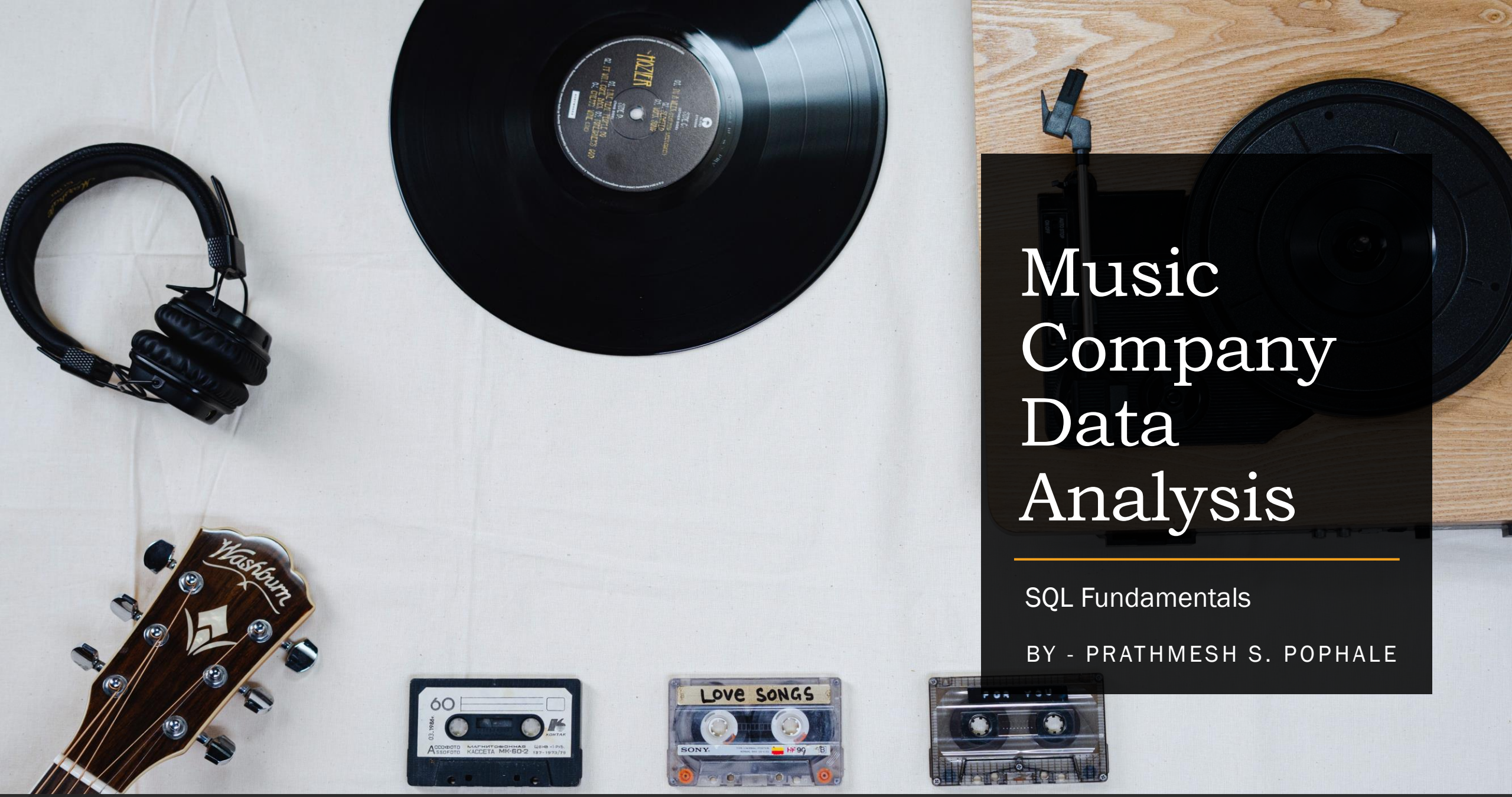


Music Company Data Analysis

SQL Fundamentals

BY - PRATHMESH S. POPHALE

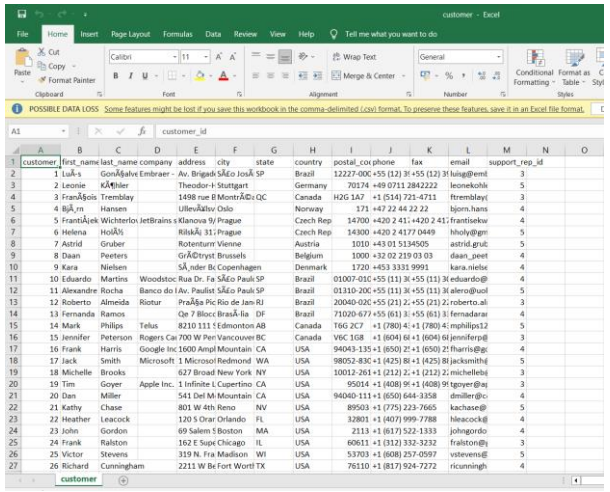
Introduction: Overview of the project and its goals

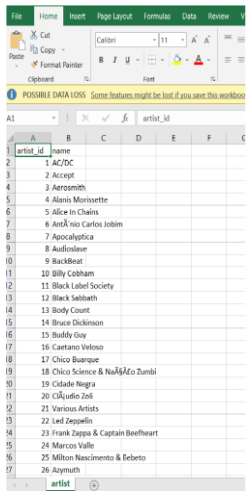
“Music Company Data Analysis” aims to help music companies make data-driven decisions by analyzing key metrics such as sales, streaming numbers, and customer demographics using MySQL. By leveraging advanced data analysis techniques, we can provide insights into which artists, albums, and genres are performing well, as well as identify potential areas for growth and improvement. Our goal is to empower music companies to make informed decisions that will help them stay competitive in an ever-changing industry, using the power and flexibility of MySQL to manage and analyze their data."

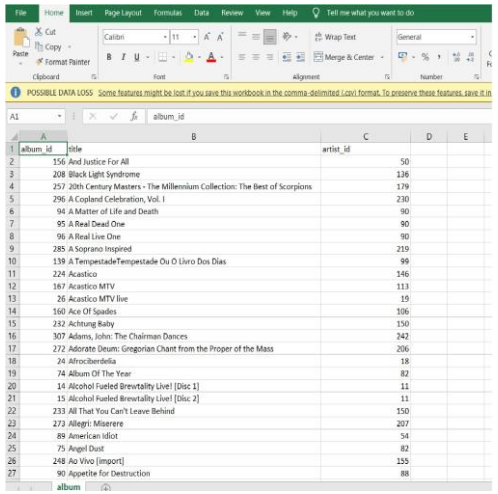
Data sources: Description of the data sources used in the project

