

## 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.



```

SELECT
    g.genre_id,
    g.name AS genre_name,
    al.album_id,
    al.title AS new_record_label,
    SUM(il.unit_price * il.quantity) AS total_genre_sales,
    DENSE_RANK() OVER (ORDER BY SUM(il.unit_price * il.quantity) DESC) AS Ranking
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
INNER JOIN album al ON t.album_id = al.album_id
WHERE c.country = 'USA'
GROUP BY g.genre_id, g.name, al.album_id, al.title
ORDER BY total_genre_sales DESC;

```

Based on my analysis following are 3 albums that should be prioritized for advertising and promotion in USA:

genre_id	genre_name	album_id	new_record_label	total_genre_sales	Ranking
1	Rock	120	Are You Experienced?	27.72	1
1	Rock	163	From The Muddy Banks Of The Wishkah [live]	27.72	1
1	Rock	214	The Doors	26.73	2
14	R&B/Soul	146	Seek And Shall Find: More Of The Best (1963-1...	25.74	3
4	Alternative & Punk	188	Green	24.75	4
6	Blues	72	The Cream Of Clapton	24.75	4
1	Rock	216	Hot Rocks, 1964-1971 (Disc 1)	24.75	4

1. "Are You Experienced?" (Rock)

- Album ID: 120
- Total Genre Sales: 27.72
- Ranking: 1

2. "From The Muddy Banks Of The Wishkah [live]" (Rock)

- Album ID: 163
- Total Genre Sales: 27.72
- Ranking: 1

### 3. **"The Doors" (Rock)**

- Album ID: 214
- Total Genre Sales: 26.73
- Ranking: 2

#### **INSIGHTS:**

- **Genre Popularity:** Rock is the dominant genre with the highest total sales figures. The top three albums all belong to this genre, indicating a strong market preference for Rock music in the USA.
- **Sales Figures:** The selected albums have the highest total genre sales, making them strong candidates for promotion.

Given the strong sales performance of these albums in the Rock genre, they should be prioritized for advertising and promotion in the USA to capitalize on their popularity.



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



```
-- Top Selling Genres in countries other than USA?
SELECT
    g.genre_id,
    g.name AS genre_name,
    c.country,
    SUM(il.quantity) AS quantity_sold
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 country <> 'USA'
GROUP BY g.genre_id, genre_name, c.country
ORDER BY quantity_sold DESC;
```

```
-- Top Selling Genres in countries in USA?
SELECT
    g.genre_id,
    g.name AS genre_name,
    c.country,
    SUM(il.quantity) AS quantity_sold
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 country = 'USA'
GROUP BY g.genre_id, genre_name, c.country
ORDER BY quantity_sold DESC;
```

### Top-Selling Genres in Countries Other Than the USA:

#### **Rock:**

Canada: 333 units sold  
 France: 211 units sold  
 Brazil: 205 units sold  
 Germany: 194 units sold  
 United Kingdom: 166 units sold  
 Czech Republic: 143 units sold  
 Portugal: 108 units sold  
 India: 102 units sold  
 Ireland: 72 units sold  
 Chile: 61 units sold  
 Sweden: 60 units sold

#### **Alternative & Punk:**

Brazil: 74 units sold

#### **Metal:**

Brazil: 73 units sold  
 Canada: 72 units sold

### COMMONALITIES:

- **Rock dominates** in almost all the countries listed, indicating a strong global preference for Rock music outside the USA.
- **Metal and Alternative & Punk are also popular in Brazil**, showing a more diverse taste in music genres in that country.
- **Rock is the top genre** in every country mentioned except for the specific instances where Metal or Alternative & Punk are highlighted.

### DIFFERENCES:

- **Brazil** stands out with **significant sales in both Alternative & Punk and Metal genres**, in addition to Rock, showing a more varied musical taste compared to other countries.
- **Metal** has a notable presence in **Canada and Brazil** but does **not appear as a top genre** in other countries listed.
- The sales volume for genres other than Rock is generally lower in all non-USA countries, indicating **Rock's dominance**.

**INSIGHTS:**

- **Rock** is the overwhelmingly **dominant genre** across most countries, indicating a universal appeal.
- **Brazil** has a **more diverse taste**, with strong sales in Alternative & Punk and Metal alongside Rock.
- The preference for **Metal** is notably **strong in Canada and Brazil**, which differs from other countries where Rock dominates.
- This analysis shows that while **Rock is a global favourite**, countries like **Brazil exhibit broader genre preferences**, which could be important for tailored marketing and promotion strategies.

country - Other Than USA		(All)	country - USA		(All)
genre_name (other than USA)	Sum of quantity_sold		genre_name (USA)	Sum of quantity_sold	
Alternative	82		Alternative	35	
Alternative & Punk	362		Alternative & Punk	130	
Blues	88		Blues	36	
Classical	43		Classical	4	
Drama	1		Easy Listening	13	
Easy Listening	61		Electronica/Dance	5	
Electronica/Dance	50		Heavy Metal	3	
Heavy Metal	5		Hip Hop/Rap	20	
Hip Hop/Rap	13		Jazz	14	
Jazz	107		Latin	22	
Latin	145		Metal	124	
Metal	495		Pop	22	
Pop	41		R&B/Soul	53	
R&B/Soul	106		Reggae	6	
Reggae	29		Rock	561	
Rock	2074		Soundtrack	2	
Soundtrack	3		TV Shows	1	
TV Shows	1		<b>Grand Total</b>	<b>1051</b>	
<b>Grand Total</b>	<b>3706</b>				

Tool : MS-EXCEL - Pivot Table 

### 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?

- To analyse customer purchasing behaviour and compare the habits of long-term customers with those of new customers, we can use the following SQL query. This query will compute the purchase frequency, average basket size, and average spending amount for both customer groups.

```

1  WITH CustomerPurchaseStats AS (
2      SELECT
3          c.customer_id,
4          COUNT(i.invoice_id) AS purchase_frequency,
5          SUM(il.quantity) AS total_items_purchased,
6          SUM(i.total) AS total_spent,
7          AVG(i.total) AS avg_order_value,
8          DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS customer_tenure_days
9      FROM customer c
10     JOIN invoice i ON c.customer_id = i.customer_id
11     JOIN invoice_line il ON i.invoice_id = il.invoice_id
12     GROUP BY c.customer_id
13  ),
14 CustomerSegments AS (
15     SELECT
16         customer_id,
17         purchase_frequency,
18         total_items_purchased,
19         total_spent,
20         avg_order_value,
21         customer_tenure_days,
22         CASE
23             WHEN customer_tenure_days >= 365 THEN 'Long-Term'
24             ELSE 'New'
25         END AS customer_segment
26     FROM CustomerPurchaseStats
27  )
28     SELECT
29         customer_segment,
30         ROUND(AVG(purchase_frequency),2) AS avg_purchase_frequency,
31         ROUND(AVG(total_items_purchased),2) AS avg_basket_size,
32         ROUND(AVG(total_spent),2) AS avg_spending_amount,
33         ROUND(AVG(avg_order_value),2) AS avg_order_value
34     FROM CustomerSegments
35     GROUP BY customer_segment;

