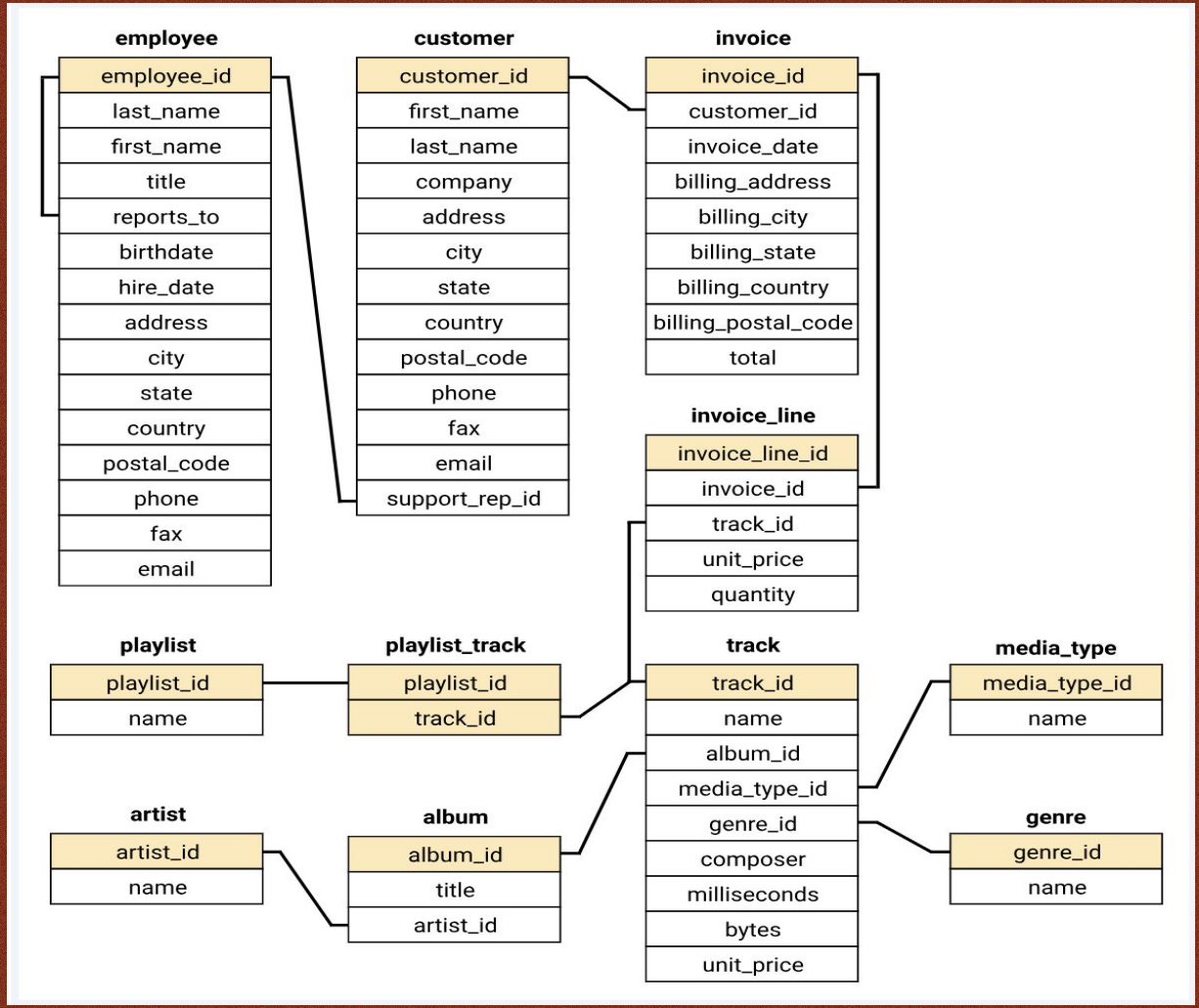
Music Store Data Analysis Using SQL

By Sherin Saji

Objective :

The research focuses on analysing the music playlist database using SQL to understand its business progress.