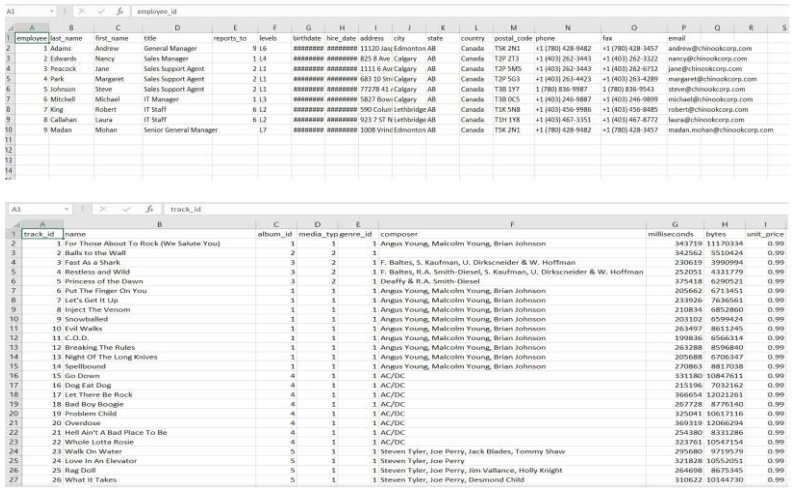
The dataset used in the "Music Company Data Analysis" project was obtained from Kaggle, a popular platform for finding and sharing datasets.

The dataset includes information such as the song title, artist name, album name, release date, customers details, and various features.









Data analysis:

To analyze the data to identify patterns, trends, and insights. This involved using MySQL to query the data and generate visualizations of the results.

One of the primary tools used in the project was MySQL Workbench, a popular database management tool that provides a range of features for creating, managing, and querying databases.

In the "Music Company Data Analysis " project, several key query questions were developed to guide the data analysis process. These questions were designed to help us identify trends and patterns in the data, and to gain insights into the music industry.

