PROJECT ON:

□ MUSIC DATA STORE :

A DEEP DIVE INTO MUSIC ANALYTICS

Analyzing the Symphony of Streaming Data

□ PROJECT BY:

Parmar Arpitaba M

□ PROJECT YEAR:

2k25-2k26

PRESENTATION OUTLINE:

01

INTRODUCTION & PROJECT GOAL

Music is more than just entertainment; it's a massive digital industry powered by data.

02

DATA LOADING & ETL PROCESS

Explain the ETL process using SQL concepts.

03

KEY QUESTIONS & ANALYSIS

A good key question for a music data project should be central to its purpose and guide the analysis

INTRODUCTION

- □ The digital music revolution: massive growth of streaming and digital sales.

- □ The need for effective data storage and analysis.

- □ The project's objective: To design and implement a robust system for storing, managing, and analyzing music data.

PROJECT GOAL

- □ Create a scalable and efficient data store for music-related information.

- □ Facilitate advanced analytics and reporting.

- □ Support the development of data-driven applications .

INTRODUCTION

- □ The digital music revolution: massive growth of streaming and digital sales.
- □ The need for effective data storage and analysis.
- □ The project's objective: To design and implement a robust system for storing, managing, and analyzing music data.

PROJECT GOAL

- □ Create a scalable and efficient data store for music-related information.
- □ Facilitate advanced analytics and reporting.
- □ Support the development of data-driven applications .

Use COPY or INSERT statements to load data from your transformed files into the database tables.

KEY QUESTION & ANALIYSIS

Q1. Who is the senior most employee based on job title?

Q2. Which countries have the most Invoices?

Q3. What are top 3 values of total invoice?

Q4. Which city has the best customers?
We would like to throw a promotional Music Festival in the city we made the most money.
Write a query that returns one city that has the highest sum of invoice totals.
Return both the city name & sum of all invoice totals.

Q5. Who is the best customer?
The customer who has spent the most money will be declared the best customer.
Write a query that returns the person who has spent the most money.

Q6. Write query to return the email, first name, last name, & Genre of all Rock Music listeners.
Return your list ordered alphabetically by email starting with A

Q7. Let's invite the artists who have written the most rock music in our dataset.
Write a query that returns the Artist name and total track count of the top 10 rock bands

Q8. Return all the track names that have a song length longer than the average song length.
Return the Name and Milliseconds for each track.
Order by the song length with the longest songs listed first

Q9. Find how much amount spent by each customer on artists?
Write a query to return customer name, artist name and total spent.

Use COPY or INSERT statements to load data from your transformed files into the database tables.

KEY QUESTION & ANALIYSIS

Q10. We want to find out the most popular music Genre for each country.
We determine the most popular genre as the genre with the highest amount of purchases.
Write a query that returns each country along with the top Genre.
For countries where the maximum number of purchases is shared return all Genres

Q11. Write a query that determines the customer that has spent the most on music for each country.
Write a query that returns the country along with the top customer and how much they spent.
For countries where the top amount spent is shared, provide all customers who spent this amount.

Q12. Who are the most popular artists?

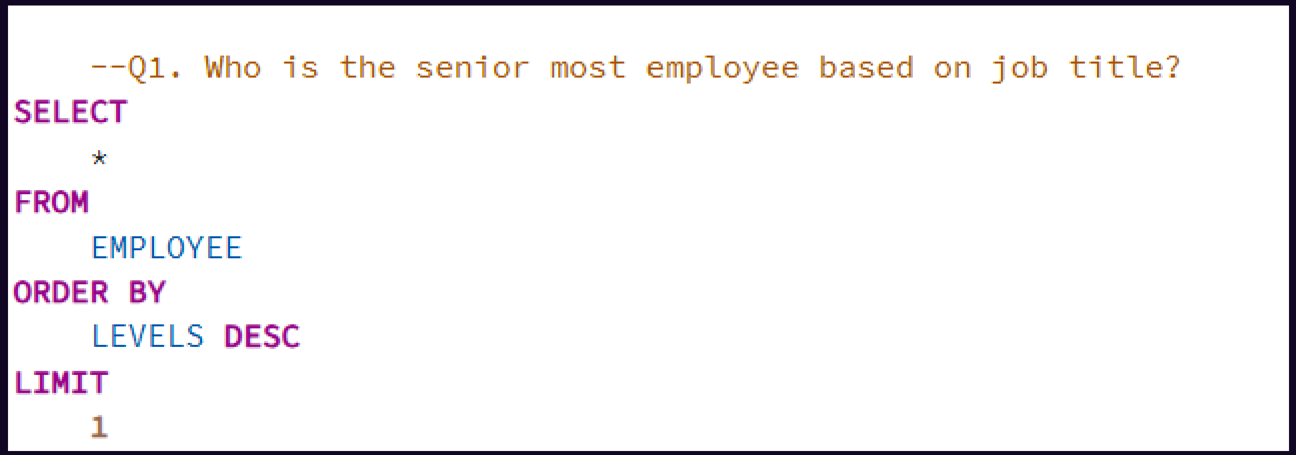
Q13. Which is the most popular song?

Q14. What are the average prices of different types of music?

Q15. What are the most popular countries for music purchases?

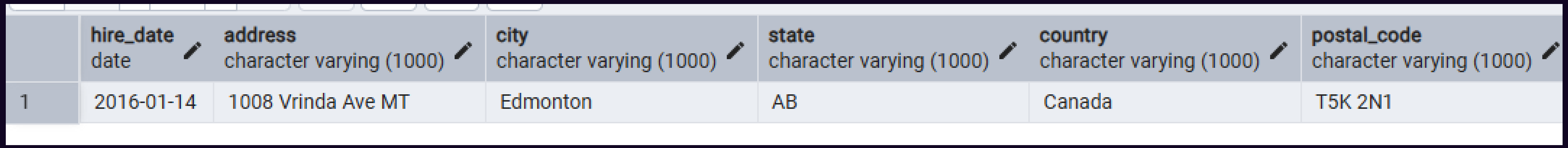
KEY QUESTION & ANALYSIS

❑ Q1.Who is the senior most employee based on job title?



```
    --Q1. Who is the senior most employee based on job title?
SELECT
    *
FROM
    EMPLOYEE
ORDER BY
    LEVELS DESC
LIMIT
    1
```

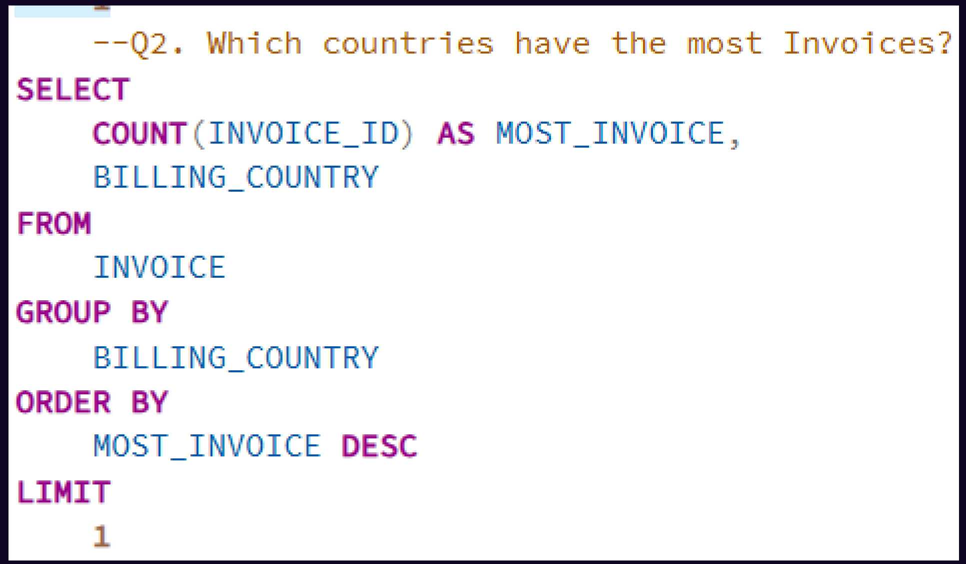
❑ Output



	hire_date date	address character varying (1000)	city character varying (1000)	state character varying (1000)	country character varying (1000)	postal_code character varying (1000)
1	2016-01-14	1008 Vrinda Ave MT	Edmonton	AB	Canada	T5K 2N1

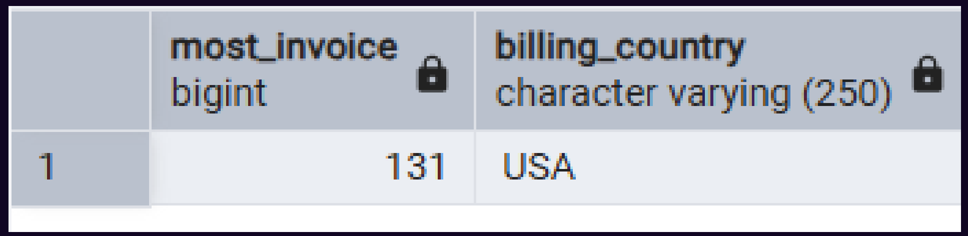
KEY QUESTION & ANALYSIS

❑ Q2. Which countries have the most Invoices?



```
    --Q2. Which countries have the most Invoices?
SELECT
    COUNT(INVOICE_ID) AS MOST_INVOICE,
    BILLING_COUNTRY
FROM
    INVOICE
GROUP BY
    BILLING_COUNTRY
ORDER BY
    MOST_INVOICE DESC
LIMIT
    1
```

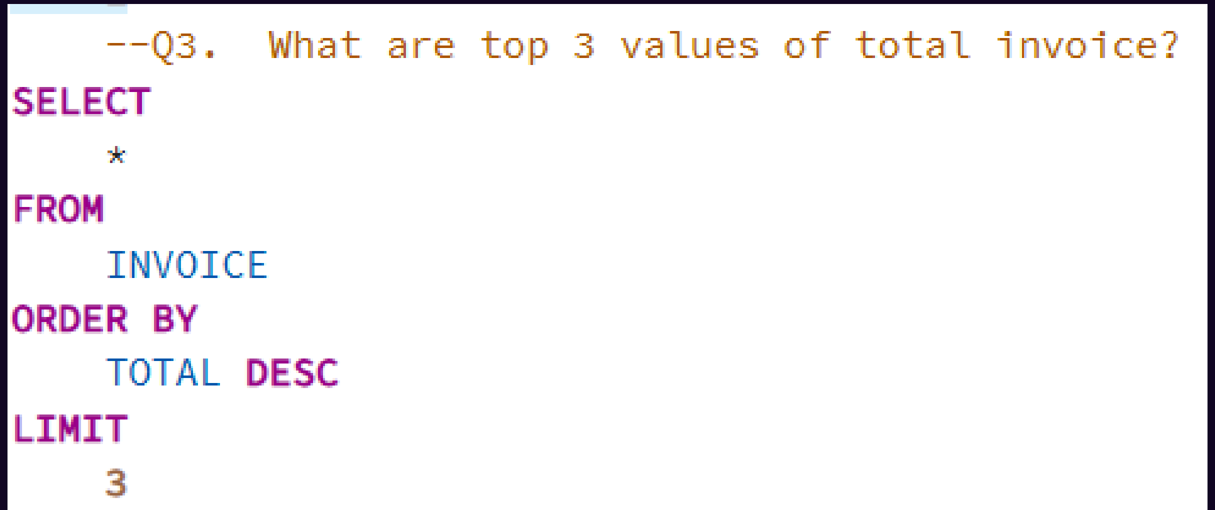
❑ Output



	most_invoice bigint	billing_country character varying (250)
1	131	USA

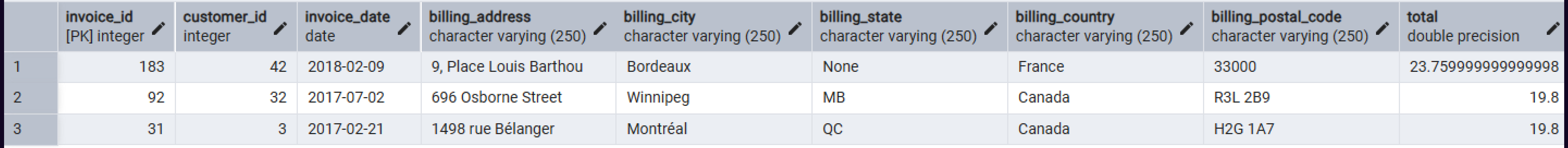
KEY QUESTION & ANALYSIS

❑ Q3. What are top 3 values of total invoice?



```
    --Q3.  What are top 3 values of total invoice?
SELECT
    *
FROM
    INVOICE
ORDER BY
    TOTAL DESC
LIMIT
    3
```

