SQL PROJECT

MUSIC STORE ANALYSIS

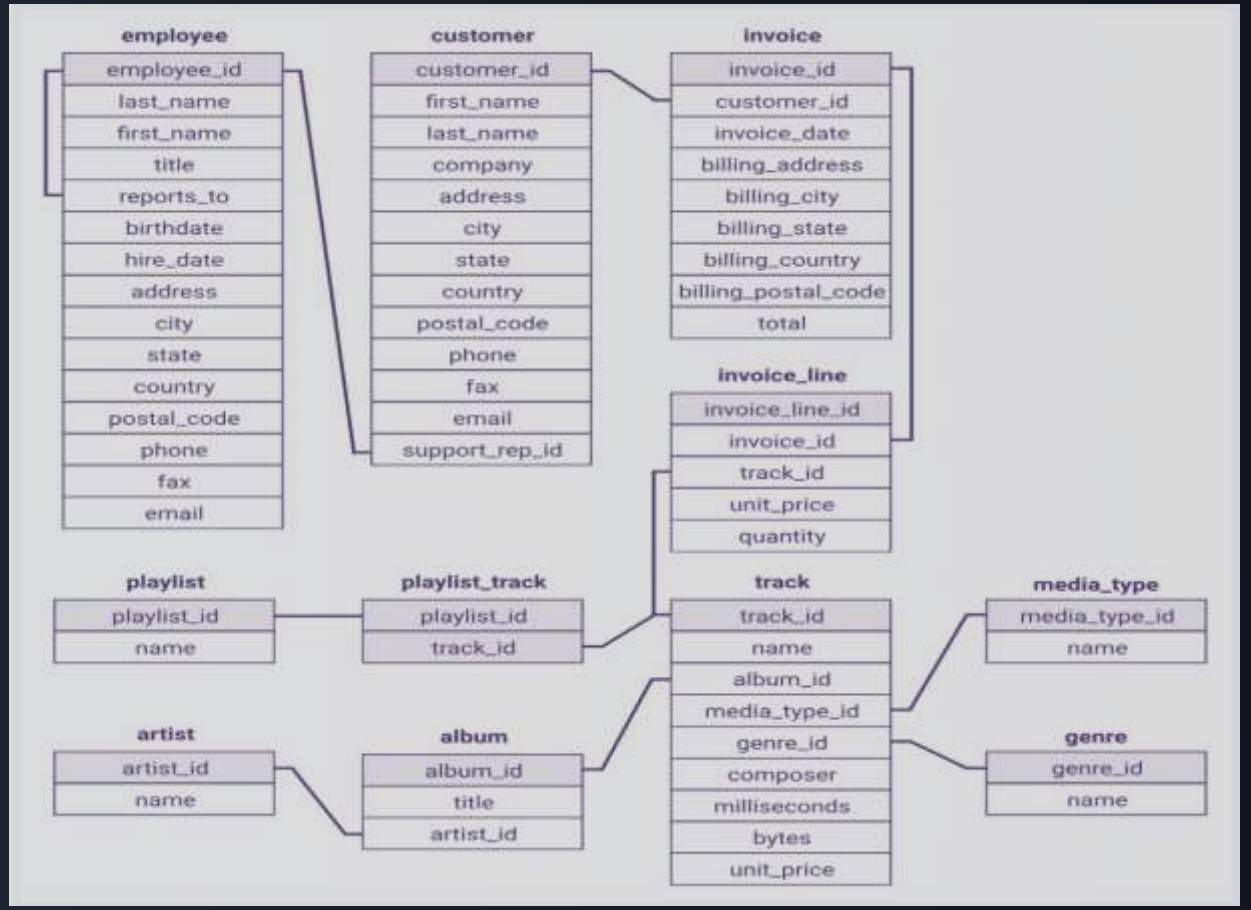
By Anish Chakravorty

OBJECTIVE

The primary objective of the music store is to achieve sustainable business growth while addressing existing challenges.

We need to examine the dataset with SQL and help the music store understand its business growth by answering simple questions.

