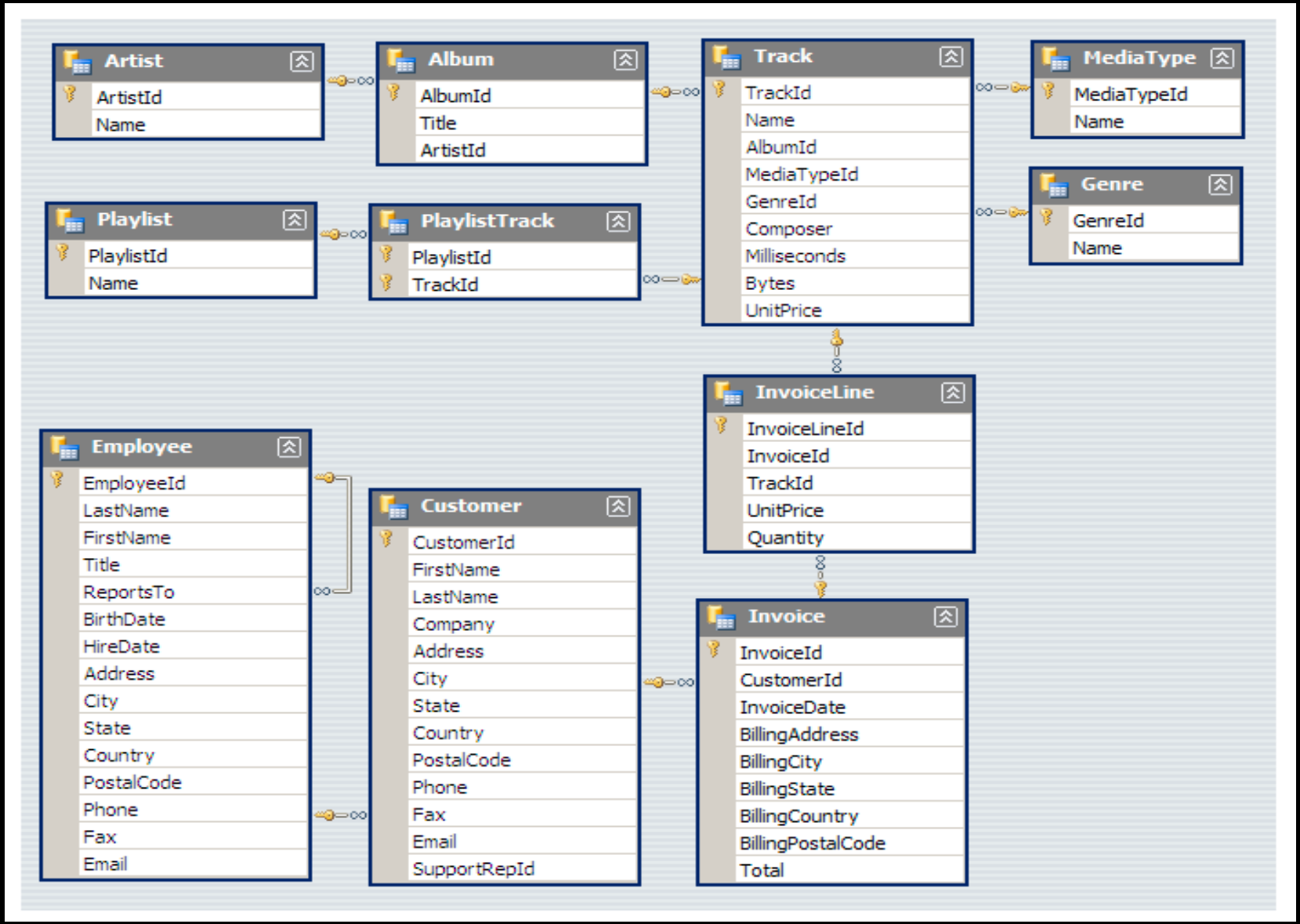
Project introduction

This presentation showcases the data analysis of music trends using SQL, integrating data from 11 tables. It highlights key insights on popular artists, trending genres, and user behaviour, demonstrating advanced SQL query techniques to uncover patterns and trends in music data.