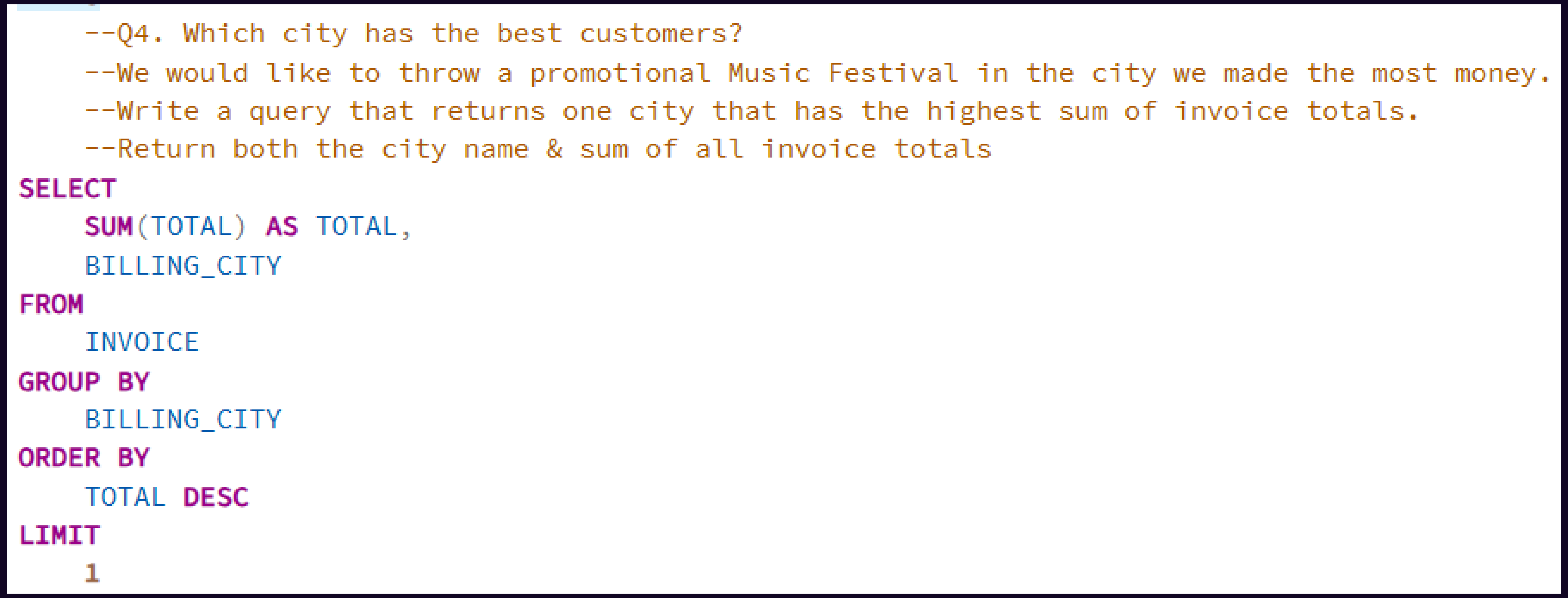
❑ Output



	invoice_id [PK] integer	customer_id integer	invoice_date date	billing_address character varying (250)	billing_city character varying (250)	billing_state character varying (250)	billing_country character varying (250)	billing_postal_code character varying (250)	total double precision
1	183	42	2018-02-09	9, Place Louis Barthou	Bordeaux	None	France	33000	23.759999999999998
2	92	32	2017-07-02	696 Osborne Street	Winnipeg	MB	Canada	R3L 2B9	19.8
3	31	3	2017-02-21	1498 rue Bélanger	Montréal	QC	Canada	H2G 1A7	19.8

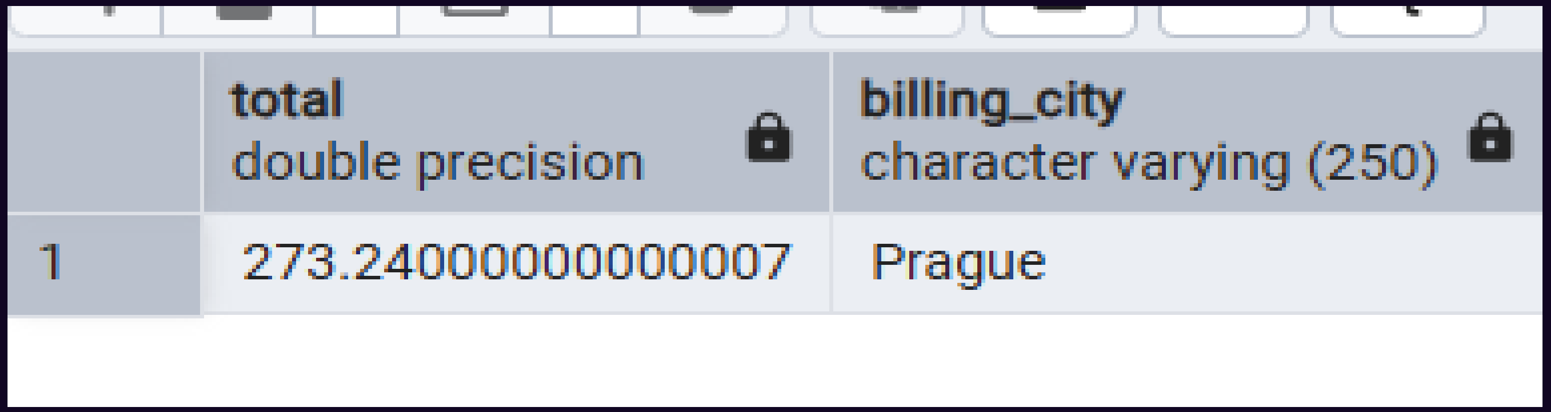
KEY QUESTION & ANALYSIS

❑ Q4. Which city has the best customers? We would like to throw a promotional Music Festival in the city we made the most money. Write a query that returns one city that has the highest sum of invoice totals. Return both the city name & sum of all invoice totals.



```sql
    --Q4. Which city has the best customers?
    --We would like to throw a promotional Music Festival in the city we made the most money.
    --Write a query that returns one city that has the highest sum of invoice totals.
    --Return both the city name & sum of all invoice totals
SELECT
    SUM(TOTAL) AS TOTAL,
    BILLING_CITY
FROM
    INVOICE
GROUP BY
    BILLING_CITY
ORDER BY
    TOTAL DESC
LIMIT
    1
```

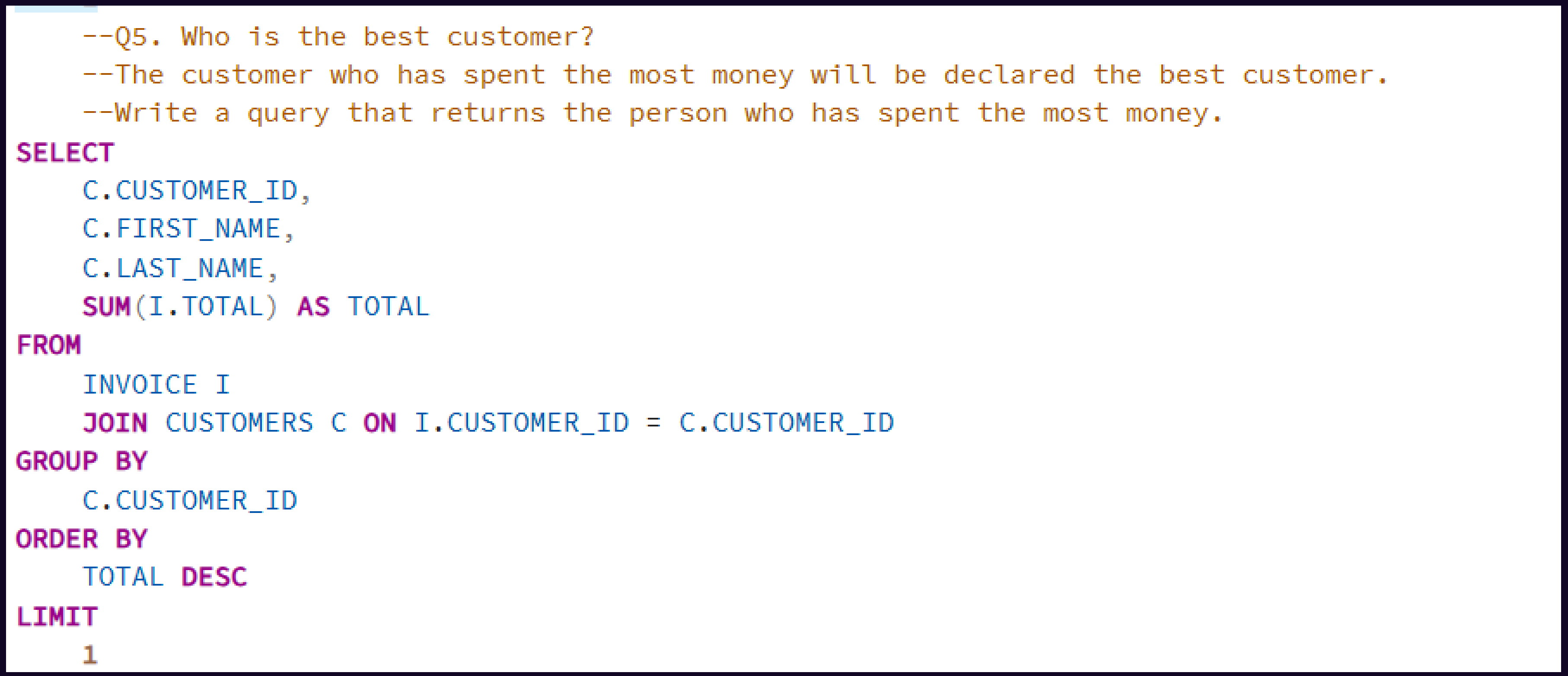
❑ Output



	total double precision	billing_city character varying (250)
1	273.24000000000007	Prague