Questions & Solving Approach:

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

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

select * from employee
order by levels Desc
limit 1;
```

Result -

	employee_id	last_name	first_name	title	reports_to	levels	birthdate	hire_date	address	city	state	country	postal_code	phone
▶	1	Adams	Andrew	General Manager	9	L6	18-02-1962 00:00	14-08-2016 00:00	11120 Jasper Ave NW	Edmonton	AB	Canada	T5K 2N1	+1 (780) 428-9482

2. Which countries have the most Invoices?

```
# Q.2. Which countries have the most Invoices?

select count(*) As Total_Invoices , billing_country from invoice
group by billing_country
order by total_invoices desc;
```

Total_Invoices	billing_country
131	USA
76	Canada
61	Brazil
50	France
41	Germany
30	Czech Republic
29	Portugal
28	United Kingdom
21	India
13	Ireland
13	Chile
11	Finland
11	Spain
10	Poland
10	Denmark
10	Australia
10	Hungary
10	Sweden
10	Netherlands

3. What are top 3 values of total invoice?

```
# Q.3. What are top 3 values of total invoice?

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

total
23.759999999999998
19.8
19.8

4.

```
# Q.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 billing_city As City , sum(total) As Total_Invoicess from invoice
group by billing_city
order by Total_Invoicess desc
limit 1;
```

City	Total_Invoicess
Prague	273.24000000000007

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 customer.customer_id, MAX(first_name) AS 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, last_name
ORDER BY total_spending DESC
LIMIT 1;
```

customer_id	first_name	last_name	total_spending
5	FrantiÅiek	WichterlovÃ¡	144.54000000000002

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

```
select distinct c.email, c.first_name , c.last_name,g.name as Genre from customer 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 genre g on t.genre_id = g.genre_id
where g.name = "Rock"
order by email;
```

email	first_name	last_name	Genre
aaronmitchell@yahoo.ca	Aaron	Mitchell	Rock
alero@uol.com.br	Alexandre	Rocha	Rock
astrid.gruber@apple.at	Astrid	Gruber	Rock
bjorn.hansen@yahoo.no	BjÃ¸rn	Hansen	Rock
camille.bernard@yahoo.fr	Camille	Bernard	Rock
daan_peeters@apple.be	Daan	Peeters	Rock
diego.gutierrez@yahoo.ar	Diego	GutiÃ©rrez	Rock
dmiller@comcast.com	Dan	Miller	Rock
dominiquelefebvre@gmail.com	Dominique	Lefebvre	Rock
edfrancis@yachoo.ca	Edward	Francis	Rock
eduardo@woodstock.com.br	Eduardo	Martins	Rock
ellie.sullivan@shaw.ca	Ellie	Sullivan	Rock
emma_jones@hotmail.com	Emma	Jones	Rock
enrique_munoz@yahoo.es	Enrique	MuÃ±oz	Rock
fernadaramos4@uol.com.br	Fernanda	Ramos	Rock

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

```sql
SELECT ar.name AS Artist, COUNT(*) AS TrackCount FROM artist ar
JOIN album2 al ON ar.artist_id = al.artist_id
JOIN track t ON al.album_id = t.album_id
JOIN genre g ON t.genre_id = g.genre_id
WHERE g.name = 'Rock'
GROUP BY ar.name
ORDER BY TrackCount DESC
LIMIT 10;
```

Artist	TrackCount
AC/DC	18
Aerosmith	15
Audioslave	14
Led Zeppelin	14
Alanis Morissette	13
Alice In Chains	12
Frank Zappa & Captain Beefheart	9
Accept	4

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

```
select track.name As Name , milliseconds As Length from track
where milliseconds > (SELECT AVG(milliseconds) FROM track)
order by milliseconds Desc;
```

Name	Length
How Many More Times	711836
Advance Romance	677694
Sleeping Village	644571
You Shook Me(2)	619467
Talkin' 'Bout Women Obviously	589531
Stratus	582086
No More Tears	555075
The Alchemist	509413
Wheels Of Confusion / The Straightener	494524
Book Of Thel	494393
You Oughta Know (Alternate)	491885
Terra	482429
Snoopy's search-Red baron	456071
Sozinho (Hitmakers Classic Mix)	436636
Master Of Puppets	436453
Stone Crazy	433397

track 82

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

