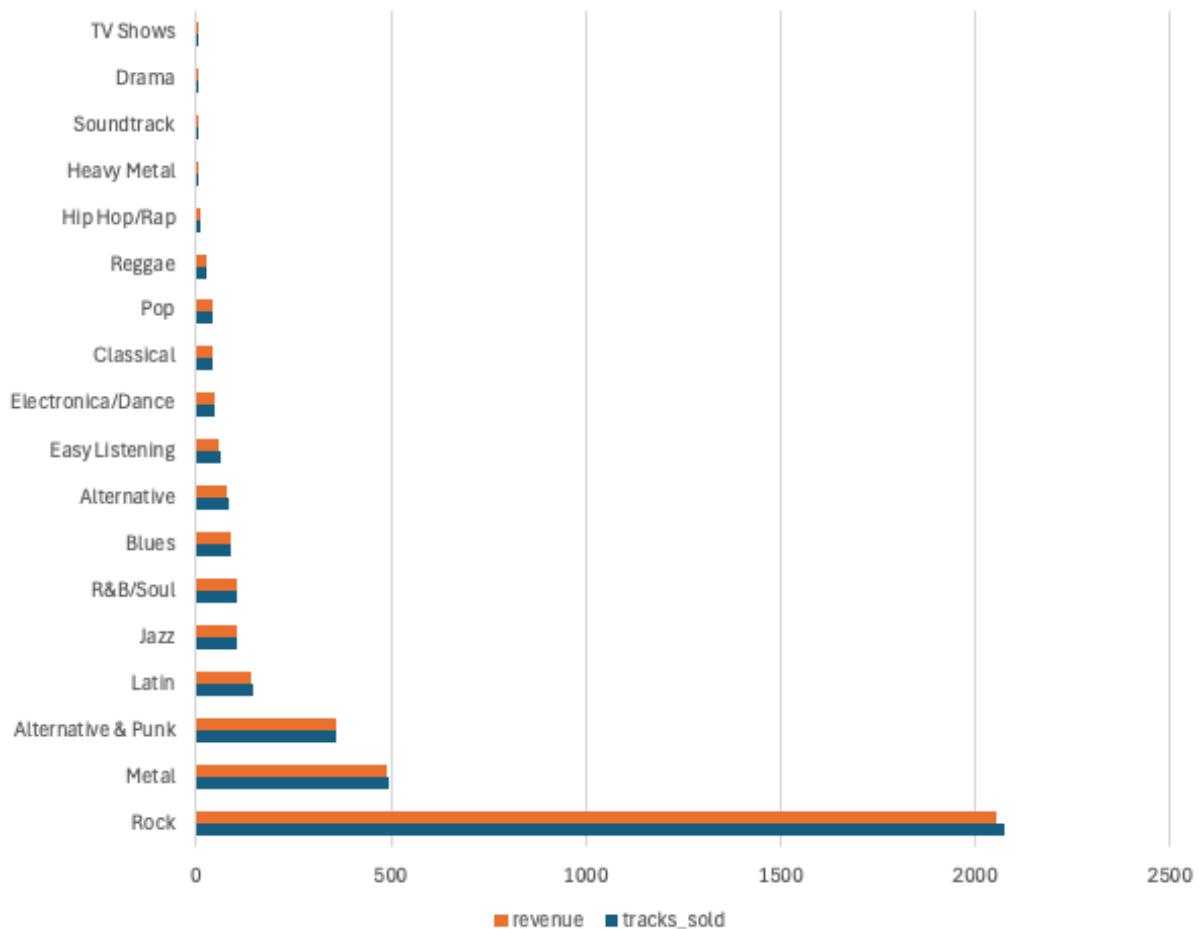
### SQL Query:

```
SELECT
    g.name AS genre_name,
    SUM(il.quantity) AS tracks_sold,
    ROUND(SUM(il.quantity * il.unit_price), 2) AS revenue,
    RANK() OVER(ORDER BY SUM(il.quantity * il.unit_price) DESC) genre_rank
FROM invoice i
JOIN invoice_line il ON i.invoice_id = il.invoice_id
JOIN track t ON il.track_id = t.track_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE i.billing_country != 'USA'
GROUP BY g.genre_id
ORDER BY tracks_sold DESC;
```

### Output:

genre_name	tracks_sold	revenue	genre_rank
Rock	2074	2053.26	1
Metal	495	490.05	2
Alternative & Punk	362	358.38	3
Latin	145	143.55	4
Jazz	107	105.93	5
R&B/Soul	106	104.94	6
Blues	88	87.12	7
Alternative	82	81.18	8

Genre Performance in other markets than USA



## B. Insights:

- Rock , Alternative & Punk and Metal remains as top 3 in both USA and other countries.
- Latin and Jazz are other top performing genre for the rest of the world. While USA has R&B/Soul and Blues in ranking

## C. Recommendations:

- Rock , Alternative Punk and Metal are performing well on USA and Global markets, so albums and artists in this genre can be prioritized for global marketing.
- To increase revenue in other markets, Jazz and Latin genre can be considered for promoting in region and culturally relevant geographical locations.
- Geography and cultural relevant recommendations can help increase in engagement and revenue across the markets.

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

#### A. Approach:

- Since we have invoice data of customers available from '2017-01-03' to '2020-12-30', to segregate long-term and new users we can consider users from (2017-01-03)-(2018-12-31) as long-term users else new. Then we can take the aggregate result to analyse the data.