```

	customer_segment	avg_purchase_frequency	avg_basket_size	avg_spending_amount	avg_order_value
▶	Long-Term	80.63	80.63	805.14	9.77

**INSIGHTS:**

- **Purchase Frequency:** By comparing avg\_purchase\_frequency, we can determine whether long-term customers make purchases more frequently than new customers.
- **Basket Size:** The avg\_basket\_size indicates whether long-term customers tend to buy more items per order compared to new customers.
- **Spending Amount:** The avg\_spending\_amount helps in understanding whether long-term customers spend more money overall, which could be a sign of loyalty and satisfaction.
- **Order Value:** The avg\_order\_value reveals whether long-term customers place higher-value orders, which could indicate trust and a preference for the product offerings.

**RECOMMENDATION:**

These insights can inform customer loyalty and retention strategies, such as offering personalized promotions to long-term customers or targeted marketing campaigns to encourage repeat purchases from new 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?



##### a) Frequently Purchased Genres Together:

```

1  -- 1. Genre Affinity Analysis
2  WITH track_combinations AS (
3      SELECT
4          il1.track_id AS track_id_1,
5          il2.track_id AS track_id_2,
6          COUNT(*) AS times_purchased_together
7      FROM invoice_line il1
8      JOIN invoice_line il2 ON il1.invoice_id = il2.invoice_id AND il1.track_id < il2.track_id
9      GROUP BY il1.track_id, il2.track_id
10  ),
11  genre_combinations AS (
12      SELECT
13          t1.genre_id AS genre_id_1,
14          t2.genre_id AS genre_id_2,
15          COUNT(*) AS times_purchased_together
16      FROM track_combinations tc
17      JOIN track t1 ON tc.track_id_1 = t1.track_id
18      JOIN track t2 ON tc.track_id_2 = t2.track_id
19      WHERE t1.genre_id <> t2.genre_id
20      GROUP BY t1.genre_id, t2.genre_id
21  )
22  SELECT
23      g1.name AS genre_1,
24      g2.name AS genre_2,
25      gc.times_purchased_together
26  FROM genre_combinations gc
27  JOIN genre g1 ON gc.genre_id_1 = g1.genre_id
28  JOIN genre g2 ON gc.genre_id_2 = g2.genre_id
29  ORDER BY gc.times_purchased_together DESC;

```

genre_1	genre_2	times_purchased_together
Metal	Rock	986
Alternative & Punk	Rock	629
Rock	Metal	613
Rock	Alternative & Punk	412
Rock	Alternative	302
Latin	Rock	251
Rock	R&B/Soul	231
Blues	Rock	180
Rock	Latin	176
Metal	Alternative & Punk	174
R&B/Soul	Rock	173
Rock	Classical	159
Jazz	Rock	158
Rock	Pop	142
Alternative & Punk	Metal	135
Easy Listening	Rock	118
Rock	Blues	96
Metal	Alternative	91
Rock	Reggae	80
Rock	Jazz	78
Metal	R&B/Soul	65
Latin	Metal	64
Rock	Easy Listening	60
Metal	Latin	59

Sum of times_purchased_together	Column Labels	Alternative	Alternative & Punk	Blues	Classical	Drama	Easy Listening	Electronica/Dance	Heavy Metal	Hip Hop/Rap	Jazz	Latin	Metal	Pop	R&B	Reggae	Rock	Soundtrack	TV Shows
Row Labels																			
Alternative					8							1	1	5	12				
Alternative & Punk				53		9	34	1	7	13	6	11	17	30	135	33	55	19	629
Blues				19	34	6		8	3	1	2	4	16	35	9	17	5	180	2
Classical															2				
Easy Listening				9	20	4			4	1	4	3	7	34	3	12	3	118	
Electronica/Dance				7	6	2						2	16	2	1	3	49		
Heavy Metal				2	4	1						3	7		22				
Hip Hop/Rap				4										1		1			
Jazz				16	35	7	8		3	3	1	4	8	35	8	14	4	158	
Latin				30	50	7	14		5	3	1	9	64	15	25	6	251		
Metal				91	174	19	45		15	16	1	11	32	59	48	65	34	986	2
Pop				4			2			2					1	1			
R&B/Soul				17		31	1	10			4		5	10	47	10	9	173	1
Reggae				9		4	1	5				1	3	1	4	2	1	30	
Rock				302	412	96	159	4	60	47	13	52	78	176	613	142	231	80	9
Soundtrack				1		3	3			1			1	1		1	18		
TV Shows																			

## INSIGHTS:

- Metal and Rock are the most frequently purchased together, indicating a strong overlap in fan base.
- Alternative & Punk and Rock also show significant co-purchase rates, suggesting shared listener preferences.
- Latin and R&B/Soul genres have notable pairings with Rock, pointing to potential opportunities for cross-genre promotions.