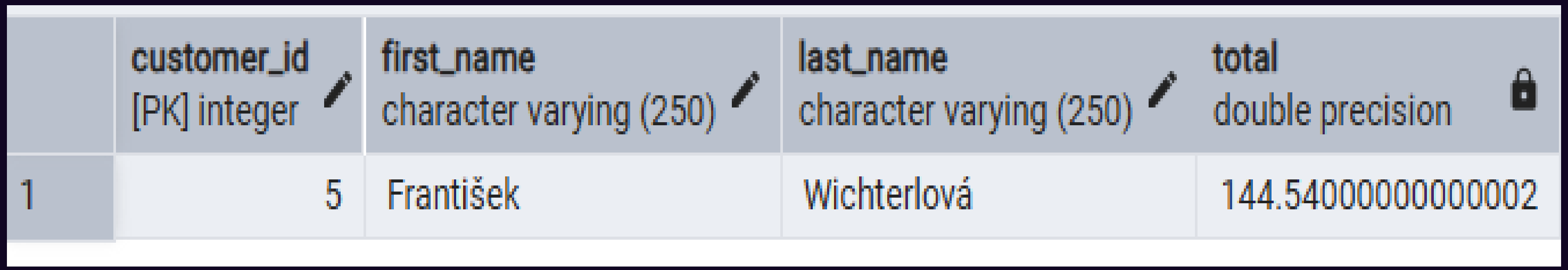
KEY QUESTION & ANALYSIS

- ❑ Q5. Who is the best customer? The customer who has spent the most money will be declared the best customer. Write a query that returns the person who has spent the most money.



```
    --Q5. Who is the best customer?
    --The customer who has spent the most money will be declared the best customer.
    --Write a query that returns the person who has spent the most money.
SELECT
    C.CUSTOMER_ID,
    C.FIRST_NAME,
    C.LAST_NAME,
    SUM(I.TOTAL) AS TOTAL
FROM
    INVOICE I
    JOIN CUSTOMERS C ON I.CUSTOMER_ID = C.CUSTOMER_ID
GROUP BY
    C.CUSTOMER_ID
ORDER BY
    TOTAL DESC
LIMIT
    1
```

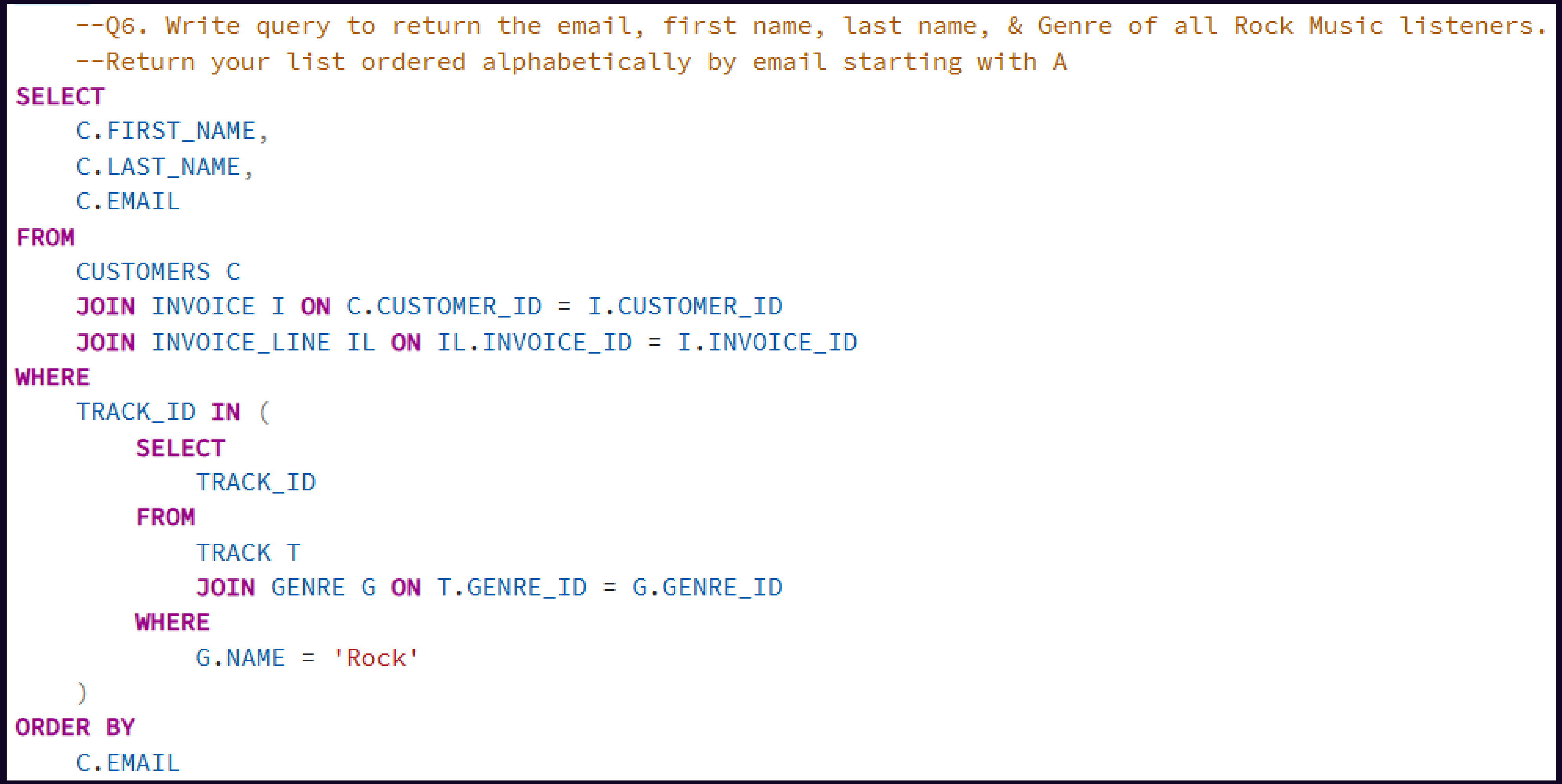
- ❑ Output



	customer_id [PK] integer	first_name character varying (250)	last_name character varying (250)	total double precision
1	5	František	Wichterlová	144.54000000000002

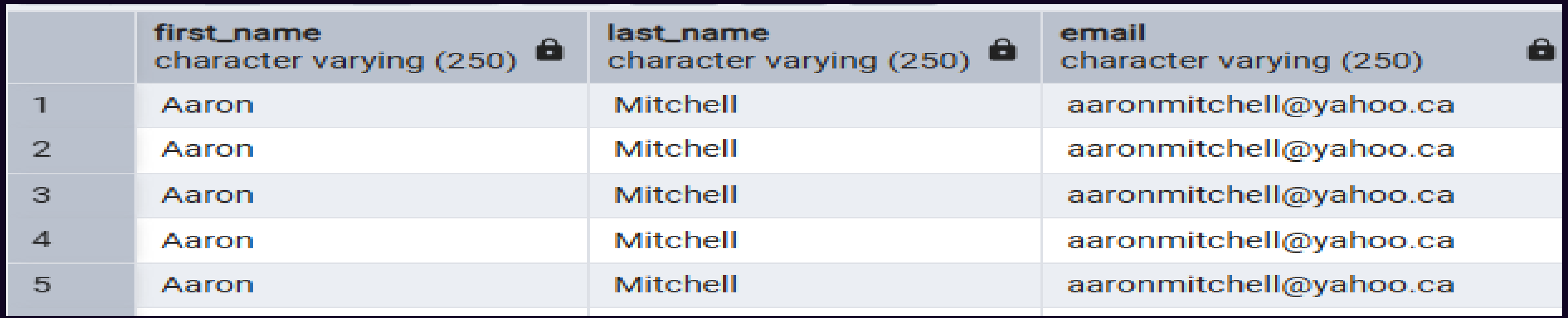
KEY QUESTION & ANALYSIS

- ❑ Q6. Write query to return the email, first name, last name, & Genre of all Rock Music listeners. Return your list ordered alphabetically by email starting with A



```
    --Q6. Write query to return the email, first name, last name, & Genre of all Rock Music listeners.
    --Return your list ordered alphabetically by email starting with A
SELECT
    C.FIRST_NAME,
    C.LAST_NAME,
    C.EMAIL
FROM
    CUSTOMERS C
    JOIN INVOICE I ON C.CUSTOMER_ID = I.CUSTOMER_ID
    JOIN INVOICE_LINE IL ON IL.INVOICE_ID = I.INVOICE_ID
WHERE
    TRACK_ID IN (
        SELECT
            TRACK_ID
        FROM
            TRACK T
            JOIN GENRE G ON T.GENRE_ID = G.GENRE_ID
        WHERE
            G.NAME = 'Rock'
    )
ORDER BY
    C.EMAIL
```

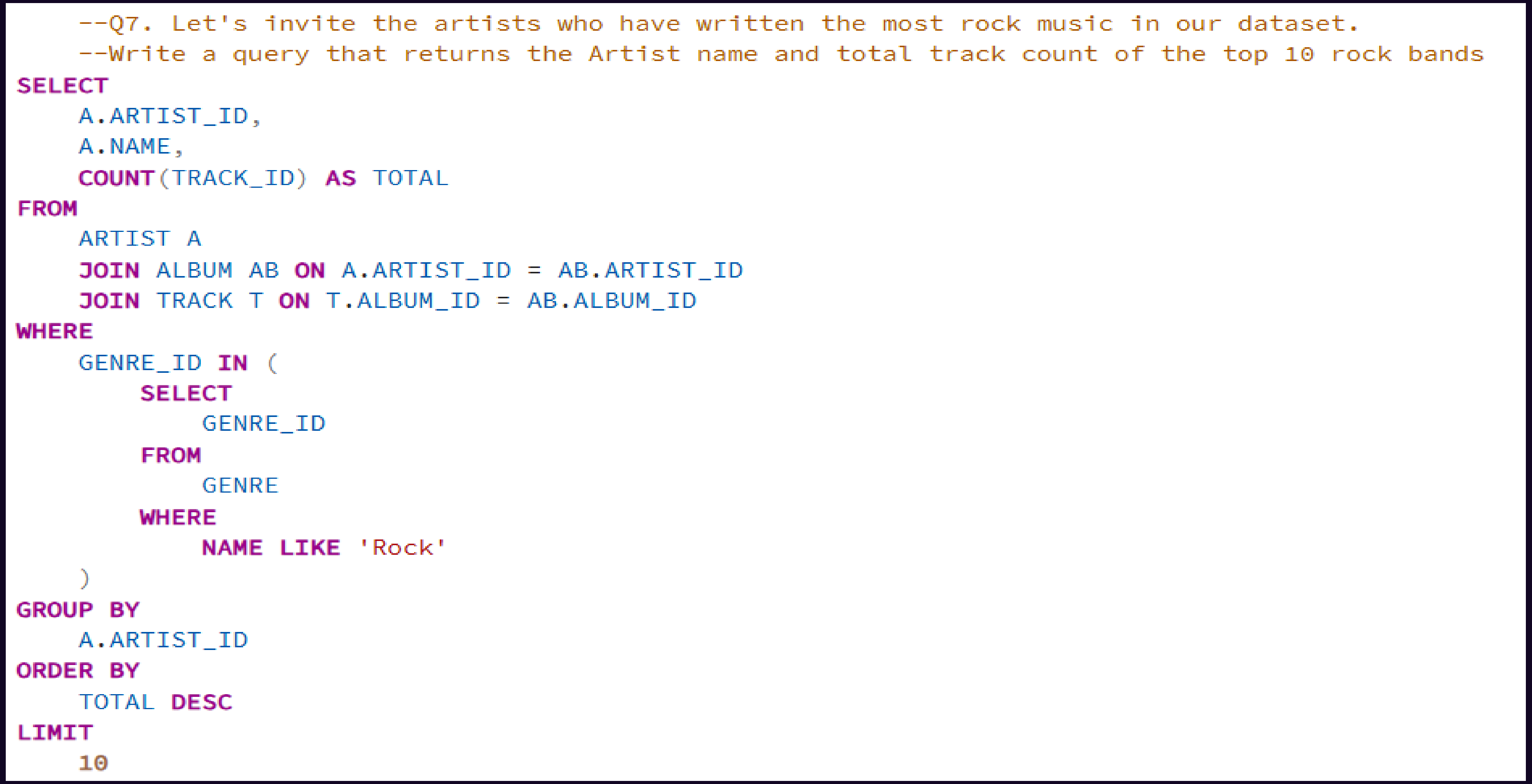
- ❑ Output



	first_name character varying (250)	last_name character varying (250)	email character varying (250)
1	Aaron	Mitchell	aaronmitchell@yahoo.ca
2	Aaron	Mitchell	aaronmitchell@yahoo.ca
3	Aaron	Mitchell	aaronmitchell@yahoo.ca
4	Aaron	Mitchell	aaronmitchell@yahoo.ca
5	Aaron	Mitchell	aaronmitchell@yahoo.ca