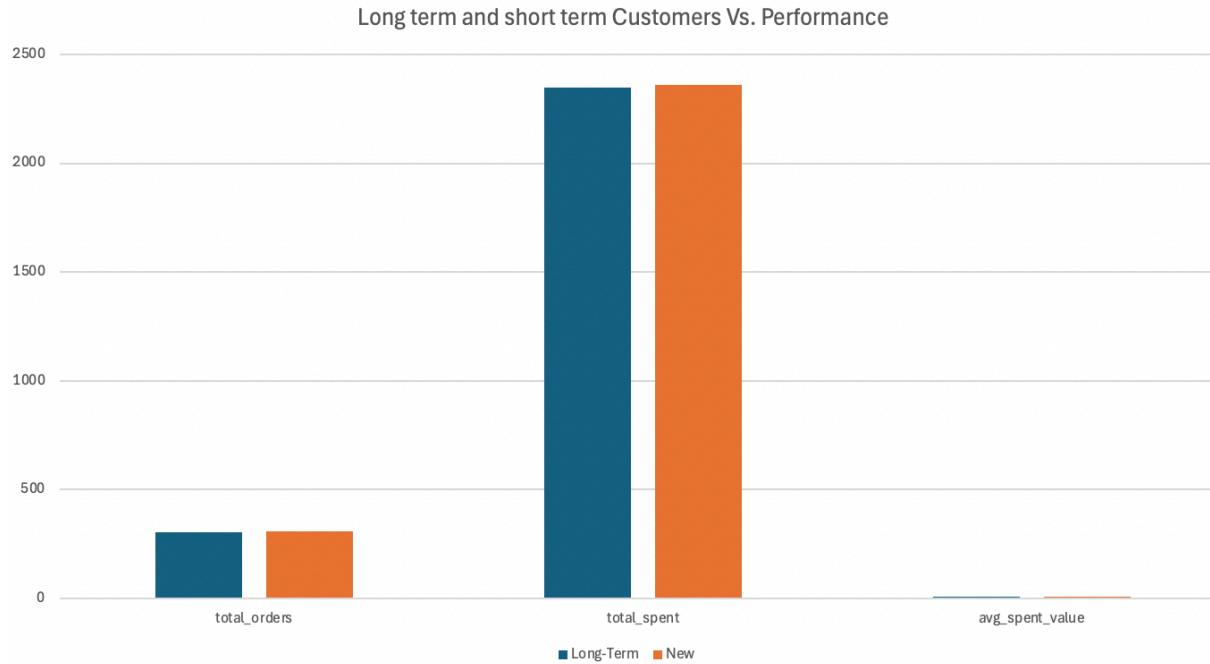
#### SQL Query:

```
select min(invoice_date),max(invoice_date) from invoice;

select case
    when year(i.invoice_date) in (2017,2018) then 'Long-Term'
    else 'New' end as customer_segment,
    count(distinct i.invoice_id) as total_orders,
    sum(total) as total_spent,
    round(avg(total),2) as avg_spent_value
from invoice i join customer c
on i.customer_id = c.customer_id
group by customer_segment;
```

#### Output:

customer_segment	total_orders	total_spent	avg_spent_value
Long-Term	305	2349.27	7.70
New	309	2360.16	7.64



### B. Insights:

- Long-Term Users have total\_orders of 305 and total spent is 2319.27 and average spent value of 7.70.
- New Users have total\_orders of 309 and total spent of 2360.16 and average spent value of 7.64.
- Since both long term and new customers perform pretty close. This could mean your customer acquisition efforts are very effective.

### C. Recommendation:

- To keep long-term users active, consider giving coupons, offers and rewards.
- Start referral programs for users to increase new users. And to motivate long term users.
- Offer exclusive access to customers.
- Focus on improve the relationship with customers.
- In conclusion both long term and new users are engaging well.

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?

### **Genre Wise:**

#### **A. Approach:**

- Affinity Analysis is a data analysis technique used to discover co-occurrence relationships among items in a dataset.
- We can do this analysis by joining tables and find combination of purchases by genres, artist and albums.

#### **SQL Query:**

```
|with affinity_cte as(
  select i.invoice_id,
         g.name genre_name,
         a1.name artist_name,
         a.title from invoice i
    join invoice_line il
      on i.invoice_id = il.invoice_id
    join track t
      on il.track_id = t.track_id
    join genre g
      on t.genre_id = g.genre_id
    join album a
      on t.album_id = a.album_id
    join artist a1
      on a.artist_id = a1.artist_id
)
```

```

select ac.genre_name genre_1,
       ac1.genre_name genre_2,
       count(*) as number_of_purchases
from affinity_cte ac
join affinity_cte ac1
on ac.invoice_id = ac1.invoice_id and ac.genre_name < ac1.genre_name
group by ac.genre_name, ac1.genre_name
order by number_of_purchases desc
limit 5;

```

## Output:

genre_1	genre_2	number_of_purchases
Metal	Rock	1622
Alternative & Punk	Rock	1056
Latin	Rock	427
R&B/Soul	Rock	407
Alternative & Punk	Metal	315

## Album Wise:

### SQL Query:

```

with affinity_cte as(
  select i.invoice_id, g.name as genre_name, a1.name as artist_name, a.title as album_name from invoice i
  join invoice_line il on i.invoice_id = il.invoice_id
  join track t
  on il.track_id = t.track_id
  join genre g
  on t.genre_id = g.genre_id
  join album a
  on t.album_id = a.album_id
  join artist a1 on a.artist_id = a1.artist_id
)
select ac1.album_name as album_1, ac2.album_name as album_2, count(*) as number_of_purchases
from affinity_cte ac1
join affinity_cte ac2 on ac1.invoice_id = ac2.invoice_id and ac1.album_name <> ac2.album_name
group by ac1.album_name, ac2.album_name
order by number_of_purchases desc
limit 5;

```

## Output:

album_1	album_2	number_of_purchases
Are You Experienced?	Mezmerize	16
Mezmerize	Are You Experienced?	16
My Generation - The Very Best Of The Who	Mezmerize	12
Vault: Def Leppard's Greatest Hits	Mezmerize	12
Mezmerize	Vault: Def Leppard's Greatest Hits	12

## Artist Wise:

### SQL Query:

```
with affinity_cte as(
select i.invoice_id,g.name as genre_name, a1.name as artist_name,a.title from invoice i
join invoice_line il on i.invoice_id = il.invoice_id
join track t on il.track_id = t.track_id
join genre g on t.genre_id = g.genre_id
join album a on t.album_id = a.album_id
join artist a1 on a.artist_id = a1.artist_id
)

select ac1.artist_name as artist_1, ac2.artist_name as artist_2, count(*) as number_of_purchases
from affinity_cte ac1
join affinity_cte ac2 on ac1.invoice_id = ac2.invoice_id and ac1.artist_name < ac2.artist_name
group by ac1.artist_name, ac2.artist_name
order by number_of_purchases desc
limit 5;
```

## Output:

artist_1	artist_2	number_of_purchases
Green Day	Led Zeppelin	24
Foo Fighters	Green Day	20
Nirvana	The Rolling Stones	19
Eric Clapton	Nirvana	19
Green Day	Metallica	18

**B. Insights:**

- Rock genre is purchased with Metal, Alternative & Punk ,Latin, R&B/Soul genres a lot, it is purchased with a lot of combinations.
- Mezmerize album is an album with highest number of purchases in combination with other albums.
- Greenday and Nirvana are the artists with high number of purchases in combinations with other artists, demonstrating high number of customer sentiments.