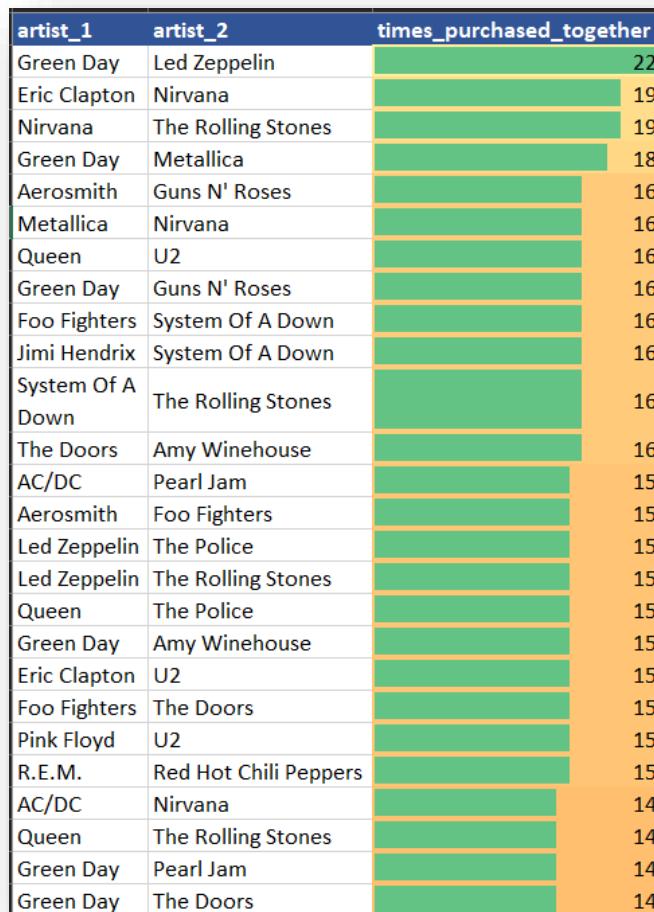
## **RECOMMENDATIONS:**

- **Bundle Offers:** Combine Metal and Rock albums, along with Alternative & Punk with Rock for popular bundles.
- **Cross-Promotions:** Promote Metal albums to Rock fans and vice versa, leveraging their shared appeal.

### b) Frequently Purchased Artists Together:

```
1  -- 2. Artist Affinity Analysis=
2  WITH track_combinations AS (
3      SELECT
4          il1.track_id AS track_id_1,
5          il2.track_id AS track_id_2,
6          COUNT(*) AS times_purchased_together
7      FROM invoice_line il1
8      JOIN invoice_line il2
9      ON il1.invoice_id = il2.invoice_id AND il1.track_id < il2.track_id
10     GROUP BY il1.track_id, il2.track_id
11 ),
12 artist_combinations AS (
13     SELECT
14         a1.artist_id AS artist_id_1,
15         a2.artist_id AS artist_id_2,
16         COUNT(*) AS times_purchased_together
17     FROM track_combinations tc
18     JOIN track t1 ON tc.track_id_1 = t1.track_id
19     JOIN album al1 ON t1.album_id = al1.album_id
20     JOIN artist a1 ON al1.artist_id = a1.artist_id
21     JOIN track t2 ON tc.track_id_2 = t2.track_id
22     JOIN album al2 ON t2.album_id = al2.album_id
23     JOIN artist a2 ON al2.artist_id = a2.artist_id
24     WHERE a1.artist_id <> a2.artist_id
25     GROUP BY a1.artist_id, a2.artist_id
26 )
27 SELECT
28     a1.name AS artist_1,
29     a2.name AS artist_2,
30     ac.times_purchased_together
31 FROM artist_combinations ac
32 JOIN artist a1 ON ac.artist_id_1 = a1.artist_id
33 JOIN artist a2 ON ac.artist_id_2 = a2.artist_id
34 ORDER BY ac.times_purchased_together DESC;
```

artist_1	artist_2	times_purchased_together
Green Day	Led Zeppelin	22
Eric Clapton	Nirvana	19
Nirvana	The Rolling Stones	19
Green Day	Metallica	18
Aerosmith	Guns N' Roses	16
Metallica	Nirvana	16
Queen	U2	16
Green Day	Guns N' Roses	16
Foo Fighters	System Of A Down	16
Jimi Hendrix	System Of A Down	16
System Of ...	The Rolling Stones	16
The Doors	Amy Winehouse	16
AC/DC	Pearl Jam	15
Aerosmith	Foo Fighters	15
Led Zeppelin	The Police	15
Led Zeppelin	The Rolling Stones	15
Queen	The Police	15
Green Day	Amy Winehouse	15
Eric Clapton	U2	15
Foo Fighters	The Doors	15
Pink Floyd	U2	15
R.E.M.	Red Hot Chili Peppers	15
AC/DC	Nirvana	14



**INSIGHTS:**

- Green Day and Led Zeppelin are frequently purchased together, indicating strong cross-genre appeal.
- Nirvana and The Rolling Stones, as well as Eric Clapton and Nirvana, show notable co-purchase patterns, suggesting these artist pairings resonate with similar audiences.
- Artists like System Of A Down and Foo Fighters, and Queen and The Police, also have significant co-purchases, highlighting popular combinations among listeners.

**RECOMMENDATIONS:**

- Artist Pairing Promotions: Offer discounts on Green Day and Led Zeppelin albums purchased together and apply similar deals to pairings like Eric Clapton and Nirvana.
- Playlist Curation: Create playlists featuring frequently paired artists to appeal to similar tastes.

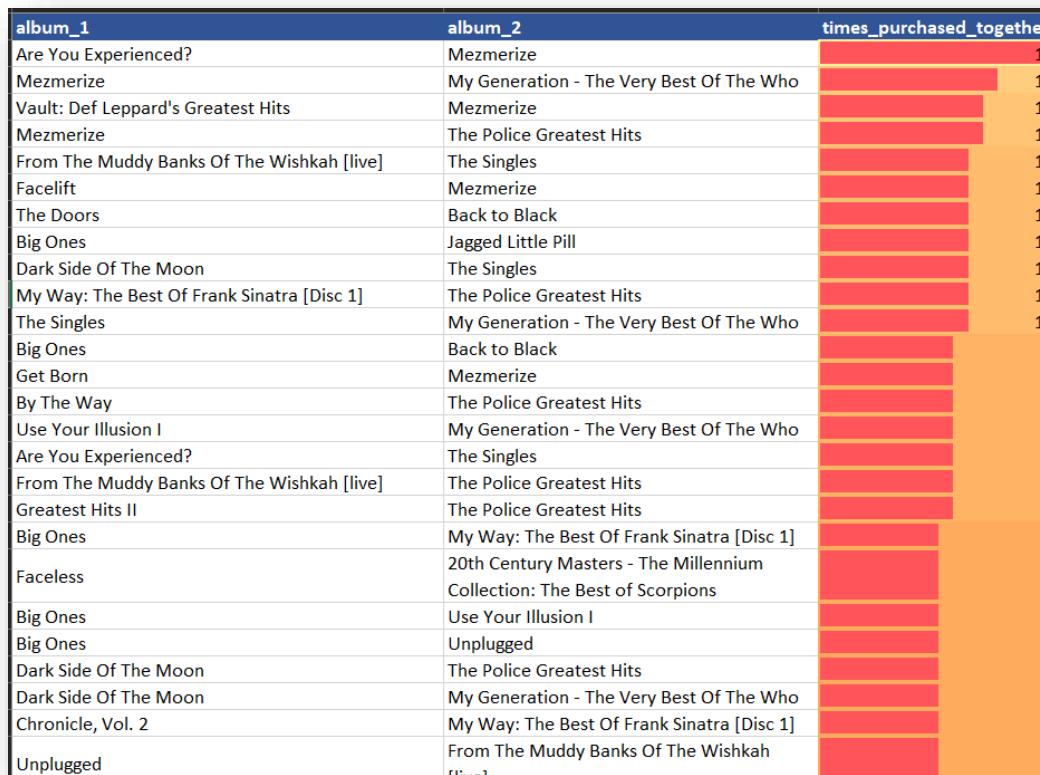
**c) Frequently Purchased Albums Together:**

```

1  -- 3. Album Affinity Analysis
2  WITH track_combinations AS (
3      SELECT
4          il1.track_id AS track_id_1,
5          il2.track_id AS track_id_2,
6          COUNT(*) AS times_purchased_together
7      FROM invoice_line il1
8      JOIN invoice_line il2
9      ON il1.invoice_id = il2.invoice_id AND il1.track_id < il2.track_id
10     GROUP BY il1.track_id, il2.track_id
11  ),
12  album_combinations AS (
13      SELECT
14          al1.album_id AS album_id_1,
15          al2.album_id AS album_id_2,
16          COUNT(*) AS times_purchased_together
17      FROM track_combinations tc
18      JOIN track t1 ON tc.track_id_1 = t1.track_id
19      JOIN album al1 ON t1.album_id = al1.album_id
20      JOIN track t2 ON tc.track_id_2 = t2.track_id
21      JOIN album al2 ON t2.album_id = al2.album_id
22      WHERE al1.album_id <> al2.album_id
23      GROUP BY al1.album_id, al2.album_id
24  )
25  SELECT
26      al1.title AS album_1,
27      al2.title AS album_2,
28      ac.times_purchased_together
29  FROM album_combinations ac
30  JOIN album al1 ON ac.album_id_1 = al1.album_id
31  JOIN album al2 ON ac.album_id_2 = al2.album_id
32  ORDER BY ac.times_purchased_together DESC;