Project Highlights

1. **Comprehensive Data Integration:** Merged data from 11 tables for a holistic view.
2. **Advanced SQL Analysis:** Employed sophisticated SQL queries to derive meaningful insights.
3. **Music Trends Identification:** Analyzed popular artists, trending genres, and user listening behavior.

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

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

Code

```
SELECT
	CONCAT(LAST_NAME, ' ', FIRST_NAME) AS "Full Name",
	LEVELS
FROM
	EMPLOYEE
ORDER BY
	2 DESC
LIMIT
	1;
```

Answer set

"Full Name"	"levels"
"Madan Mohan"	"L7"

Q2: Which countries have the most Invoices?

Code

```
SELECT
        BILLING_COUNTRY,
        COUNT(*)
FROM
        INVOICE
GROUP BY
       BILLING_COUNTRY
ORDER BY
      2 DESC
LIMIT
     1;
```

Answer set

"billing_country"	"count"
"USA"	131

Q3: What are top 3 values of total invoice?

Code

```
SELECT
       TO_CHAR(TOTAL, '999,999.00') AS "Total"
FROM
      INVOICE
ORDER BY
     1 DESC
LIMIT
    3;
```

Answer set

"Total"

"23.76"

"19.80"

" 19.80"

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

Code

```
SELECT
        BILLING_CITY,
        TO_CHAR(SUM(TOTAL), '999,999.00') AS "Invoice Total"
FROM
        INVOICE
GROUP BY
       BILLING_CITY
ORDER BY
      2 DESC
LIMIT
     1;
```

Answer set

"billing_city"	" Invoice Total "
"Prague"	273.24

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

Code

```
SELECT
	CUSTOMER.CUSTOMER_ID,   FIRST_NAME, LAST_NAME,
	SUM(TOTAL) AS TOTAL_SPENDING
FROM
	CUSTOMER
	JOIN INVOICE ON CUSTOMER.CUSTOMER_ID = INVOICE.CUSTOMER_ID GROUP
BY
	CUSTOMER.CUSTOMER_ID
ORDER BY
	TOTAL_SPENDING DESC
LIMIT
	1;
```

Answer set

"customer_id"	"first_name"	"last_name"	"total_spending"
5	"R"	"Madhav"	144.54

Q6: 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.

Code

```
SELECT
       ARTIST.ARTIST_ID, ARTIST.NAME,
       COUNT(ARTIST.ARTIST_ID) AS NUMBER_OF_SONGS
FROM
      TRACK
      JOIN ALBUM ON ALBUM.ALBUM_ID = TRACK.ALBUM_ID
      JOIN ARTIST ON ARTIST.ARTIST_ID = ALBUM.ARTIST_ID
      JOIN GENRE ON GENRE.GENRE_ID = TRACK.GENRE_ID
WHERE
      GENRE.NAME LIKE 'Rock'
GROUP BY
      ARTIST.ARTIST_ID
ORDER BY
     NUMBER_OF_SONGS DESC
LIMIT
    10;
```

Answer set

"artist_id"	"name"	"number_of_songs"
"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
"142"	"The Rolling Stones"	41
"76"	"Creedence Clearwater Revival"	40
"52"	"Kiss"	35

Q7: 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.

Code

```
SELECT
	NAME,
	MILLISECONDS
FROM
	TRACK
WHERE
	MILLISECONDS > (
		SELECT
		    AVG(MILLISECONDS)
		FROM
		  TRACK
 )
ORDER BY MILLISECONDS DESC;
```

Answer set

"name"	"milliseconds"
"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
"Take the Celestra"	2927677
"Fire In Space"	2926593
"The Long Patrol"	2925008
"The Magnificent Warriors"	2924716
"The Living Legend, Pt. 1"	2924507
"The Gun On Ice Planet Zero, Pt. 2"	2924341
"The Hand of God"	2924007
"Experiment In Terra"	2923548
"War of the Gods, Pt. 2"	2923381
"The Living Legend, Pt. 2"	2923298
"War of the Gods, Pt. 1"	2922630
"Lost Planet of the Gods, Pt. 1"	2922547