**C. Recommendations:**

- Highlight the popularity of Mezmerize album on homepage and recommendation pages.
  - Promote combinations of albums ,artist on genres, like “Fans of Mezmerize also bought”.
  - Create bundles with best combinations.
  - Create album pairing discount sales.
  - Do Artist featuring campaigns featuring both Green day and Nirvana.
  - Create merchandise bundles and playlists.
5. Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?

**A. Approach:**

- To analyse the purchasing behaviour and churn rates across different geographic regions / locations we can do an analysis based on invoice data in country basis.
- To analys the churn rate, we can segmentize the customers Active and Churned using a 180 day period.

## SQL Query:

```
WITH latest_date AS (
    SELECT MAX(invoice_date) AS max_date FROM invoice
),
customer_detailed_metrics AS (
    SELECT
        c.customer_id,
        c.country,
        COUNT(i.invoice_id) AS total_orders,
        SUM(i.total) AS total_revenue,
        MAX(i.invoice_date) AS last_purchase_date,
        MIN(i.invoice_date) AS first_purchase_date,
        DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS customer_tenure_days,
        CASE
            WHEN DATEDIFF((SELECT max_date FROM latest_date), MAX(i.invoice_date)) <= 180 THEN 'Active'
            ELSE 'Churned'
        END AS status
    FROM customer c
    JOIN invoice i ON c.customer_id = i.customer_id
    GROUP BY c.customer_id, c.country
),
country_analysis AS (
    SELECT
        country,
        -- Customer counts
        COUNT(customer_id) AS total_customers,
        SUM(CASE WHEN status = 'Active' THEN 1 ELSE 0 END) AS active_users,
        SUM(CASE WHEN status = 'Churned' THEN 1 ELSE 0 END) AS churned_users,

        -- Order metrics
        SUM(total_orders) AS total_orders,
        ROUND(AVG(total_orders), 1) AS avg_orders_per_customer,

        -- Revenue metrics
        ROUND(SUM(total_revenue), 2) AS total_revenue,
        ROUND(SUM(total_revenue) / COUNT(customer_id), 2) AS avg_customer_value,
        ROUND(SUM(total_revenue) / SUM(total_orders), 2) AS avg_order_value,

        -- Tenure metrics
        ROUND(AVG(customer_tenure_days), 0) AS avg_customer_tenure_days,

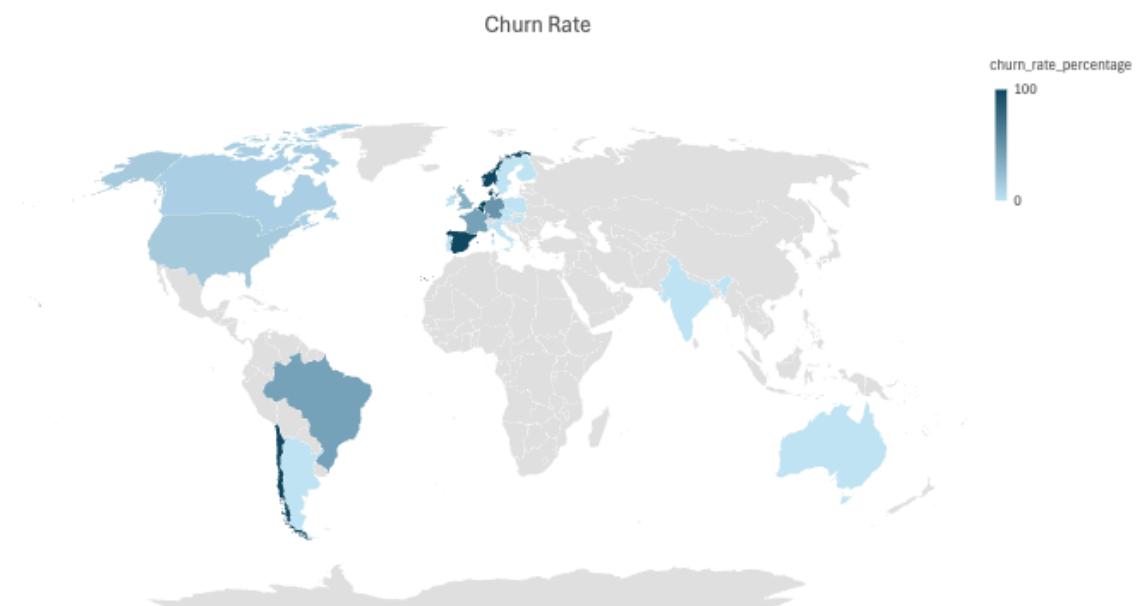
        -- Churn rate
        ROUND(SUM(CASE WHEN status = 'Churned' THEN 1 ELSE 0 END) * 100.0 / COUNT(customer_id), 2) AS churn_rate_percent
    FROM customer_detailed_metrics
    GROUP BY country
)
```

```
SELECT
```