MUSIC DATABASE SCHEME



DIVISION OF QUESTIONS

EASY

Queries Include:

SELECT, GROUP BY, ORDER BY , LIMIT, DESC

MODERATE

Queries Include:

JOINS, GROUP BY, ORDER BY , LIMIT, DESC

ADVANCE

Queries Include:

CTE (COMMON TABLE EXPRESSIONS), WINDOWS FUCTIONS

EASY QUESTIONS

Question-1

Who is the senior most employee based on job title?

```
select title, first_name,last_name, levels
from employee
order by levels desc
limit 1;
```

title character varying (50)	first_name character	last_name character	levels character varying (10)
Senior General Manager	Mohan	Madan	L7

EASY QUESTIONS

Question-2

Which countries have the most Invoices?

```
select count(*) as c, billing_country
from invoice
group by billing_country
order by c desc;
```

c bigint	billing_country character varying (30)
131	USA
76	Canada
61	Brazil
50	France
41	Germany
30	Czech Republic
29	Portugal
28	United Kingdom

EASY QUESTIONS

Question-3

What are top 3 values of total invoice?

```
select total from invoice
order by total desc
limit 3;
```

total double precision
23.759999999999998
19.8
19.8

EASY QUESTIONS

Question-4

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 invoice_total, billing_city
from invoice
group by billing_city
order by invoice_total desc
```

invoice_total double precision	billing_city character varying (30)
273.24000000000007	Prague

EASY QUESTIONS

Question-5

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(a.total) as total
from customer as c
join invoice as a on c.customer_id = a.customer_id
group by c.customer_id
order by total desc
limit 1;
```

customer_id [PK] integer	first_name character	last_name character	total double precision
5	R ...	Madhav	144.54000000000002

MODERATE QUESTIONS

Question-1

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 DISTINCT email AS Email,first_name AS FirstName, last_name AS LastName, genre.name AS Name
FROM customer
JOIN invoice ON invoice.customer_id = customer.customer_id
JOIN invoice_line ON invoice_line.invoice_id = invoice.invoice_id
JOIN track ON track.track_id = invoice_line.track_id
JOIN genre ON genre.genre_id = track.genre_id
WHERE genre.name LIKE 'Rock'
ORDER BY email;
```

email character varying (50)	firstname character	lastname character	name character varying (120)
aaronmitchell@yahoo.ca	Aaron	Mitchell	Rock
alero@uol.com.br	Alexandre	Rocha	Rock
astrid.gruber@apple.at	Astrid	Gruber	Rock
bjorn.hansen@yahoo.no	Bjørn	Hansen	Rock
camille.bernard@yahoo.fr	Camille	Bernard	Rock
daan_peeters@apple.be	Daan	Peeters	Rock
diego.gutierrez@yahoo.ar	Diego	Gutiérrez	Rock

MODERATE QUESTIONS

Question-2

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(a.artist_id) as number_of_songs
from track as t
join album as b on t.album_id = b.album_id
join artist as a on a.artist_id = b.artist_id
join genre as g on g.genre_id = t.genre_id
where g.name like 'Rock'
group by a.artist_id
order by number_of_songs desc
limit 10;
```

artist_id [PK] character varying (50)	name character varying (120)	number_of_songs bigint
22	Led Zeppelin	114
150	U2	112
58	Deep Purple	92
90	Iron Maiden	81
118	Pearl Jam	54
152	Van Halen	52
51	Queen	45

