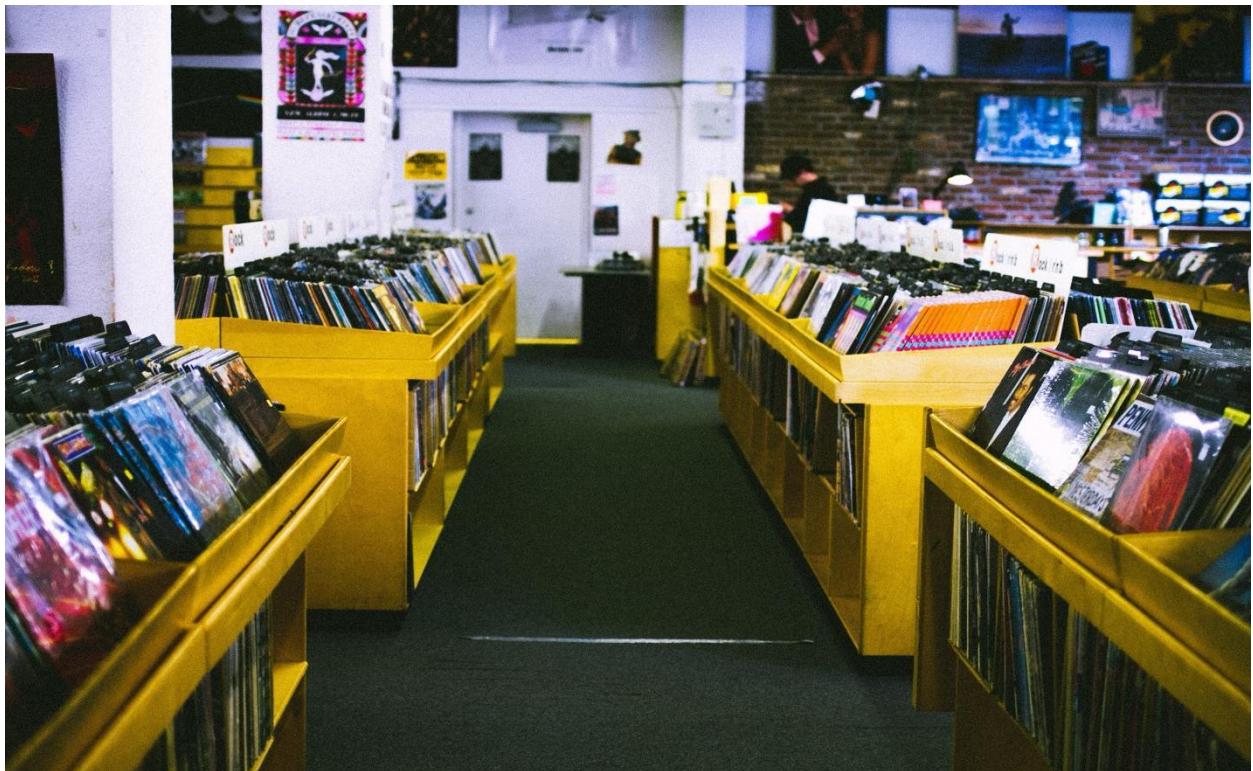


## CHINOOK MUSIC STORE



SUMANTH RAJ S  
6-5-25

**TABLE OF  
CONTENTS**

- 1.Objective questions
- 2.Subjective questions

## Objective Questions

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

Yes, there are missing values and duplicates in the original dataset

Approach:

- To check if there are missing values in the tables, for each table the null check should be conducted using SQL Queries with WHERE and OR conditions
- The following code was used across the key tables  

```
SELECT *
FROM table_name
WHERE column_name IS NULL;
```
- To check if there were any duplicate values, COUNT function was used across various tables

Examples of queries used to check null and duplicates :

Similar queries are used across tables

(FILE : CHINOOK\_QUERIES.sql)

```

-- 1.Checking for null values and duplicates

-- *** ***
-- album table
select * from album ORDER BY album_id;

-- null check
SELECT *
FROM album
WHERE album_id IS NULL or title IS NULL OR artist_id IS NULL;

-- duplicates check

SELECT title,artist_id, COUNT(*) AS count
FROM album
GROUP BY title,artist_id
HAVING COUNT(*) > 1;

```

```

-- employee table
select * from employee;

-- duplicate check
SELECT first_name, last_name, email, COUNT(*) as dup_count
FROM employee
GROUP BY first_name, last_name, email
HAVING COUNT(*) > 1;

-- null check for employee -- (only CEO has null)
SELECT *
FROM employee
WHERE employee_id IS NULL
    OR last_name IS NULL
    OR first_name IS NULL
    OR title IS NULL
    OR reports_to IS NULL
    OR birthdate IS NULL
    OR hire_date IS NULL
    OR address IS NULL
    OR city IS NULL
    OR state IS NULL
    OR country IS NULL
    OR postal_code IS NULL
    OR phone IS NULL
    OR fax IS NULL
    OR email IS NULL;

```

RESULTS :

2

TABLE	NULL values	DUPLICATE values
album	NO	NO
artist	NO	NO
employee	YES  (CEO of the organisation reports to none hence the value is null in reports_to column, this requires no modification ,hence null kept as it is )	NO
genre	NO	NO
invoice	NO	NO
invoice_line	NO	NO
media_type	NO	NO
playlist	NO	YES  column name has duplicates Music , Movies , TV shows are repeated twice with different playlist_id  Handled using DELETE Query  Attached query in the document below
playlist_track	NO	YES  playlist_id's which are associated with same playlist_names are duplicates

		<p>along with same track _id 's for MUSIC and TV SHOWS category Handled using DELETE QUERY referencing the duplicated id's i.e ( 8,10 ) for MUSIC and TV SHOWS</p> <p>Attached query in the document below</p>
track	<p>YES</p> <p>composer names have null values Used UPDATE QUERY to update the NULL values using artist names ,which makes it meaningful in this context rather than filling unknown/NA</p> <p>Attached query in the document below</p>	NO
customer	<p>YES</p> <p>1.COMPANY names with null values are updated with "Not Available"</p> <p>2.PHONE - only 1 null value for HUNGARY is updated using the country code appended by placeholder</p> <p>3.FAX - null values replaced by "Not Available"</p> <p>4.POSTAL CODES : strategy: based on the country and city</p>	NO

	<p>combination names, picked up values from web (verified) based on non null data values</p> <p>5.STATE : recognised the pattern of state codes based on non null values and replaced accordingly with update query using web inputs</p> <p>Attached query in the document below</p>	
--	--	--

#### QUERIES USED TO HANDLE NULLS AND DUPLICATES :

1.Playlist table with duplicates handled using DELETE query keeping the 1<sup>st</sup> instances of the duplicates.

```
-- playlist table
select * from playlist;

-- null check
select * from playlist where name is null ;

-- duplicate check
select name ,count(*) as dup_cnt from
playlist
group by name
having count(*) >1 ;

-- deleting duplicates :
SET SQL_SAFE_UPDATES = 0;

DELETE p1
FROM playlist p1
JOIN playlist p2 ON p1.name = p2.name AND p1.playlist_id > p2.playlist_id;
```

2.playlist\_track table with duplicates handled as below

```
-- playlist_track table
select * from playlist_track;

-- null check
select * from playlist_track where track_id is null ;

-- duplicate check (music and tv shows have been duplicated with track id's hence deleting those id's)

DELETE from playlist_track where playlist_id in (8,10);
```

3. track table with null values for composers handled as below by fetching artist names

- `UPDATE track t  
JOIN album al ON t.album_id = al.album_id  
JOIN artist ar ON al.artist_id = ar.artist_id  
SET t.composer = ar.name  
WHERE t.composer IS NULL;`

4.Customer table with null values across different columns handled as below

```
-- postal codes with null as values
-- strategy: based on the country and city name picked up values from web (verified) based on non null data values

-- Update Portuguese postal codes
UPDATE customer
SET postal_code = '1100-012'
WHERE customer_id = 34; -- Lisbon (Rua da Assunção 53)

UPDATE customer
SET postal_code = '4200-450'
WHERE customer_id = 35; -- Porto (Rua dos Campeões Europeus de Viena)

-- Update Irish postal code (Eircode format)
UPDATE customer
SET postal_code = 'D02 KX84'
WHERE customer_id = 46; -- Dublin (3 Chatham Street)

-- Update Chilean postal code
UPDATE customer
SET postal_code = '8320000'
WHERE customer_id = 57; -- Santiago (Calle Lira)

-- null handling

-- company with value null can be individuals
UPDATE customer
SET company = 'Not Available'
WHERE company IS NULL;

-- phone value is null for one record where country is Hungary ,handled as below
UPDATE customer
SET phone = '+36 XXXX-XXXX' -- Hungarian country code + placeholder
WHERE country = 'Hungary' AND phone IS NULL;

-- fax with null as value
UPDATE customer SET fax = 'Not Available' WHERE fax IS NULL;
```

```

-- state with null values
-- strategy : analysed the pattern of the non null values and handled it as below
UPDATE customer
SET state = CASE
    -- Germany
    WHEN city = 'Stuttgart' THEN 'BW'      -- Baden-Württemberg
    WHEN city = 'Berlin' THEN 'BE'         -- Berlin (city-state)
    WHEN city = 'Frankfurt' THEN 'HE'       -- Hesse

    -- France (department codes)
    WHEN city = 'Paris' THEN '75'          -- Paris department
    WHEN city = 'Lyon' THEN '69'           -- Rhône department
    WHEN city = 'Bordeaux' THEN '33'        -- Gironde department
    WHEN city = 'Dijon' THEN '21'          -- Côte-d'Or department

    -- Other European cities
    WHEN city = 'Prague' THEN 'PRG'        -- Prague (capital)
    WHEN city = 'Brussels' THEN 'BRU'       -- Brussels-Capital Region
    WHEN city = 'Vienna' THEN 'W'          -- Vienna (Wien)
    WHEN city = 'Budapest' THEN 'BU'        -- Budapest (capital)
    WHEN city = 'Warsaw' THEN 'MZ'         -- Masovian Voivodeship
    WHEN city = 'Madrid' THEN 'M'          -- Madrid autonomous community
    WHEN city = 'Lisbon' THEN 'LIS'         -- Lisbon district
    WHEN city = 'Porto' THEN 'POR'         -- Porto district
    WHEN city = 'Stockholm' THEN 'AB'       -- Stockholm County

    -- Unitary countries (country codes)
    WHEN city = 'London' THEN 'GB'         -- United Kingdom
    WHEN city = 'Edinburgh' THEN 'GB'       -- United Kingdom
    WHEN city = 'Oslo' THEN 'NO'           -- Norway
    WHEN city = 'Copenhagen' THEN 'DK'      -- Denmark
    WHEN city = 'Helsinki' THEN 'FI'        -- Finland

    -- Rest of world
    WHEN city = 'Buenos Aires' THEN 'BA'     -- Buenos Aires
    WHEN city = 'Santiago' THEN 'RM'         -- Santiago Metropolitan Region
    WHEN city = 'Delhi' THEN 'DL'           -- Delhi NCT
    WHEN city = 'Bangalore' THEN 'KA'        -- Karnataka

    -- Default: First 2 letters of city (UPPERCASE)
    ELSE UPPER(SUBSTRING(city, 1, 2))
END
WHERE state IS NULL;

```

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

Approach: By computing total\_quantity , total\_revenue and ordering the values in the descending order to find the top picks

- a.identifying top 5 tracks in USA – WAR PIGS tops
- b.top artist in USA – Van Halen
- c.top genre of the top artist – Van Halen's top genre is ROCK

a.

```
369 -- top 5 tracks in USA :  
370 • SELECT  
371     t.track_id,  
372     t.name,  
373     SUM(il.quantity) AS total_quantity,  
374     SUM(il.quantity * il.unit_price) AS total_revenue  
375   FROM invoice i  
376   JOIN invoice_line il ON i.invoice_id = il.invoice_id  
377   JOIN track t ON il.track_id = t.track_id  
378   WHERE i.billing_country = 'USA'  
379   GROUP BY t.track_id, t.name  
380   ORDER BY total_revenue DESC, total_quantity DESC  
381   LIMIT 5;  
382
```

100% 12:372 1 error found

Result Grid Filter Rows: Search Export: Fetch rows:

track_id	name	total_quantity	total_revenue
3336	War Pigs	6	5.94
3465	You Know I'm No Good (feat. Ghostface Killah)	5	4.95
2560	Violent Pornography	4	3.96
13	Night Of The Long Knives	4	3.96
1995	Scentless Apprentice	4	3.96

```

382
383      -- top artist in USA
384 •   SELECT
385          a.artist_id,
386          a.name AS artist_name,
387          SUM(il.quantity) AS total_quantity,
388          SUM(il.quantity * il.unit_price) AS total_revenue
389      FROM invoice i
390      JOIN invoice_line il ON i.invoice_id = il.invoice_id
391      JOIN track t ON il.track_id = t.track_id
392      JOIN album alb ON t.album_id = alb.album_id
393      JOIN artist a ON alb.artist_id = a.artist_id
394      WHERE i.billing_country = 'USA'
395      GROUP BY a.artist_id, a.name
396      ORDER BY total_revenue DESC, total_quantity DESC
397      LIMIT 1;
398
100%  9:381 | 1 error found

```

Result Grid   Filter Rows:   Search   Export:   Fetch rows:

artist_id	artist_name	total_quant...	total_revenue
152	Van Halen	43	42.57

b.

```

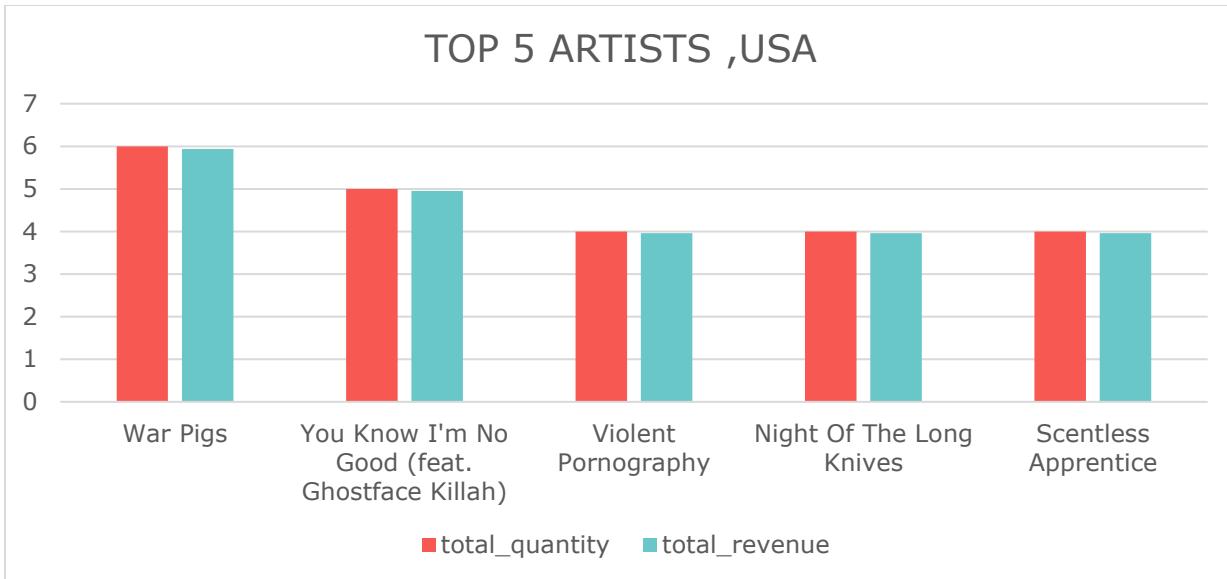
399      -- top genre of the top artist
400 •   SELECT
401          g.genre_id,
402          g.name AS genre_name,
403          SUM(il.quantity) AS total_quantity,
404          SUM(il.quantity * il.unit_price) AS total_revenue
405      FROM invoice i
406      JOIN invoice_line il ON i.invoice_id = il.invoice_id
407      JOIN track t ON il.track_id = t.track_id
408      JOIN album alb ON t.album_id = alb.album_id
409      JOIN artist a ON alb.artist_id = a.artist_id
410      JOIN genre g ON t.genre_id = g.genre_id
411      WHERE i.billing_country = 'USA'
412      AND a.artist_id = 152 -- top artist in USA as per previous query
413      GROUP BY g.genre_id, g.name
414      ORDER BY total_revenue DESC, total_quantity DESC;
415
100%  68:416 | 1 error found

```

Result Grid   Filter Rows:   Search   Export:   Fetch rows:

genre_id	genre_name	total_quant...	total_revenue
1	Rock	43	42.57

c.