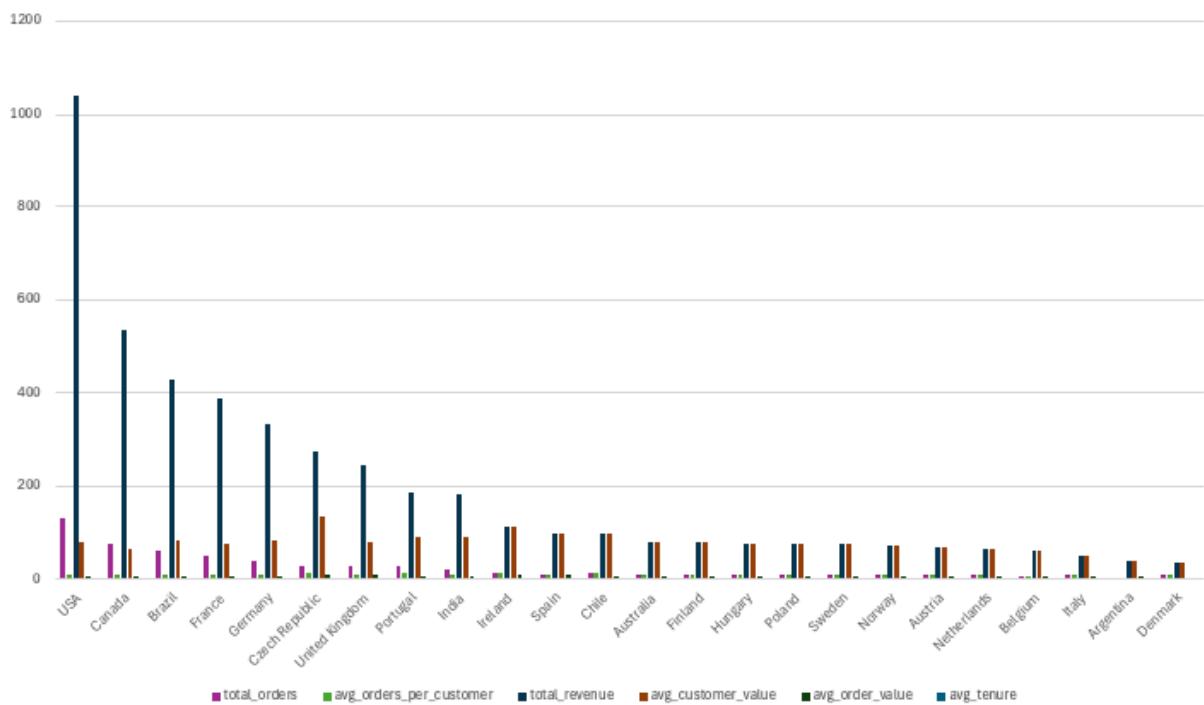
```
    country,  
    total_customers,  
    active_users,  
    churned_users,  
    churn_rate_percentage,  
    total_orders,  
    avg_orders_per_customer,  
    total_revenue,  
    avg_customer_value,  
    avg_order_value,  
    concat(avg_customer_tenure_days, " days") AS avg_tenure  
FROM country_analysis  
ORDER BY total_revenue DESC;
```

## Output:

country	total_customers	active_users	churned_users	churn_rate_p...	total_orders	avg_orders_per...	total_revenue	avg_customer_v...	avg_order_value	avg_tenure
USA	13	10	3	23.08	131	10.1	1040.49	80.04	7.94	1272 days
Canada	8	7	1	12.50	76	9.5	535.59	66.95	7.05	1284 days
Brazil	5	2	3	60.00	61	12.2	427.68	85.54	7.01	1230 days
France	5	3	2	40.00	50	10.0	389.07	77.81	7.78	1235 days
Germany	4	1	3	75.00	41	10.3	334.62	83.66	8.16	1221 days
Czech Repu...	2	2	0	0.00	30	15.0	273.24	136.62	9.11	1206 days
United Kingd...	3	2	1	33.33	28	9.3	245.52	81.84	8.77	1237 days
Portugal	2	2	0	0.00	29	14.5	185.13	92.57	6.38	1376 days
India	2	1	1	50.00	21	10.5	183.15	91.58	8.72	1338 days
Ireland	1	1	0	0.00	13	13.0	114.84	114.84	8.83	1393 days



### Country Wise Analysis



### B. Insights:

- Churn rate is 100% on countries like Spain, Chile, Norway, Austria, Netherlands, Belgium, Argentina, Denmark. The number of customer and total revenue is so low in these countries.
- Countries like Czech Republic, Portugal, Ireland, Australia..etc have 0 percent churnrate with high number of orders from existing customers.
- Countries like USA, Canada and Brazil have high revenue from customer with very low percent in churn rates.
- We can see consistency in Average order value across countries.

### C. Recommendations:

- Try to identify the trends in high performing countries with low churn rate and high revenue and implement it on other markets.
- Try customer retention campaigns on countries with high churn rate.
- Add more investments into countries like Canada and USA they provide significant income and active customers.
- When a churn rate in a country is increase try more outreach programs and discount programs to reengage customers.

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?

#### A. Approach:

- o Since chinook db doesn't have age or gender data on customer , we will use location and purchase history for customer risk profiling. We can consider the average of customers country wise and segmentize the customers in to high risk and low risk.

