Music Store Data Analysis

By Muhammad Waqas

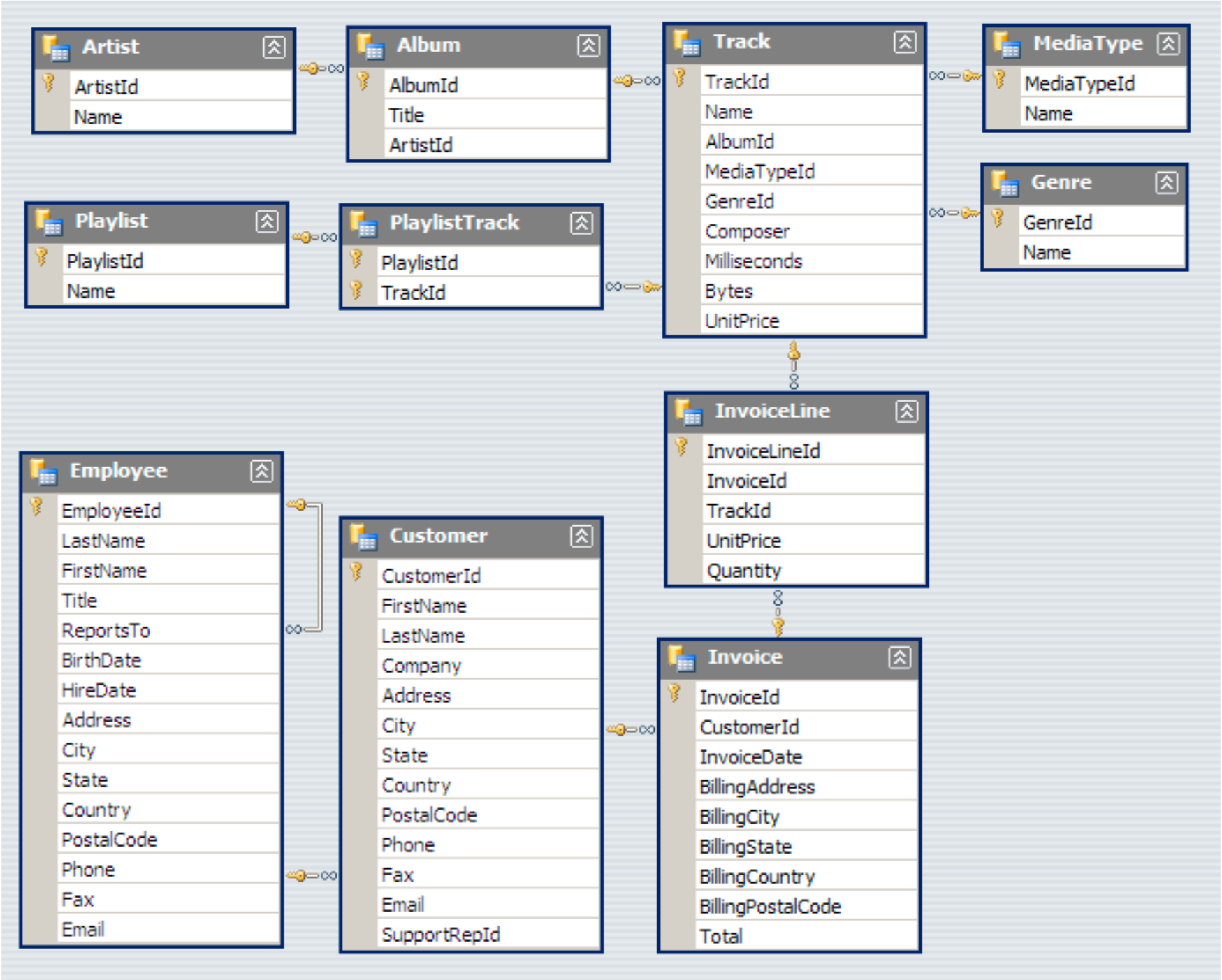
Hello Everyone

I'm Muhammad Waqas

Welcome to my data analysis project!

In this project, I have showcased my SQL skills by performing an in-depth analysis on a music store dataset using PostgreSQL. The analysis covers a range of questions from basic to advanced levels, demonstrating my expertise in SQL and data analysis.

Schema


Artist
ArtistId
Name
Album
AlbumId
Title
ArtistId
Track
TrackId
Name
AlbumId
MediaTypeId
GenreId
Composer
Milliseconds
Bytes
UnitPrice
MediaType
MediaTypeId
Name
Genre
GenreId
Name
Playlist
PlaylistId
Name
PlaylistTrack
PlaylistId
TrackId
InvoiceLine
InvoiceLineId
InvoiceId
TrackId
UnitPrice
Quantity
Employee
EmployeeId
LastName
FirstName
Title
ReportsTo
BirthDate
HireDate
Address
City
State
Country
PostalCode
Phone
Fax
Email
Customer
CustomerId
FirstName
LastName
Company
Address
City
State
Country
PostalCode
Phone
Fax
Email
SupportRepId
Invoice
InvoiceId
CustomerId
InvoiceDate
BillingAddress
BillingCity
BillingState
BillingCountry
BillingPostalCode
Total

All Questions

- **Question Set 1 - Easy**
- 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.