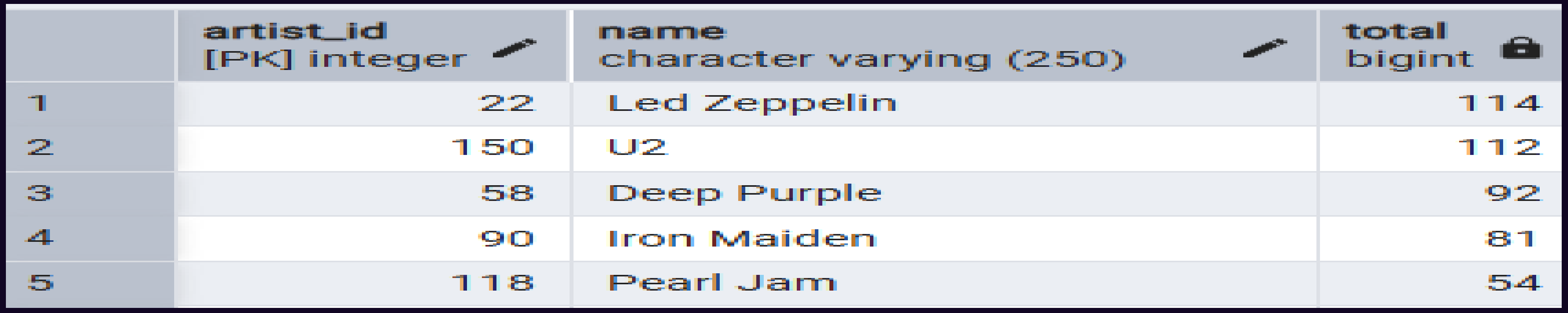
KEY QUESTION & ANALYSIS

❑ Q7. Let's invite the artists who have written the most rock music in our dataset. Write a query that returns the Artist name and total track count of the top 10 rock bands



```
    --Q7. Let's invite the artists who have written the most rock music in our dataset.
    --Write a query that returns the Artist name and total track count of the top 10 rock bands
SELECT
    A.ARTIST_ID,
    A.NAME,
    COUNT(TRACK_ID) AS TOTAL
FROM
    ARTIST A
    JOIN ALBUM AB ON A.ARTIST_ID = AB.ARTIST_ID
    JOIN TRACK T ON T.ALBUM_ID = AB.ALBUM_ID
WHERE
    GENRE_ID IN (
        SELECT
            GENRE_ID
        FROM
            GENRE
        WHERE
            NAME LIKE 'Rock'
    )
GROUP BY
    A.ARTIST_ID
ORDER BY
    TOTAL DESC
LIMIT
    10
```

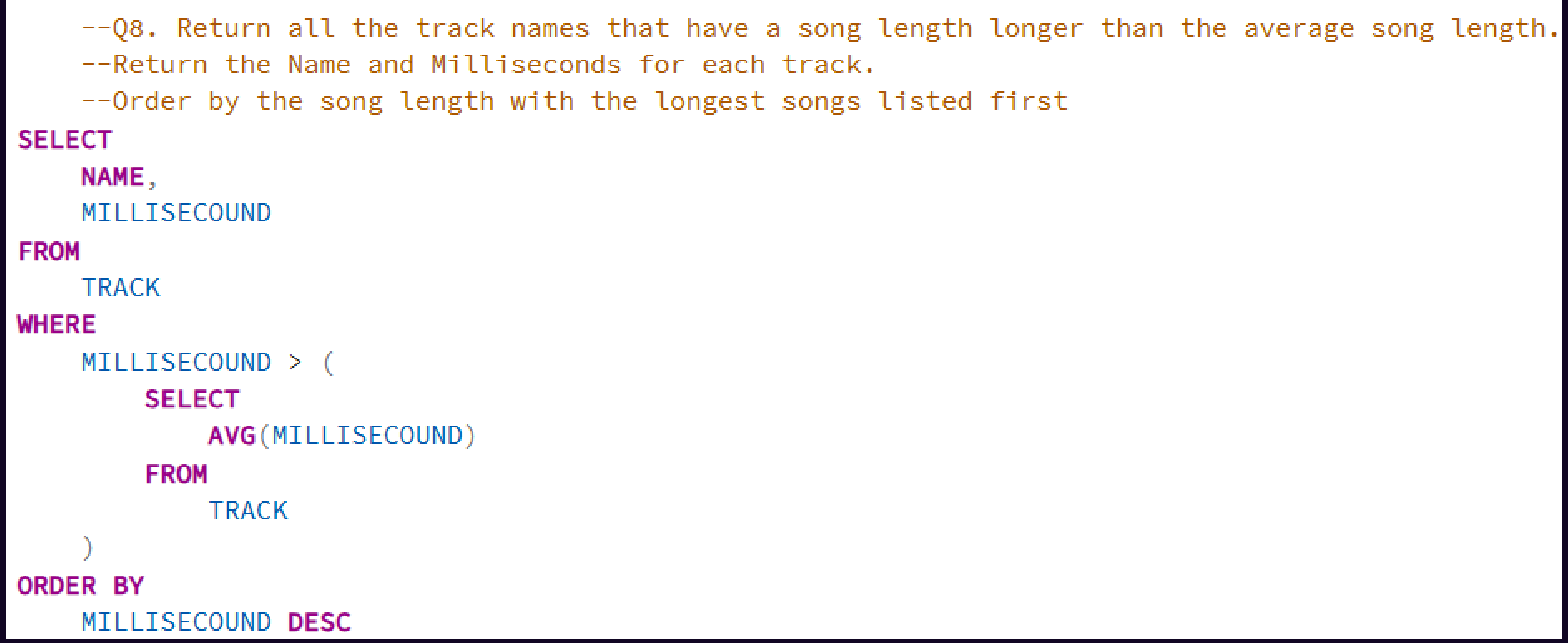
❑ Output



	artist_id [PK] integer	name character varying (250)	total bigint
1	22	Led Zeppelin	114
2	150	U2	112
3	58	Deep Purple	92
4	90	Iron Maiden	81
5	118	Pearl Jam	54

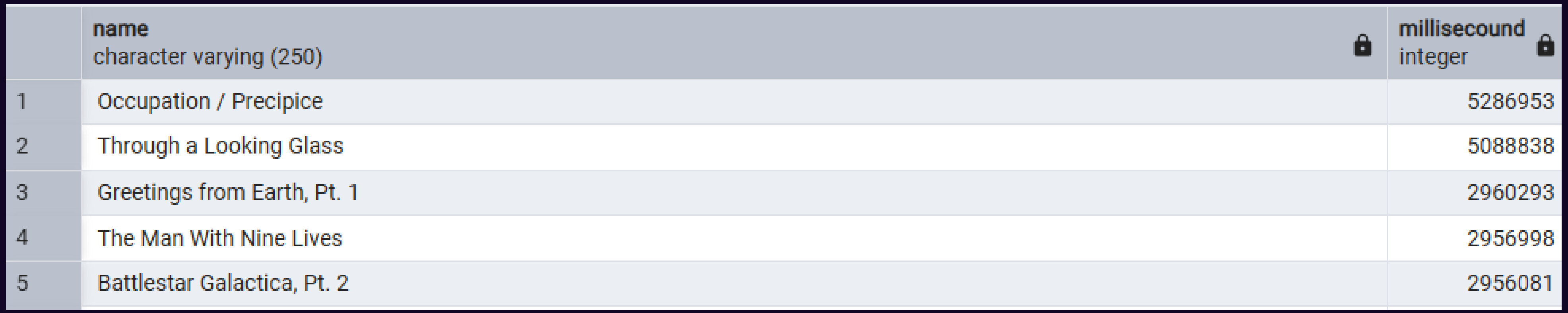
KEY QUESTION & ANALYSIS

❑ Q8. Return all the track names that have a song length longer than the average song length. Return the Name and Milliseconds for each track. Order by the song length with the longest songs listed first



```
    --Q8. Return all the track names that have a song length longer than the average song length.
    --Return the Name and Milliseconds for each track.
    --Order by the song length with the longest songs listed first
SELECT
    NAME,
    MILLISECOUND
FROM
    TRACK
WHERE
    MILLISECOUND > (
        SELECT
            AVG(MILLISECOUND)
        FROM
            TRACK
    )
ORDER BY
    MILLISECOUND DESC
```

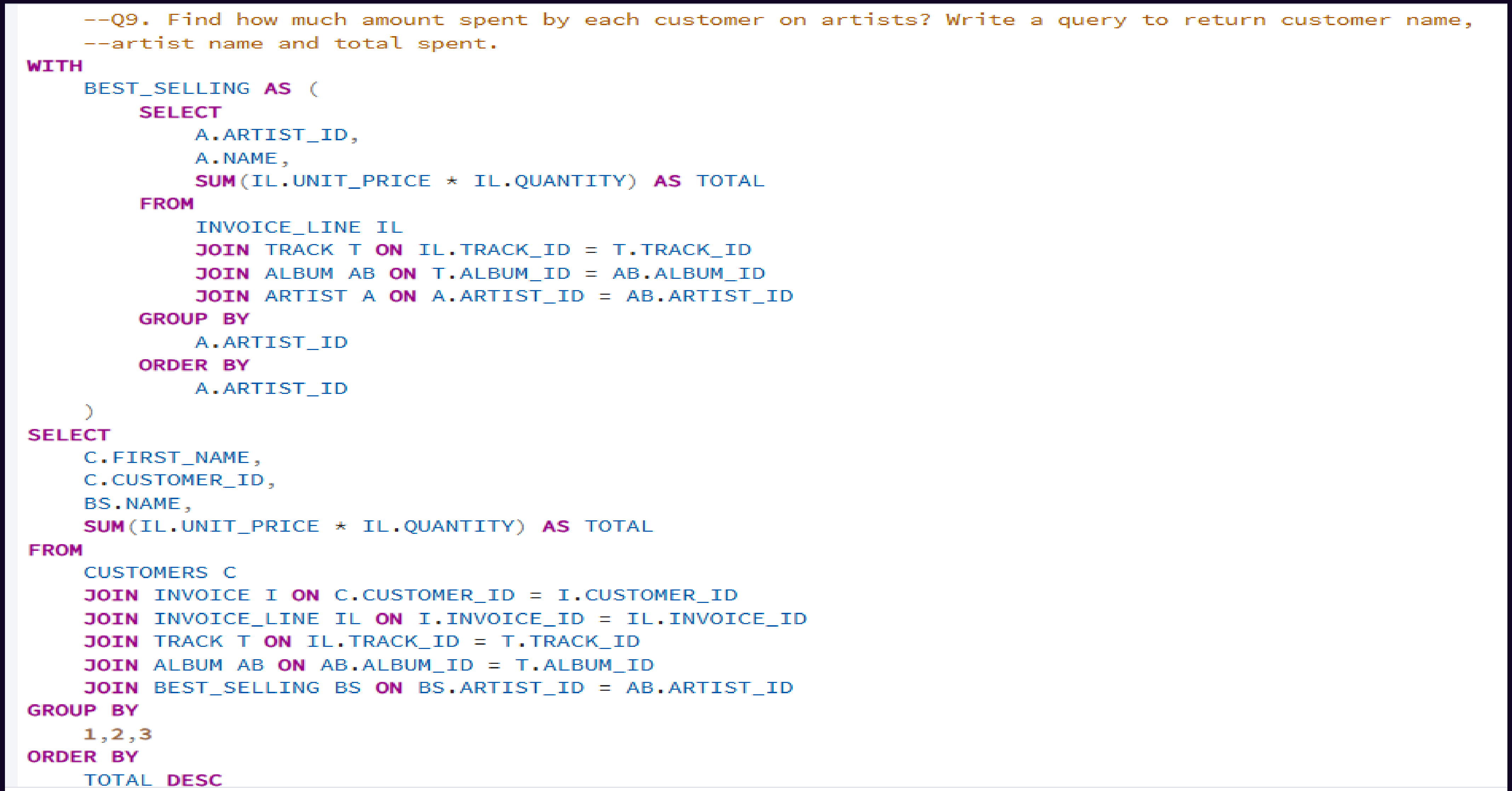
❑ Output



	name character varying (250)	millisecound integer
1	Occupation / Precipice	5286953
2	Through a Looking Glass	5088838
3	Greetings from Earth, Pt. 1	2960293
4	The Man With Nine Lives	2956998
5	Battlestar Galactica, Pt. 2	2956081