#### SQL Query:

```

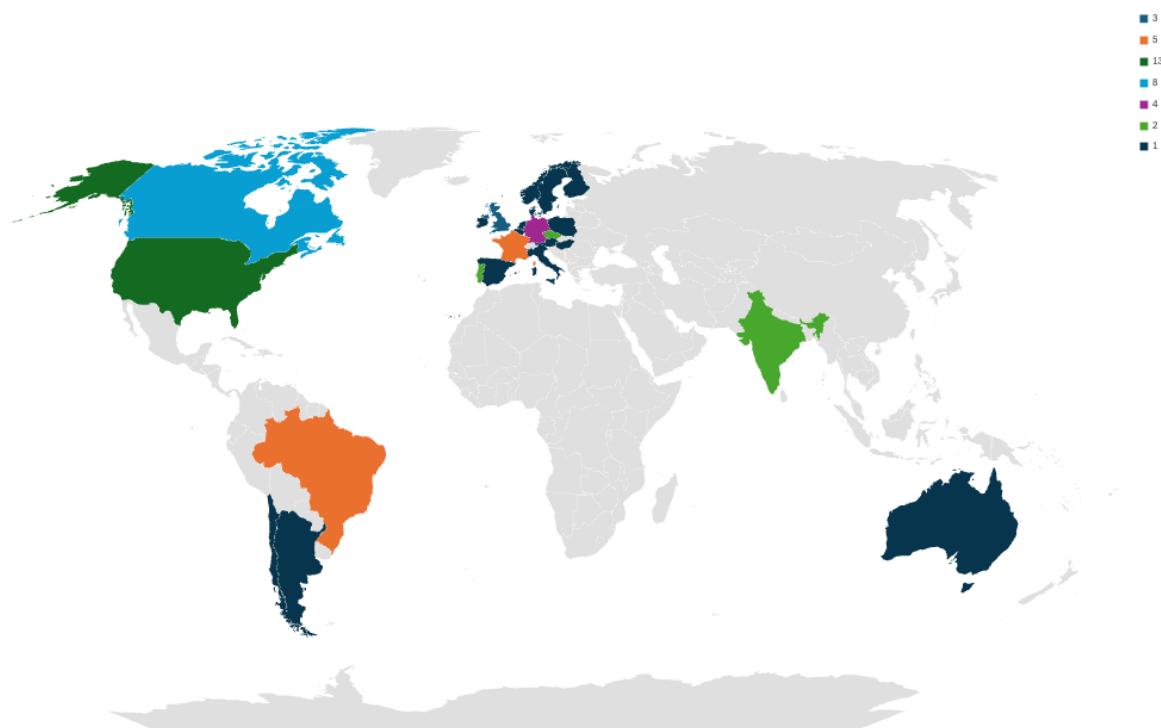
) WITH customer_risk AS (
  SELECT
    c.customer_id,
    c.first_name,
    c.country,
    COALESCE(SUM(i.total), 0) as total_spent,
  CASE
    WHEN COALESCE(SUM(i.total), 0) < AVG(COALESCE(SUM(i.total), 0)) OVER (PARTITION BY c.country)
      THEN 'High Risk to Churn'
      ELSE 'Low Risk to Churn'
  END as churn_risk_segment
  FROM customer c
  LEFT JOIN invoice i ON c.customer_id = i.customer_id
  GROUP BY c.customer_id, c.first_name, c.country
)
SELECT
  country,
  COUNT(*) as total_customers,
  SUM(CASE WHEN churn_risk_segment = 'High Risk to Churn' THEN 1 ELSE 0 END) as high_risk_count,
  SUM(CASE WHEN churn_risk_segment = 'Low Risk to Churn' THEN 1 ELSE 0 END) as low_risk_count,
  ROUND(SUM(CASE WHEN churn_risk_segment = 'High Risk to Churn' THEN 1 ELSE 0 END) * 100.0 / COUNT(*), 2) as high_ris
FROM customer_risk
GROUP BY country
ORDER BY high_risk_percentage DESC, high_risk_count DESC;

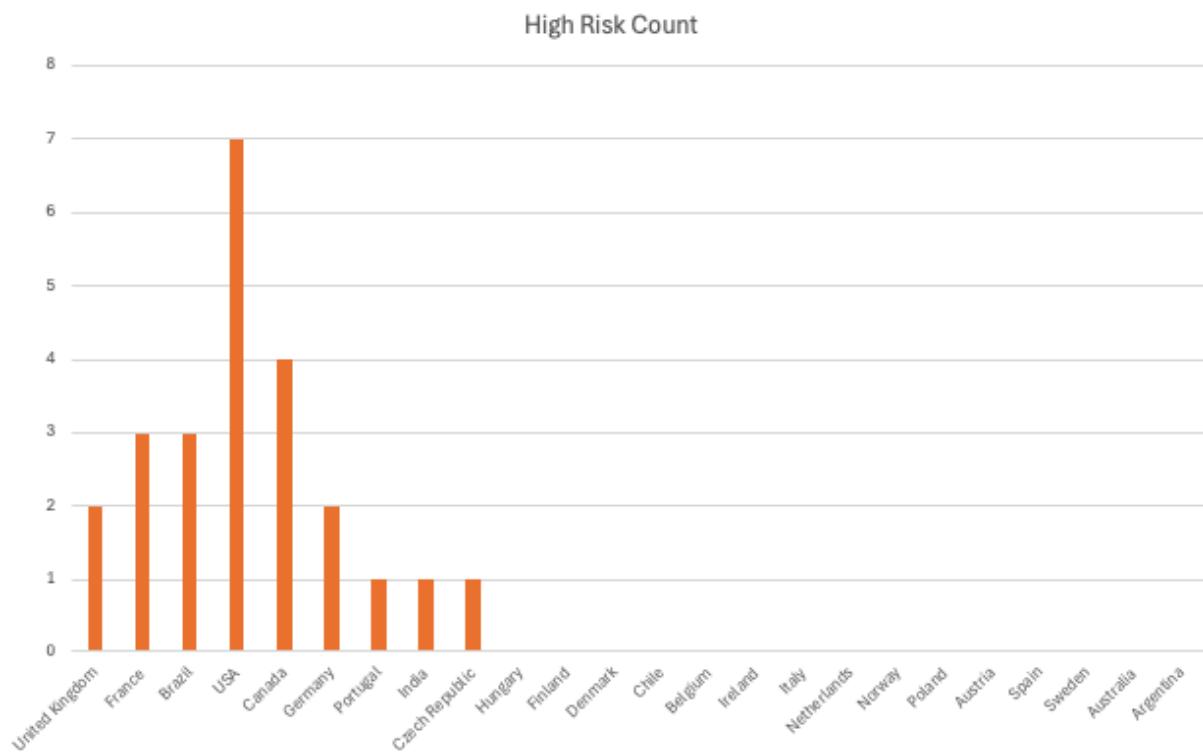
```

#### Output:

country	total_customers	high_risk_count	low_risk_count	high_risk_percentage
United Kingd...	3	2	1	66.67
France	5	3	2	60.00
Brazil	5	3	2	60.00
USA	13	7	6	53.85
Canada	8	4	4	50.00
Germany	4	2	2	50.00
Portugal	2	1	1	50.00
India	2	1	1	50.00
Czech Repu...	2	1	1	50.00

Count of Customer At Risk





### **B. Insight:**

- Countries like UK ,France ,Brazil and USA have more than 50 of customers at risk of churning.
- Countries like India, Portugal, Germany, Czech and Hungary lack the number of customers and have a higher chance of risk of churning.