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

NOTE : data related to age , gender is absent across dataset provided ,hence analysis is done based on the location (country , city , state)

Approach : By computing customer\_count and percentage of customers compared to total customers across locations, ordering the values in descending order to pick top customer base respectively

- a.countrywise customer base – USA tops
- b.citywise customer base – Berlin tops
- c.statewise customer base – SP (São Paulo) tops
- d.combined granular analysis – (Brazil,SP,Sao Paulo) tops

a.

```
421      -- Top countries
422 •  SELECT
423      country,
424      COUNT(*) AS customer_count,
425      ROUND(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM customer), 1) AS percentage
426  FROM customer
427  GROUP BY country
428  ORDER BY customer_count DESC
429  LIMIT 5;
430
431  ◇ 13:423 | 1 error found
```

Result Grid Filter Rows:  Search Export: Fetch rows:

country	customer_count	percentage
USA	13	22.0
Canada	8	13.6
Brazil	5	8.5
France	5	8.5
Germany	4	6.8

b.

```
430
431  -- Top cities globally (regardless of country/state)
432 •  SELECT
433      city,
434      COUNT(*) AS customer_count,
435      ROUND(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM customer), 1) AS percentage
436  FROM customer
437  WHERE city IS NOT NULL AND city != ''
438  GROUP BY city
439  ORDER BY customer_count DESC
440  LIMIT 5;
441
442  ◇ 38:437 | 1 error found
```

Result Grid Filter Rows:  Search Export: Fetch rows:

city	customer_count	percentage
Berlin	2	3.4
Mountain View	2	3.4
Paris	2	3.4
Prague	2	3.4
São Paulo	2	3.4

C.

```
442 -- Top states globally (regardless of country/city)
443 • SELECT
444     state,
445     COUNT(*) AS customer_count,
446     ROUND(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM customer), 1) AS percentage
447 FROM customer
448 WHERE state IS NOT NULL AND city != ''
449 GROUP BY state
450 ORDER BY customer_count DESC
451 LIMIT 5;
452
453
```

00% 1:441 | 1 error found

Result Grid Filter Rows: Search Export: Fetch rows:

state	customer_count	percentage
SP	3	5.1
CA	3	5.1
ON	2	3.4
PRG	2	3.4
AB	2	3.4

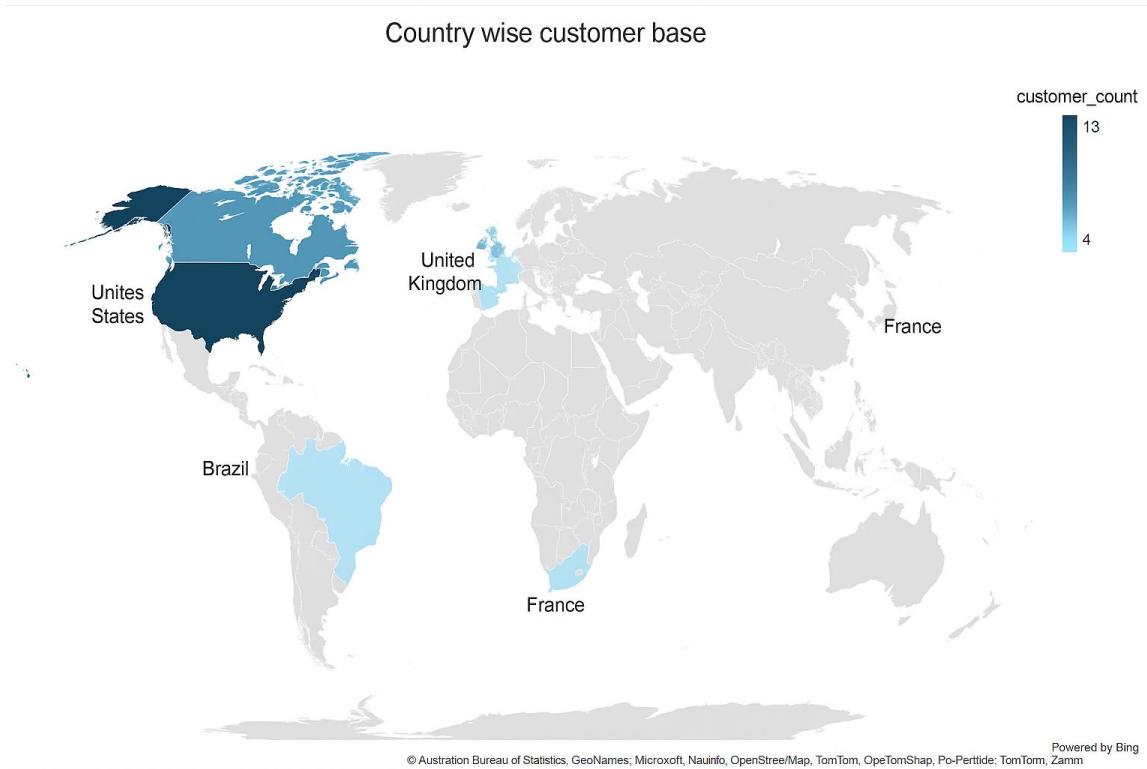
d.Granular analysis across locations:

```
453
454     -- combined demographic analysis :
455     -- Granular view (city within state within country)
456 • SELECT
457     country,
458     state,
459     city,
460     COUNT(*) AS customer_count
461 FROM customer
462 WHERE
463     city IS NOT NULL AND city != '' AND
464     state IS NOT NULL AND state != '' AND
465     country IS NOT NULL AND country != ''
466 GROUP BY country, state, city
467 ORDER BY customer_count DESC, country
468 LIMIT 5 ;
469
470
471
```

100% 40:463 | 1 error found

Result Grid Filter Rows: Search Export: Fetch rows:

country	state	city	customer_count
Brazil	SP	São Paulo	2
Czech Republic	PRG	Prague	2
France	75	Paris	2
Germany	BE	Berlin	2
United Kingdom	GB	London	2



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

Approach :total\_revenue ,total\_invoice\_count is computed based using SUM and COUNT functions as mentioned in the queries below

- a.country based total revenue and invoice count -- USA tops with 131 invoices and 1040 revenue
- b.city based total revenue and invoice count – Prague tops with 30 invoices and 273 as revenue
- c.state based total revenue and invoice count – PRG(state code ) tops i.e Prague is considered a city-state
- d.granular analysis – (Czech Republic,PRG,Prague) tops overall

```

473      -- country based total revenue and invoice count
474  ●  SELECT
475      c.country,
476      COUNT(i.invoice_id) AS total_invoices,
477      SUM(i.total) AS total_revenue
478  FROM customer c
479  JOIN invoice i ON c.customer_id = i.customer_id
480  WHERE c.country IS NOT NULL AND c.country != ''
481  GROUP BY c.country
482  ORDER BY total_revenue DESC;
483
484      -- state based total revenue and invoice count
485  ●

```

75% 16:478 | 1 error found

**Result Grid** Filter Rows: Search Export:

country	total_invoices	total_revenue
USA	131	1040.49
Canada	76	535.59
Brazil	61	427.68
France	50	389.07
Germany	41	334.62
Czech Republic	30	273.24
United Kingdom	28	245.52
Portugal	29	185.13
India	21	183.15
Ireland	13	114.84
Spain	11	98.01
Chile	13	97.02
Australia	10	81.18
Finland	11	79.20
Hungary	10	78.21
Poland	10	76.23
Sweden	10	75.24
Norway	9	72.27
Austria	9	69.30
Netherlands	10	65.34
Belgium	7	60.39
Italy	9	50.49
Argentina	5	39.60
Denmark	10	37.62

a.

```

495      -- city based total revenue and invoice count
496
497  ●  SELECT
498      c.city,
499      COUNT(i.invoice_id) AS total_invoices,
500      SUM(i.total) AS total_revenue
501  FROM customer c
502  JOIN invoice i ON c.customer_id = i.customer_id
503  WHERE c.city IS NOT NULL AND c.city != ''
504  GROUP BY c.city
505  ORDER BY total_revenue DESC
506  LIMIT 10;
507
508
509

```

75% 48:502 | 1 error found

**Result Grid** Filter Rows: Search Export: Fetch rows:

city	total_invoices	total_revenue
Prague	30	273.24
Mountain View	20	169.29
London	19	166.32
Berlin	20	158.40
Paris	18	151.47
São Paulo	22	129.69
Dublin	13	114.84
Delhi	13	111.87
São José dos Campos	13	108.90
Brasília	15	106.92

b.

```

484      -- state based total revenue and invoice count
485  •  SELECT
486      c.state,
487      COUNT(i.invoice_id) AS total_invoices,
488      SUM(i.total) AS total_revenue
489  FROM customer c
490  JOIN invoice i ON c.customer_id = i.customer_id
491  WHERE c.state IS NOT NULL AND c.state != ''
492  GROUP BY c.state
493  ORDER BY total_revenue DESC;
494

```

75% 43:487 | 1 error found

Result Grid Filter Rows: Search Export:

state	total_invoices	total_revenue
PRG	30	273.24
SP	35	238.59
CA	29	223.74
GB	19	166.32
BE	20	158.40
75	18	151.47
RM	22	147.51
ON	17	131.67
Dublin	13	114.84
DL	13	111.87
DF	15	106.92
AB	20	104.94
LIS	13	102.96
33	11	99.99
QC	9	99.99
M	11	98.01

C.

```

509      -- granular ( country , state , city ) :
510  •  SELECT
511      c.country,
512      c.state,
513      c.city,
514      COUNT(i.invoice_id) AS total_invoices,
515      SUM(i.total) AS total_revenue,
516      ROUND(SUM(i.total) / COUNT(i.invoice_id), 2) AS avg_order_value
517  FROM customer c
518  JOIN invoice i ON c.customer_id = i.customer_id
519  WHERE
520      c.country IS NOT NULL AND c.country != '' AND
521      c.state IS NOT NULL AND c.state != '' AND
522      c.city IS NOT NULL AND c.city != ''
523  GROUP BY c.country, c.state, c.city
524  ORDER BY total_revenue DESC;
525
526

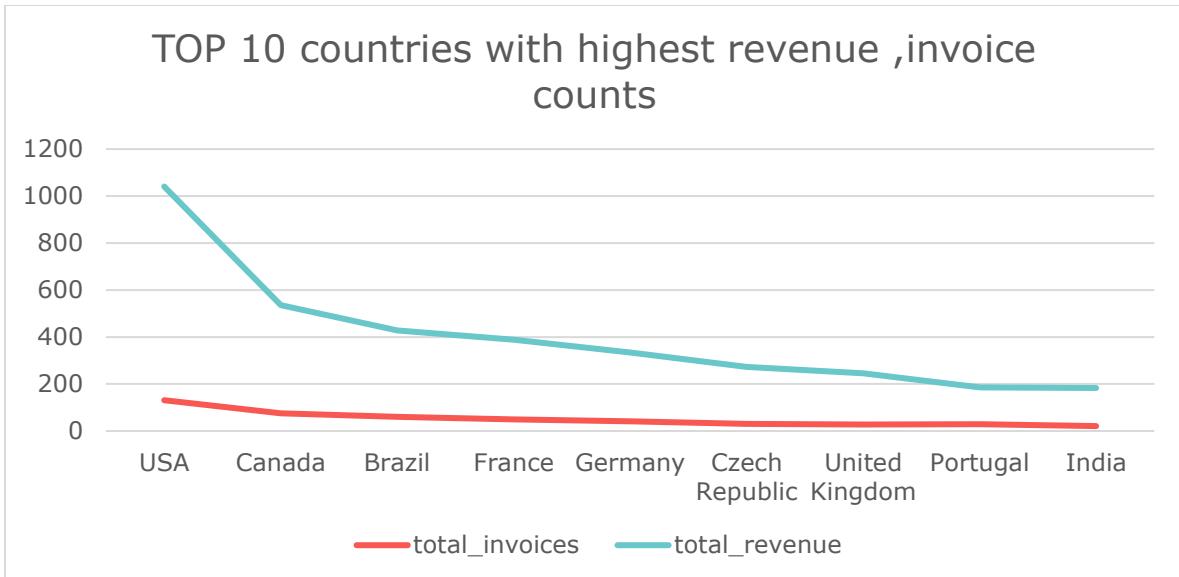
```

75% 12:513 | 1 error found

Result Grid Filter Rows: Search Export:

country	state	city	total_invoices	total_revenue	avg_order_value
Czech Republic	PRG	Prague	30	273.24	9.11
USA	CA	Mountain View	20	169.29	8.46
United Kingdom	GB	London	19	166.32	8.75
Germany	BE	Berlin	20	158.40	7.92
France	75	Paris	18	151.47	8.42
Brazil	SP	São Paulo	22	129.69	5.90
Ireland	Dublin	Dublin	13	114.84	8.83
India	DL	Delhi	13	111.87	8.61
Brazil	SP	São José dos Campos	13	108.90	8.38
Brazil	DF	Brasília	15	106.92	7.13
Portugal	LIS	Lisbon	13	102.96	7.92
France	33	Bordeaux	11	99.99	9.09
Canada	QC	Montréal	9	99.99	11.11
Spain	M	Madrid	11	98.01	8.91
USA	WA	Redmond	12	98.01	8.17

d.