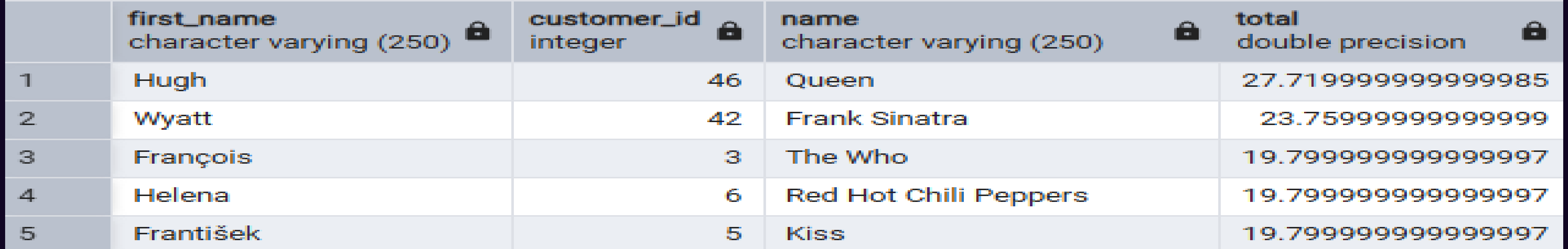
KEY QUESTION & ANALYSIS

❑ Q9. Find how much amount spent by each customer on artists? Write a query to return customer name, artist name and total spent



```
    --Q9. Find how much amount spent by each customer on artists? Write a query to return customer name,
    --artist name and total spent.
WITH
    BEST_SELLING AS (
        SELECT
            A.ARTIST_ID,
            A.NAME,
            SUM(IL.UNIT_PRICE * IL.QUANTITY) AS TOTAL
        FROM
            INVOICE_LINE IL
            JOIN TRACK T ON IL.TRACK_ID = T.TRACK_ID
            JOIN ALBUM AB ON T.ALBUM_ID = AB.ALBUM_ID
            JOIN ARTIST A ON A.ARTIST_ID = AB.ARTIST_ID
        GROUP BY
            A.ARTIST_ID
        ORDER BY
            A.ARTIST_ID
    )
SELECT
    C.FIRST_NAME,
    C.CUSTOMER_ID,
    BS.NAME,
    SUM(IL.UNIT_PRICE * IL.QUANTITY) AS TOTAL
FROM
    CUSTOMERS C
    JOIN INVOICE I ON C.CUSTOMER_ID = I.CUSTOMER_ID
    JOIN INVOICE_LINE IL ON I.INVOICE_ID = IL.INVOICE_ID
    JOIN TRACK T ON IL.TRACK_ID = T.TRACK_ID
    JOIN ALBUM AB ON AB.ALBUM_ID = T.ALBUM_ID
    JOIN BEST_SELLING BS ON BS.ARTIST_ID = AB.ARTIST_ID
GROUP BY
    1,2,3
ORDER BY
    TOTAL DESC
```

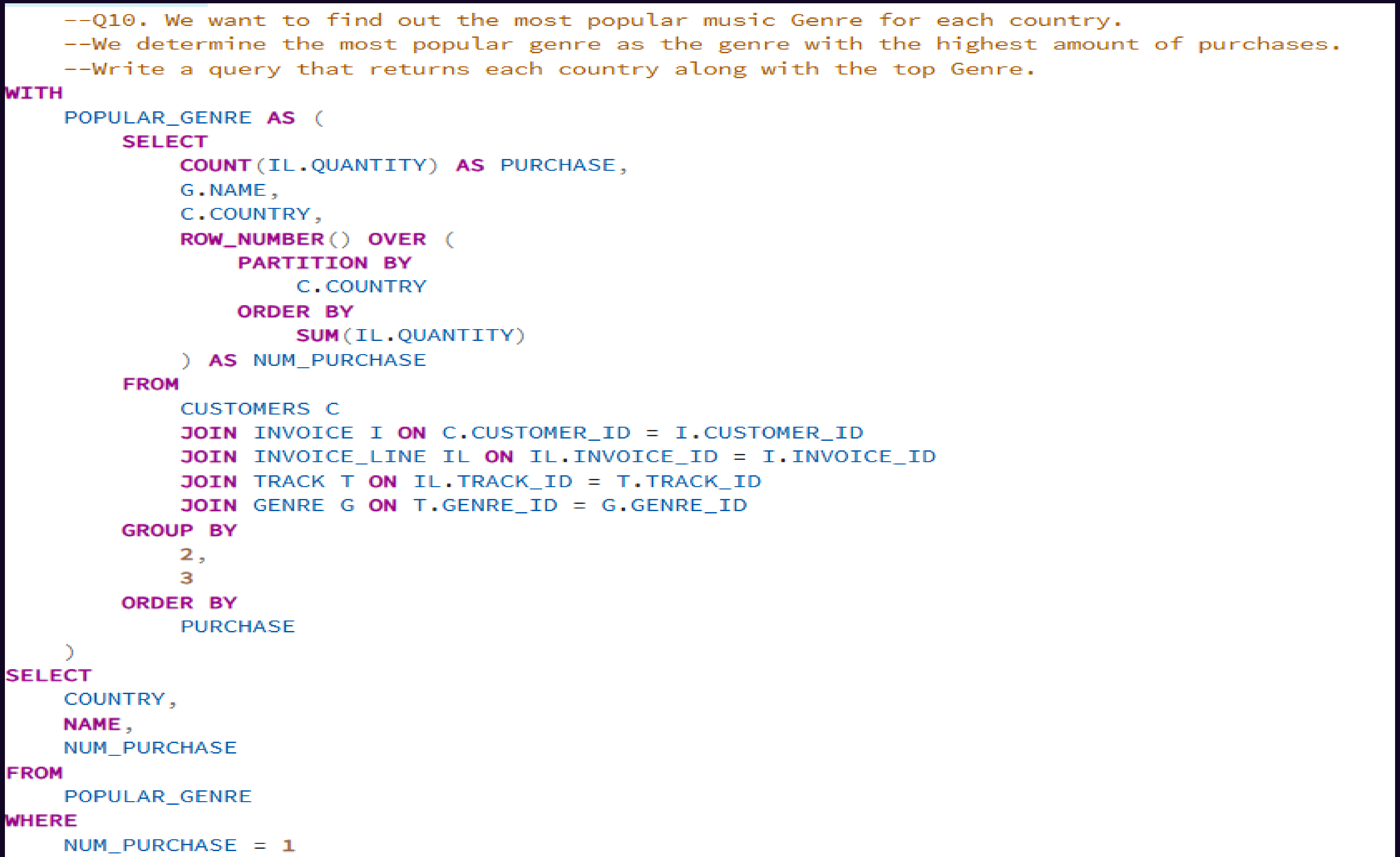
❑ Output



	first_name character varying (250)	customer_id integer	name character varying (250)	total double precision
1	Hugh	46	Queen	27.719999999999985
2	Wyatt	42	Frank Sinatra	23.75999999999999
3	François	3	The Who	19.799999999999997
4	Helena	6	Red Hot Chili Peppers	19.799999999999997
5	František	5	Kiss	19.799999999999997

KEY QUESTION & ANALYSIS

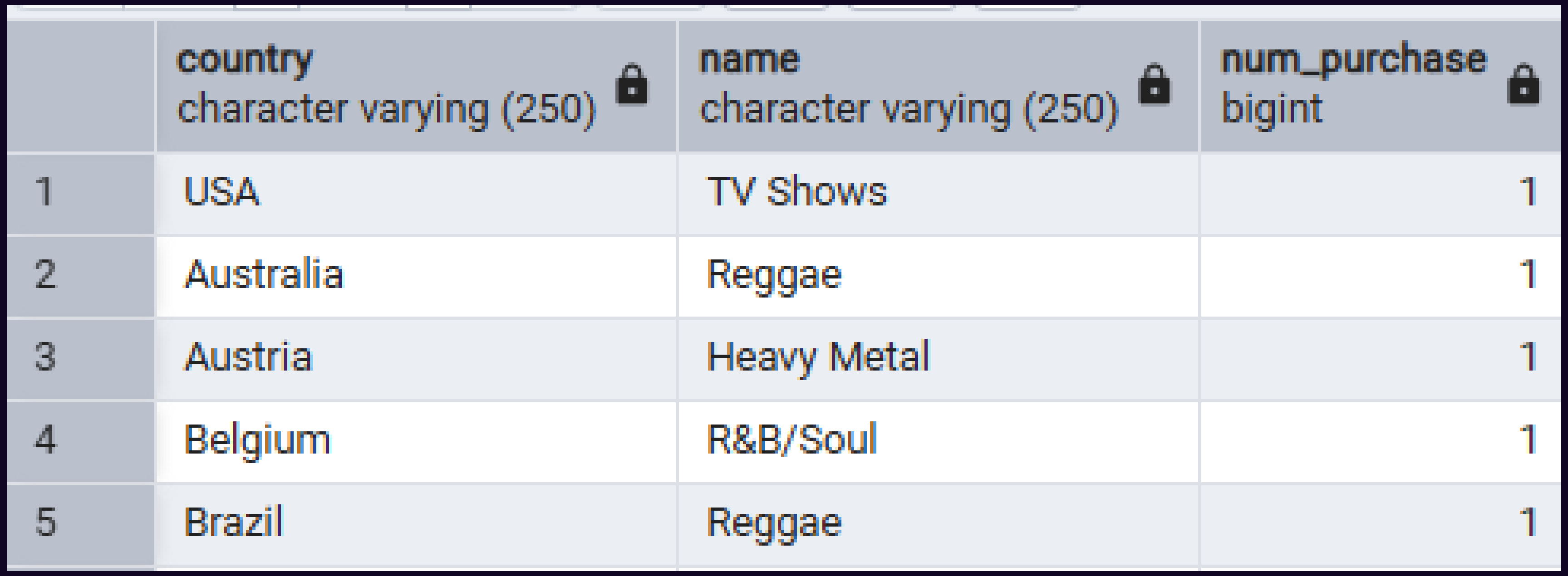
- ❑ Q10. We want to find out the most popular music Genre for each country. We determine the most popular genre as the genre with the highest amount of purchases. Write a query that returns each country along with the top Genre. For countries where the maximum number of purchases is shared return all Genres.



```
    --Q10. We want to find out the most popular music Genre for each country.
    --We determine the most popular genre as the genre with the highest amount of purchases.
    --Write a query that returns each country along with the top Genre.
WITH
    POPULAR_GENRE AS (
        SELECT
            COUNT(IL.QUANTITY) AS PURCHASE,
            G.NAME,
            C.COUNTRY,
            ROW_NUMBER() OVER (
                PARTITION BY
                    C.COUNTRY
                ORDER BY
                    SUM(IL.QUANTITY)
            ) AS NUM_PURCHASE
        FROM
            CUSTOMERS C
            JOIN INVOICE I ON C.CUSTOMER_ID = I.CUSTOMER_ID
            JOIN INVOICE_LINE IL ON IL.INVOICE_ID = I.INVOICE_ID
            JOIN TRACK T ON IL.TRACK_ID = T.TRACK_ID
            JOIN GENRE G ON T.GENRE_ID = G.GENRE_ID
        GROUP BY
            2,
            3
        ORDER BY
            PURCHASE
    )
SELECT
    COUNTRY,
    NAME,
    NUM_PURCHASE
FROM
    POPULAR_GENRE
WHERE
    NUM_PURCHASE = 1
```