### **C. Recommendations:**

- Promote and market the store to reach to more customer at low customer regions.
- Start offers and discounts to make the customer engaging and loyalty.
- Launch recovery programs for churned customer, like emailing, calling..etc.
- Monitor churn rate of regions and plan retention of customers by regional activities, like hosting a show or meetup.

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?

**A. Approach:**

- First we can categorise the customers who are active and churned, then we can analyse the CLVM of customers in both category customers.

**SQL Query:**

```
WITH latest_date AS (
    SELECT MAX(invoice_date) as max_date FROM invoice
),
```

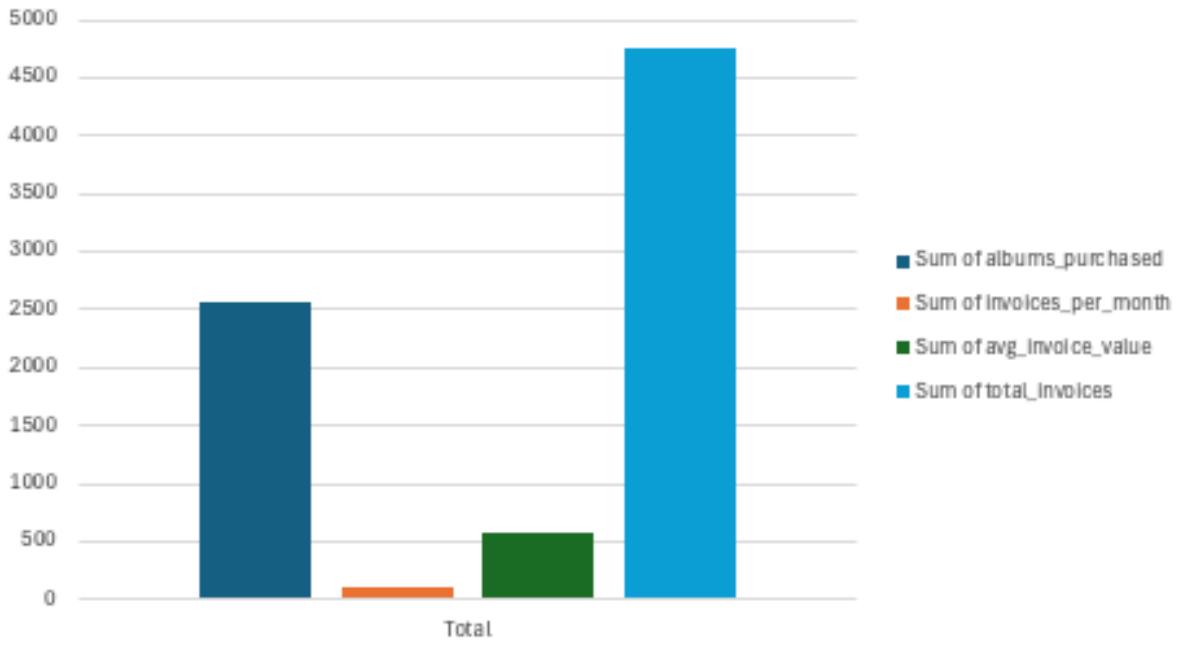
```
customer_analysis AS (
    SELECT
        c.customer_id,
        c.first_name,
        -- Calculate customer metrics
        COUNT(i.invoice_id) as total_invoices,
        SUM(i.total) as lifetime_value,
        AVG(i.total) as avg_invoice_value,
        MIN(i.invoice_date) as first_purchase_date,
        MAX(i.invoice_date) as last_purchase_date,
        DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) as tenure_days,
        -- Days since last purchase from latest date
        DATEDIFF((SELECT max_date FROM latest_date), MAX(i.invoice_date)) as days_since_last_purchas
        -- Purchase frequency metrics
        COUNT(i.invoice_id) / NULLIF(DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) / 30.0, 0) a
        -- Engagement metrics
        COUNT(il.track_id) as total_tracks_purchased,
        COUNT(DISTINCT il.track_id) as unique_tracks_purchased,
        COUNT(DISTINCT t.genre_id) as genres_purchased,
        COUNT(DISTINCT t.album_id) as albums_purchased,
        COUNT(DISTINCT a.artist_id) as artists_followed,
        CASE
            WHEN DATEDIFF((SELECT max_date FROM latest_date), MAX(i.invoice_date)) > 90
            THEN 'Churned'
            ELSE 'Active'
        END as churn_status
    FROM customer c
    JOIN invoice i ON c.customer_id = i.customer_id
    JOIN invoice_line il ON i.invoice_id = il.invoice_id
    JOIN track t ON il.track_id = t.track_id
    JOIN album a ON t.album_id = a.album_id
    GROUP BY c.customer_id, c.first_name
)
```

```
SELECT
```

```
    customer_id,  
    first_name,  
    churn_status,  
    total_invoices,  
    ROUND(lifetime_value, 2) as lifetime_value,  
    ROUND(avg_invoice_value, 2) as avg_invoice_value,  
    tenure_days,  
    days_since_last_purchase,  
    ROUND(invoices_per_month, 2) as invoices_per_month,  
    total_tracks_purchased,  
    unique_tracks_purchased,  
    genres_purchased,  
    albums_purchased,  
    artists_followed  
FROM customer_analysis  
ORDER BY churn_status, lifetime_value DESC;
```

## Output:

first_name	churn_stat...	total_inv...	lifetime_value	avg_invoic...	tenure_days	days_since...	invoices_pe...	total_tracks_...	unique_tracks...	genres_purch...	albums_pur...	artists
František	Active	146	1683.00	11.53	1263	48	3.47	146	146	13	62	52
Helena	Active	130	1500.84	11.54	1149	68	3.39	130	130	11	72	54
Hugh	Active	116	1433.52	12.36	1393	18	2.50	116	115	12	47	42
Wyatt	Active	101	1248.39	12.36	1406	33	2.16	101	101	12	56	43
João	Active	104	1144.44	11.00	1326	78	2.35	104	103	10	44	40
Fernanda	Active	108	1025.64	9.50	1290	32	2.51	108	107	12	59	49
Phil	Active	99	1024.65	10.35	1420	0	2.09	99	99	8	34	29
Heather	Active	93	1002.87	10.78	1321	61	2.11	93	93	13	45	38
Dan	Active	96	966.24	10.07	1431	1	2.01	96	96	9	51	37
Patrick	Active	85	943.47	11.10	1378	75	1.85	85	85	9	44	39
Mark	Active	82	940.50	11.47	1405	9	1.75	82	82	11	37	35
Aaron	Active	71	862.29	12.14	1212	65	1.76	71	71	9	38	34
Kathy	Active	92	841.50	9.15	1360	21	2.03	92	92	10	46	35



