

# Mentoring Week 2 - SQL Query

SQL and Relational Database - Job Preparation Program - Pacmann AI



Playing

source: [freepik](https://www.freepik.com)

# Task Description

Assume you are a data analyst in a **digital media store**. You are asked by the manager to

- Give some **insights of current transaction histories**.
- Give information for the Product team or marketing team related to music genres, artist, track, ect

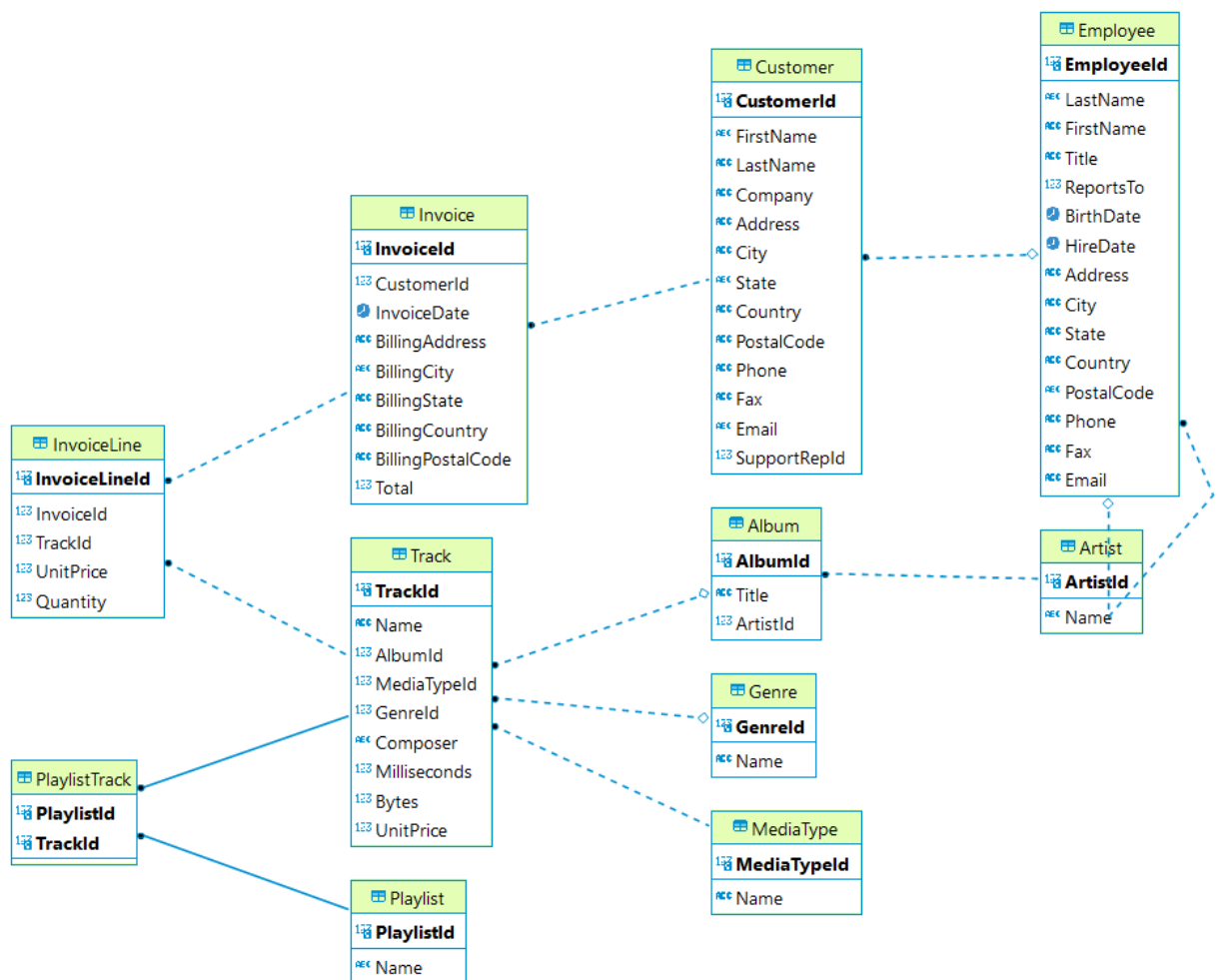
## The Dataset

- Download Dataset: [week 2](#) (please use this link to download the dataset)
- This database provides an insight into the digital music shop, including artist details, album catalogs, album and song sales information, and information on customers and employees associated with a digital media store. This data can be used for various analytical purposes, including understanding sales trends, customer preferences, and developments in the digital music industry.
- List of Table

Tables	Descriptions
"Album"	This table contains information about music albums. Each record in this table represents one album and includes details such as the album title, the artist who created it, and the release year.
"Artist"	This table lists music artists. Each record in the table includes information about artists, such as name, country of origin, and more.
"Customer"	This table contains data about customers of the digital media store. Customer information, such as name, address, and email, is stored in this table.
"Employee"	This table contains data about employees of the digital media store. Staff information, such as name, position, hiring date, and more, can be found in this table.
"Genre"	This table holds a list of music genres available in the digital store's catalog. Each record includes details about genres, such as name and description.
"Invoice"	This table contains information about sales invoices. This data includes the invoice date, total payment, and customer information associated with the invoice.