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

Approach : computing the total revenue , using DENSE\_RANK() window function to find the accurate top 5 customers within each country

Sample output explanation of analysis from the below attached output and query images :

1. 'Diego Gutiérrez' – ranks 1<sup>st</sup> in Argentina and he is the only customer from Argentina
2. 'Francois Tremblay' – ranks 1<sup>st</sup> in Canada among the top 5 customers

```
-- 5.Top 5 customers by total revenue in each country

529 • ( WITH customer_revenue AS (
530     SELECT
531         c.country,
532         CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
533         SUM(i.total) AS total_revenue,
534         DENSE_RANK() OVER (
535             PARTITION BY c.country
536             ORDER BY SUM(i.total) DESC
537         ) AS revenue_rank,
538         COUNT(*) OVER (PARTITION BY c.country) AS customers_in_country
539     FROM customer c
540     JOIN invoice i ON c.customer_id = i.customer_id
541     WHERE c.country IS NOT NULL AND c.country != ''
542     GROUP BY c.country, c.customer_id, c.first_name, c.last_name
543 )
544     SELECT
545         customer_name,
546         country,
547         total_revenue,
548         revenue_rank,
549         customers_in_country
550     FROM customer_revenue
551     WHERE revenue_rank <= 5
552     ORDER BY
553         customers_in_country ASC, -- Countries with most customers i.e 5 at the end
554         country,
555         revenue_rank;
556
```

customer_name	country	total_revenue	revenue_rank	customers_in_coun...
Diego Gutiérrez	Argentina	39.60	1	1
Mark Taylor	Australia	81.18	1	1
Astrid Gruber	Austria	69.30	1	1
Daan Peeters	Belgium	60.39	1	1
Luis Rojas	Chile	97.02	1	1
Kara Nielsen	Denmark	37.62	1	1
Terhi Hämäläinen	Finland	79.20	1	1
Ladislav Kovács	Hungary	78.21	1	1
Hugh O'Reilly	Ireland	114.84	1	1
Lucas Mancini	Italy	50.49	1	1
Johannes Van der Berg	Netherla...	65.34	1	1
Bjørn Hansen	Norway	72.27	1	1
Stanisław Wójcik	Poland	76.23	1	1
Enrique Muñoz	Spain	98.01	1	1
Joakim Johansson	Sweden	75.24	1	1
František Wichterlová	Czech...	144.54	1	2
Helena Holý	Czech...	128.70	2	2
Manoj Pareek	India	111.87	1	2
Puja Srivastava	India	71.28	2	2
João Fernandes	Portugal	102.96	1	2
Madalena Sampaio	Portugal	82.17	2	2
Phil Hughes	United...	98.01	1	3
Steve Murray	United...	79.20	2	3
Emma Jones	United...	68.31	3	3
Fynn Zimmermann	Germany	94.05	1	4
Hannah Schneider	Germany	85.14	2	4
Leonie Köhler	Germany	82.17	3	4
Niklas Schröder	Germany	73.26	4	4
Luís Gonçalves	Brazil	108.90	1	5
Fernanda Ramos	Brazil	106.92	2	5
Roberto Almeida	Brazil	82.17	3	5
Alexandre Rocha	Brazil	69.30	4	5
Eduardo Martins	Brazil	60.39	5	5
Wyatt Girard	France	99.99	1	5
Camille Bernard	France	79.20	2	5
Isabelle Mercier	France	73.26	3	5
Dominique Lefebvre	France	72.27	4	5
Marc Dubois	France	64.35	5	5
François Tremblay	Canada	99.99	1	8
Edward Francis	Canada	91.08	2	8
Ellie Sullivan	Canada	75.24	3	8
Aaron Mitchell	Canada	70.29	4	8
Jennifer Peterson	Canada	66.33	5	8
Jack Smith	USA	98.01	1	13
Don Miller	USA	65.04	2	12

6. Identify the top-selling track for each customer

Approach: computing number of times a particular track was purchased and total amount spent for that track for each customer , and using ROW\_NUMBER() to pick the tracks ranked 1<sup>st</sup> (row number 1 ) ,among all the tracks purchased by each customer based on number of units bought and money spent

Sample output explanation of the analysis from the images attached below:

- 1.Francois Tremblay – Sting Me ( top track )
- 2.Luis Goncalves – My generation ( top track )

```
-- 6.The top-selling track for each customer:
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,
        COUNT(*) AS times_purchased,
        SUM(il.unit_price * il.quantity) AS total_spent,
        ROW_NUMBER() OVER (PARTITION BY c.customer_id ORDER BY SUM(il.unit_price * il.quantity) DESC, COUNT(*) DESC
        ) AS track_rank
    FROM customer c
    JOIN invoice inv ON c.customer_id = inv.customer_id
    JOIN invoice_line il ON inv.invoice_id = il.invoice_id
    JOIN track t ON il.track_id = t.track_id
    GROUP BY c.customer_id, c.first_name, c.last_name, t.track_id, t.name
)
SELECT
    customer_id,
    customer_name,
    track_name,
    times_purchased,
    total_spent
FROM customer_track_purchases
WHERE track_rank = 1
ORDER BY customer_id;
```

customer_id	customer_name	track_name	times_purchased	total_spe...
1	Luís Gonçalves	My Generation	1	0.99
2	Leonie Köhler	End Of The Night	1	0.99
3	François Tremblay	Sting Me	2	1.98
4	Bjørn Hansen	Welcome to the Jungle	1	0.99
5	František Wichterlová	Rapidamente	1	0.99
6	Helena Holý	Foxy Lady	1	0.99
7	Astrid Gruber	Who Are You (Single Edit Version)	1	0.99
8	Daan Peeters	Wake Up	1	0.99
9	Kara Nielsen	Mr. Moon	1	0.99
10	Eduardo Martins	Like A Bird	2	1.98
11	Alexandre Rocha	That's Life	1	0.99
12	Roberto Almeida	Love And Marriage	2	1.98
13	Fernanda Ramos	24 Caprices, Op. 1, No. 24, for S...	2	1.98
14	Mark Philips	Us And Them	1	0.99
15	Jennifer Peterson	Dead And Broken	1	0.99
16	Frank Harris	Pop Song 89	1	0.99
17	Jack Smith	Scentless Apprentice	1	0.99
18	Michelle Brooks	Right Next Door to Hell	1	0.99
19	Tim Goyer	Message in a Bottle (new classic...)	1	0.99
20	Dan Miller	I Like Dirt	1	0.99
21	Kathy Chase	The Other Side	1	0.99
22	Heather Leacock	You Know I'm No Good (feat. Gh...)	1	0.99
23	John Gordon	The Small Hours	1	0.99
24	Frank Ralston	War Pigs	1	0.99
25	Victor Stevens	The Wrong Child	1	0.99
26	Richard Cunningham	Wherever I Lay My Hat	1	0.99
27	Patrick Gray	Bitter Suite: Brief Encounter / Lo...	1	0.99
28	Julia Barnett	Get What You Need	2	1.98
29	Robert Brown	Let There Be Rock	1	0.99
30	Edward Francis	Waiting On A Friend	1	0.99
31	Martha Silk	Boris The Spider	1	0.99
32	Aaron Mitchell	Please Please Please	1	0.99
33	Ellie Sullivan	Most High	1	0.99
34	João Fernandes	Train In Vain	2	1.98
35	Madalena Sampaio	Superhuman	1	0.99
36	Hannah Schneider	Behind The Wall Of Sleep	1	0.99
37	Fynn Zimmermann	Radio/Video	2	1.98
38	Niklas Schröder	English Civil War	1	0.99

Result 221

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

NOTE: payment\_method data is not available

Approach: 4 different analytical approaches to address this question

- a.purchase frequency of customers – using min, max , count ,DATEDIFF functions
- b.average order value – using AVG() function
- c.customers year wise purchase frequency with total order value and avg order value – to understand the trend of purchasing behaviour of customers over years and cater accordingly
- d.countrywise order value distribution – helps to choose countries for further

a. Frantisek Wichterlova has made maximum purchases in lesser span of days (indicates he is the most frequent buyer )

```

592    -- purchase frequency per customer :
593 •  SELECT
594     c.customer_id,
595     CONCAT(c.first_name, " ",c.last_name) as customer_name,
596     COUNT(i.invoice_id) AS total_purchases,
597     MIN(i.invoice_date) AS first_purchase,
598     MAX(i.invoice_date) AS last_purchase,
599     DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS span_days
600   FROM customer c
601   JOIN invoice i ON c.customer_id = i.customer_id
602   GROUP BY c.customer_id,customer_name
603   ORDER BY total_purchases desc,span_days asc;
604
605

```

75%

46:603 | 1 error found

Result Grid



Filter Rows:

Search

Export:

customer_id	customer_name	total_purchas...	first_purchase	last_purchase	span_days
5	František Wichterlová	18	2017-05-29 00:00:00	2020-11-12 00:00:00	1263
35	Madalena Sampalo	16	2017-01-22 00:00:00	2020-12-17 00:00:00	1425
13	Fernanda Ramos	15	2017-05-18 00:00:00	2020-11-28 00:00:00	1290
57	Luis Rojas	13	2017-03-25 00:00:00	2020-06-09 00:00:00	1172
58	Manoj Pareek	13	2017-02-21 00:00:00	2020-07-15 00:00:00	1240
1	Luís Gonçalves	13	2017-01-26 00:00:00	2020-07-24 00:00:00	1275
34	João Fernandes	13	2017-02-25 00:00:00	2020-10-13 00:00:00	1326
46	Hugh O'Reilly	13	2017-02-18 00:00:00	2020-12-12 00:00:00	1393
30	Edward Francis	13	2017-01-03 00:00:00	2020-10-31 00:00:00	1397
17	Jack Smith	12	2017-07-23 00:00:00	2020-09-11 00:00:00	1146
6	Helena Holy	12	2017-08-31 00:00:00	2020-10-23 00:00:00	1149
43	Isabelle Mercier	12	2017-02-06 00:00:00	2020-05-02 00:00:00	1181
10	Eduardo Martins	12	2017-02-06 00:00:00	2020-06-25 00:00:00	1235
22	Heather Leacock	12	2017-03-19 00:00:00	2020-10-30 00:00:00	1321
26	Richard Cunningham	12	2017-04-13 00:00:00	2020-12-10 00:00:00	1337
33	Ellie Sullivan	12	2017-02-23 00:00:00	2020-12-27 00:00:00	1403
20	Dan Miller	12	2017-01-28 00:00:00	2020-12-29 00:00:00	1431
36	Hannah Schneider	11	2017-03-24 00:00:00	2020-04-07 00:00:00	1110
50	Enrique Muñoz	11	2017-03-17 00:00:00	2020-06-10 00:00:00	1181
19	Ilaria Kähler	11	2017-05-09 00:00:00	2020-11-21 00:00:00	1282

Result 228

b.In terms of average order value alongside lifetime value Francois Tremblay tops the list with avg order value of 11.11

```

604
605      -- average order value, lifetime_value per customer
606 •   SELECT
607          c.customer_id,
608          CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
609          ROUND(AVG(i.total), 2) AS avg_order_value,
610          ROUND(SUM(i.total), 2) AS lifetime_value,
611          COUNT(*) AS total_orders
612      FROM customer c
613      JOIN invoice i ON c.customer_id = i.customer_id
614      GROUP BY c.customer_id, customer_name
615      ORDER BY avg_order_value DESC;
616
617

```

75% 16:612 | 1 error found

**Result Grid** Filter Rows: Search Export:

customer_id	customer_name	avg_order_value	lifetime_value	total_orders
3	François Tremblay	11.11	99.99	9
6	Helena Holý	10.73	128.70	12
29	Robert Brown	10.15	40.59	4
18	Michelle Brooks	9.90	79.20	8
37	Fynn Zimmermann	9.41	94.05	10
27	Patrick Gray	9.35	84.15	9
16	Frank Harris	9.28	74.25	8
42	Wyatt Girard	9.09	99.99	11
53	Phil Hughes	8.91	98.01	11
59	Puja Srivastava	8.91	71.28	8
50	Enrique Muñoz	8.91	98.01	11
24	Frank Ralston	8.91	71.28	8
46	Hugh O'Reilly	8.83	114.84	13
54	Steve Murray	8.80	79.20	9
39	Camille Bernard	8.80	79.20	9
32	Aaron Mitchell	8.79	70.29	8
8	Daan Peeters	8.63	60.39	7
58	Manoj Pareek	8.61	111.87	13
52	Emma Jones	8.54	68.31	8
1	Luis Goncalves	8.38	108.00	12

Result 229

c. Customers year wise purchase frequency with total and average order value

Considering a customer for analysis explanation :

Luis Goncalves we can clearly see that his purchase count has significantly reduced in the later years (2019,2020) compared to early years (2017,2018) – implies he might end up as a churned customer in the coming years ,hence providing us insights on whom to reach out for feedback and consider improvements

```

617
618      -- customers year wise purchase frequency with total order value and avg order value
619  •  SELECT c.customer_id, concat(c.first_name, ' ', c.last_name) as customer_name,
620        year(i.invoice_date) as year, count(i.invoice_id) as purchase_count,
621        sum(i.total) as tot_revenue, avg(i.total) as avg_ord_value
622    from customer c
623    join invoice i on c.customer_id = i.customer_id
624  group by c.customer_id, customer_name, year(i.invoice_date)
625  order by c.customer_id, customer_name, year(i.invoice_date);
626
627

```

75% 16:622 1 error found

Result Grid Filter Rows:  Search Export:

customer_id	customer_name	year	purchase_count	tot_revenue	avg_ord_value
1	Luís Gonçalves	2017	4	37.62	9.405000
1	Luís Gonçalves	2018	5	48.51	9.702000
1	Luís Gonçalves	2019	2	7.92	3.960000
1	Luís Gonçalves	2020	2	14.85	7.425000
2	Leonia Köhler	2017	3	24.75	8.250000
2	Leonia Köhler	2018	3	21.78	7.260000
2	Leonia Köhler	2019	1	5.94	5.940000
2	Leonia Köhler	2020	4	29.70	7.425000
3	François Tremblay	2017	3	46.53	15.510000
3	François Tremblay	2018	2	14.85	7.425000
3	François Tremblay	2019	3	29.70	9.900000
3	François Tremblay	2020	1	8.91	8.910000
4	Bjørn Hansen	2017	4	30.69	7.672500
4	Bjørn Hansen	2018	4	34.65	8.662500
4	Bjørn Hansen	2020	1	6.93	6.930000
5	František Wichterlová	2017	2	10.89	5.445000
5	František Wichterlová	2018	2	17.82	8.910000
5	František Wichterlová	2019	6	39.60	6.600000
5	František Wichterlová	2020	8	76.23	9.528750
6	Helena Holý	2017	2	27.72	13.860000
6	Helena Holý	2018	4	38.61	9.652500
6	Helena Holý	2019	3	29.70	9.900000
6	Helena Holý	2020	3	32.67	10.890000

Result 230

d.country wise order value distribution : considering the combination of avg order value , no of orders , total revenue generated  
--USA , CANADA , Brazil stand as top 3 countries , indicating us to focus more on these countries in terms of sales , investment ,introducing new tracks , etc

```

627      -- additional analysis : country wise order value distribution (avg, total , no's)
628
629 •   SELECT
630     billing_country,
631     ROUND(AVG(total), 2) AS avg_order_value,
632     ROUND(SUM(total), 2) AS total_revenue,
633     COUNT(*) AS num_orders
634   FROM invoice
635   GROUP BY billing_country
636   ORDER BY total_revenue DESC;
637
638

```

75% 1:628 | 1 error found

**Result Grid** Filter Rows: Search Export:

billing_country	avg_order_value	total_revenue	num_orders
USA	7.94	1040.49	131
Canada	7.05	535.59	76
Brazil	7.01	427.68	61
France	7.78	389.07	50
Germany	8.16	334.62	41
Czech Republic	9.11	273.24	30
United Kingdom	8.77	245.52	28
Portugal	6.38	185.13	29
India	8.72	183.15	21
Ireland	8.83	114.84	13
Spain	8.91	98.01	11
Chile	7.46	97.02	13
Australia	8.12	81.18	10
Finland	7.20	79.20	11
Hungary	7.82	78.21	10
Poland	7.62	76.23	10
Sweden	7.52	75.24	10
Norway	8.03	72.27	9
Austria	7.70	60.20	10

## 8.Customer churn rate :

Answer : 1.6949 , that is (1/59)

Approach:

The first step to approach to question is to understand what is churn rate

Churn rate is the percentage of customers who stop using a product or service over a specific period. It's commonly used in subscription-based businesses to measure customer retention.

Formula:

$$\text{Churn Rate} = \frac{\text{Customers Lost in a Period}}{\text{Total Customers at the Start of the Period}} \times 100$$

```

638
639      -- 8. Customer churn rate:
640
641  * with cte as (
642      select max(invoice_date) as recent_inovice_date from invoice
643      ),last_year as (
644      select date_sub(recent_inovice_date,INTERVAL 1 YEAR) as last_year_dt from cte
645      ),churn_customers as
646  (
647      select c.customer_id
648      from customer c join invoice i on c.customer_id=i.customer_id
649      group by customer_id
650      having max(invoice_date)<(select last_year_dt from last_year)
651  )
652
653      select (select count(*) from churn_customers)/(select count(*) from customer) * 100 as churn_rate;
654
655
656

```

75% 1:637 | 1 error found

Result Grid Filter Rows: Search Export:

churn_rate
1.6949

Logical assumption: A customer is considered churned if they have not made any purchases in the last 1 year from the most recent invoice date.