## B. Insights:

- Almost all of churned customers have a last purchase date is longer than 100 days, Active customers have day since last purchase less than 100.
- Total number of invoices are consistent across all type of customers.
- Average lifetime value of Churned is 765.59 and for active its 828.66.
- Some of the churned customers has low tenure days indicating new customers are not satisfied with the service.
- Artists followed are comparatively less by churned customers than active customers.
- Analysis shows customers like Dan who is a long term customer and is consistent on purchases.

## C. Recommendations:

- Stratagize to bring back churned customers.
- Analys the long term loyal customer with high lifetime value and there behaviour to playlists and recommendation and take there survey about store to implement changes and push the advantages.
- As per CLVM , categorize customers and create separate recommendation models and marketing campaign and analys the genre that could potentially improve the lifetime value of the customers category wise.

- Prepare subscription plans for a particular period of discount to maintain customer loyalty and engagement. Subscription plans could motivate customers to stay loyal and engaging to the service.
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?

**A. Approach:**

- To measure campaign impact, I would compare key metrics before, during, and after each campaign. Track new customer acquisitions, purchase frequency, average order values, and retention rates.

**B. Insights:**

- Effective campaigns show sustained increases in new customers and higher customer lifetime values. They also boost repeat purchase rates and engagement among existing customers. Poor campaigns only create temporary sales spikes or attract one-time buyers who don't return. The best campaigns balance both acquisition and retention benefits.

**C. Recommendation:**

- Focus acquisition campaigns on high-value customer profiles identified from historical data.
- For retention, create follow-up campaigns to keep new customers engaged. Always track long-term customer value, not just immediate sales.
- Use tiered rewards that encourage both initial purchases and ongoing loyalty.

9. How would you approach this problem, if the objective and subjective questions weren't given?
- If no question were given I would start by understanding the DB schemas and data sets, then check for the null values and duplicates in data.
  - Then I would understand the business objective which here

is to analyze music record sales data to gain insights and make recommendations for the company's strategy in the physical music market.

- Next, I would research the music retail industry to understand common business challenges and success metrics. Based on this context, I'd generate hypotheses about customer behaviour, sales patterns, and operational efficiency. Common starting points include customer segmentation, product performance, and seasonal trends that are relevant to most retail businesses.
  - I'd identify the most significant patterns and potential business opportunities. This includes detecting customer segments with distinct behaviors, products with unusual performance characteristics, time periods with notable trends, and operational patterns that suggest optimization opportunities.
  - Finally, I would prioritize the identified patterns based on potential business impact and actionable. The most valuable insights would be those that suggest clear business decisions, have significant financial implications, and can be implemented with reasonable effort. This ensures the analysis delivers practical recommendations rather than just interesting observations.
  - This approach ensures comprehensive analysis while remaining flexible enough to adapt to whatever patterns the data reveals, ultimately focusing on insights that drive meaningful business decisions.
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?

A. Before altering:

Field	Type	Null	Key	Default	Extra
album_id	int	NO	PRI	NULL	
title	varchar(160)	NO		NULL	
artist_id	int	NO		NULL	

B. To add a new Column named ReleaseYear with type INTEGER I can use the DDL command ALTER.

**SQL Query:**

```
ALTER TABLE album
ADD ReleaseYear INT NOT NULL;

desc album;
```

**Output:**

Field	Type	Null	Key	Default	Extra
album_id	int	NO	PRI	NULL	
title	varchar(160)	NO		NULL	
artist_id	int	NO		NULL	
ReleaseYear	int	YES		NULL	

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

**A. Approach:**

- First I would group customers and get their invoice aggregates, then I would group the country to get the sums and averages of customer in that country.