All Questions

Question Set 2 – Moderate

Q1: 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.

Q2: 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.

Q3: 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.

All Questions

Question Set 3 – Advance

Q1: Find how much amount spent by each customer on artists? Write a query to return customer name, artist name and total spent.

Q2: 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.

Q3: 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.

Set No. 01

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

Madan Mohan is the senior most employee as a Senior General Manager.

Query | Query History

```
1  SELECT *
2  FROM employee
3  ORDER BY levels DESC
4  LIMIT 1
```

	employee_id [PK] character varying (50)	last_name character (50)	first_name character (50)	title character varying (50)	reports_to character varying (30)	levels character varying (
1	9	Madan	Mohan	Senior General Manager	[null]	L7

Q2: Which countries have the most Invoices?

USA have the most number of invoices.

Query | Query History

```
1  SELECT billing_country,count(billing_country) AS invoice_number
2  FROM invoice
3  GROUP BY billing_country
4  ORDER BY invoice_number DESC
5  LIMIT 1
```

SQL

	billing_country character varying (30)	invoice_number bigint
1	USA	131

Q3: What are top 3 values of total invoice?

23.75999, 19.8 and 18.81 are top 3 values of total invoices.

Query | Query History

```
1 SELECT total
2 FROM invoice
3 ORDER BY total DESC
```

	total double precision
1	23.759999999999998
2	19.8
3	19.8
4	19.8
5	19.8
6	18.81
7	17.82
8	17.82
9	17.82

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

Prague have best customers, We should throw Music Festival there.

Query | Query History

```
1  SELECT billing_city, sum(total) AS tot_invoice
2  FROM invoice
3  GROUP BY billing_city
4  ORDER BY tot_invoice DESC
```

	billing_city character varying (30)	tot_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 totals

Customer of id 5 is the best customer.

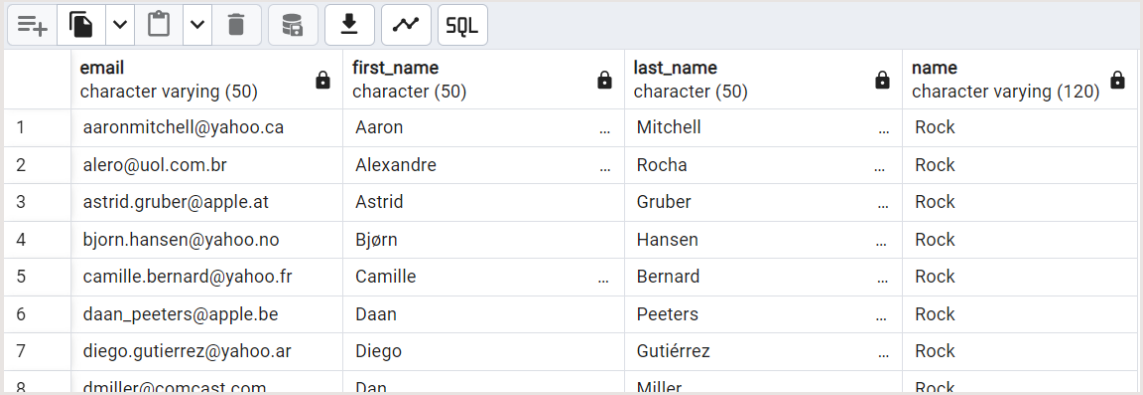
Query | Query History

```
1  SELECT customer_id, SUM(total) AS total_by_customer
2  FROM invoice
3  GROUP BY customer_id
4  ORDER BY total_by_customer DESC
```

	customer_id [PK] integer	first_name character (50)	last_name character (50)	total_spending double precision
1	5	R ...	Madhav	144.54000000000002

Set No. 02

Q1: 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.



	email character varying (50)	first_name character (50)	last_name character (50)	name character varying (120)
1	aaronmitchell@yahoo.ca	Aaron	Mitchell	Rock
2	alero@uol.com.br	Alexandre	Rocha	Rock
3	astrid.gruber@apple.at	Astrid	Gruber	Rock
4	bjorn.hansen@yahoo.no	Bjørn	Hansen	Rock
5	camille.bernard@yahoo.fr	Camille	Bernard	Rock
6	daan_peeters@apple.be	Daan	Peeters	Rock
7	diego.gutierrez@yahoo.ar	Diego	Gutiérrez	Rock
8	dmiller@comcast.com	Dan	Miller	Rock

Query | Query History

```
1  SELECT DISTINCT email,first_name, last_name, genre.name
2  FROM customer
3  JOIN invoice ON invoice.customer_id = customer.customer_id
4  JOIN invoice_line ON invoice_line.invoice_id = invoice.invoice_id
5  JOIN track ON track.track_id = invoice_line.track_id
6  JOIN genre ON genre.genre_id = track.genre_id
7  WHERE genre.name LIKE 'Rock'
8  ORDER BY email;
```