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

Approach:

Computing the total sales per genre using SUM window function, ordering them in descending order based on sales

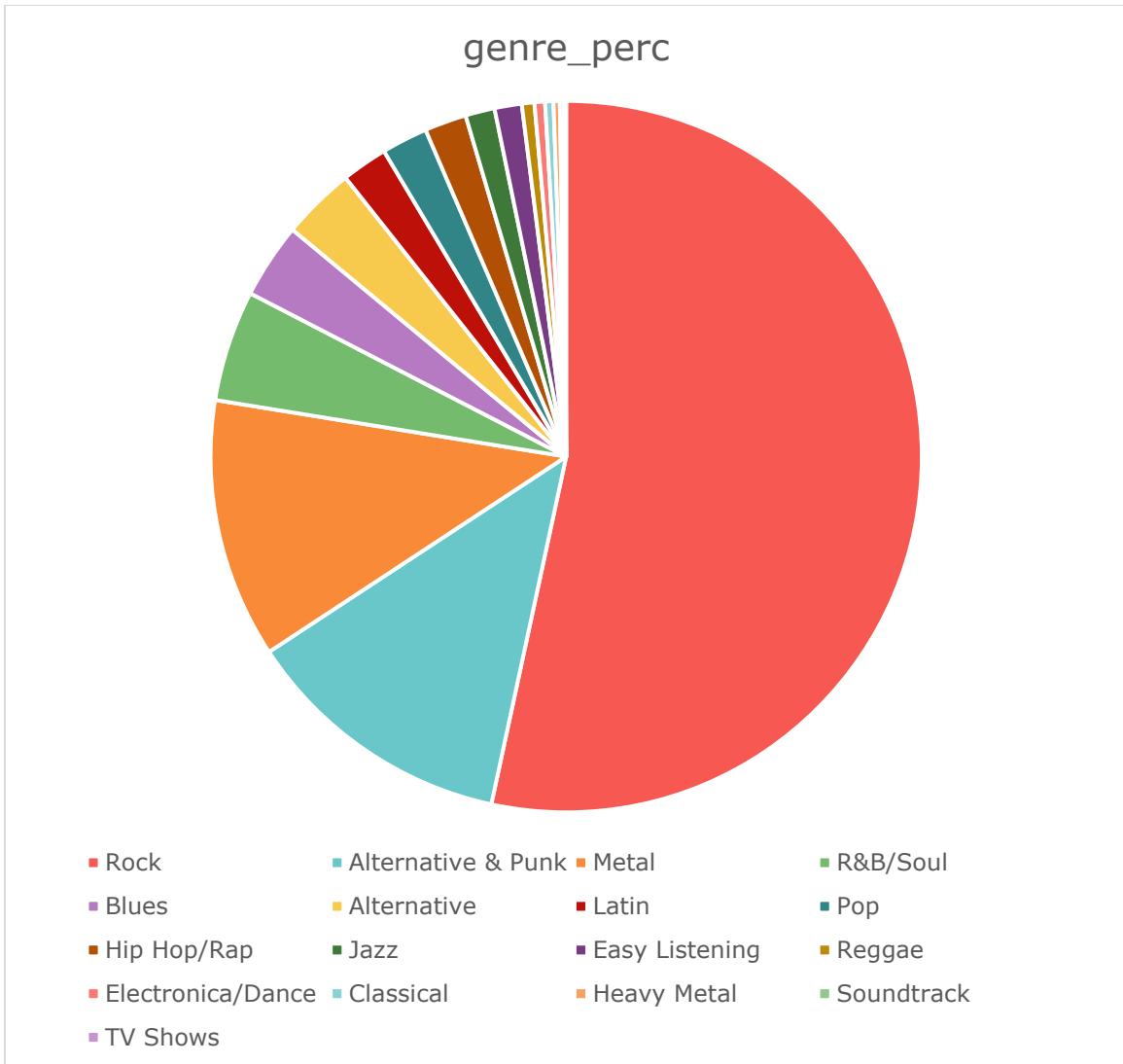
```
656 -- 9 . The percentage of total sales contributed by each genre in the USA and the best-selling genres and artists
657
658
659 *   select distinct g.name,
660       sum(il.quantity) over(partition by g.name) as genre_tot_sales,
661       ( sum(il.quantity) over(partition by g.name) *100/sum(il.quantity) over() ) as genre_perc
662   from invoice i
663   join invoice_line il on i.invoice_id=il.invoice_id
664   join track t on t.track_id=il.track_id
665   join genre g on g.genre_id=t.genre_id
666   where i.billing_country='USA'
667   order by genre_tot_sales desc;
668
```

75% 1:655 1 error found

Result Grid Filter Rows: Search Export:

name	genre_tot_sales	genre_perc
Rock	561	53.3777
Alternative & Punk	130	12.3692
Metal	124	11.7983
R&B/Soul	53	5.0428
Blues	36	3.4253
Alternative	35	3.3302
Latin	22	2.0932
Pop	22	2.0932
Hip Hop/Rap	20	1.9029
Jazz	14	1.3321
Easy Listening	13	1.2369
Reggae	6	0.5709
Electronica/Dance	5	0.4757
Classical	4	0.3806
Heavy Metal	3	0.2854
Soundtrack	2	0.1903
TV Shows	1	0.0951

Rock genre contributes 53% of total\_sales making it the most selling genre in USA



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

Approach: computing count of tracks purchased and filtering data ,using having clause for the aggregate count()  $\geq 3$

```

670      -- 10. Customers who have purchased tracks from at least 3 different genres
671
672 •      select concat(c.first_name, ' ', c.last_name) as customer_name,
673             count(distinct g.genre_id) as genres_purchased
674     from customer c
675     join invoice i on c.customer_id = i.customer_id
676     join invoice_line il on i.invoice_id = il.invoice_id
677     join track t on il.track_id = t.track_id
678     join genre g on t.genre_id = g.genre_id
679     group by customer_name
680     having count(distinct g.genre_id)>=3
681     order by genres_purchased desc;
682
683
684

```

75% 23:679 | 1 error found

Result Grid Filter Rows: Search Export:

customer_name	genres_purchas...
Leonie Köhler	14
František Wichterlová	13
Terhi Hämäläinen	13
Madalena Sampaio	13
Edward Francis	13
Heather Leacock	13
Niklas Schröder	12
Michelle Brooks	12
Marc Dubois	12
Julia Barnett	12
John Gordon	12
Wyatt Girard	12
Jack Smith	12
Hugh O'Reilly	12
Fernanda Ramos	12
Johannes Van der Berg	11
Steve Murray	11
Enrique Muñoz	11
Helene Holtz	11

Result 234

- Leonie Kohler tops the list with 14 tracks among the customers who purchased at least 3 tracks

11. Rank genres based on their sales performance in the USA  
 Approach : Using DENSE\_RANK() on total\_revenue ordered in descending order

```

683      --
684      -- 11. Rank genres based on their sales performance in the USA
685
686 *   select g.name,
687         sum(il.quantity*il.unit_price) as total_revenue,
688         dense_rank() over ( order by sum(il.quantity) desc) as rnk
689     from invoice i
690     join invoice_line il on i.invoice_id = il.invoice_id
691     join track t on il.track_id = t.track_id
692     join genre g on t.genre_id = g.genre_id
693     where i.billing_country='USA'
694     group by g.name
695     order by rnk;
696
697

```

75%    16:694 | 1 error found

**Result Grid**



Filter Rows:

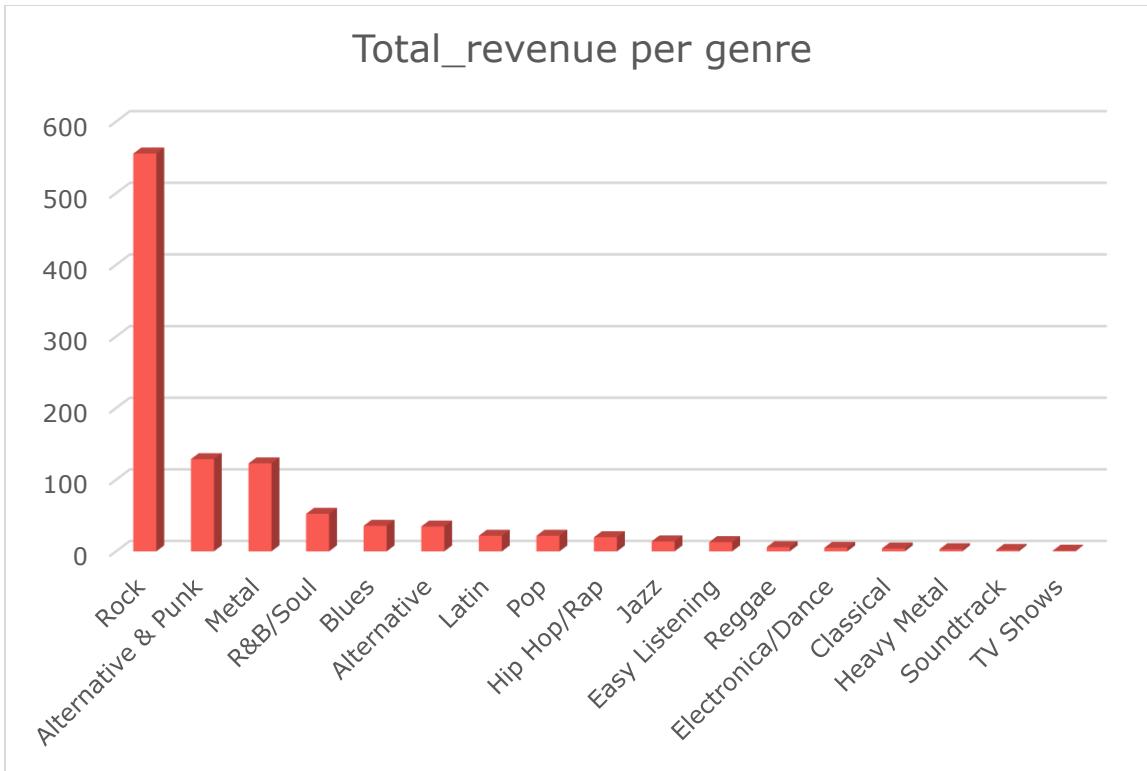
Search

Export:



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

Rock genre contributes the most in terms of total revenue in USA



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

Approach : calculating the latest purchase date of the entire dataset, DATE\_ADD() to subtract interval of 3 months from last date of purchase to set of the cut off date , performing required joins to fetch the right data

NOTE :

CURRENT\_DATE() is not used for any calculations as it is not a live dataset

"The most recent date in the dataset serves as our cutoff point, representing the final day when purchase data was recorded (effectively acting as the 'current date' for our analysis)."

```
698  + 12.Customers who have not made a purchase in the last 3 months
699
700  ● with latest_billday as(
701      select date_add( max(date(invoice_date)),interval -3 month) as recent_day
702      from invoice)
703
704      select distinct c.customer_id,concat(c.first_name,' ',c.last_name) as customer_name
705      from customer c
706      join invoice i on c.customer_id=i.customer_id
707      group by customer_id
708      having max(i.invoice_date)<(select * from latest_billday)
709      order by c.customer_id;
710
```

75% | 2:698 | 1 error found

Result Grid    Filter Rows:    Search    Export:

customer_id	customer_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

## 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.

Approach : computing total sales for each album ranking them using DENSE\_RANK() and picking top 3 using limit .

```
733
734 with cte as(
735     select g.name as genre_name, al.title as album_title,
736             sum(il.unit_price * il.quantity) as album_sales,
737             dense_rank() over (partition by g.name order by sum(il.unit_price*il.quantity) desc) rnk
738     from genre g
739     join track t on g.genre_id = t.genre_id
740     join invoice_line il on t.track_id = il.track_id
741     join invoice i on il.invoice_id = i.invoice_id
742     join album al on t.album_id = al.album_id
743     where i.billing_country = 'USA'
744     group by g.name, al.title
745     having sum(il.quantity) >4 ) -- considering only albums with more than 4 units sold
746
747     select *
748     from cte
749     where rnk<=3 --- selecting top 3 albums per genre
750     order by album_sales desc;
751
```

75% 1:751 | 1 error found

Result Grid Filter Rows: Search Export:

genre_name	album_title	album_sales	rnk
Rock	Are You Experienced?	27.72	1
Rock	From The Muddy Banks Of The Wishkah [live]	27.72	1
Rock	The Doors	26.73	2
R&B/Soul	Seek And Shall Find: More Of The Best (1963-1...	25.74	1
Alternative & Punk	Green	24.75	1
Blues	The Cream Of Clapton	24.75	1
Rock	Hot Rocks, 1964-1971 (Disc 1)	24.75	3
Metal	Mezmerize	21.78	1
Metal	Faceless	19.80	2
Hip Hop/Rap	House of Pain	19.80	1
Alternative & Punk	American Idiot	19.80	2
Alternative & Punk	Minha História	18.81	3
Alternative	Carry On	18.81	1
Metal	Black Sabbath	17.82	3
R&B/Soul	Back to Black	15.84	2
Pop	Frank	13.86	1
Easy Listening	My Way: The Best Of Frank Sinatra [Disc 1]	12.87	1

## Analysis:

Considering only albums with more than 4 units sold ;

1.TOP 3 genres in USA as per the objective analysis :

Rock ,Alternative & Punk , Metal

name	genre_tot_sal...	genre_perc
Rock	561	53.3777
Alternative & Punk	130	12.3692
Metal	124	11.7983
R&B/Soul	53	5.0428

2.TOP 3 albums across these genres would be :

- a.Are you Experienced , From the muddy banks of the Wishkah, The Doors (ROCK)
- b.Green (Alternative and Punk)
- c.Metal (Mesmerize,Faceless )

## Multi-Channel Campaigns

- Advertise top genres (Rock, Alternative/Punk, Metal) via TV, billboards, radio, and social media for maximum reach.

## Leverage Popular Albums for Engagement

- Use bestsellers (The Doors, Green, Mesmerize) in merchandise, concerts, and digital bundles to drive sales.

## Localized Strategy for Niche Albums

- Promote medium/low-selling albums in region-specific markets where they perform well, balancing inventory and marketing spend.

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

Approach : computing total sales for each genre for every country , ROW\_NUMBER() to pick top 3 genres within each country ,filtering non USA countries ,finally comparing the output with the USA based genre analysis done earlier

a.Countries except USA :

```
219
220 •Θ with genresales as (
221     select i.billing_country,g.name as genre_name,sum(il.quantity * t.unit_price) as total_sales,
222         row_number() over(partition by i.billing_country order by sum(il.quantity * t.unit_price) desc)
223             as rnk from invoice_line il
224     join track t
225     on il.track_id = t.track_id
226     join genre g
227     on t.genre_id = g.genre_id
228     join invoice i
229     on il.invoice_id = i.invoice_id
230     where i.billing_country != 'usa'
231     group by i.billing_country, g.name
232 )
233     select billing_country,genre_name,total_sales from genresales
234     where rnk <= 3;
235
```

Result Grid   Filter Rows:  Search

billing_country	genre_name	total_sales
Argentina	Alternative & Punk	16.83
Argentina	Rock	10.89
Argentina	Latin	1.98
Australia	Rock	33.66
Australia	Alternative & Punk	21.78
Australia	Metal	13.86
Austria	Rock	39.60
Austria	Jazz	14.85
Austria	Metal	6.93
Belgium	Rock	25.74
Belgium	Metal	18.81
Belgium	Alternative & Punk	6.93
Brazil	Rock	202.95
Brazil	Alternative & Punk	73.26
Brazil	Metal	72.27
Canada	Rock	329.67
Canada	Metal	71.28
Canada	Alternative & Punk	30.69
Chile	Rock	60.39
Chile	Metal	10.89
Chile	Alternative & Punk	7.92
Czech Repu...	Rock	141.57
Czech Repu...	Alternative & Punk	41.58
Czech Repu...	Metal	22.77
Denmark	Rock	23.76
Denmark	Metal	4.95
Denmark	Jazz	3.96
Finland	Rock	45.54
Finland	Metal	9.90
Finland	Alternative & Punk	5.94
France	Rock	208.89
France	Metal	53.46
France	Easy Listening	28.71
Germany	Rock	192.06
Germany	Metal	43.56
Germany	Jazz	17.82
Hungary	Rock	43.56
Hungary	Metal	18.81
Hungary	Alternative & Punk	7.92
India	Rock	100.98
India	Alternative & Punk	24.75
India	Metal	21.78
Ireland	Rock	71.28
Ireland	Latin	19.80
Ireland	Alternative & Punk	6.93
Ireland	Blues	1.98
Ireland	R&B/Soul	32.67
Ireland	Metal	4.95
Italy	Rock	34.65
Italy	Alternative & Punk	4.95
Italy	Metal	3.96
Netherlands	Rock	11.88
Netherlands	Metal	4.95
Norway	Rock	39.60
Norway	Metal	16.83
Norway	Alternative & Punk	9.90
Norway	Blues	12.87
Poland	Rock	39.60
Poland	Metal	8.91
Portugal	Rock	106.92
Portugal	Metal	27.72
Portugal	Alternative & Punk	12.87
Spain	Rock	45.54
Spain	Jazz	15.84
Spain	Metal	12.87
Sweden	Rock	59.40
Sweden	Metal	6.93
Sweden	Latin	1.98
United Kingdom	Rock	164.34
United Kingdom	Metal	30.69
United Kingdom	Alternative & Punk	23.76

b.In USA :

name	genre_tot_sal...	genre_perc
Rock	561	53.3777
Alternative & Punk	130	12.3692
Metal	124	11.7983
R&B/Soul	53	5.0428

**ANALYSIS:**

Comparing the outputs :

a.3 Key Commonalities

1. Rock Dominance

- #1 genre in all listed countries (50-80% of sales in most markets).
- Example: USA (53%), Brazil (203 units), UK (164 units).