```

album_1	album_2	times_purchased_together
Are You Experienced?	Mezmerize	16
Mezmerize	My Generation - The Very Best Of The Who	12
Vault: Def Leppard's Greatest Hits	Mezmerize	11
Mezmerize	The Police Greatest Hits	11
From The Muddy Banks Of The Wishkah [live]	The Singles	10
Facelift	Mezmerize	10
The Doors	Back to Black	10
Big Ones	Jagged Little Pill	10
Dark Side Of The Moon	The Singles	10
My Way: The Best Of Frank Sinatra [Disc 1]	The Police Greatest Hits	10
The Singles	My Generation - The Very Best Of The Who	10
Big Ones	Back to Black	9
Get Born	Mezmerize	9
By The Way	The Police Greatest Hits	9
Use Your Illusion I	My Generation - The Very Best Of The Who	9
Are You Experienced?	The Singles	9
From The Muddy Banks Of The Wishkah [live]	The Police Greatest Hits	9
Greatest Hits II	The Police Greatest Hits	9
Big Ones	My Way: The Best Of Frank Sinatra [Disc 1]	8
Faceless	20th Century Masters - The Millennium Coll...	8
Big Ones	Use Your Illusion I	8

**INSIGHTS:**

- "Mezmerize" is a popular album paired with several others, including "Are You Experienced?" and "The Police Greatest Hits," indicating it resonates well with fans of diverse classic rock albums.

- Albums like "The Police Greatest Hits" and "My Generation - The Very Best Of The Who" frequently appear together, suggesting a strong preference for classic rock compilations.
- "Big Ones" and "Back to Black" also show notable co-purchase rates, highlighting popular cross-genre interests among listeners.

### **RECOMMENDATIONS:**

- **Album Pair Deals:** Offer combined deals on album pairs like "Are You Experienced?" and "Mezmerize."
- **Special Editions:** Create box sets with frequently paired albums to attract collectors.

### **GENERAL CROSS-SELLING INITIATIVES:**

- **Curated Collections:** Create and market "Perfect Pairings" or "Ultimate Combo Packs" based on genre and artist affinities.
- **Personalized Recommendations:** Suggest related albums (e.g., Metal or Alternative & Punk) when a customer buys a Rock album.
- **Targeted Marketing:** Use email, ads, or notifications to promote synergies between popular genres, artists, and albums.

## 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?

### a) Customer Purchasing Behaviours:

```
1 -- Customer Purchasing Behaviors by Region
2 WITH purchase_frequency AS (
3     SELECT
4         customer_id,
5         COUNT(invoice_id) AS total_purchase_freq,
6         SUM(total) AS total_spending,
7         AVG(total) AS avg_order_value
8     FROM invoice
9     GROUP BY customer_id
10    ),
11 customer_region_summary AS (
12     SELECT
13         c.customer_id,
14         c.country,
15         COALESCE(c.state,'N.A') as state,
16         c.city,
17         pf.total_purchase_freq,
18         pf.total_spending,
19         pf.avg_order_value
20     FROM customer c
21     JOIN purchase_frequency pf ON c.customer_id = pf.customer_id
22    ),
23 regional_summary AS (
24     SELECT
25         country,
26         state,
27         city,
28         ROUND(COUNT(DISTINCT customer_id),2) AS total_customers,
29         ROUND(SUM(total_purchase_freq),2) AS total_purchases,
30         ROUND(SUM(total_spending),2) AS total_spending,
31         ROUND(AVG(avg_order_value),2) AS avg_order_value,
32         ROUND(AVG(total_purchase_freq),2) AS avg_purchase_frequency
33     FROM customer_region_summary
34     GROUP BY country, state, city
35    )
36     SELECT
37         country,
38         state,
39         city,
40         total_customers,
41         total_purchases,
42         total_spending,
43         avg_order_value,
44         avg_purchase_frequency
45     FROM regional_summary
46     ORDER BY total_spending DESC;
```

country	state	city	total_customers	total_purchases	total_spending	avg_order_value	avg_purchase_frequency
Czech Republic	N.A	Prague	2	30	273.24	9.38	15.00
USA	CA	Mountain View	2	20	169.29	8.60	10.00
United Kingdom	N.A	London	2	19	166.32	8.72	9.50
Germany	N.A	Berlin	2	20	158.40	7.94	10.00
France	N.A	Paris	2	18	151.47	8.42	9.00
Brazil	SP	São Paulo	2	22	129.69	5.98	11.00
Ireland	Dublin	Dublin	1	13	114.84	8.83	13.00
India	N.A	Delhi	1	13	111.87	8.61	13.00
Brazil	SP	São José dos Campos	1	13	108.90	8.38	13.00
Brazil	DF	Brasília	1	15	106.92	7.13	15.00
Portugal	N.A	Lisbon	1	13	102.96	7.92	13.00
Canada	QC	Montréal	1	9	99.99	11.11	9.00
France	N.A	Bordeaux	1	11	99.99	9.09	11.00
Spain	N.A	Madrid	1	11	98.01	8.91	11.00
USA	WA	Redmond	1	12	98.01	8.17	12.00
Chile	N.A	Santiago	1	13	97.02	7.46	13.00
Germany	N.A	Frankfurt	1	10	94.05	9.41	10.00
USA	FL	Orlando	1	12	92.07	7.67	12.00
Canada	ON	Ottawa	1	13	91.08	7.01	13.00



- **High Average Order Value & Purchase Frequency:**

- Prague, Czech Republic: \$9.38 average order value, frequency 15.
- Mountain View, USA: \$8.60 average order value, frequency 10.
- London, UK: \$8.72 average order value, frequency 9.5.