MODERATE QUESTIONS

Question-3

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,milliseconds
FROM track
WHERE milliseconds > (
    SELECT AVG(milliseconds) AS avg_track_length
    FROM track )
ORDER BY milliseconds DESC;
```

name character varying (150)	milliseconds integer
Occupation / Precipice	5286953
Through a Looking Glass	5088838
Greetings from Earth, Pt. 1	2960293
The Man With Nine Lives	2956998
Battlestar Galactica, Pt. 2	2956081
Battlestar Galactica, Pt. 1	2952702
Murder On the Rising Star	2935894
Battlestar Galactica, Pt. 3	2927802

ADVANCE QUESTIONS

Question-1

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_artist AS (
    SELECT a.artist_id AS artist_id, a.name AS artist_name, SUM(il.unit_price*il.quantity) AS total_sales
    FROM invoice_line as il
    JOIN track as t ON t.track_id = il.track_id
    JOIN album as b ON b.album_id = t.album_id
    JOIN artist as a ON a.artist_id = b.artist_id
    GROUP BY 1
    ORDER BY 3 DESC
    LIMIT 1
)
SELECT c.customer_id, c.first_name, c.last_name, bsa.artist_name, SUM(il.unit_price*il.quantity) AS amount_spent
FROM invoice i
JOIN customer as c ON c.customer_id = i.customer_id
JOIN invoice_line as il ON il.invoice_id = i.invoice_id
JOIN track as t ON t.track_id = il.track_id
JOIN album as b ON b.album_id = t.album_id
JOIN best_selling_artist as bsa ON bsa.artist_id = b.artist_id
GROUP BY 1,2,3,4
ORDER BY 5 DESC;
```

customer_id integer	first_name character	last_name character	artist_name character varying (120)	amount_spent double precision
46	Hugh	O'Reilly	Queen	27.719999999999985
38	Niklas	Schröder	Queen	18.81
3	François	Tremblay	Queen	17.82
34	João	Fernandes	Queen	16.830000000000002
53	Phil	Hughes	Queen	11.88
41	Marc	Dubois	Queen	11.88
47	Lucas	Mancini	Queen	10.89
33	Ellie	Sullivan	Queen	10.89

ADVANCE QUESTIONS

Question-2

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 amounts 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.

```
WITH popular_genre AS
(
    SELECT COUNT(il.quantity) AS purchases, c.country, g.name, g.genre_id,
    ROW_NUMBER() OVER(PARTITION BY c.country ORDER BY COUNT(il.quantity) DESC) AS RowNo
    FROM invoice_line as il
    JOIN invoice as i ON il.invoice_id = i.invoice_id
    JOIN customer as c ON c.customer_id = i.customer_id
    JOIN track as t ON t.track_id = il.track_id
    JOIN genre as g ON g.genre_id = t.genre_id
    GROUP BY 2,3,4
    ORDER BY 2 ASC, 1 DESC
)
SELECT * FROM popular_genre WHERE RowNo <= 1
```

purchases bigint	country character varying (50)	name character varying (120)	genre_id character varying (50)	rowno bigint
17	Argentina	Alternative & Punk	4	1
34	Australia	Rock	1	1
40	Austria	Rock	1	1
26	Belgium	Rock	1	1
205	Brazil	Rock	1	1
333	Canada	Rock	1	1
61	Chile	Rock	1	1
143	Czech Republic	Rock	1	1

ADVANCE QUESTIONS

Question-3

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.

```
WITH Customer_with_country AS (
        SELECT c.customer_id,c.first_name,c.last_name,i.billing_country,SUM(i.total) AS total_spending,
        ROW_NUMBER() OVER(PARTITION BY billing_country ORDER BY SUM(total) DESC) AS RowNo
        FROM invoice as i
        JOIN customer as c ON c.customer_id = i.customer_id
        GROUP BY 1,2,3,4
        ORDER BY 4 ASC,5 DESC)
SELECT * FROM Customer_with_country WHERE RowNo <= 1
```

customer_id integer	first_name character	last_name character	billing_country character varying (30)	total_spending double precision	rowno bigint
56	Diego	Gutiérrez	Argentina	39.6	1
55	Mark	Taylor	Australia	81.18	1
7	Astrid	Gruber	Austria	69.3	1
8	Daan	Peeters	Belgium	60.38999999999999	1
1	Luis	Gonçalves	Brazil	108.89999999999998	1
3	François	Tremblay	Canada	99.99	1
57	Luis	Rojas	Chile	97.02000000000001	1
5	R	Madhav	Czech Republic	144.54000000000002	1

THANK YOU