2. Alternative & Punk + Metal Consistency

- Both genres appear in top 3 for 90% of countries.
- Exception: Austria (Jazz over Metal), Netherlands (R&B over Metal).

3. Local Genres Fill the Gaps

- Most markets have one niche genre in top 3 (e.g., Latin in Argentina/Ireland, Jazz in Austria/Germany).
- 

b.3 Key Differences

1. Market Size Variation

- Brazil/Canada/France have 2-3x higher Rock sales than smaller markets (e.g., Denmark, Norway).
- Action: Scale inventory in high-volume regions.

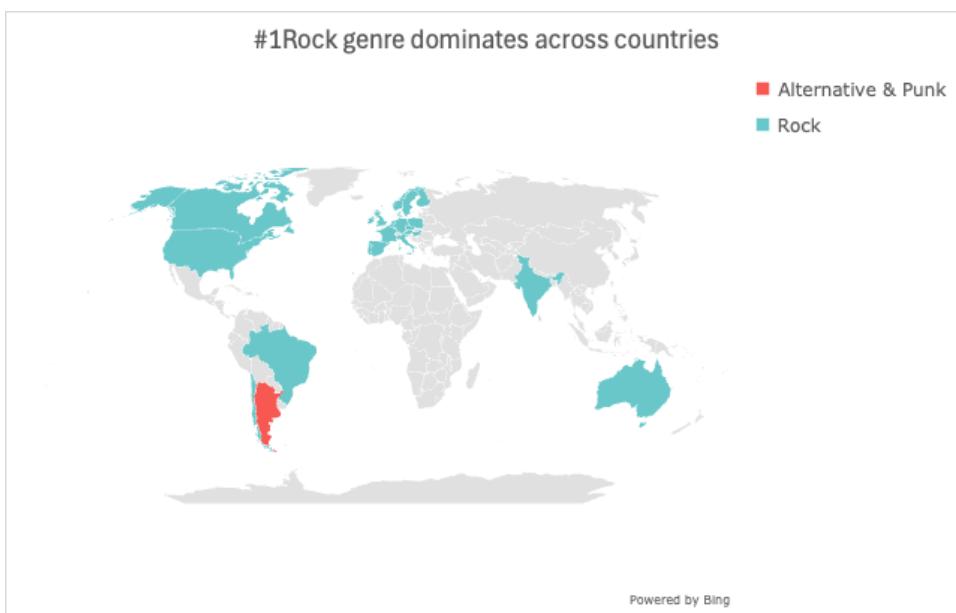
2. Genre Ranking Shifts

- Metal is #2 in Germany/Portugal but absent in Netherlands' top 3 (replaced by R&B).

- Latin thrives in Argentina/Ireland but negligible elsewhere.
3. Cultural Niche Preferences
- Jazz in Austria/Germany vs. Blues in Poland.
  - Easy Listening in France (unique in dataset).
- 

### Strategic Takeaways

1. Global: Push Rock + Alt-Punk/Metal as core inventory.
2. Localized: Add 1-2 niche genres per market (e.g., Latin for Argentina, Jazz for Austria).
3. Risk: Avoid overstocking genres like Easy Listening (France-only demand).



3. 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?

Approach :

Long term customers : Customer Tenure , with at least 1050 days are considered under this bracket

DATEDIFF() used to compute this

avg\_purchase\_frequency, avg\_basket\_size, avg\_spending\_amount, avg\_order\_value are the key factors considered for long term and short term customer analysis

```

778
779      -- Analysis based on the total number of days the customer has stayed active with purchases , as long term customers
780  with customer_details as (
781      select c.customer_id, count(distinct i.invoice_id) as tot_purchases, sum(il.quantity) as basket_size,
782          sum(i.total) as total_spent, round(avg(i.total),2) as avg_order_value,
783          (
784              case
785                  when datediff(max(i.invoice_date), min(i.invoice_date)) >=1050 then 'long-term customers'
786                  else 'new customers'
787              end
788          )as customer_category
789      from customer c
790      join invoice i on c.customer_id = i.customer_id
791      join invoice_line il on i.invoice_id = il.invoice_id
792      group by c.customer_id
793  )
794
795      select
796          customer_category,
797          round(avg(tot_purchases),2) as avg_purchase_frequency,
798          round(avg(basket_size),2) as avg_basket_size,
799          round(avg(total_spent),2) as avg_spending_amount,
800          round(avg(avg_order_value),2) as avg_order_value
801      from customer_details
802      group by customer_category
803
804
    75%  1:779  1 error found

```

**Result Grid** Filter Rows:  Search Export:

customer_category	avg_purchase_frequency	avg_basket_size	avg_spending_amount	avg_order_value
long-term customers	10.64	82.55	831.92	9.86
new customers	8.33	63.67	568.59	8.99

## Analysis:

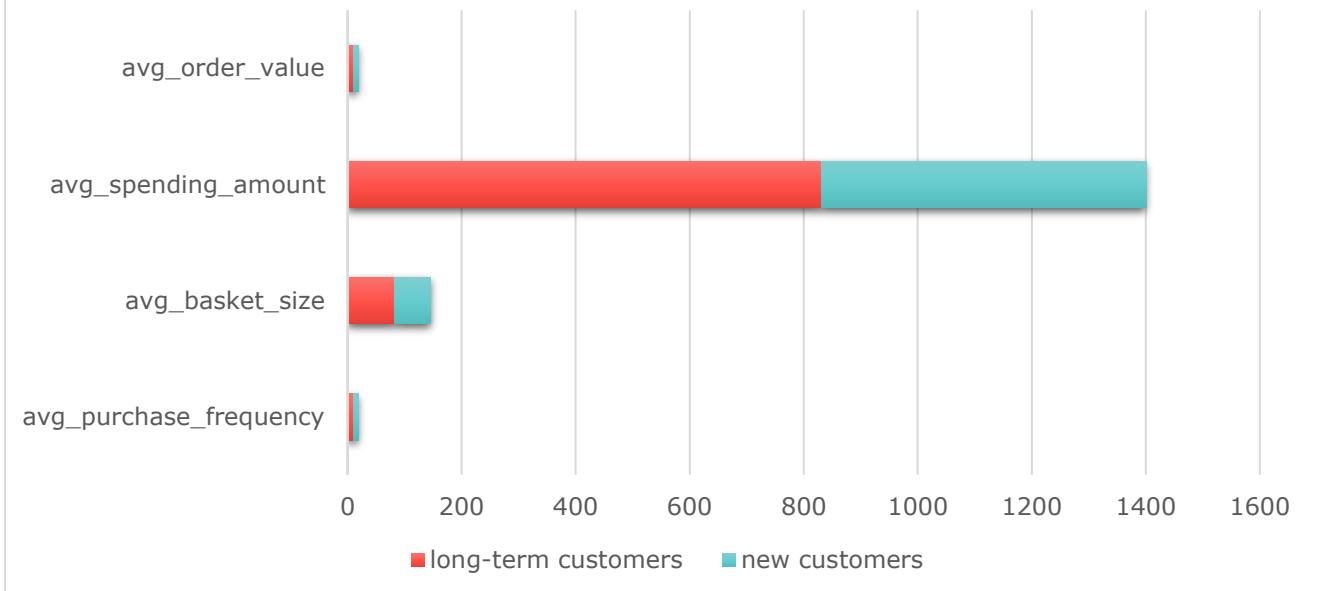
### Insights on Customer Loyalty and Retention

1. Long-term customers spend more:  
Long-term customers have a significantly higher total spend and basket size. This indicates strong customer lifetime value and shows that these customers are more engaged and profitable.
  2. Higher purchase frequency:  
Long-term customers purchase more often, which suggests they are consistent buyers. This indicates a level of trust or preference for the brand or platform.
  3. Larger basket sizes:  
They tend to buy more items per transaction. This may imply they are more confident in their purchasing decisions or are shopping for broader needs.
  4. Slightly higher average order value:  
While the difference in order value is not very large, the total contribution from long-term customers is significantly higher due to increased purchase frequency and basket size.
- 

### Strategic Implications

1. Retention focus(RETENTION > ACQUISITION)  
Long-term customers are more valuable. Businesses should target them with loyalty programs, exclusive offers, or early access incentives to keep them engaged.
2. Convert new customers to long-term  
Since new customers already show decent average order value, businesses can focus on personalized onboarding journeys, promotional offers, and product bundling to encourage repeat purchases and increased basket size.

## Long term vs Short term customer analysis



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?

Approach :

Affinity analysis :

- 1.genre – using joins to form unique combinations of genres from required tables
- 2.artist - using joins to form unique combinations of artists from required tables
- 3.album - using joins to form unique combinations of albums from required tables

## 1.genre affinity analysis :

```

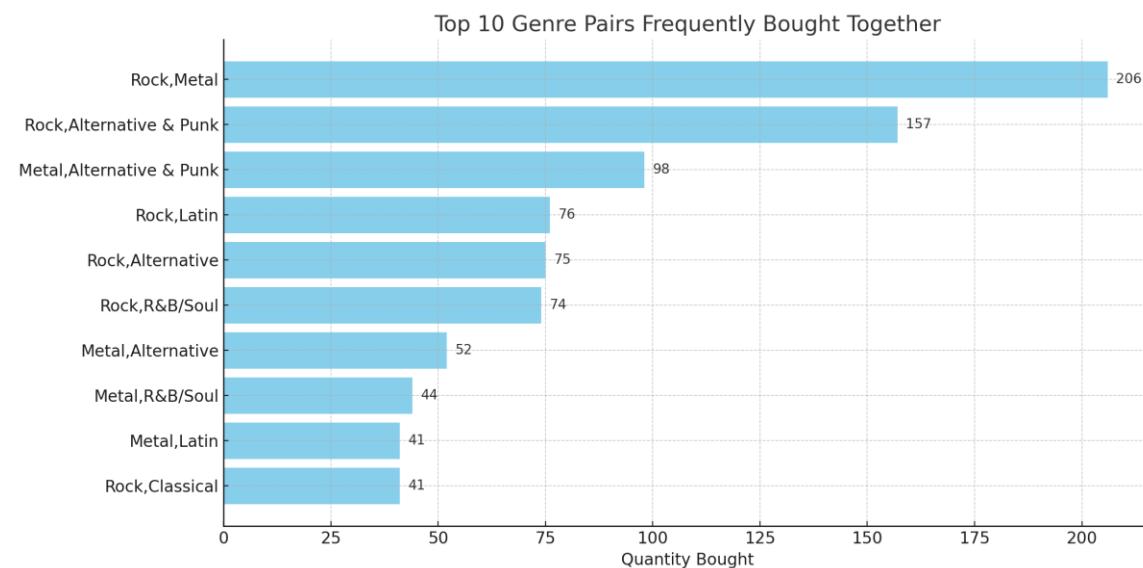
898 -- Genre Affinity Analysis:
899
900 •   select
901     distinct g1.name as genre01,
902     g2.name as genre02,
903     count(distinct il1.invoice_id) as quantity_bought
904     from invoice_line il1
905     join invoice_line il2 on il1.invoice_id = il2.invoice_id and il1.track_id < il2.track_id
906     join track t1 on il1.track_id = t1.track_id
907     join track t2 on il2.track_id = t2.track_id
908     join genre g1 on t1.genre_id = g1.genre_id
909     join genre g2 on t2.genre_id = g2.genre_id where g1.genre_id < g2.genre_id
910     group by genre01,genre02
911     order by quantity_bought desc ;
912

```

75% 44:906 | 1 error found

Result Grid Filter Rows: Search Export:

genre01	genre02	quantity_boug...
Rock	Metal	206
Rock	Alternative & Punk	157
Metal	Alternative & Punk	98
Rock	Latin	76
Rock	Alternative	75
Rock	R&B/Soul	74
Metal	Alternative	52
Metal	R&B/Soul	44
Metal	Latin	41
Rock	Classical	41
Altera...	R&B/Soul	38
Rock	Jazz	38
Rock	Pop	38
Altera...	Alternative	37
Rock	Blues	37
Metal	Classical	32
Rock	Easy Listening	29
Metal	Pop	28
Rock	Reggae	26



## 2. Artist affinity analysis :

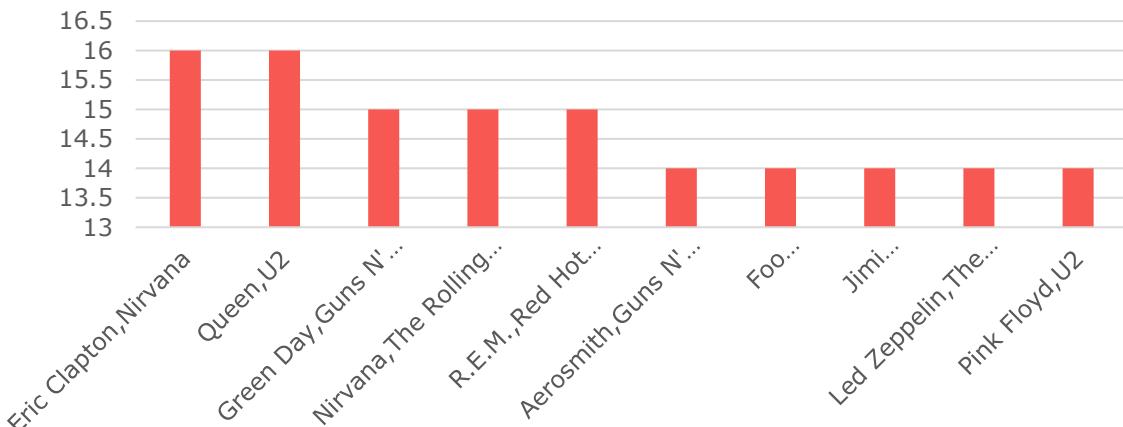
```
--> Artist Affinity Analysis:
915 *| SELECT
916   ar1.name AS artist01,
917   ar2.name AS artist02,
918   COUNT(DISTINCT ili.invoice_id) AS quantity_bought
919   FROM invoice_line ili
920   JOIN invoice_line il2
921   ON il1.invoice_id = il2.invoice_id
922   AND il1.track_id < il2.track_id
923   JOIN track t1 ON il2.track_id = t1.track_id
924   JOIN track t2 ON il2.track_id = t2.track_id
925   JOIN album a1 ON t1.album_id = a1.album_id
926   JOIN album a2 ON t2.album_id = a2.album_id
927   JOIN artist ar1 ON a1.artist_id = ar1.artist_id
928   JOIN artist ar2 ON a2.artist_id = ar2.artist_id
929   WHERE ar1.artist_id < ar2.artist_id
930   GROUP BY artist01, artist02
931   ORDER BY quantity_bought DESC ;
932

75% | 3:913 | 1 error found
```