**Insight:** Target affluent regions with premium products and tailored marketing.

- **Moderate Average Order Value & Purchase Frequency:**

- São Paulo, Brazil: \$5.98 average order value, frequency 11.
- Berlin, Germany: \$7.94 average order value, frequency 10.

**Insight:** Focus on customer retention and repeat purchase strategies.

- **Low Average Order Value & Purchase Frequency:**

- Edmonton, Canada: \$2.97 average order value, frequency 10.
- Copenhagen, Denmark: \$3.76 average order value, frequency 10.

**Insight:** Use promotions to boost order values and frequency; consider local economic factors.

### b) Churn Rates:

```
● ● ●
1  -- Churn Rate by Region
2  WITH last_purchase AS (
3      SELECT
4          c.customer_id,
5          c.country,
6          COALESCE(c.state,'N.A') AS state,
7          c.city,
8          MAX(i.invoice_date) AS last_purchase_date
9      FROM customer c
10     JOIN invoice i ON c.customer_id = i.customer_id
11    GROUP BY c.customer_id, c.country, c.state, c.city
12  ),
13  churned_customers AS (
14      SELECT
15          country,
16          state,
17          city,
18          COUNT(customer_id) AS churned_customers
19      FROM last_purchase
20      WHERE last_purchase_date < DATE_SUB(CURDATE(), INTERVAL 1 YEAR)
21      GROUP BY country, state, city
22  )
23  SELECT
24      lc.country,
25      lc.state,
26      lc.city,
27      lc.churned_customers,
28      COUNT(c.customer_id) AS total_customers,
29      (lc.churned_customers / COUNT(c.customer_id)) * 100 AS churn_rate
30  FROM churned_customers lc
31  JOIN customer c
32  ON lc.country = c.country
33      AND lc.state = c.state
34      AND lc.city = c.city
35  GROUP BY lc.country, lc.state, lc.city
36  ORDER BY churn_rate DESC;
```

	country	state	city	churned_customers	total_customers
▶	Brazil	SP	São José dos Campos	1	1
	Canada	QC	Montréal	1	1
	Brazil	SP	São Paulo	2	2
	Brazil	RJ	Rio de Janeiro	1	1
	Brazil	DF	Brasília	1	1
	Canada	AB	Edmonton	1	1
	Canada	BC	Vancouver	1	1
	USA	CA	Mountain View	2	2
	USA	WA	Redmond	1	1
	USA	NY	New York	1	1
	USA	CA	Cupertino	1	1
	USA	NV	Reno	1	1
	USA	FL	Orlando	1	1
	USA	MA	Boston	1	1
	USA	IL	Chicago	1	1
	USA	WI	Madison	1	1
	USA	TX	Fort Worth	1	1
	USA	AZ	Tucson	1	1
	USA	UT	Salt Lake City	1	1

- **High Churn Regions:** São José dos Campos, Montréal, São Paulo, Rio de Janeiro, Brasília, Edmonton, Mountain View, and Redmond.

**Insight:** High churn suggests a need for better customer retention strategies, such as loyalty programs and enhanced support.

### CORRELATION WITH LOCAL FACTORS:

- **Affluent Areas:** High spending and frequency suggest higher income and better economic conditions.
- **Emerging Markets:** Moderate spending with high frequency points to budget-conscious middle-class consumers.
- **High Churn in Competitive Areas:** Competition and high living costs may affect loyalty.

### RECOMMENDATIONS:

- **High-Value Areas:** Offer exclusive products and premium pricing.
- **Moderate Areas:** Enhance engagement with loyalty programs and value offers.
- **High Churn Areas:** Address churn with discounts, improved service, and feedback collection.

**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?



```

1  WITH customer_profile AS (
2      SELECT
3          c.customer_id,
4          c.country,
5          COALESCE(c.state,'N.A') AS state,
6          c.city,
7          MAX(i.invoice_date) AS last_purchase_date,
8          SUM(i.total) AS total_spending,
9          COUNT(i.invoice_id) AS purchase_frequency,
10         AVG(i.total) AS avg_order_value
11     FROM customer c
12     LEFT JOIN invoice i ON c.customer_id = i.customer_id
13     GROUP BY c.customer_id
14   ),
15   churn_risk AS (
16     SELECT
17         cp.customer_id,
18         cp.country,
19         cp.state,
20         cp.city,
21         cp.total_spending,
22         cp.purchase_frequency,
23         cp.avg_order_value,
24         CASE
25             WHEN cp.last_purchase_date < DATE_SUB(CURDATE(), INTERVAL 1 YEAR)
26                 THEN 'High Risk'
27             WHEN cp.total_spending < 100
28                 THEN 'Medium Risk'
29             ELSE 'Low Risk'
30         END AS risk_profile
31     FROM customer_profile cp
32   ),

```

```

33   risk_summary AS (
34     SELECT
35         country,
36         state,
37         city,
38         risk_profile,
39         ROUND(COUNT(customer_id),2) AS num_customers,
40         ROUND(AVG(total_spending),2) AS avg_total_spending,
41         ROUND(AVG(purchase_frequency),2) AS avg_purchase_frequency,
42         ROUND(AVG(avg_order_value),2) AS avg_order_value
43     FROM churn_risk
44     GROUP BY country, state, city, risk_profile
45   )
46   SELECT
47     country,
48     state,
49     city,
50     risk_profile,
51     num_customers,
52     avg_total_spending,
53     avg_purchase_frequency,
54     avg_order_value
55   FROM risk_summary
56   ORDER BY risk_profile DESC, avg_total_spending DESC;