KEY QUESTION & ANALYSIS

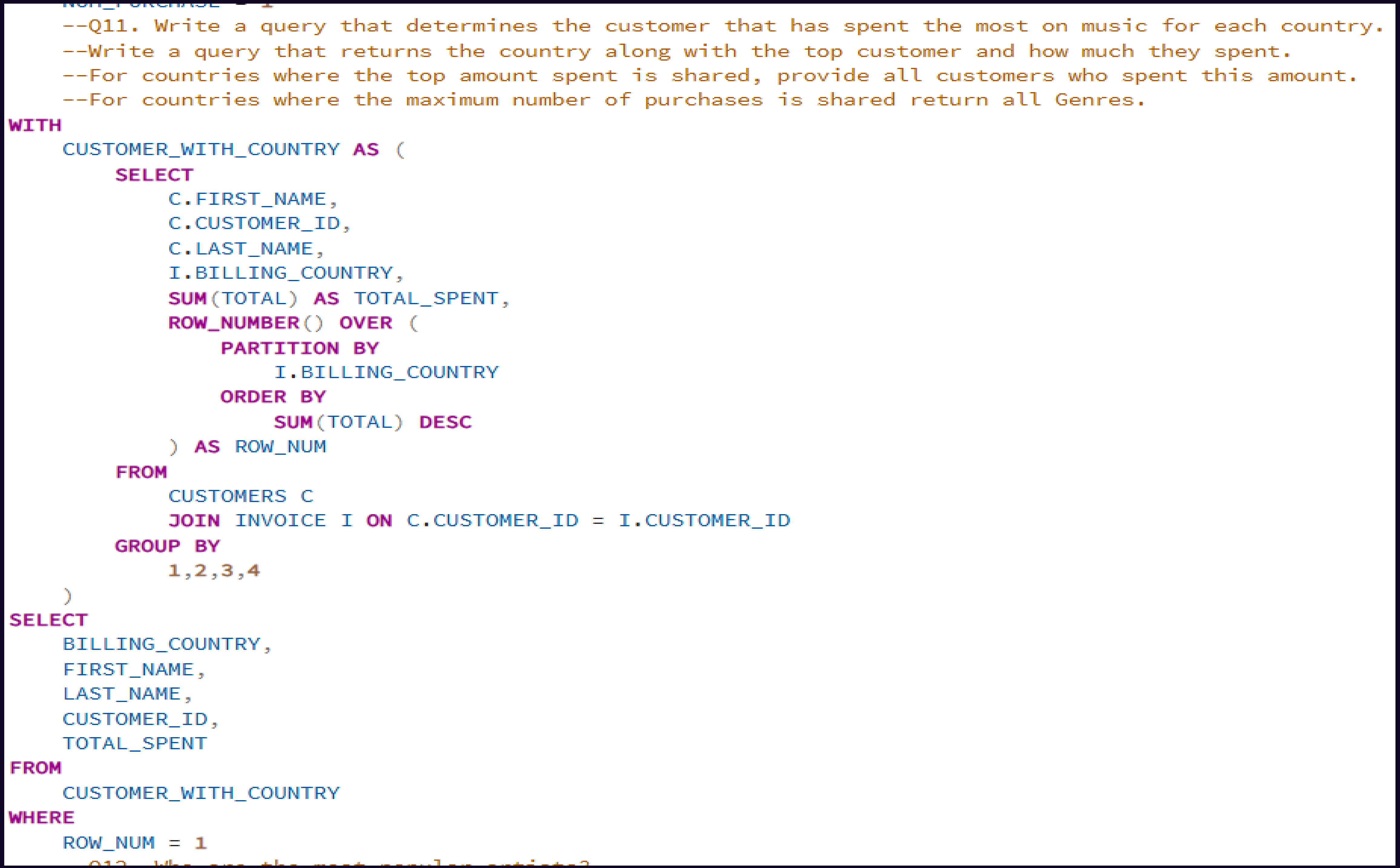
- ❑ Output



	country character varying (250)	name character varying (250)	num_purchase bigint
1	USA	TV Shows	1
2	Australia	Reggae	1
3	Austria	Heavy Metal	1
4	Belgium	R&B/Soul	1
5	Brazil	Reggae	1

KEY QUESTION & ANALYSIS

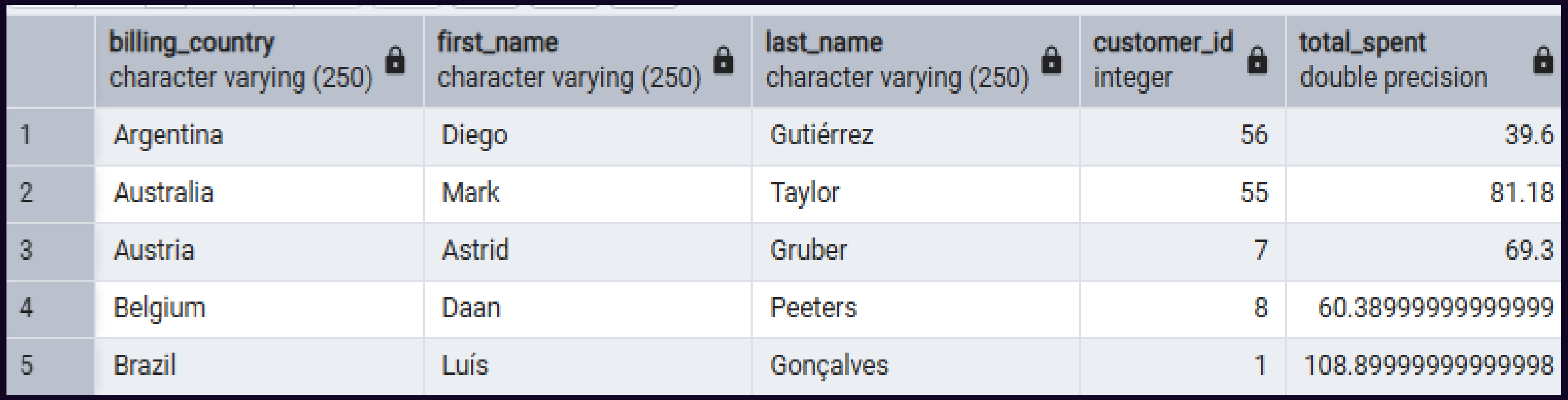
❑ Q11. Write a query that determines the customer that has spent the most on music for each country. Write a query that returns the country along with the top customer and how much they spent. For countries where the top amount spent is shared, provide all customers who spent this amount..



```
    NUM_PURCHASE = 1
    --Q11. Write a query that determines the customer that has spent the most on music for each country.
    --Write a query that returns the country along with the top customer and how much they spent.
    --For countries where the top amount spent is shared, provide all customers who spent this amount.
    --For countries where the maximum number of purchases is shared return all Genres.
WITH
    CUSTOMER_WITH_COUNTRY AS (
        SELECT
            C.FIRST_NAME,
            C.CUSTOMER_ID,
            C.LAST_NAME,
            I.BILLING_COUNTRY,
            SUM(TOTAL) AS TOTAL_SPENT,
            ROW_NUMBER() OVER (
                PARTITION BY
                    I.BILLING_COUNTRY
                ORDER BY
                    SUM(TOTAL) DESC
            ) AS ROW_NUM
        FROM
            CUSTOMERS C
            JOIN INVOICE I ON C.CUSTOMER_ID = I.CUSTOMER_ID
        GROUP BY
            1,2,3,4
    )
SELECT
    BILLING_COUNTRY,
    FIRST_NAME,
    LAST_NAME,
    CUSTOMER_ID,
    TOTAL_SPENT
FROM
    CUSTOMER_WITH_COUNTRY
WHERE
    ROW_NUM = 1
    --Q12. Who are the most popular artists?
```

KEY QUESTION & ANALYSIS

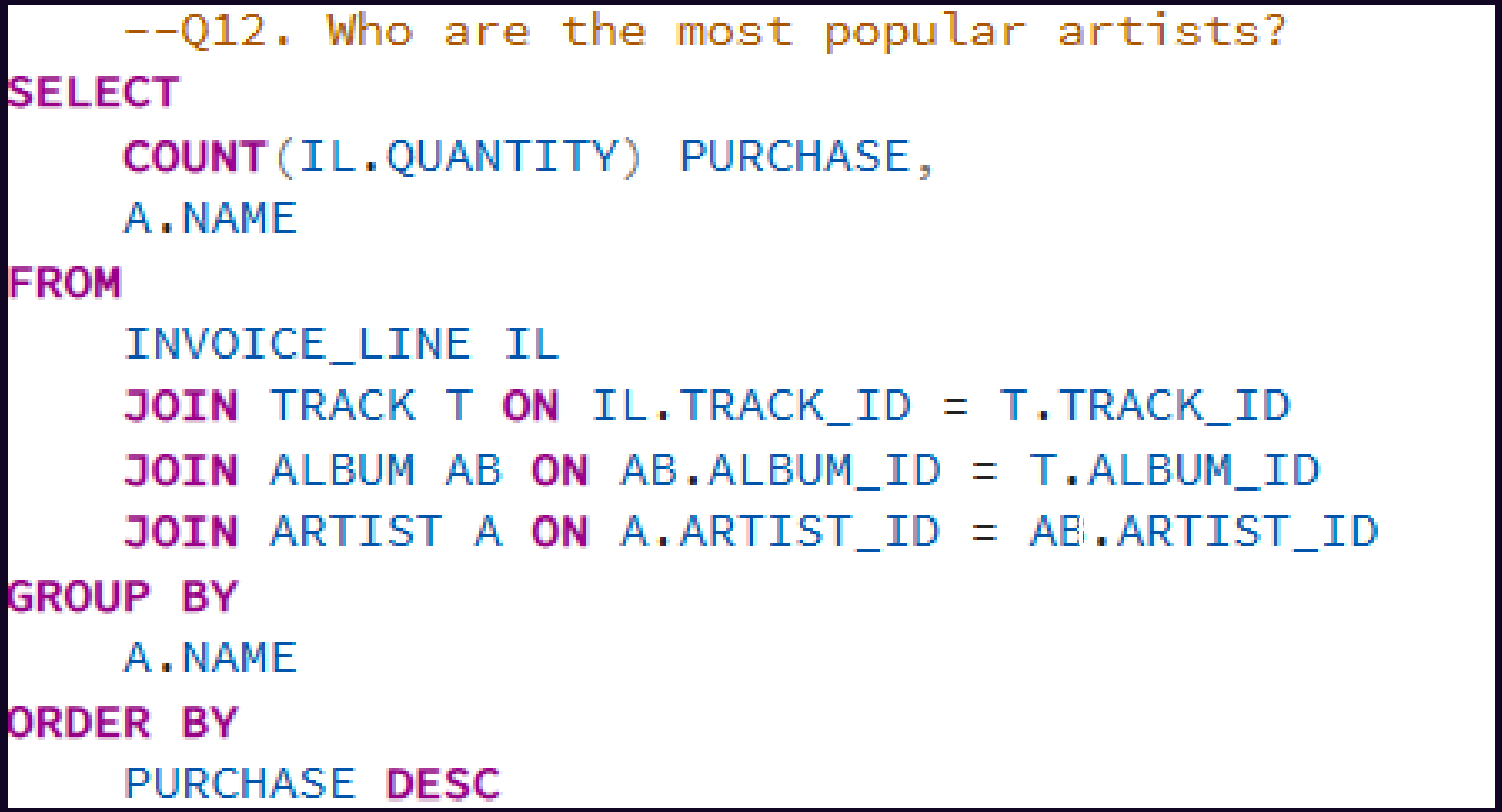
❑ Output



	billing_country character varying (250)	first_name character varying (250)	last_name character varying (250)	customer_id integer	total_spent double precision
1	Argentina	Diego	Gutiérrez	56	39.6
2	Australia	Mark	Taylor	55	81.18
3	Austria	Astrid	Gruber	7	69.3
4	Belgium	Daan	Peeters	8	60.38999999999999
5	Brazil	Luís	Gonçalves	1	108.89999999999998