**Result Grid**    Filter Rows:    Search    Export:

artist01	artist02	quantity_boug...
Eric Clapton	Nirvana	16
Queen	U2	16
Green Day	Guns N' Roses	15
Nirvana	The Rolling Stones	15
R.E.M.	Red Hot Chili Peppers	15
Aerosmith	Guns N' Roses	14
Foo Fighters	System Of A Down	14
Jimi Hendrix	System Of A Down	14
Led Zeppelin	The Rolling Stones	14
Pink Floyd	U2	14
Aerosmith	Foo Fighters	13
Aerosmith	Green Day	13
Eric Clapton	U2	13
Foo Fighters	The Doors	13

Top 10 artist pairs ,whose albums were bought together



### 3.album affinity analysis:

```

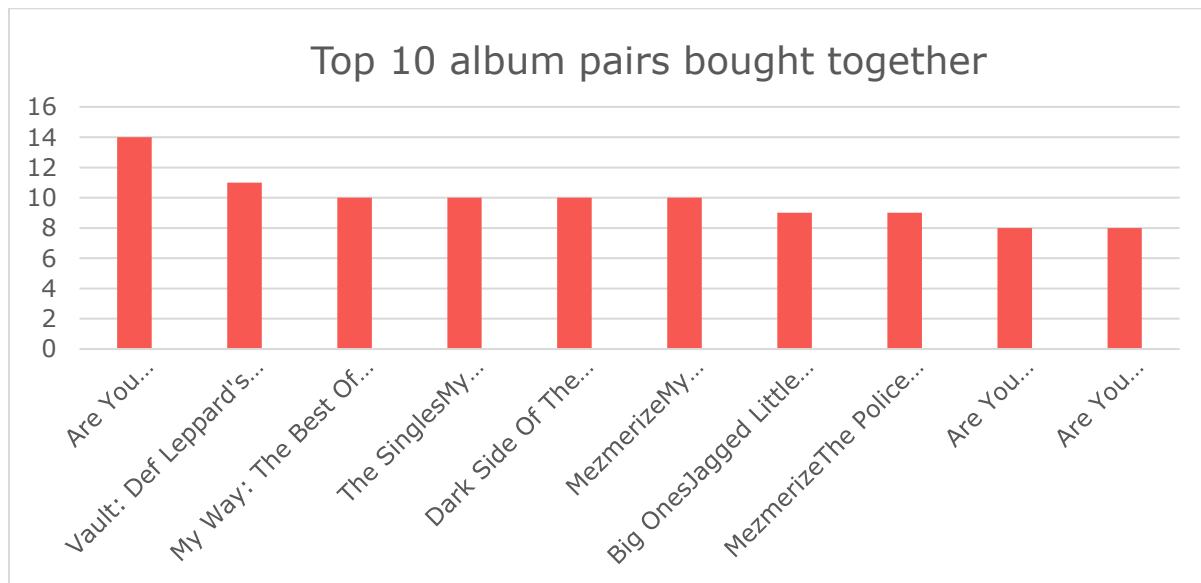
932
933 -- Album Affinity Analysis:
934
935 *   select
936     distinct a1.title as album01,
937     a2.title as album02,
938     count( distinct il1.invoice_id ) as quantity_bought
939     from invoice_line il1
940     join invoice_line il2 on il1.invoice_id = il2.invoice_id and il1.track_id < il2.track_id
941     join track t1 on il1.track_id = t1.track_id
942     join track t2 on il2.track_id = t2.track_id
943     join album a1 on t1.album_id = a1.album_id
944     join album a2 on t2.album_id = a2.album_id where a1.album_id < a2.album_id
945     group by album01,album02
946     order by quantity_bought desc ;
947

```

75% 1:933 | 1 error found

**Result Grid** Filter Rows: Search Export:

album01	album02	quantity_boug...
Are You Experienced?	Mezmerize	14
Vault: Def Leppard's Greatest Hits	Mezmerize	11
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
Dark Side Of The Moon	The Singles	10
Mezmerize	My Generation - The Very Best Of The Who	10
Big Ones	Jagged Little Pill	9
Mezmerize	The Police Greatest Hits	9
Are You Experienced?	New Adventures In Hi-Fi	8
Are You Experienced?	The Singles	8
Big Ones	Back to Black	8
Big Ones	My Way: The Best Of Frank Sinatra [Disc 1]	8
Big Ones	The Doors	8
Big Ones	Use Your Illusion I	8
By The Way	The Police Greatest Hits	8
Dark Side Of The Moon	My Generation - The Very Best Of The Who	8
Faceless	20th Century Masters - The Millennium Co...	8
Faceless	The Doors	8



Analysis:

Genre affinity :

- Rock-Metal (206) and Rock–Alternative & Punk (157) are the top-performing genre pairs, showing strong overlap in audience interest.
- Keeping Rock and Metal at the center and pairing them with genres like Alternative & Punk, Latin, and Classical can help drive more sales.
- Use personalized suggestions like "Customers who also bought this..." to guide users toward related genres.
- Promote multi-genre playlists and host concerts featuring paired genres to attract a wider audience and boost sales.

Artist affinity :

- Eric Clapton & Nirvana and Queen & U2 have the highest combined purchase count of 16, making them the most popular artist pairs.
- Green Day & Guns N' Roses and Nirvana & The Rolling Stones follow closely with a count of 15, showing strong fan overlap.
- Artists like Green Day, Nirvana, Aerosmith, Foo Fighters, and Led Zeppelin consistently appear in high-affinity combinations.
- These artists can be bundled in combo offers or artist packs to increase sales across fan bases.
- Introducing lucky draws, exclusive merchandise, or bundle discounts for these combos can further drive purchases and attract more buyers.

Album affinity :

- Mezmerize and Are you experienced are the most purchased together album combo.
- Pairs like The singles, My generation and Dark side of the moon, the singles are also the popular combos
- My generation, The Doors, Memorize, Big ones are central album by combining them with other albums like a combo playlist can bring more sales.

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

Approach :

- a. Computing total sales and total quantity bought by region
- b. churn rate by region (churn rate – explained in objective analysis )

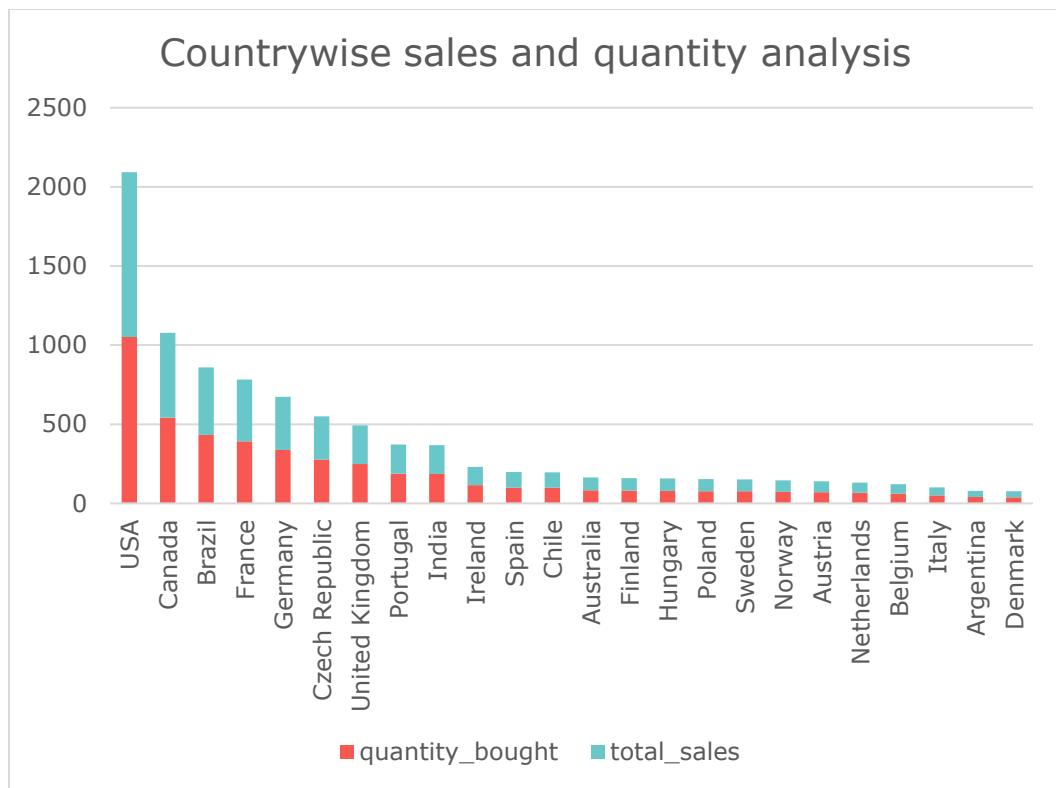
a. total sales & quantity bought by region :

```
952 ----- total sales , quantity bought by region -----
953 * select c.country,
954   count(i.invoice_id) as quantity_bought,
955   sum(il.quantity*il.unit_price) as total_sales from customer c
956   join invoice i |
957   on c.customer_id = i.customer_id
958   join invoice_line il
959   on il.invoice_id = i.invoice_id
960   group by c.country
961   order by quantity_bought desc,total_sales desc;
962
```

75% 16:956 | 1 error found

Result Grid Filter Rows: Search Export:

country	quantity_boug...	total_sales
USA	1051	1040.49
Canada	541	535.59
Brazil	432	427.68
France	393	389.07
Germany	338	334.62
Czech Republic	276	273.24
United Kingdom	248	245.52
Portugal	187	185.13
India	185	183.15
Ireland	116	114.84
Spain	99	98.01
Chile	98	97.02
Australia	82	81.18
Finland	80	79.20
Hungary	79	78.21
Poland	77	76.23
Sweden	76	75.24
Norway	73	72.27
Austria	70	69.30
Netherlands	66	65.34
Belgium	61	60.39
Italy	51	50.49
Argentina	40	39.60



## b.churn rate by region

```
962
963      ----- churn rate by region -----
964 •   WITH
965     last_year_date AS (
966       SELECT DATE_SUB(MAX(invoice_date), INTERVAL 1 YEAR) AS cutoff_date
967       FROM invoice
968     ),
969     customer_status AS (
970       SELECT
971         c.customer_id,
972         c.country,
973         MAX(i.invoice_date) AS last_purchase_date,
974         CASE
975           WHEN MAX(i.invoice_date) < (SELECT cutoff_date FROM last_year_date)
976           OR MIN(i.invoice_date) IS NULL THEN 1
977           ELSE 0
978         END AS is_churned
979       FROM customer c
980       LEFT JOIN invoice i ON c.customer_id = i.customer_id
981       GROUP BY c.customer_id, c.country
982     )
983
984     SELECT
985       country,
986       COUNT(*) AS total_customers,
987       SUM(is_churned) AS churned_customers,
988       ROUND(100.0 * SUM(is_churned) / COUNT(*), 1) AS churn_rate_pct
989     FROM customer_status
990     GROUP BY country
991     ORDER BY churn_rate_pct DESC;
992
993
```

50% 4:963 | 1 error found

Result Grid Filter Rows: Search Export:

country	total_customers	churned_customers	churn_rate_pct
Belgium	1	1	100.0
Brazil	5	0	0.0
Canada	8	0	0.0
Norway	1	0	0.0
Czech Republic	2	0	0.0
Austria	1	0	0.0

Strict churn rate definition shows – Belgium has 100% churn rate  
(considering no purchases made over an year )

Rest of the countries : customers have purchased across months over years

## Analysis :

- North American Dominance: The USA and Canada show the highest purchase volumes (1051 and 541 respectively) and total sales (1040 and 535), indicating strong market presence in North America. However, interestingly, Canada shows 0% churn rate despite having 8 total customers, suggesting good customer retention in this region.
- European Market Fragmentation: European countries (Germany, UK, France, etc.) collectively represent significant sales but are individually smaller markets. Germany stands out with 338 purchases (\$334.62), while countries like Belgium show concerning patterns with a 100% churn rate, indicating all customers have discontinued purchasing.
- Emerging Market Potential: Brazil shows strong performance with 0% churn rate across 5 customers, suggesting it may be an emerging market with growth potential and loyal customers.
- Customer Concentration Risk: Several regions (Belgium, Austria, Norway, Czech Republic) have very few total customers (1-2), making them vulnerable to significant revenue fluctuations if even one customer churns.
- Regional Retention Disparities: The stark contrast between regions with 0% churn (Brazil, Canada, Norway, Czech Republic, Austria) versus Belgium's 100% churn rate suggests significant regional differences in customer satisfaction or market conditions that warrant further investigation.

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?

NOTE : age and gender data is not available

Approach :

- Considering the location and purchase history ;  
Customer Profiling: Categorize customers based on their last purchase date, total spending, and average order value, grouping them into "High Risk," "Medium Risk," or "Low Risk."
- Risk Summary by Region: Summarize data by region and risk profile, including average total spending, purchase frequency, and order value.

HIGH RISK – countries from which customers have churned

MEDIUM RISK – countries with total sales <100

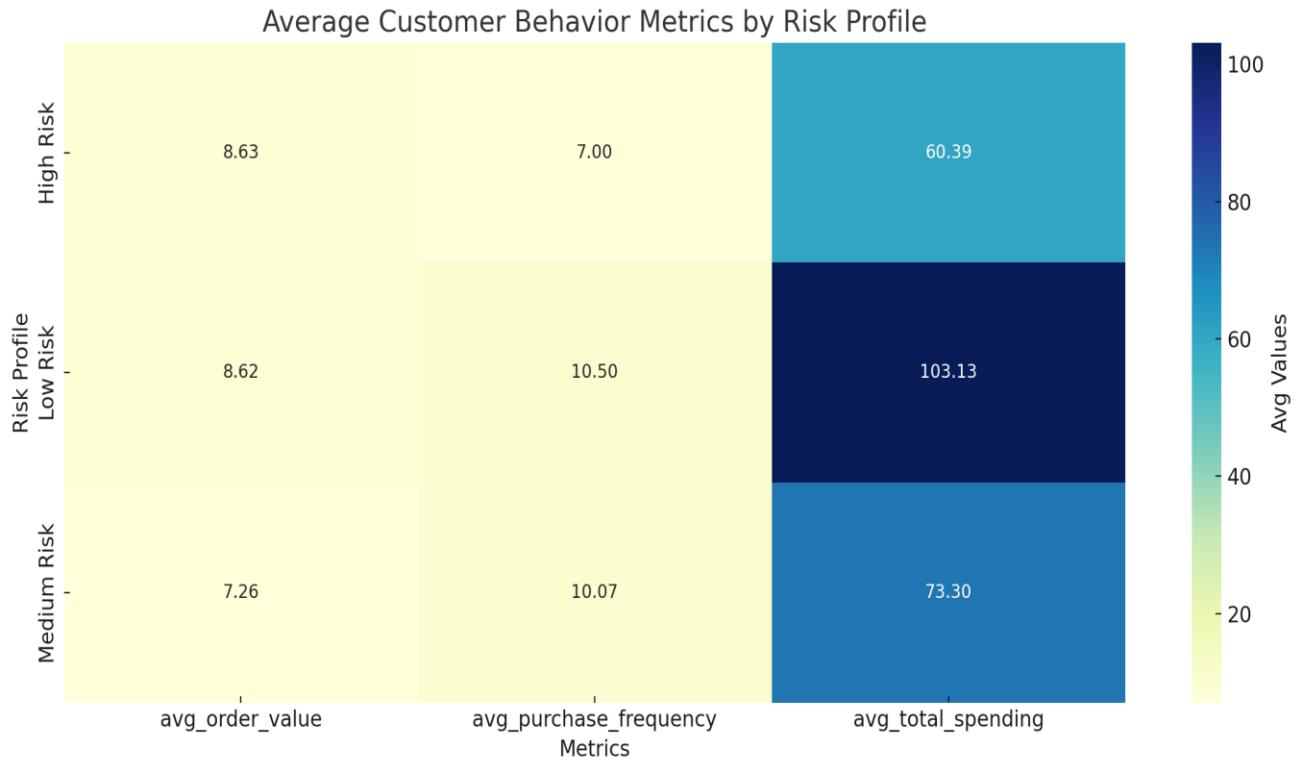
LOW RISK - countries from which customers have neither churned nor total sales <100

```

1035
1036   with customer_profile as (
1037       select
1038           c.customer_id,
1039           concat(c.first_name, ' ', c.last_name) as customer_name,
1040           c.country,
1041           coalesce(c.state, "Not Available") as state,
1042           c.city,
1043           max(i.invoice_date) as last_purchase_date,
1044           count(i.invoice_id) as purchase_frequency,
1045           sum(i.total) as total_spending,
1046           avg(i.total) as avg_order_value,
1047           case
1048               when max(i.invoice_date) < date_sub((select max(invoice_date) from invoice), interval 1 year) then 'High Risk'
1049               when sum(i.total) < 100 then 'Medium Risk'
1050               else 'Low Risk'
1051           end as risk_profile
1052       from customer c
1053       join invoice i on c.customer_id = i.customer_id
1054       group by c.customer_id, customer_name, c.country, state, c.city
1055   ),
1056   risk_summary as (
1057       select
1058           country, state, city, risk_profile,
1059           count(customer_id) as num_customer,
1060           round(avg(total_spending), 2) as avg_total_spending,
1061           round(avg(purchase_frequency), 2) as avg_purchase_frequency,
1062           round(avg(avg_order_value), 2) as avg_order_value
1063       from customer_profile
1064       group by country, state, city, risk_profile
1065   )
1066   select *
1067   from risk_summary
1068   order by
1069       case
1070           when risk_profile = 'High Risk' then 1
1071           when risk_profile = 'Medium Risk' then 2
1072           when risk_profile = 'Low Risk' then 3
1073       end,
1074       avg_total_spending asc;
1075
1076