Schema


employee
employee_id
last_name
first_name
title
reports_to
birthdate
hire_date
address
city
state
country
postal_code
phone
fax
email
customer
customer_id
first_name
last_name
company
address
city
state
country
postal_code
phone
fax
email
support_rep_id
invoice
invoice_id
customer_id
invoice_date
billing_address
billing_city
billing_state
billing_country
billing_postal_code
total
invoice_line
invoice_line_id
invoice_id
track_id
unit_price
quantity
playlist
playlist_id
name
playlist_track
playlist_id
track_id
track
track_id
name
album_id
media_type_id
genre_id
composer
milliseconds
bytes
unit_price
media_type
media_type_id
name
genre
genre_id
name
artist
artist_id
name
album
album_id
title
artist_id

All Questions

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

2. Which countries have the most Invoices?

3. What are top 3 values of total invoice?

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

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

6. 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

7. 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

All Questions

8. 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

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

10. 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

11. 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.

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

```
select title,max(extract(year from age(birthdate))) as max_age
from employee
group by title;
```

	title character varying (50)	max_age numeric
1	Sales Manager	65
2	Sales Support Agent	76
3	General Manager	62
4	Senior General Manager	63
5	IT Manager	51
6	IT Staff	56

Q2. Which countries have the most Invoices?

	billing_country character varying (30)	max_invoice double precision
1	USA	1040.4899999999998
2	Canada	535.5900000000001
3	Brazil	427.68000000000006
4	France	389.0699999999999
5	Germany	334.62
6	Czech Republic	273.24000000000007
7	United Kingdom	245.52
8	Portugal	185.13000000000002
9	India	183.14999999999998
10	Ireland	114.83999999999997
11	Spain	98.01
12	Chile	97.02000000000001
13	Australia	81.18
14	Finland	79.2

```
select temp.billing_country,max(Max_total) as max_invoice
from (select billing_country,sum(total)as Max_total
from invoice
group by billing_country) as temp
group by temp.billing_country
order by max(Max_total) desc;
```

Q3. What are top 3 values of total invoice?

```
select billing_country,count(*)
from invoice
group by billing_country
order by count(*) desc
limit 5;
```

	billing_country character varying (30)	count bigint
1	USA	131
2	Canada	76
3	Brazil	61
4	France	50
5	Germany	41

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 temp.billing_city,max(Max_total) as max_invoice
from (select billing_city,sum(total)as Max_total
from invoice
group by billing_city) as temp
group by temp.billing_city
order by max(Max_total) desc
limit 1;
```

	billing_city character varying (30)	max_invoice double precision
1	Prague	273.24000000000007

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 cu.first_name,sum(inv.total) as Total_invoice
from customer as cu
inner join invoice as inv on cu.customer_id=inv.customer_id
group by cu.first_name
order by Total_invoice desc
limit 1;
```

	first_name character	total_invoice double precision
1	Frank	145.53000000000003

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 distinct first_name,last_name,email
from customer as cu
join invoice as inv on cu.customer_id=inv.customer_id
join invoice_line as invl on inv.invoice_id=invl.invoice_id
where track_id in ( select track_id
                    from track as tr
                    join genre as gen on tr.genre_id=gen.genre_id
                    where gen.name like '%Rock%' )
order by email ;
```

	first_name character	last_name character	email character varying (50)
1	Aaron	Mitchell	aaronmitchell@yahoo.ca
2	Alexandre	Rocha	alero@uol.com.br
3	Astrid	Gruber	astrid.gruber@apple.at
4	Bjørn	Hansen	bjorn.hansen@yahoo.no
5	Camille	Bernard	camille.bernard@yahoo.fr
6	Daan	Peeters	daan_peeters@apple.be
7	Diego	Gutiérrez	diego.gutierrez@yahoo.ar
8	Dan	Miller	dmiller@comcast.com
9	Dominique	Lefebvre	dominiquelefebvre@gmail.c...
10	Edward	Francis	edfrancis@yachoo.ca
11	Eduardo	Martins	eduardo@woodstock.com.br

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

```sql
select ar.name,count(ar.artist_id) as no_of_songs
from artist as ar
join album as al on ar.artist_id=al.artist_id
join track as tr on al.album_id=tr.album_id
where tr.genre_id::int=1
group by ar.artist_id
order by no_of_songs desc
limit 10;
```

	name character varying (120)	no_of_songs bigint
1	Led Zeppelin	114
2	U2	112
3	Deep Purple	92
4	Iron Maiden	81
5	Pearl Jam	54
6	Van Halen	52
7	Queen	45
8	The Rolling Stones	41
9	Creedence Clearwater Revival	40
10	Kiss	35

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 track.name,milliseconds as song_length
from track
where milliseconds>(select avg(milliseconds)
                    from track)
order by song_length desc
```

	name character varying (150)	song_length 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
6	Battlestar Galactica, Pt. 1	2952702
7	Murder On the Rising Star	2935894
8	Battlestar Galactica, Pt. 3	2927802
9	Take the Celestra	2927677
10	Fire In Space	2926593
11	The Long Patrol	2925008