```

country	state	city	risk_profile	num_customers	avg_total_spending	avg_purchase_frequency	avg_order_value
Czech Republic	N.A	Prague	High Risk	2	136.62	15.00	9.38
Ireland	Dublin	Dublin	High Risk	1	114.84	13.00	8.83
India	N.A	Delhi	High Risk	1	111.87	13.00	8.61
Brazil	SP	São José dos Campos	High Risk	1	108.90	13.00	8.38
Brazil	DF	Brasília	High Risk	1	106.92	15.00	7.13
Portugal	N.A	Lisbon	High Risk	1	102.96	13.00	7.92
Canada	QC	Montréal	High Risk	1	99.99	9.00	11.11
France	N.A	Bordeaux	High Risk	1	99.99	11.00	9.09
Spain	N.A	Madrid	High Risk	1	98.01	11.00	8.91
USA	WA	Redmond	High Risk	1	98.01	12.00	8.17
Chile	N.A	Santiago	High Risk	1	97.02	13.00	7.46
Germany	N.A	Frankfurt	High Risk	1	94.05	10.00	9.41
USA	FL	Orlando	High Risk	1	92.07	12.00	7.67
USA	NV	Reno	High Risk	1	91.08	11.00	8.28
Canada	ON	Ottawa	High Risk	1	91.08	13.00	7.01
USA	TX	Fort Worth	High Risk	1	86.13	12.00	7.18
USA	CA	Mountain View	High Risk	2	84.65	10.00	8.60
USA	AZ	Tucson	High Risk	1	84.15	9.00	9.35
United Kingdom	N.A	London	High Risk	2	83.16	9.50	8.72



Below are the key insights and contributing factors to this risk:

#### **1. High Spending with High Frequency (Potential Churn Risk):**

- Regions: Prague, Czech Republic; Delhi, India; Brasília, Brazil.
- Risk Factors: High expectations and increased competition may lead to churn if service or deals decline.

#### **2. Moderate Spending with Moderate Frequency:**

- Regions: Mountain View, USA; Stuttgart, Germany.
- Risk Factors: Dependence on satisfaction and economic sensitivity could lead to reduced spending or churn if conditions change.

#### **3. Low Spending with High Frequency:**

- Regions: Yellowknife, Canada; Stockholm, Sweden.
- Risk Factors: Value-seeking and promotional dependence make these customers sensitive to price increases or reduced deals.

#### **4. Low Spending with Low Frequency (High Churn Risk):**

- Regions: Edmonton, Canada; Copenhagen, Denmark.
- Risk Factors: Low engagement and price sensitivity increase churn risk if offerings aren't compelling or prices rise.

#### **FACTORS CONTRIBUTING TO HIGH RISK:**

Economic instability, competitive marketplaces, declining service/product quality, and unmet demographic needs.

#### **RECOMMENDATIONS:**

- **Enhanced Engagement:** Use loyalty programs and personalized offers for high-spending customers.
- **Value-Based Marketing:** Highlight value for money and maintain promotions for price-sensitive regions.
- **Service Improvement:** Ensure high-quality service to retain frequent buyers.

- **Churn Prediction:** Implement predictive analytics to identify and address at-risk customers early.

These strategies will help reduce churn and maximize customer lifetime value.

## 7. Customer Lifetime Value Modelling: 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?

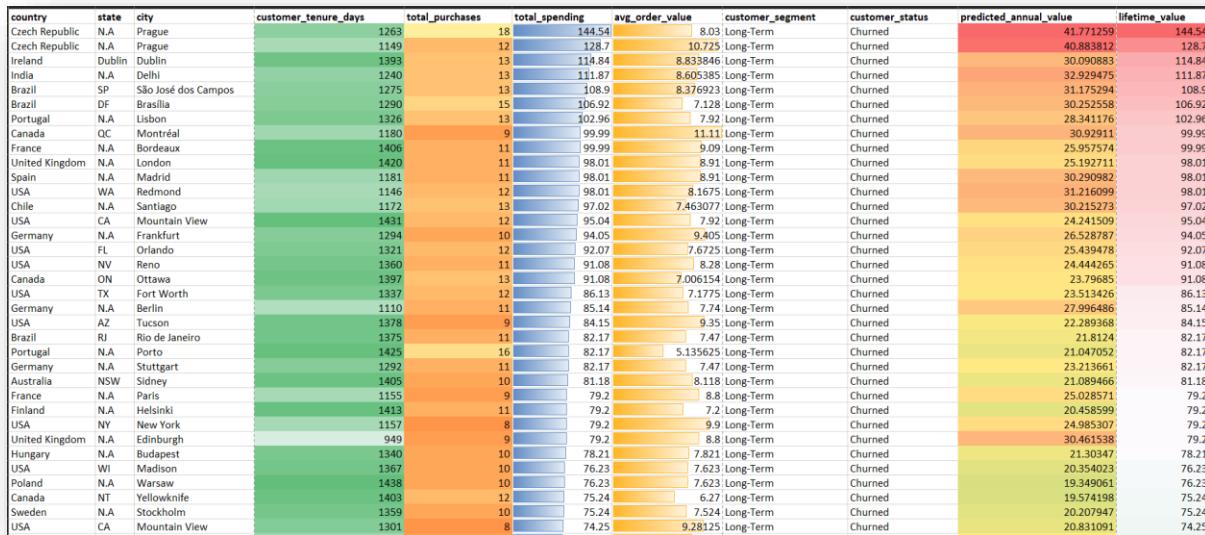


```

1 WITH customer_profile AS (
2     SELECT
3         c.customer_id,
4         CONCAT(c.first_name, ' ', c.last_name) AS customers,
5         c.country,
6         COALESCE(c.state,'N.A') AS state,
7         c.city,
8         MIN(i.invoice_date) AS first_purchase_date,
9         MAX(i.invoice_date) AS last_purchase_date,
10        DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS customer_tenure_days,
11        COUNT(i.invoice_id) AS total_purchases,
12        SUM(i.total) AS total_spending,
13        AVG(i.total) AS avg_order_value
14    FROM customer c
15    LEFT JOIN invoice i ON c.customer_id = i.customer_id
16    GROUP BY c.customer_id
17 ),
18 customer_lifetime_value AS (
19     SELECT
20         cp.customer_id,
21         cp.customers,
22         cp.country,
23         cp.state,
24         cp.city,
25         cp.customer_tenure_days,
26         cp.total_purchases,
27         cp.total_spending,
28         cp.avg_order_value,
29         CASE
30             WHEN cp.customer_tenure_days >= 365 THEN 'Long-Term'
31             ELSE 'Short-Term'
32         END AS customer_segment,
33         CASE
34             WHEN cp.last_purchase_date < DATE_SUB(CURDATE(), INTERVAL 1 YEAR) THEN 'Churned'
35             ELSE 'Active'
36         END AS customer_status,
37         (cp.total_spending / GREATEST(cp.customer_tenure_days, 1)) * 365 AS predicted_annual_value,
38         cp.total_spending AS lifetime_value
39     FROM customer_profile cp
40 ),
41 segment_analysis AS (
42     SELECT
43         customer_segment,
44         customer_status,
45         COUNT(customer_id) AS num_customers,
46         AVG(customer_tenure_days) AS avg_tenure_days,
47         AVG(total_spending) AS avg_lifetime_value,
48         AVG(predicted_annual_value) AS avg_predicted_annual_value
49     FROM customer_lifetime_value
50     GROUP BY customer_segment, customer_status
51 ),
52 churn_analysis AS (
53     SELECT
54         country,
55         state,
56         city,
57         customer_segment,
58         COUNT(customer_id) AS churned_customers,
59         AVG(total_spending) AS avg_lifetime_value
60     FROM customer_lifetime_value
61     WHERE customer_status = 'Churned'
62     GROUP BY country, state, city, customer_segment
63 )
64 SELECT *
65 FROM customer_lifetime_value
66 ORDER BY lifetime_value DESC;