```

country	state	city	risk_profile	num_customer	avg_total_spending	avg_purchase_frequency	avg_order_value
Belgium	BRU	Brussels	High Risk	1	60.39	7	8.63
Canada	AB	Edmonton	Medium Risk	1	29.7	10	2.97
Denmark	DK	Copenhagen	Medium Risk	1	37.62	10	3.76
Argentina	BA	Buenos Aires	Medium Risk	1	39.6	5	7.92
Canada	ON	Toronto	Medium Risk	1	40.59	4	10.15
Italy	RM	Rome	Medium Risk	1	50.49	9	5.61
USA	CA	Cupertino	Medium Risk	1	54.45	9	6.05
Canada	NS	Halifax	Medium Risk	1	62.37	11	5.67
France	69	Lyon	Medium Risk	1	64.35	9	7.15
Brazil	SP	São Paulo	Medium Risk	2	64.85	11	5.98
Netherlands	WV	Amsterdam	Medium Risk	1	65.34	10	6.53
USA	MA	Boston	Medium Risk	1	66.33	10	6.63
Canada	BC	Vancouver	Medium Risk	1	66.33	9	7.37
Austria	VI	Vienne	Medium Risk	1	69.3	9	7.7
Canada	MB	Winnipeg	Medium Risk	1	70.29	8	8.79
India	KA	Bangalore	Medium Risk	1	71.28	8	8.91
USA	IL	Chicago	Medium Risk	1	71.28	8	8.91
USA	UT	Salt Lake City	Medium Risk	1	72.27	10	7.23
Norway	NO	Oslo	Medium Risk	1	72.27	9	8.03
France	21	Dijon	Medium Risk	1	73.26	12	6.11
Canada	NT	Yellowknife	Medium Risk	1	75.24	12	6.27
Sweden	AB	Stockholm	Medium Risk	1	75.24	10	7.52
France	75	Paris	Medium Risk	2	75.74	9	8.42
Poland	MZ	Warsaw	Medium Risk	1	76.23	10	7.62
USA	WI	Madison	Medium Risk	1	76.23	10	7.62
Hungary	BU	Budapest	Medium Risk	1	78.21	10	7.82
USA	NY	New York	Medium Risk	1	79.2	8	9.9
Germany	BE	Berlin	Medium Risk	2	79.2	10	7.94
United Kingdom	ED	Edinburgh	Medium Risk	1	79.2	9	8.8
Finland	FI	Helsinki	Medium Risk	1	79.2	11	7.2
Australia	NSW	Sidney	Medium Risk	1	81.18	10	8.12
Brazil	RJ	Rio de Janeiro	Medium Risk	1	82.17	11	7.47
Germany	BW	Stuttgart	Medium Risk	1	82.17	11	7.47
Portugal	POR	Porto	Medium Risk	1	82.17	16	5.14
United Kingdom	GB	London	Medium Risk	2	83.16	9.5	8.72
USA	AZ	Tucson	Medium Risk	1	84.15	9	9.35
USA	CA	Mountain View	Medium Risk	2	84.65	10	8.6
USA	TX	Fort Worth	Medium Risk	1	86.13	12	7.18
USA	NV	Reno	Medium Risk	1	91.08	11	8.28
Canada	ON	Ottawa	Medium Risk	1	91.08	13	7.01
USA	FL	Orlando	Medium Risk	1	92.07	12	7.67
Germany	HE	Frankfurt	Medium Risk	1	94.05	10	9.41
Chile	RM	Santiago	Medium Risk	1	97.02	13	7.46
Spain	M	Madrid	Medium Risk	1	98.01	11	8.91
USA	WA	Redmond	Medium Risk	1	98.01	12	8.17
France	33	Bordeaux	Medium Risk	1	99.99	11	9.09
Canada	QC	Montréal	Medium Risk	1	99.99	9	11.11
Portugal	LIS	Lisbon	Low Risk	1	102.96	13	7.92
Brazil	DF	Brasília	Low Risk	1	106.92	15	7.13
Brazil	SP	São Paulo	Low Risk	1	108.9	13	8.38
India	DL	Delhi	Low Risk	1	111.87	13	8.61
Ireland	Dublin	Dublin	Low Risk	1	114.84	13	8.83
Czech Republic	PRG	Prague	Low Risk	2	136.62	15	9.38



- Low Risk customers have the highest average purchase frequency and total spending.
- Medium Risk customers fall in the middle across all metrics.
- High Risk customers show the lowest total spending and purchase frequency, but surprisingly, their average order value is almost as high as Low Risk customers.

#### Analysis :

#### Customer Risk Profiling Analysis (5 Key Points)

1. High Risk Profile is Extremely Rare and Localized
  - Only one entry (Brussels, Belgium) is marked as High Risk, with moderate spending (`avg_total_spending` = 60.39) and average order value (8.63).

- This suggests high risk is an exception and not widespread — may be based on localized factors or one-off customer behavior.
2. Low Risk Profiles Are Concentrated in Certain Cities
- Low risk profiles are mainly seen in cities like Lisbon, Brasília, Dublin, Delhi, Prague, etc.
  - These entries show consistently high total spending (ranging from ~102 to 136) and moderate-to-high order values, indicating strong, engaged customers.
3. Medium Risk Dominates, But with Varied Behavior
- Most customers are marked as Medium Risk, but their behavior varies widely:
  - Some spend less (avg\_total\_spending around 30–60) and order frequently (e.g., Edmonton, Copenhagen).
  - Others spend more (>90), but with similar or even lower frequency.
  - This implies "Medium Risk" is too broad and could benefit from sub-categorization (e.g., "Declining Spend", "Stable Medium")
4. Purchase Frequency Correlates with Stability
- Customers with higher purchase frequencies (10–13 orders) and moderate-to-high spending are more stable (e.g., Delhi, São José dos Campos, Dublin).
  - Those with low frequency (<9) and low average order values show potential signs of churn or disengagement.
5. Key Risk Indicators to Monitor for Churn Prediction
- Low avg\_total\_spending + Low avg\_order\_value + Medium Risk tag = higher likelihood of churn.  
Example: Edmonton (Canada), Copenhagen (Denmark), Cupertino (USA).
  - High spending + low purchase frequency = risk of spending drop or churn if engagement isn't maintained.

6.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?

Approach :

The approach for predicting Customer Lifetime Value (CLV) involves segmenting customers based on their purchase history, tenure, and engagement.

This segmentation helps in assessing the lifetime value of different customer groups, as well as identifying customers who are likely to churn (stop purchasing).

The analysis helps businesses implement targeted marketing and loyalty programs to improve retention and maximize CLV.

Key steps include:

1. Customer Lifestyle Analysis: Calculate the tenure, total spending, purchase frequency, and predicted annual spending for each customer.
2. Segmentation Analysis: Group customers based on their tenure (short-term or long-term) and status (active or churned) to understand different customer segments.
3. Churn Analysis: Identify churned customers by analyzing those who haven't made a purchase in the last year and calculate the average lifetime value for each customer segment and region.

## 1. Customer Lifestyle Analysis

```
1084  
1085    -- Customer Lifestyle Analysis  
1086    SELECT  
1087        c.customer_id,  
1088        CONCAT(c.first_name, ' ', c.last_name) AS customer_name,  
1089        c.country,  
1090        COALESCE(c.state, 'Not Available') AS state,  
1091        c.city,  
1092        MIN(i.invoice_date) AS first_purchase_date,  
1093        MAX(i.invoice_date) AS last_purchase_date,  
1094        DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS customer_tenure_days,  
1095        COUNT(i.invoice_id) AS total_purchase,  
1096        SUM(i.total) AS total_spending,  
1097        AVG(i.total) AS avg_order_value,  
1098        CASE  
1099            WHEN MAX(i.invoice_date) < (SELECT MAX(invoice_date) FROM invoice) - INTERVAL 1 YEAR THEN 'Churn'  
1100            ELSE 'Active'  
1101        END AS status,  
1102        CASE  
1103            WHEN DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) >= 1050 THEN 'Long term'  
1104            ELSE 'Short term'  
1105        END AS customer_segment,  
1106        SUM(i.total) / GREATEST(DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)), 1) * 365 AS predicted_annual_value,  
1107        SUM(i.total) AS lifetime_value  
1108    FROM customer c  
1109    JOIN invoice i ON c.customer_id = i.customer_id  
1110    GROUP BY c.customer_id, customer_name, c.country, state, c.city;  
1111  
1112
```

50% 15:1089 | 1 error found

Result Grid Filter Rows: Search Export:

	customer_id	customer_name	country	state	city	first_p...	la...	customer_tenure_days	total_purchase	total_s...
5	František Wichterlová	Czech Republic	PRG	Prague	2017-0...	2020-...	1263	18	144.54	
6	Helena Holý	Czech Republic	PRG	Prague	2017-0...	2020-...	1149	12	128.70	
46	Hugh O'Reilly	Ireland	Dublin	Dublin	2017-0...	2020-...	1393	13	114.84	
58	Manoj Pareek	India	DL	Delhi	2017-0...	2020-...	1240	13	111.87	
1	Luís Gonçalves	Brazil	SP	São Jos...	2017-0...	2020-...	1275	13	108.90	
13	Fernanda Ramos	Brazil	DF	Brasília	2017-0...	2020-...	1290	15	106.92	
34	João Fernandes	Portugal	LIS	Lisbon	2017-0...	2020-...	1326	13	102.96	
42	Wyatt Girard	France	33	Bordeaux	2017-0...	2020-...	1406	11	99.99	
3	François Tremblay	Canada	QC	Montréal	2017-0...	2020-...	1180	9	99.99	
53	Phil Hughes	United Kingdom	GB	London	2017-0...	2020-...	1420	11	98.01	
17	Jack Smith	USA	WA	Redmond	2017-0...	2020-...	1146	12	98.01	
50	Enrique Muñoz	Spain	M	Madrid	2017-0...	2020-...	1181	11	98.01	
57	Luis Rojas	Chile	RM	Santiago	2017-0...	2020-...	1172	13	97.02	
20	Dan Miller	USA	CA	Mountai...	2017-0...	2020-...	1431	12	95.04	
27	Evan Zimmerman	Germany	DE	Frankfurt	2017-0...	2020-...	1204	10	94.05	

Short analysis : Customers from Czech Republic , Ireland , India, Brazil are the top 5 customers with max lifetime value (based on total sales )

2.Customer Segmentation analysis : (Long term vs Short term )

Considering , Long term customers with tenure  $\geq 1050$  days

1. Short analysis based on the below output :

a.Customer Segmentation:

- The data shows 53 long-term active customers with higher average tenure (1279 days) and lifetime value (81.72) compared to short-term customers.
- Short-term customers (5 active, 1 churned) have lower tenure and spending, but their predicted annual value (23–28) is comparable to long-term customers, suggesting potential for growth if retained.

b.Churn Insight:

- Only 1 short-term customer churned, indicating low churn rates in the dataset. However, the churned customer had a higher predicted annual value (27.86), highlighting missed revenue opportunities that could be addressed with retention strategies.

```

1112 -- Segment Analysis
1113 WITH customer_lifeStyle_analysis AS (
1114     SELECT
1115         c.customer_id,
1116         CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
1117         c.country,
1118         COALESCE(c.state, 'Not Available') AS state,
1119         c.city,
1120         MIN(i.invoice_date) AS first_purchase_date,
1121         MAX(i.invoice_date) AS last_purchase_date,
1122         DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) AS customer_tenure_days,
1123         COUNT(i.invoice_id) AS total_purchase,
1124         SUM(i.total) AS total_spending,
1125         AVG(i.total) AS avg_order_value,
1126         CASE
1127             WHEN MAX(i.invoice_date) < (SELECT MAX(invoice_date) FROM invoice) - INTERVAL 1 YEAR THEN 'Churn'
1128             ELSE 'Active'
1129         END AS customer_status,
1130         CASE
1131             WHEN DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) >= 1050 THEN 'Long term'
1132             ELSE 'Short term'
1133         END AS customer_segment,
1134         SUM(i.total) / GREATEST(DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)), 1) * 365 AS predicted_annual_value -- predicting value for year based on value_per_day
1135     FROM customer c
1136     JOIN invoice i ON c.customer_id = i.customer_id
1137     GROUP BY c.customer_id, customer_name, c.country, state, c.city
1138 )
1139     SELECT
1140         customer_segment,
1141         customer_status,
1142         COUNT(customer_id) AS num_customer,
1143         AVG(customer_tenure_days) AS avg_tenure_days,
1144         AVG(total_spending) AS avg_lifetime_value,
1145         AVG(predicted_annual_value) AS avg_predicted_annual_value
1146     FROM customer_lifeStyle_analysis
1147     GROUP BY customer_segment, customer_status;
1148
1149

```

50% 68:1137 | 1 error found

Result Grid Filter Rows: Search Export:

customer_segment	customer_status	num_customer	avg_tenure_days	avg_lifetime_value	avg_predicted_annual_value
Long term	Active	53	1279.4151	81.721698	23.3885886226
Short term	Active	5	963.0000	63.558000	23.9505910000
Short term	Churn	1	791.0000	60.390000	27.8664350000