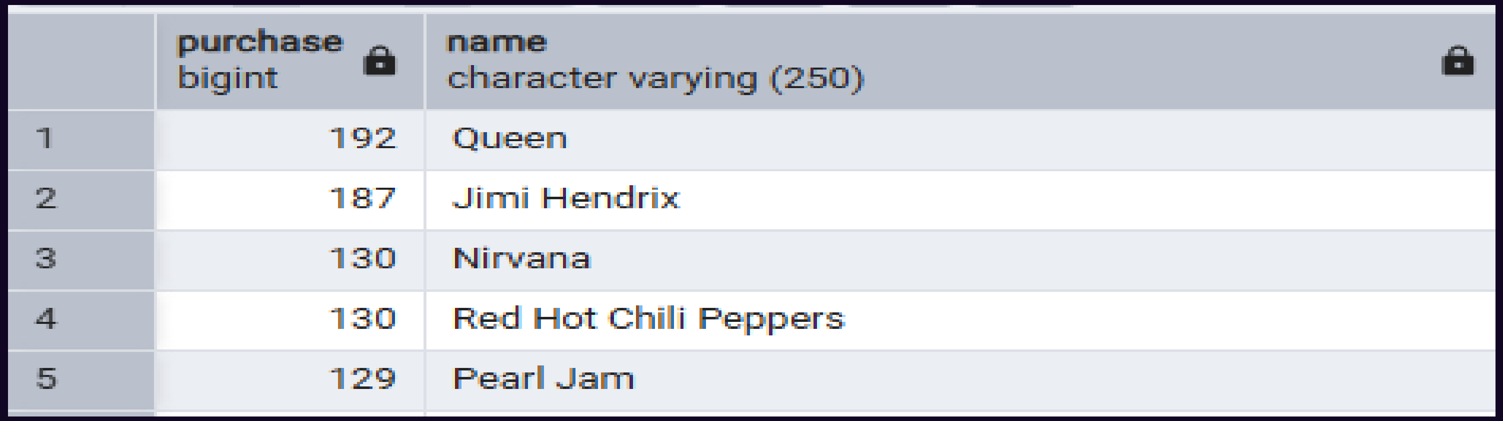
KEY QUESTION & ANALYSIS

❑ Q12. Who are the most popular artists?



```
    --Q12. Who are the most popular artists?
SELECT
    COUNT(IL.QUANTITY) PURCHASE,
    A.NAME
FROM
    INVOICE_LINE IL
    JOIN TRACK T ON IL.TRACK_ID = T.TRACK_ID
    JOIN ALBUM AB ON AB.ALBUM_ID = T.ALBUM_ID
    JOIN ARTIST A ON A.ARTIST_ID = AB.ARTIST_ID
GROUP BY
    A.NAME
ORDER BY
    PURCHASE DESC
```

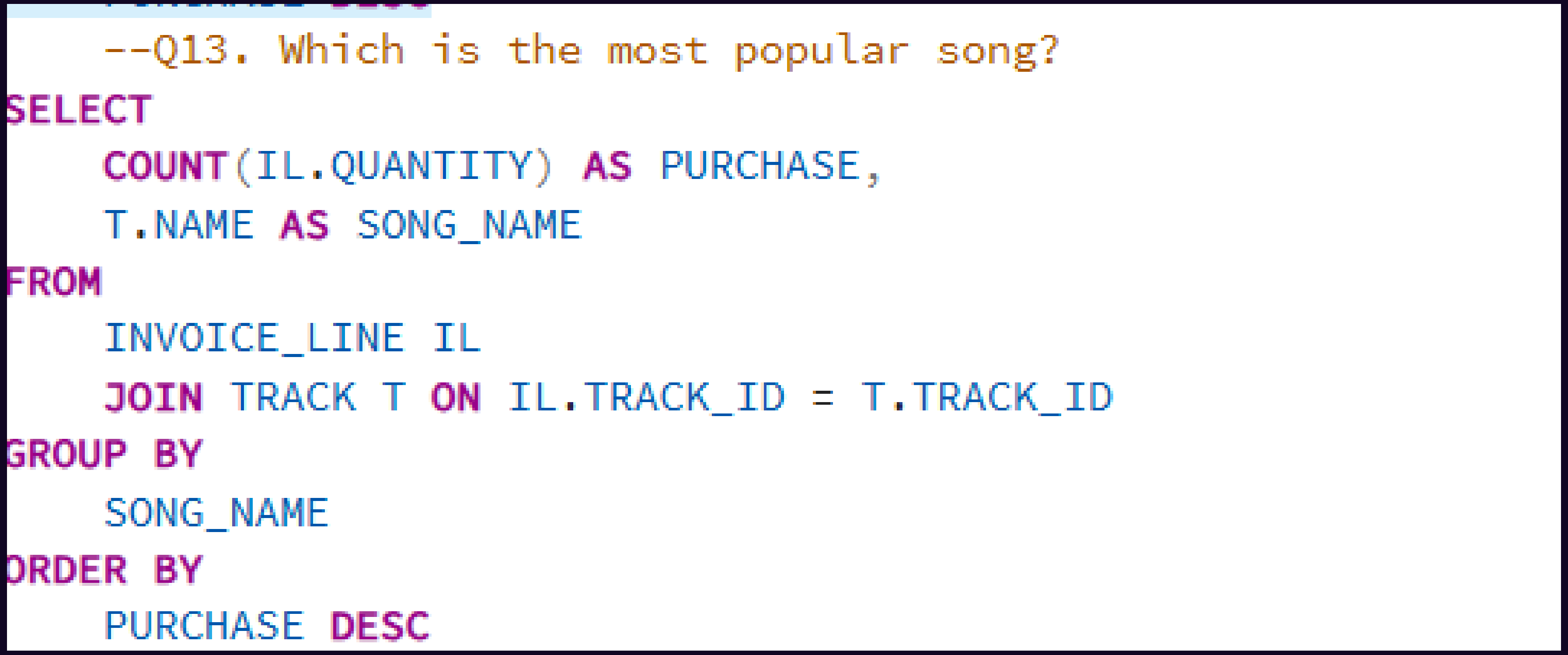
❑ Output



	purchase	name
	bigint	character varying (250)
1	192	Queen
2	187	Jimi Hendrix
3	130	Nirvana
4	130	Red Hot Chili Peppers
5	129	Pearl Jam

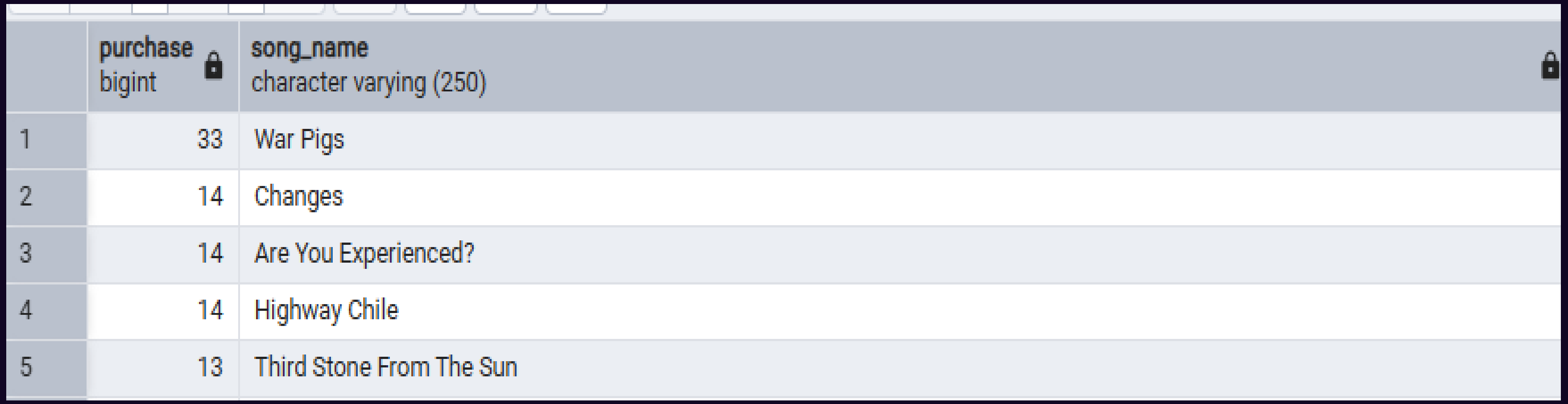
KEY QUESTION & ANALYSIS

- ❑ Q13. Which is the most popular song?



```
    --Q13. Which is the most popular song?
SELECT
    COUNT(IL.QUANTITY) AS PURCHASE,
    T.NAME AS SONG_NAME
FROM
    INVOICE_LINE IL
    JOIN TRACK T ON IL.TRACK_ID = T.TRACK_ID
GROUP BY
    SONG_NAME
ORDER BY
    PURCHASE DESC
```