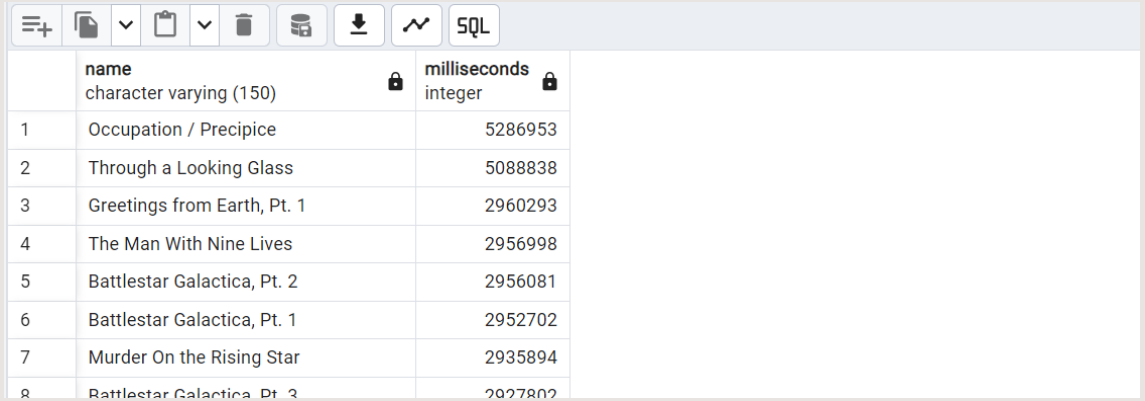
Q2: Let's invite the artists who have written the most rock music in our dataset.
Write a query that returns the Artist name and total track count of the top 10 rock bands.

	artist_id [PK] character varying (50)	name character varying (120)	number_of_songs 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
6	152	Van Halen	52
7	51	Queen	45
8	142	The Rolling Stones	41

Query | Query History

```
1  SELECT artist.artist_id, artist.name,COUNT(art
2  FROM track
3  JOIN album ON album.album_id = track.album_id
4  JOIN artist ON artist.artist_id = album.artist
5  JOIN genre ON genre.genre_id = track.genre_id
6  WHERE genre.name LIKE 'Rock'
7  GROUP BY artist.artist_id
8  ORDER BY number_of_songs DESC
9  LIMIT 10;
```

Q3: 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.



	name character varying (150)	milliseconds 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

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

Set No. 03

Q1: Find how much amount spent by each customer on artists? Write a query to return customer name, artist name and total spent

```sql
WITH best_selling_artist AS (
    SELECT artist.artist_id AS artist_id, artist.name AS artist_name, SUM(invoice_
    FROM invoice_line
    JOIN track ON track.track_id = invoice_line.track_id
    JOIN album ON album.album_id = track.album_id
    JOIN artist ON artist.artist_id = album.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_pri
FROM invoice i
JOIN customer c ON c.customer_id = i.customer_id
JOIN invoice_line il ON il.invoice_id = i.invoice_id
JOIN track t ON t.track_id = il.track_id
JOIN album alb ON alb.album_id = t.album_id
JOIN best_selling_artist bsa ON bsa.artist_id = alb.artist_id
GROUP BY 1,2,3,4
ORDER BY 5 DESC;
```

	customer_id integer	first_name character (50)	last_name character (50)	artist_name character varying (120)	amount_spent 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

***Q2: 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.***

Query | Query History

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

Data Output | Messages | Notifications

	purchases bigint	country character varying (50)	name character varying (120)	genre_id character varying (50)	rowno bigint
1	17	Argentina	Alternative & Punk	4	1
2	34	Australia	Rock	1	1
3	40	Austria	Rock	1	1

Q3: 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. return all Genres.

Query | Query History

```
1  WITH Customter_with_country AS (
2          SELECT customer.customer_id,first_name,last_name,billing_country,SUM(total
3          ROW_NUMBER() OVER(PARTITION BY billing_country ORDER BY SUM(total) DESC) A
4          FROM invoice
5          JOIN customer ON customer.customer_id = invoice.customer_id
6          GROUP BY 1,2,3,4
7          ORDER BY 4 ASC,5 DESC)
8  SELECT * FROM Customter_with_country WHERE RowNo <= 1
```

	customer_id integer	first_name character (50)	last_name character (50)	billing_country character varying (30)	total_spending double precision	rowno bigint
1	56	Diego	Gutiérrez	Argentina	39.6	1
2	55	Mark	Taylor	Australia	81.18	1
3	7	Astrid	Gruber	Austria	69.3	1
4	8	Daan	Peeters	Belgium	60.38999999999999	1
5	1	Luís	Gonçalves	Brazil	108.89999999999998	1
6	3	François	Tremblay	Canada	99.99	1

Thank For Your Time