Total rows: 494 of 494 Query complete 00:00:00.394

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

```
with my_cte as(
select ar.artist_id,ar.name,sum(inv.unit_price*inv.quantity) as total_Sales
from invoice_line as inv
join track as tr on tr.track_id=inv.track_id
join album as al on al.album_id=tr.album_id
join artist as ar on ar.artist_id=al.artist_id
group by 1,2
order by 3 desc
limit 1
)

select cu.customer_id,cu.first_name,cu.last_name,mt.name,sum(invl.unit_price*invl.quantity)
from customer as cu
join invoice as inv on inv.customer_id=cu.customer_id
join invoice_line as invl on invl.invoice_id=inv.invoice_id
join track as tr on tr.track_id=invl.track_id
join album as al on al.album_id=tr.album_id
join my_cte as mt on mt.artist_id=al.artist_id
group by 1,2,3,4
order by 5 desc
```

	customer_id integer	first_name character	last_name character	name character varying (120)	sum double precision
1	46	Hugh	O'Reilly	Queen	27.719999999999985
2	38	Niklas	Schröder	Queen	18.81
3	3	François	Tremblay	Queen	17.82
4	34	João	Fernandes	Queen	16.830000000000002
5	53	Phil	Hughes	Queen	11.88
6	41	Marc	Dubois	Queen	11.88
7	47	Lucas	Mancini	Queen	10.89
8	33	Ellie	Sullivan	Queen	10.89
9	20	Dan	Miller	Queen	3.96
10	5	R	Madhav	Queen	3.96
11	23	John	Gordon	Queen	2.9699999999999998

Total rows: 43 of 43 Query complete 00:00:00.674

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

```
with most_popular_genre as
(select count(invl.quantity) as purchases,cu.country,gen.name,row_number()
over(partition by cu.country order by count(invl.quantity) desc ) as row_no
from invoice as inv
join customer as cu on cu.customer_id=inv.customer_id
join invoice_line as invl on invl.invoice_id=inv.invoice_id
join track as tr on tr.track_id=invl.track_id
join genre as gen on gen.genre_id=tr.genre_id
group by 2,3
order by 2)

select mpg.country,mpg.name
from most_popular_genre as mpg
where mpg.row_no=1
```

	country character varying (50)	name character varying (120)
1	Argentina	Alternative & Punk
2	Australia	Rock
3	Austria	Rock
4	Belgium	Rock
5	Brazil	Rock
6	Canada	Rock
7	Chile	Rock
8	Czech Republic	Rock
9	Denmark	Rock

Total rows: 24 of 24 Query complete 00:00:00.553

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

```
with customer_spend_per_country as
(
select cu.customer_id,cu.country,cu.first_name,sum(invl.unit_price*invl.quantity),
row_number() over(partition by cu.country order by sum(invl.unit_price*invl.quantity) desc ) as row_no
from customer as cu
join invoice as inv on inv.customer_id=cu.customer_id
join invoice_line as invl on invl.invoice_id=inv.invoice_id
group by 1,2,3
order by 1
)
select customer_spend_per_country.customer_id,customer_spend_per_country.country,
customer_spend_per_country.first_name
from customer_spend_per_country
where row_no=1
order by customer_spend_per_country.country
```

	customer_id [PK] integer	country character varying (50)	first_name character
1	56	Argentina	Diego
2	55	Australia	Mark
3	7	Austria	Astrid
4	8	Belgium	Daan
5	1	Brazil	Luís
6	3	Canada	François
7	57	Chile	Luis
8	5	Czech Republic	R

Total rows: 24 of 24 Query complete 00:00:00.324

Insights :

When compared to other stores in other cities, the **Prague store** has earned the greatest invoice, totaling **$273**, making it the best.

The song that has sold the most, **Queen**, has earned us a total of **$190**, making her the most well-known singer among our patrons.

Rock is the most popular genre among our consumers in **23 of 24 countries**, accounting for the majority of transactions.

R has been named the music enthusiast and best customer since he has spent the most money on music.