```

customer_id	customers	country	state	city	customer_tenure_days	total_purchases	total_spending	avg_order_value	customer_segment	customer_status	predicted_annual_value	lifetime_value
5	František Wichterlová	Czech Republic	N.A	Prague	1263	18	144.54	8.030000	Long-Term	Churned	41.771259	144.54
6	Helena Holý	Czech Republic	N.A	Prague	1149	12	128.70	10.725000	Long-Term	Churned	40.883812	128.70
46	Hugh O'Reilly	Ireland	Dublin	Dublin	1393	13	114.84	8.833846	Long-Term	Churned	30.090883	114.84
58	Manoj Pareek	India	N.A	Delhi	1240	13	111.87	8.605385	Long-Term	Churned	32.924975	111.87
1	Luís Gonçalves	Brazil	SP	São José dos Campos	1275	13	108.90	8.376923	Long-Term	Churned	31.175294	108.90
13	Fernanda Ramos	Brazil	DF	Brasília	1290	15	106.92	7.128000	Long-Term	Churned	30.252558	106.92
34	João Fernandes	Portugal	N.A	Lisbon	1326	13	102.96	7.920000	Long-Term	Churned	28.341176	102.96
3	François Tremblay	Canada	QC	Montréal	1180	9	99.99	11.110000	Long-Term	Churned	30.929110	99.99
42	Wyatt Girard	France	N.A	Bordeaux	1406	11	99.99	9.090000	Long-Term	Churned	25.957574	99.99
53	Phil Hughes	United Kingdom	N.A	London	1420	11	98.01	8.910000	Long-Term	Churned	25.192711	98.01
50	Enrique Muñoz	Spain	N.A	Madrid	1181	11	98.01	8.910000	Long-Term	Churned	30.290982	98.01
17	Jack Smith	USA	WA	Redmond	1146	12	98.01	8.167500	Long-Term	Churned	31.216099	98.01
57	Luis Rojas	Chile	N.A	Santiago	1172	13	97.02	7.463077	Long-Term	Churned	30.215273	97.02
20	Dan Miller	USA	CA	Mountain View	1431	12	95.04	7.920000	Long-Term	Churned	24.241509	95.04
37	Fynn Zimmermann	Germany	N.A	Frankfurt	1294	10	94.05	9.405000	Long-Term	Churned	26.528787	94.05
22	Heather Leacock	USA	FL	Orlando	1321	12	92.07	7.672500	Long-Term	Churned	25.439478	92.07
21	Kathy Chase	USA	NV	Reno	1360	11	91.08	8.280000	Long-Term	Churned	24.444265	91.08



we can analyse the customer lifetime value (CLV) and identify patterns using conditional formatting among customers who have churned.

## INSIGHTS:

### 1. Customer Segments and Predicted Lifetime Value:

- High Lifetime Value Customers: Example: František Wichterlová from Prague. High spending and average order value. Need targeted retention strategies.

### 2. Low Lifetime Value Customers: Example: Mark Philips from Edmonton. Lower spending and order values. More price-sensitive or less engaged.

### 3. Common Characteristics of Churned Customers:

- Long-Term Customers: Churn can occur even with long tenure. Example: Helena Holý from Prague.
- Moderate Spending: Customers with moderate spending may churn due to better alternatives. Example: Camille Bernard from Paris.
- High Risk, Low Engagement: High predicted lifetime value but low engagement. Example: Robert Brown from Toronto.

**4. Observed Purchase Patterns:**

- Higher Engagement: Frequent purchases reduce churn risk.
- Lower Order Value: Customers with low average order values but long tenure may be at risk if they don't perceive value.

**5. Targeted Marketing and Loyalty Strategies:**

- Personalized Offers: For high CLV customers, use targeted promotions.
- Loyalty Programs: Reward moderate spenders to increase engagement.
- Reactivation Campaigns: For long-term, low-spending customers, offer incentives to boost spending.

**RECOMMENDATIONS:**

Identifying churn patterns and characteristics helps in creating effective retention strategies, enhancing customer loyalty, and reducing churn.

**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?**

- we don't have direct access to data related to promotional campaigns such as discounts, events, or email marketing.
- However, if such data were available, the impact of promotional campaigns on customer acquisition, retention, and overall sales could be measured using the following approach:

**APPROACH TO MEASURE IMPACT:****a) Identify Promotional Periods:**

- If we had a table or column that tracked when promotions occurred, we would first identify the dates or periods during which promotions were active.

**b) Segment Data:**

- Segment the data into two groups: periods with active promotions and periods without promotions.

**c) Customer Acquisition:**

- Calculate the number of new customers acquired during the promotional periods versus non-promotional periods.

**d) Customer Retention:**

- Analyse the retention rate by comparing the number of repeat purchases made by customers during and after promotional periods.

**e) Sales Performance:**

- Compare total sales, average order value, and basket size during promotional periods with non-promotional periods.

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

If no predefined questions are given, the approach to solving the problem involves a systematic, exploratory analysis of available data:

### 1. Understand Data Schema

- Identify entities and relationships (e.g., Employee, Customer, Invoice).
- Determine key metrics like sales performance and customer demographics.

### 2. Data Exploration and Quality Assessment

- Check for missing values and duplicates.
- Clean data by imputing or removing incomplete records.