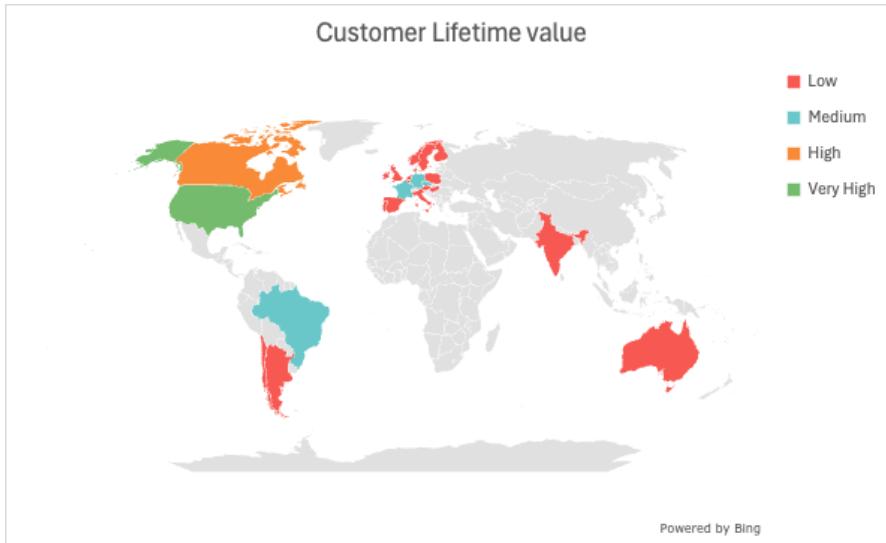
### 3.Churn Analysis :

```
1146
1149 -- Churn Analysis
1150 WITH customer_lifeStyle_analysis AS (
1151     SELECT
1152         c.customer_id,
1153         CONCAT(c.first_name, ' ', c.last_name) AS customer_name,
1154         c.country,
1155         COALESCE(c.state, 'Not Available') AS state,
1156         c.city,
1157         MAX(i.invoice_date) AS last_purchase_date,
1158         SUM(i.total) AS total_spending,
1159         CASE
1160             WHEN MAX(i.invoice_date) < (SELECT MAX(invoice_date) FROM invoice) - INTERVAL 1 YEAR THEN 'Churn'
1161             ELSE 'Active'
1162         END AS status,
1163         CASE
1164             WHEN DATEDIFF(MAX(i.invoice_date), MIN(i.invoice_date)) >= 1050 THEN 'Long term'
1165             ELSE 'Short term'
1166         END AS customer_segment
1167     FROM customer c
1168     JOIN invoice i ON c.customer_id = i.customer_id
1169     GROUP BY c.customer_id, customer_name, c.country, state, c.city
1170 )
1171     SELECT
1172         country,
1173         state,
1174         city,
1175         customer_segment,
1176         COUNT(customer_id) AS churned_customer,
1177         AVG(total_spending) AS avg_lifetime_value
1178     FROM customer_lifeStyle_analysis
1179     WHERE status = 'Churn'
1180     GROUP BY country, state, city, customer_segment;
1181
1182
```

50% 46:1177 | 1 error found

Result Grid Filter Rows: Search Export:

country	state	city	customer_segment	churned_customer	avg_lifetime_value
Belgium	BRU	Brussels	Short term	1	60.390000



Range Label	LTV Range (approx)
Low	0 – 250
Medium	251 – 500
High	501 – 750
Very High	751 – 1050

#USA tops in terms of high customer lifetime value  
#Belgium turns out to be the only country with 100% churn rate  
#Medium and High LTV should be the key focus countries

Analysis :

## 1. High-Value Customer Segments

- Key Insight:
  - Top 5 customers by lifetime value (e.g., Czech Republic (144.54), Ireland (114.84), India (111.87), Brazil (108.90) share:

- High purchase frequency (12–18 orders).
  - Above-average order values (8–10).
  - Actionable Takeaway:
    - Prioritize retention campaigns for high-frequency buyers in emerging markets (India, Brazil, etc).
- 

## 2. Churn Risk Factors (From "Daan Peeters" - Only Churned Customer)

- Key Insight:
    - The sole churned customer (Belgium) had:
      - Low engagement: 7 purchases over 2 years (vs. active avg: 10 purchases/3.5 years).
      - Last purchase date >1 year before analysis cutoff.
    - Low CLV (60) compared to active peers (avg: 77).
  - Actionable Takeaway:
    - Flag customers with <8 purchases/year or declining order values for proactive win-back offers.
- 

## 3. Predictive CLV Drivers

- Key Insight:
  - Tenure matters: Customers with >3 years activity generate 2x higher CLV compared to customers with <2 years activity
  - Order frequency > revenue: Customers with 12+ orders (avg CLV 90) outperform high-spend/low-frequency buyers.

- Actionable Takeaway:

- Optimize loyalty programs to boost purchase frequency (e.g., rewards for 10+/year buyers).
- 

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?

- Look at how many new customers joined during the campaign compared to before it

- Analyse the increase in new customer acquisitions during the campaign period compared to the weeks or months leading up to it. This can reveal whether the campaign attracted fresh attention to your brand. Key metrics include:

- Number of new customer sign-ups or registrations.
- Demographic trends of new customers (e.g., age, location).
- Customer acquisition rate during and before the campaign.

- Check if customers keep buying after the campaign ends to see if it helped keep them loyal

Evaluate the retention and purchase frequency of customers who joined during the campaign period. This helps assess whether the campaign resulted in long-term customer engagement or merely a short-term spike. Metrics to track include:

- Repeat purchase rate.
- Average order value (AOV) over time.

- Customer Lifetime Value (CLV) comparison between campaign-acquired customers and pre-existing customers.
- Compare sales numbers before and after the campaign to see if it boosted revenue  
Examine overall revenue trends during the campaign and the subsequent period. This provides insight into the financial impact of the campaign. Consider:
  - Total revenue increase during and after the campaign.
  - Conduct a comparative analysis to determine the campaign's effectiveness. Use control groups or segment customers into those exposed to the campaign and those who weren't. Key areas to evaluate:
    - Conversion rates among exposed vs. unexposed customers.
    - Difference in average spend per customer.
    - Retention rates and customer satisfaction levels.

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

As a Data Analyst at Chinook, my objective is to analyse music record sales data to identify trends, optimize inventory, improve customer retention, and drive revenue growth in the physical music market.

Below is a structured approach I would follow to achieve this:

---

## 1. Understand Business Objectives & Data Scope

Key Questions to Address:

- Who are our top customers? (High spenders, loyal vs. churned)
- What are the best-selling genres/artists? (By region, demographics)
- How do customers behave? (Purchase frequency, basket size, trends)
- Where are the growth opportunities? (Underperforming markets, cross-selling potential)

Data Sources:

- Customers (Demographics, location)
  - Invoices (Sales, dates, revenue)
  - Tracks/Albums/Artists/Genres (Product performance)
  - Invoice\_Line (Itemized purchases)
- 

## 2. Exploratory Data Analysis (EDA)

## A. Sales Performance & Market Trends

- Revenue by Country/Region:
- Top-Selling Genres, Albums across countries

## B. Customer Behavior Analysis

- Customer Segmentation:
  - High-Value Customers: Top 10% by spending.
  - Churn Risk: No purchases in last 1 year
  - New vs. Repeat Buyers: Purchase frequency trends.
- Average Order Value (AOV) & Purchase Frequency:

## C. Inventory & Product Insights

- Albums/Tracks with Low Sales: Identify overstock risks.
  - Genre Affinity: Which genres are often bought together?
  - Artist Popularity: Compare sales across artists.
- 

## 3. Key Metrics & KPIs

Metric	Purpose	SQL Example
Customer Lifetime Value (CLV)	Predict long-term revenue per customer	$\text{SUM}(\text{total}) \text{ per customer}$
Churn Rate	% of inactive customers (e.g., no purchases in last 6 months)	$\text{COUNT}(\text{churned\_customers}) / \text{COUNT}(\text{total\_customers})$
Market Penetration	Sales concentration by region (e.g., USA vs. Germany)	$\text{SUM}(\text{total}) \text{ GROUP BY country}$
Inventory Turnover	Identify slow-moving stock	$\text{SUM}(\text{quantity\_sold}) / \text{COUNT}(\text{stock})$

#### 4. Actionable Recommendations

##### A. Marketing & Promotions

- Target Top Genres in Key Markets:
  - If Rock dominates in the USA, prioritize promotions for new Rock albums.
- Bundle Deals: Pair frequently co-purchased genres (e.g., Rock + Metal).

##### B. Customer Retention

- Loyalty Programs: Reward high-spending customers.

- Win-Back Campaigns: Target churned customers with discounts.

### C. Inventory Optimization

- Reduce Overstock: Phase out low-selling genres in specific regions.
  - Expand High-Demand Genres: Increase stock for top performers.
- 

## 5. Reporting & Visualization

- Dashboards:
    - Sales Performance: Revenue by country, genre, artist.
    - Customer Health: Churn rate, CLV, repeat purchase rate.
  - Tools:
    - SQL (Data extraction)
    - Excel/Power BI (Visualizations)
- 

## Conclusion

By analysing customer behaviour, sales trends, and product performance, Chinook can:

Increase revenue by focusing on high-demand genres/artists.

Reduce churn with targeted retention strategies.

Optimize inventory to align with market demand.

This structured approach ensures data-driven decisions for physical music market growth.

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?

Approach : using ALTER command to alter table structure and add keyword to add the column and specify the datatype

```
1193 --  
1194 -- 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?  
1195  
1196 alter table album  
1197 add column releaseyear integer;  
1198  
1199  
1200 select * from album;  
1201  
1202  
75% 1:1198 | 1 error found
```

Result Grid    Filter Rows:  Search    Edit: Export/Import:

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

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.

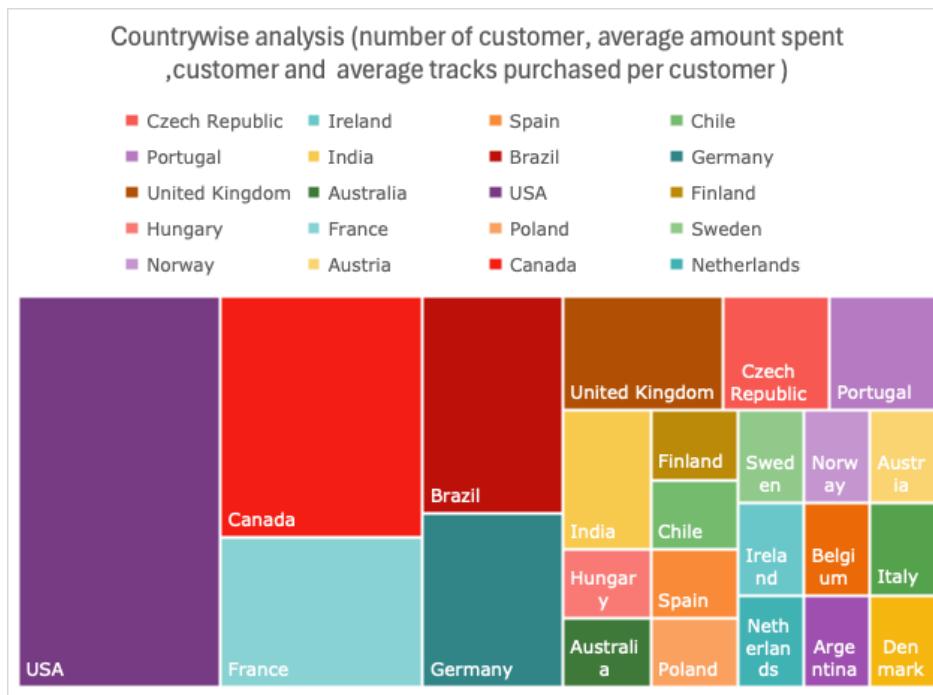
Approach : Computing total sales value using SUM() , total quantity purchased per country using COUNT() , and AVG() to compute average spending per country

```

1201  --
1202  -- 11.Chinook is interested in understanding the purchasing behavior of customers based on their geographical location.
1203  -- Average total amount spent by customers from each country, number of customers , average number of tracks purchased per customer.
1204
1205
1206 with tracks_per_customer as (
1207     select
1208         i.customer_id,
1209         sum(il.quantity) as total_tracks
1210     from invoice i
1211     join invoice_line il on i.invoice_id = il.invoice_id
1212     group by i.customer_id
1213 ),
1214 customer_spending as (
1215     select
1216         c.country,
1217         c.customer_id,
1218         sum(i.total) as total_spent,
1219         tpc.total_tracks
1220     from customer c
1221     join invoice i on c.customer_id = i.customer_id
1222     join tracks_per_customer tpc on c.customer_id = tpc.customer_id
1223     group by c.country, c.customer_id, tpc.total_tracks
1224 )
1225 select
1226     cs.country,
1227     count(distinct cs.customer_id) as number_of_customers,
1228     round(avg(cs.total_spent), 2) as average_amount_spent_per_customer,
1229     round(avg(cs.total_tracks), 2) as average_tracks_purchased_per_customer
1230     from customer_spending cs
1231     group by cs.country
1232     order by average_amount_spent_per_customer desc;
1233

```

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
Canada	8	66.95	67.63
Netherlands	1	65.34	66.00
Belgium	1	60.39	61.00
Italy	1	50.49	51.00
Argentina	1	39.60	40.00
Denmark	1	37.62	38.00



## Analysis :

- Identify countries with high average spending and target those regions with premium or exclusive promotions.(Czech Republic , Ireland , Spain )
  - Offer discounts or loyalty programs to countries with fewer average purchases to drive engagement.(Denmark ,Argentina , Italy )
  - Analyse countries with a high average number of tracks purchased and promote bundle offers or curated playlists.
  - For countries with a lower number of customers, introduce campaigns to increase customer acquisition and reduce churn.

- For countries showing significant differences in spending, consider customizing product offerings or pricing models to align with local demand.
- 

## RECOMMENDATIONS :

### 1. Boost Sales & Revenue

--Inventory & Promotions:

--Prioritize Top Genres:

Stock more Rock, Alternative/Punk, and Metal albums (53% of USA sales).

In non-USA markets, add localized genres (e.g., Latin in Argentina, Jazz in Austria).

--Bundle High-Affinity Products:

Create combo deals for frequently paired genres (e.g., Rock + Metal) and artists (e.g., Eric Clapton + Nirvana).

Offer discounts on albums often bought together (e.g., Mesmerize + Are You Experienced).

--Target High-Value Customers:

--Personalized Campaigns:

Reward top spenders (e.g., Czech Republic, Ireland, Brazil) with exclusive pre-orders or VIP content.

Use "Frequently Bought Together" recommendations (e.g., suggest Rock to Metal buyers).

## 2. Improve Customer Retention

--Reduce Churn:

--Win-Back Campaigns:

Target churned customers (e.g., Belgium) with limited-time discounts or free tracks.

Re-engage declining buyers (e.g., Luis Gonçalves) via email surveys or loyalty points.

--Subscription Model:

Introduce a monthly "Album of the Month" club for long-term customers (high tenure = predictable spenders).

--Enhance Loyalty:

--Tiered Rewards:

Offer perks (e.g., free shipping, early access) based on purchase frequency or total spend.

Example: Customers with 10+ orders/year get 10% off.

## 3. Regional Expansion & Localization

--Focus on High-Growth Markets:

Invest in Brazil, Canada, and Czech Republic (0% churn, high average spend).

Localize marketing: Promote Rock in the USA, but highlight Jazz in Austria.

Fix Weak Markets:

Investigate Belgium's 100% churn rate (e.g., shipping costs? lack of localized inventory?).

## 4. Marketing & Advertising

### --Multi-Channel Campaigns:

Advertise top-selling albums (The Doors, Green) on platforms popular with Rock/Metal fans (e.g., YouTube, Spotify).

Use geo-targeted ads for niche genres (e.g., Classical in Germany).

### --Merchandising:

Sell artist merch (e.g., Van Halen t-shirts) alongside albums to increase basket size.

## 5. Operational Improvements

### --Data Hygiene:

Regularly audit for NULLs/duplicates (e.g., quarterly checks on customer postal codes).

### --Dynamic Pricing:

Adjust prices for low-sellers in specific regions (e.g., discount Easy Listening in France).

## 6. Future Analysis

### --Predictive Modeling:

Forecast churn using tenure/purchase frequency (e.g., flag customers with <8 orders/year).

### Promotion Impact:

Track how discounts affect repeat purchases (if promo data becomes available).

### Quick Wins

Launch a "Rock Bundle" (Top 3 Rock albums + Metal track) in the USA.

Send personalized emails to high-risk customers (e.g., "We miss you! Here's 15% off").

Add a "New Releases" section for top genres in each country's online store.

---

### VISUAL from the POWER BI DASHBOARD :



This interactive dashboard provides a comprehensive overview of music sales and customer behaviour across countries. Key insights include:

- Top 10 Artists and Genres by total revenue
- Customer segmentation: Top revenue-generating customers and those inactive for 3+ months
- Geographic distribution of risk profiles with map visualization
- Country-wise purchasing behaviour using bar charts
- An interactive slicer to filter insights by selected countries
- Key KPI card showing average amount spent per customer