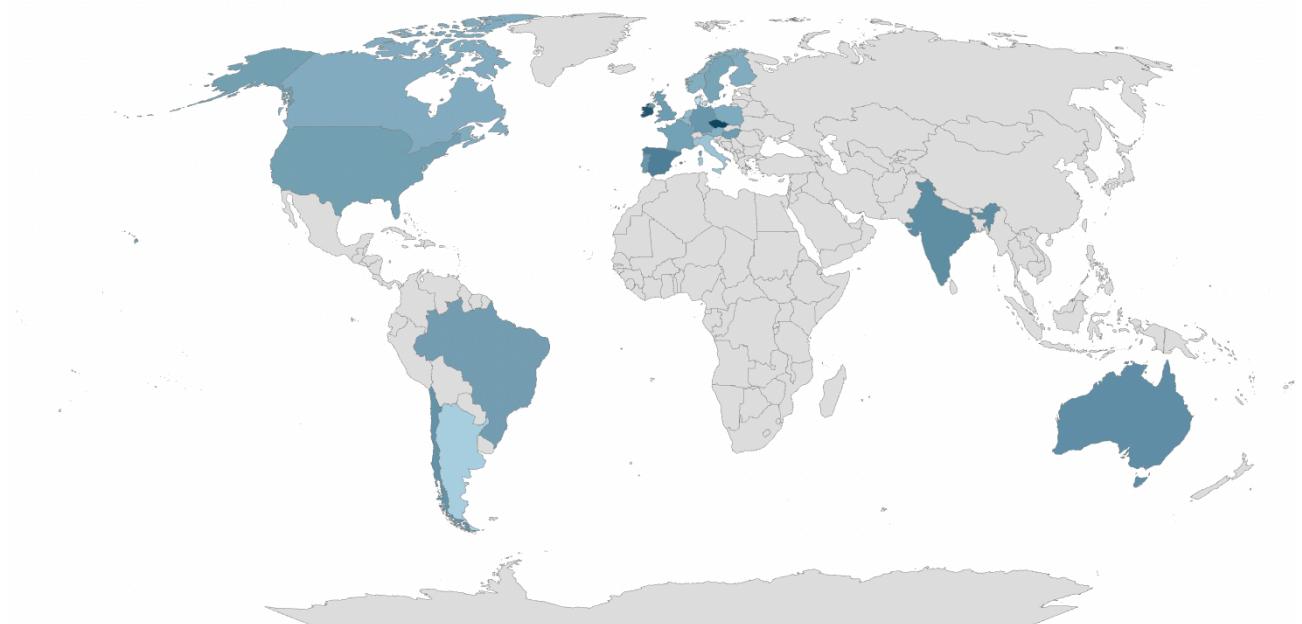
**SQL Query:**

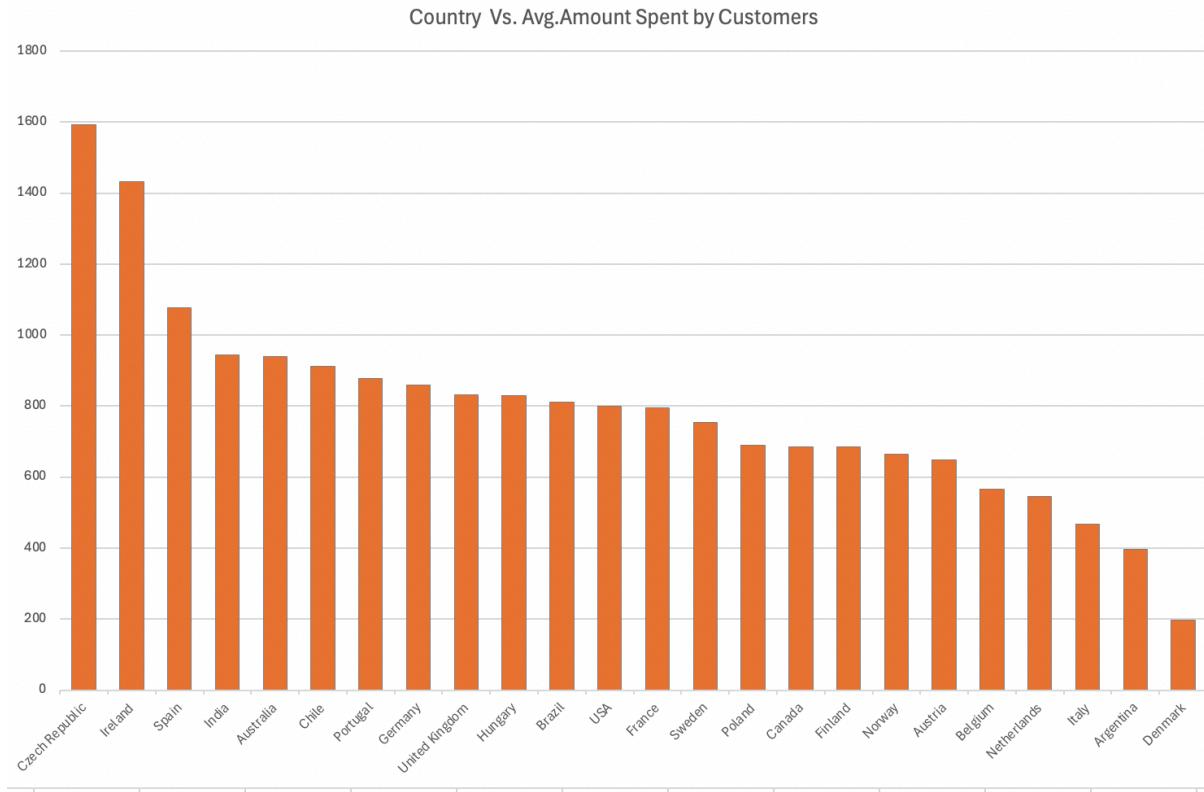
```
SELECT
    c.country,
    COUNT(DISTINCT c.customer_id) AS total_customers,
    COUNT(i.invoice_id) AS total_invoices,
    SUM(i.total) AS total_revenue,
    ROUND(SUM(i.total) / COUNT(DISTINCT c.customer_id), 2) AS revenue_per_customer,
    ROUND(AVG(i.total), 2) AS avg_invoice_value
FROM customer c
LEFT JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY c.country
ORDER BY total_revenue DESC;
```

## Output:

country	total_customers	total_inv...	total_revenue	revenue_per_customer
USA	13	131	1040.49	80.04
Canada	8	76	535.59	66.95
Brazil	5	61	427.68	85.54
France	5	50	389.07	77.81
Germany	4	41	334.62	83.66
Czech Republic	2	30	273.24	136.62
United Kingdom	3	28	245.52	81.84
Portugal	2	29	185.13	92.57
India	2	21	183.15	91.58
Ireland	1	13	114.84	114.84
Spain	1	11	98.01	98.01
Chile	1	13	97.02	97.02
Australia	1	10	81.18	81.18
Finland	1	11	79.20	79.20
Hungary	1	10	78.21	78.21
Poland	1	10	76.23	76.23
Sweden	1	10	75.24	75.24
Norway	1	9	72.27	72.27
Austria	1	9	69.30	69.30
Netherlands	1	10	65.34	65.34
Belgium	1	7	60.39	60.39
Italy	1	9	50.49	50.49
Argentina	1	5	39.60	39.60
Denmark	1	10	37.62	37.62

Country Vs. Avg.Amount Spent by Customers





### C. Insights:

- Czech Republic has highest number of average amount spent per customer followed by Ireland and Spain.
- Denmark has lowest average amount spent by customer.
- USA lies in middle with average amount spent by high in number of customers.

### D. Recommendation:

- Launch focused marketing campaigns in the Czech Republic and Ireland to acquire more customers, leveraging their high spending propensity.
- In the USA, Canada, and the UK, introduce upselling strategies such as curated albums, exclusive content, or loyalty programs to increase the average spend per customer.
- Since track quantity correlates with spending, promote bundled offers or subscriptions in mid-performing markets like Germany and France to boost both metrics.

- Explore reasons behind lower spending in countries like Denmark and Argentina, whether due to economic factors, localization issues, or product relevance and adjust pricing offerings accordingly.