Q8: Find how much amount spent by each customer on artists? Write a query to return customer name, artist name and total spent

Code

```
WITH        BEST_SELLING_ARTIST AS (
        SELECT ARTIST.ARTIST_ID AS ARTIST_ID, ARTIST.NAME AS ARTIST_NAME,

                SUM(INVOICE_LINE.UNIT_PRICE * INVOICE_LINE.QUANTITY) AS TOTAL_SALES
        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_PRICE * IL.QUANTITY) AS AMOUNT_SPENT
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;
```

Answer set

"customer_id"	"first_name"	"last_name"	"artist_name"	"amount_spent"
46	Hugh	O'Reilly	Queen	27.71
38	Niklas	Schroder	Queen	18.81
3	François	Tremblay	Queen	17.82
34	João	Fernandes	Queen	16.83
53	Phil	Hughes	Queen	11.88
41	Marc	Dubois	Queen	11.88
47	Lucas	Mancini	Queen	10.89
33	Ellie	Sullivan	Queen	10.89
20	Dan	Miller	Queen	3.96
5	R	Madhav	Queen	3.96

Q9: 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.

Code

```
WITH popular_genre AS
(
   SELECT COUNT(invoice_line.quantity) AS purchases, customer.country,
genre.name, genre.genre_id,
        ROW_NUMBER() OVER(PARTITION BY customer.country ORDER BY
COUNT(invoice_line.quantity) DESC) AS RowNo
   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
```

Answer set

"purchases"	"country"	"name"	"genre_id"	"rowno"
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
24	"Denmark"	"Rock"	"1"	1
46	"Finland"	"Rock"	"1"	1
211	"France"	"Rock"	"1"	1
194	"Germany"	"Rock"	"1"	1
44	"Hungary"	"Rock"	"1"	1
102	"India"	"Rock"	"1"	1
72	"Ireland"	"Rock"	"1"	1
35	"Italy"	"Rock"	"1"	1
33	"Netherlands"	"Rock"	"1"	1
40	"Norway"	"Rock"	"1"	1
40	"Poland"	"Rock"	"1"	1
108	"Portugal"	"Rock"	"1"	1

Q10: 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.

Code

```
WITH Customter_with_country AS (
        SELECT customer.customer_id, first_name,last_name, billing_country,
        to_char(SUM(total),'999,999.99') AS total_spending,
        ROW_NUMBER() OVER(PARTITION BY billing_country ORDER BY SUM(total) DESC) AS RowNo
        FROM invoice
        JOIN customer ON customer.customer_id = invoice.customer_id
        GROUP BY 1,2,3,4
        ORDER BY 4 ASC,5 DESC)
SELECT * FROM Customter_with_country WHERE RowNo <= 1
```

Answer set

"customer_id"	"first_name"	"last_name"	"billing_country"	"total_spending"	"rowno"
56	"Diego"	"Gutiérrez"	"Argentina"	"39.60"	1
55	"Mark"	"Taylor"	"Australia"	"81.18"	1
7	"Astrid"	"Gruber"	"Austria"	"69.30"	1
8	"Daan"	"Peeters"	"Belgium"	"60.39"	1
1	"Luís"	"Gonçalves"	"Brazil"	"108.90"	1
3	"François"	"Tremblay"	"Canada"	"99.99"	1
57	"Luis"	"Rojas"	"Chile"	"97.02"	1
5	"R"	"Madhav"	"Czech Republic"	"144.54"	1
9	"Kara"	"Nielsen"	"Denmark"	"37.62"	1
44	"Terhi"	"Hamalainen"	"Finland"	"79.20"	1
42	"Wyatt"	"Girard"	"France"	"99.99“	1
37	"Fynn"	"Zimmermann"	"Germany"	"94.05"	1
45	"Ladislav"	"Kovacs"	"Hungary“	"78.21"	1
58	"Manoj"	"Pareek"	"India"	"111.87"	1
46	"Hugh"	"O'Reilly"	"Ireland"	"114.84"	1
47	"Lucas"	"Mancini"	"Italy"	"50.49"	1