```
SELECT customer.first_name  As Customer_FName, customer.last_name AS Customer_Name, artist.name AS Artist_Name, SUM(invoice.total) AS Total_Spent
FROM customer
JOIN invoice ON customer.customer_id = invoice.customer_id
JOIN invoice_line ON invoice.invoice_id = invoice_line.invoice_id
JOIN track ON invoice_line.track_id = track.track_id
JOIN album2 ON track.album_id = album2.album_id
JOIN artist ON album2.artist_id = artist.artist_id
GROUP BY Customer_FName, Customer_Name, Artist_Name
ORDER BY Total_Spent DESC;
```

Customer_FName	Customer_Name	Artist_Name	Total_Spent
Jennifer	Peterson	Aerosmith	222.74999999999994
Mark	Taylor	Aerosmith	222.74999999999994
Fernanda	Ramos	AntÃ´nio Carlos Jobim	194.04000000000002
Leonie	KÃ¶hler	Audioslave	194.04000000000002
Edward	Francis	Alanis Morissette	167.31000000000003
Emma	Jones	Alanis Morissette	167.31000000000003
Steve	Murray	AC/DC	162.36
JoÃ£o	Fernandes	Alanis Morissette	144.53999999999996
Victor	Stevens	Alice In Chains	142.55999999999997
StanisÅ‚aw	WÃ³jcik	Buddy Guy	119.79
Phil	Hughes	AC/DC	108.90000000000002
Kathy	Chase	AC/DC	93.06000000000002
StanisÅ‚aw	WÃ³jcik	AC/DC	89.09999999999998
FranÃ§ois	Tremblay	Apocalyptica	63.36000000000001

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

```
# using common table expression (CTE)
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;
```

purchases	country	name	genre_id	RowNo
1	Argentina	Rock	1	1
18	Australia	Rock	1	1
6	Austria	Rock	1	1
5	Belgium	Rock	1	1
26	Brazil	Rock	1	1
57	Canada	Rock	1	1
7	Chile	Rock	1	1
14	Czech Republic	Rock	1	1
6	Denmark	Rock	1	1
6	Finland	Rock	1	1
26	France	Rock	1	1
28	Germany	Rock	1	1
4	Hungary	Rock	1	1
13	India	Rock	1	1
2	Ireland	Rock	1	1
3	Italy	Rock	1	1
6	Netherlands	Rock	1	1
2	Norway	Metal	3	1

Result 85

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

```
WITH Customter_with_country AS (
        SELECT billing_country As Country ,customer.customer_id,first_name,last_name,SUM(total) 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
```

Country	customer_id	first_name	last_name	total_spending	RowNo
Portugal	34	JoÃ£o	Fernandes	102.96000000000001	1
France	42	Wyatt	Girard	99.99	1
Brazil	1	LuÃ-s	GonÃ§alves	108.89999999999998	1
Austria	7	Astrid	Gruber	69.3	1
Argentina	56	Diego	GutiÃ©rrez	39.6	1
Finland	44	Terhi	HÃ¤mÃ¤lÃ¤inen	79.2	1
Norway	4	BjÃ¸rn	Hansen	72.27000000000001	1
United Kingdom	53	Phil	Hughes	98.01	1
Sweden	51	Joakim	Johansson	75.24	1
Hungary	45	Ladislav	KovÃ¡cs	78.21	1
Italy	47	Lucas	Mancini	50.49	1
Spain	50	Enrique	MuÃ±oz	98.01	1
Denmark	9	Kara	Nielsen	37.61999999999999	1
Ireland	46	Hugh	O'Reilly	114.83999999999997	1
India	58	Manoj	Pareek	111.86999999999999	1

Result 86

Some of the key insights that we gained from our analysis included:

1. The music industry is highly competitive, with a small number of artists and labels dominating the market.

2. Album sales are an important indicator of success in the music industry, but other factors such as genre popularity and streaming data are also important.

3. The popularity of different genres of music can vary significantly over time, and understanding these trends is important for predicting future success.

4. Successful artists and labels tend to have a consistent track record of success over time, suggesting that long-term planning and strategy are important in the music industry.

Conclusion

The "Music Company Data Analysis " project provided a valuable opportunity to explore the music industry and gain insights into the factors that contribute to success in the industry. Through a structured approach to developing research questions and analyzing the data, we were able to identify key trends and patterns in the data, and to gain a deep understanding of the complex dynamics of the music industry.