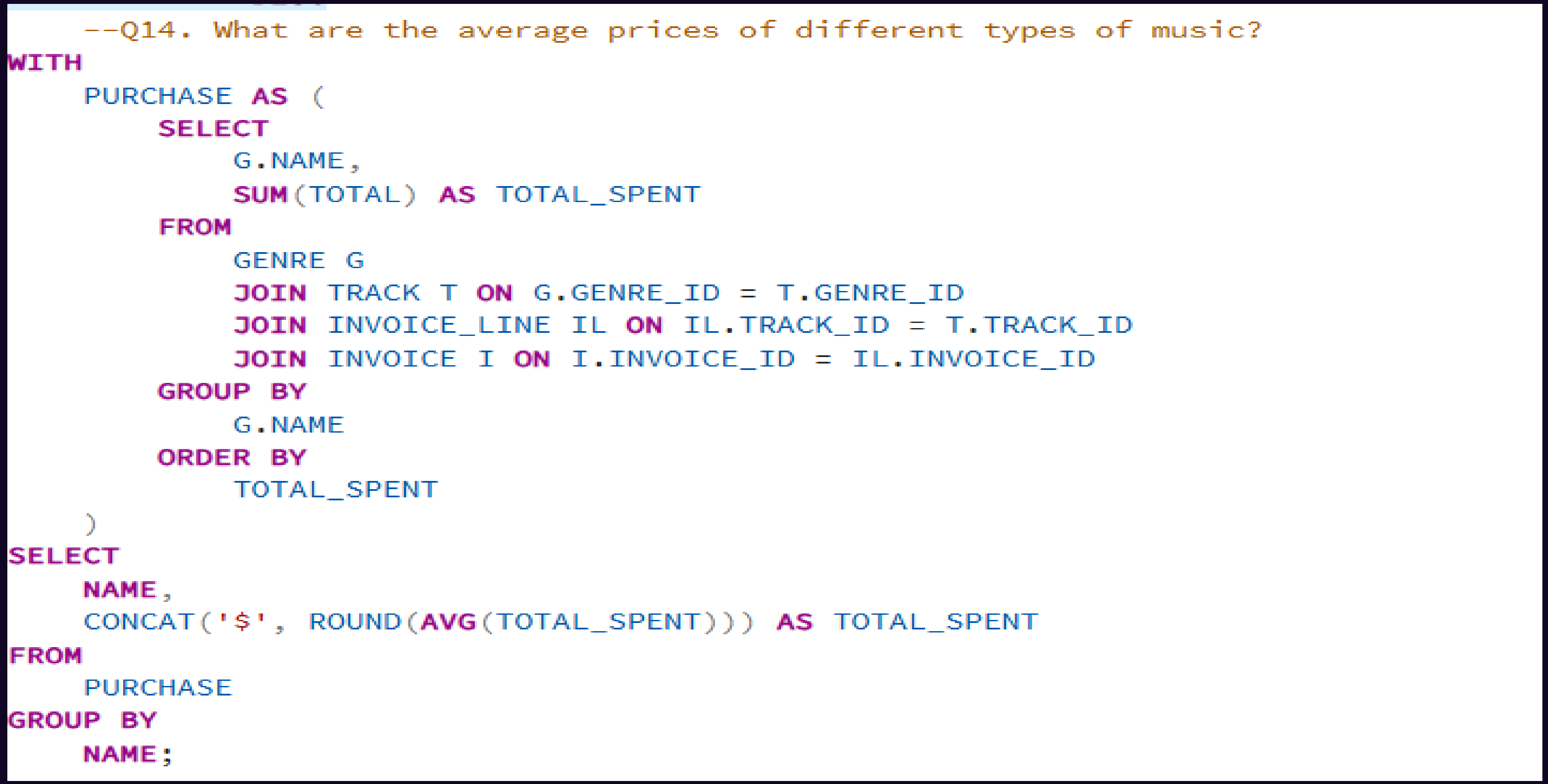
- ❑ Output



	purchase bigint	song_name character varying (250)
1	33	War Pigs
2	14	Changes
3	14	Are You Experienced?
4	14	Highway Chile
5	13	Third Stone From The Sun

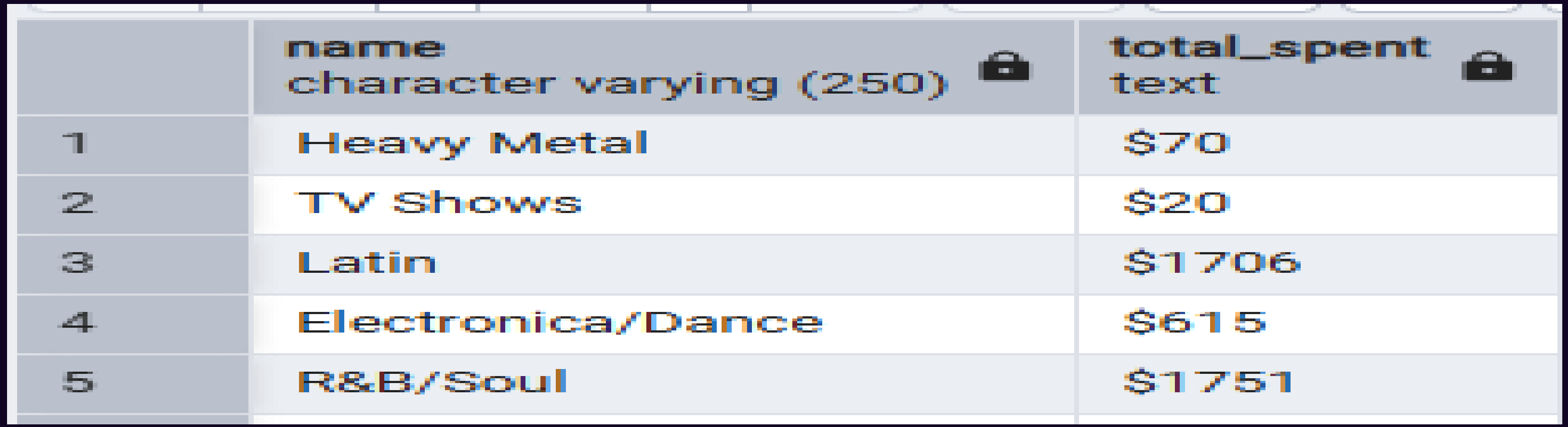
KEY QUESTION & ANALYSIS

❑ Q14. What are the average prices of different types of music?



```
    --Q14. What are the average prices of different types of music?
WITH
    PURCHASE AS (
        SELECT
            G.NAME,
            SUM(TOTAL) AS TOTAL_SPENT
        FROM
            GENRE G
            JOIN TRACK T ON G.GENRE_ID = T.GENRE_ID
            JOIN INVOICE_LINE IL ON IL.TRACK_ID = T.TRACK_ID
            JOIN INVOICE I ON I.INVOICE_ID = IL.INVOICE_ID
        GROUP BY
            G.NAME
        ORDER BY
            TOTAL_SPENT
    )
SELECT
    NAME,
    CONCAT('$', ROUND(AVG(TOTAL_SPENT))) AS TOTAL_SPENT
FROM
    PURCHASE
GROUP BY
    NAME;
```

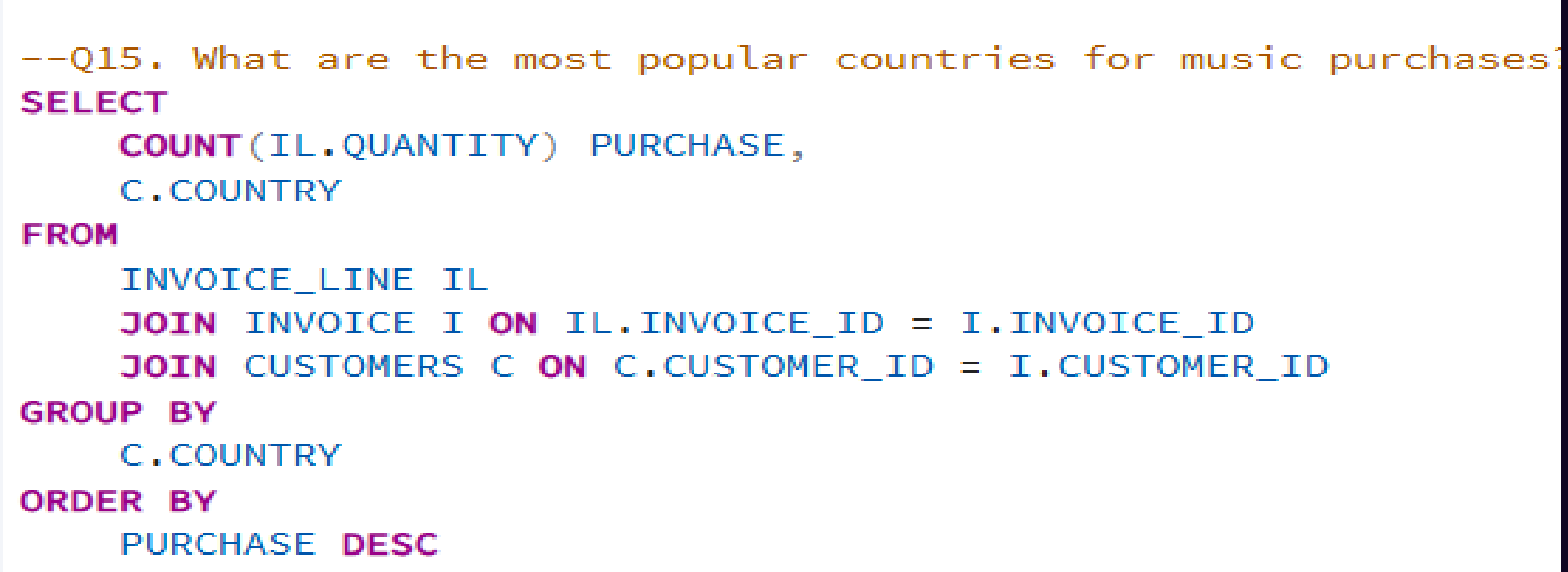
❑ Output



	name character varying (250) 🔒	total_spent text 🔒
1	Heavy Metal	$70
2	TV Shows	$20
3	Latin	$1706
4	Electronica/Dance	$615
5	R&B/Soul	$1751

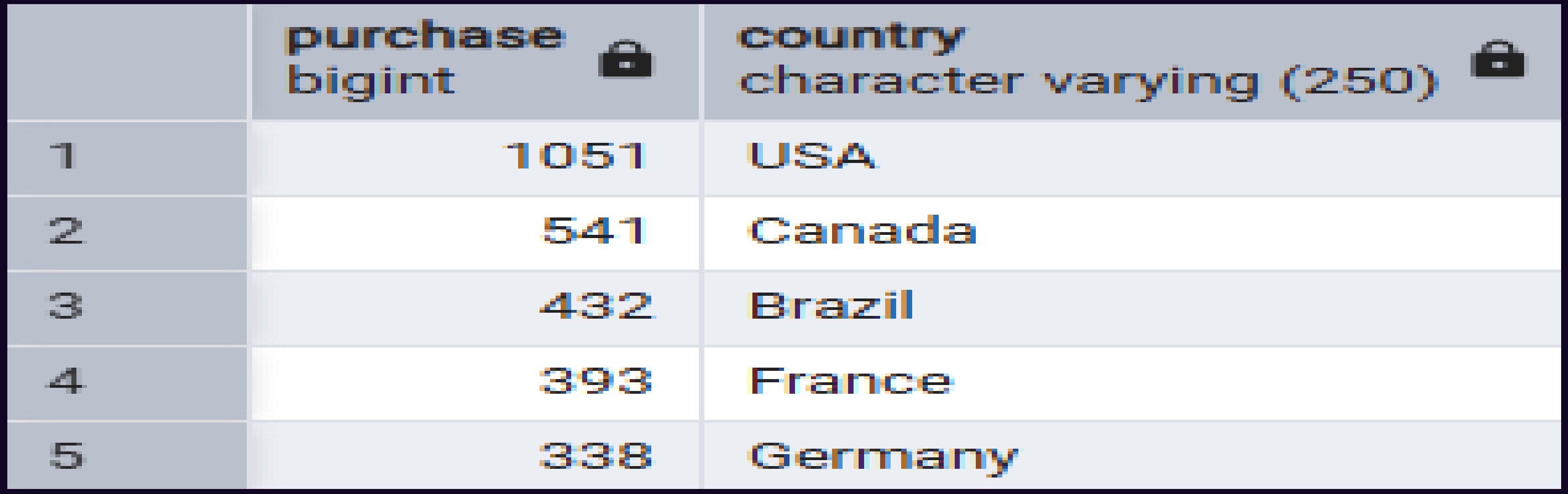
KEY QUESTION & ANALYSIS

❑ Q15. What are the most popular countries for music purchases?



```
--Q15. What are the most popular countries for music purchases
SELECT
    COUNT(IL.QUANTITY) PURCHASE,
    C.COUNTRY
FROM
    INVOICE_LINE IL
    JOIN INVOICE I ON IL.INVOICE_ID = I.INVOICE_ID
    JOIN CUSTOMERS C ON C.CUSTOMER_ID = I.CUSTOMER_ID
GROUP BY
    C.COUNTRY
ORDER BY
    PURCHASE DESC
```

❑ Output



	purchase bigint	country character varying (250)
1	1051	USA
2	541	Canada
3	432	Brazil
4	393	France
5	338	Germany