### 3. Sales Performance Analysis

- Analyse total sales, average order value, and top-selling tracks across time, geography, and customer segments.

### 4. Customer Analysis

- Analyse demographics and segment customers by purchase history.
- Calculate churn rates and model Customer Lifetime Value (CLV).

### 5. Market and Regional Analysis

- Assess sales performance across regions and correlate with external factors like demographics.
- Product Affinity and Cross-Selling
- Identify frequently purchased product combinations for cross-selling opportunities.

### 6. Promotional Impact Analysis

- Measure the impact of promotions on sales and customer behaviour.

### 7. Strategic Insights and Recommendations

- Identify growth opportunities, optimize pricing, and develop targeted marketing strategies.

### 8. Documentation and Reporting

- Summarize findings with visualizations and provide actionable recommendations.

## CONCLUSION:

This approach involves comprehensive data exploration to uncover insights that guide strategic decisions, enhance customer retention, optimize sales, and identify growth opportunities. The process is iterative and evolves with new insights.

**10. How can you alter the "Albums" table to add a new column named "Release Year" of type INTEGER to store the release year of each album?**



```
ALTER TABLE Album
ADD COLUMN ReleaseYear INT;
select * from Album;

UPDATE album
SET ReleaseYear = 2017
WHERE album_id = 1;

UPDATE album
SET ReleaseYear = 2017
WHERE album_id = 2;

UPDATE album
SET ReleaseYear = 2017
WHERE album_id = 3;

UPDATE album
SET ReleaseYear = 2017
WHERE album_id = 4;

UPDATE album
SET ReleaseYear = 2017
WHERE album_id = 5;
```

```
UPDATE album
SET ReleaseYear = 2018
WHERE album_id = 6;

UPDATE album
SET ReleaseYear = 2018
WHERE album_id = 7;

UPDATE album
SET ReleaseYear = 2018
WHERE album_id = 8;

UPDATE album
SET ReleaseYear = 2018
WHERE album_id = 9;

UPDATE album
SET ReleaseYear = 2018
WHERE album_id = 10;
```

album_id	title	artist_id	ReleaseYear
1	For Those About To Rock We Salute You	1	2017
2	Balls to the Wall	2	2017
3	Restless and Wild	2	2017
4	Let There Be Rock	1	2017
5	Big Ones	3	2017
6	Jagged Little Pill	4	2018
7	Facelift	5	2018
8	Warner 25 Anos	6	2018
9	Plays Metallica By Four Cellos	7	2018
10	Audioslave	8	2018

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.

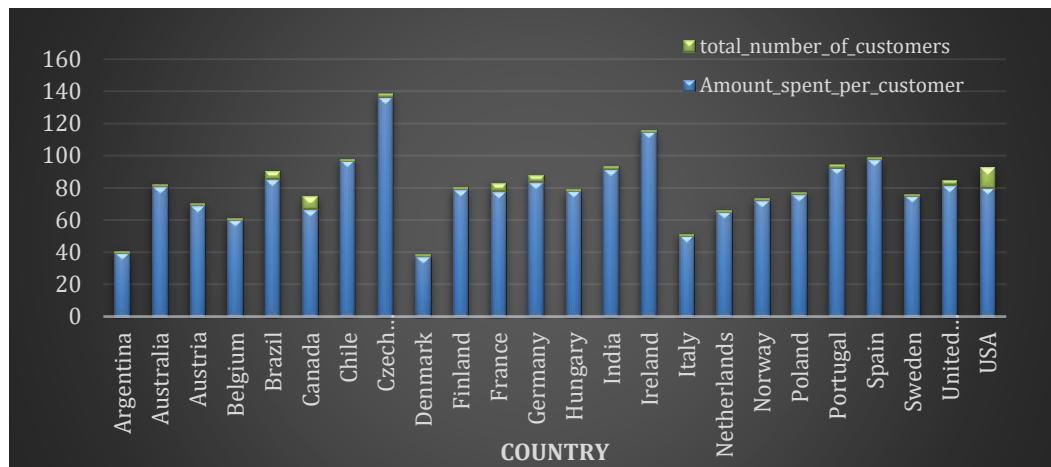


```

WITH tracks_per_customer AS (
    SELECT
        i.customer_id,
        SUM(il.quantity) AS total_tracks
    FROM invoice i
    JOIN invoice_line il ON i.invoice_id = il.invoice_id
    GROUP BY i.customer_id
),
customer_spending AS (
    SELECT
        c.country,
        c.customer_id,
        SUM(i.total) AS total_spent,
        tpc.total_tracks
    FROM customer c
    JOIN invoice i ON c.customer_id = i.customer_id
    JOIN tracks_per_customer tpc ON c.customer_id = tpc.customer_id
    GROUP BY c.country, c.customer_id, tpc.total_tracks
)
SELECT
    cs.country,
    COUNT(DISTINCT cs.customer_id) AS number_of_customers,
    ROUND(AVG(cs.total_spent),2) AS average_amount_spent_per_customer,
    ROUND(AVG(cs.total_tracks),2) AS average_tracks_purchased_per_customer
FROM customer_spending cs
GROUP BY cs.country
ORDER BY average_amount_spent_per_customer DESC;

```

	country	number_of_customers	average_amount_spent_per_customer	average_tracks_purchased_per_customer
▶	Czech Republic	2	136.62	138.00
	Ireland	1	114.84	116.00
	Spain	1	98.01	99.00
	Chile	1	97.02	98.00
	Portugal	2	92.57	93.50
	India	2	91.58	92.50
	Brazil	5	85.54	86.40
	Germany	4	83.66	84.50
	United Kingdom	3	81.84	82.67
	Australia	1	81.18	82.00
	USA	13	80.04	80.85
	Finland	1	79.20	80.00
	Hungary	1	78.21	79.00
	France	5	77.81	78.60
	Poland	1	76.23	77.00
	Sweden	1	75.24	76.00
	Norway	1	72.27	73.00
	Austria	1	69.30	70.00

**INSIGHTS:**

- The Czech Republic has the highest average amount spent per customer at \$136.62, indicating high customer value in this region.
- The USA has the largest customer base with 13 customers but a lower average spend of \$80.04, suggesting a broad but less intense engagement.
- Customers from India, Portugal, and Brazil show moderate average spending, ranging from \$85.54 to \$92.57, with varied customer bases.
- Countries like Argentina and Denmark have the lowest average spend per customer, around \$40, reflecting limited purchasing activity.
- Countries with only one customer, like Ireland and Spain, have relatively high average spends, highlighting significant individual contributions.

===== END OF SUBJECTIVE QUESTIONS =====