"invoice_line"	This table details the items on each sales invoice. Information such as the tracks purchased, quantity, and individual item prices can be found here.
"media_type"	This table contains information about the types of media used to store songs, such as MP3 or FLAC.
"Playlist"	This table contains data about playlists created by customers. Each record includes the playlist name and information about the playlist owner.
"playlist_track"	This table acts as a link between playlists and the songs included in them, connecting songs to specific playlists.
"Track"	This table lists the available songs in the digital store. Information such as song title, artist, album, genre, and price is available in this table.

- ERD



# The Tools

**PostgreSQL** is our primary tool during the class and mentoring session.

## Your Tasks

- Let's think as a data analyst, Help the manager to answer the questions!
- As a data analyst, you can help the manager answer the questions by providing **SQL queries (syntax), screenshots of the query results, and the information obtained from those queries**. Here's an example of how you can structure your response:

**Task 1:** How much data in the album table?

*Answer here*

- **SQL Query Syntax (TEXT NOT IMAGE)**

```
SELECT count(*) FROM "Album"
```

- **Screenshot of Query Results:**

[Insert Screenshot Here]

- **Description of Query Results:**

In this query, I retrieved a total of 347 data from the album table.

### Basic Information about transactional data

1. (10 point ) Determine which countries have the most number of invoices (top 10). Order them by the number of invoices in descending order and if there are the same number of invoices, sort them by country name in ascending order. Show Country Name and total number of invoices.

*Answer here*

- **SQL Query Syntax:**

```
SELECT
    "BillingCountry",
    COUNT("InvoiceId") AS number_of_invoices
FROM
```

```

        "Invoice"
GROUP BY
    1
ORDER BY
    COUNT("InvoiceId") DESC,
    "BillingCountry" ASC
LIMIT 10;

```

○ **Screenshot of Query Results:**

	BillingCountry character varying (40) 🔒	number_of_invoices bigint 🔒
1	USA	91
2	Canada	56
3	Brazil	35
4	France	35
5	Germany	28
6	United Kingdom	21
7	Czech Republic	14
8	Portugal	14
9	India	13
10	Argentina	7

○ **Description of Query Results:**

The results display the top 10 countries with the highest number of invoices, ordered by the number of invoices in descending order and by country name in ascending order. The highest country is USA

2. (10 point) The top 10 genres by total sales in the database. The total sales are obtained by multiplying the quantity of items sold by their respective prices. Shows Genre Name and Total Sales

*Answer here*

○ **SQL Query Syntax**

```

SELECT
    "Genre"."Name" AS "Genre_Music",
    SUM("InvoiceLine"."Quantity" * "InvoiceLine"."UnitPrice") AS "Total_Sales"
FROM "Genre"
JOIN "Track" USING("GenreId")
JOIN "InvoiceLine" USING("TrackId")
GROUP BY 1
ORDER BY 2 DESC, 1 ASC
LIMIT 10;

```

○ **Screenshot of Query Results:**

	Name character varying (120) 🔒	Total_Sales numeric 🔒
1	Rock	826.65
2	Latin	382.14
3	Metal	261.36
4	Alternative & Punk	241.56
5	TV Shows	93.53
6	Jazz	79.20
7	Blues	60.39
8	Drama	57.71
9	Classical	40.59
10	R&B/Soul	40.59

- **Description of Query Results:**

The results display the top 10 Genre name with the highest of total sales, ordered by total sales and genre's name. The highest total sales is Rock

3. (10 point ) Who are the top 10 customers by their total spending? Shows Customer Name (consist of first name and last name), Email, and Total Spending

*Answer here*

- **SQL Query Syntax**

```
SELECT
    CONCAT("Customer"."FirstName",' ','Customer"."LastName") AS "FullName",
    "Customer"."Email",
    SUM("Invoice"."Total") AS "TotalSpending"
FROM
    "Customer"
JOIN "Invoice" USING("CustomerId")
GROUP BY 1,2
ORDER BY 3 DESC
LIMIT 10
```

- **Screenshot of Query Results:**

	FullName text	Email character varying (60)	TotalSpending numeric
1	Helena Holy	hholy@gmail.com	49.62
2	Richard Cunningham	ricunningham@hotmail....	47.62
3	Luis Rojas	luisrojas@yahoo.cl	46.62
4	Hugh O'Reilly	hughoreilly@apple.ie	45.62
5	Ladislav Kovács	ladislav_kovacs@apple.hu	45.62
6	Fynn Zimmermann	fzimmermann@yahoo.de	43.62
7	Frank Ralston	fralston@gmail.com	43.62
8	Julia Barnett	jubarnett@gmail.com	43.62
9	Astrid Gruber	astrid.gruber@apple.at	42.62
10	Victor Stevens	vstevens@yahoo.com	42.62

- **Description of Query Results:**

The result of rank of total spending customer. The highest total spending is Helena Holy with total spending 49.62

4. (10 point ) In the results list of countries in number 1, which city has the most number of invoices? Show Country Name, City Name and total number of invoices.

*Answer here*

- **SQL Query Syntax :**

```
SELECT
    "BillingCountry" AS "CountryName",
    "BillingCity" AS "CityName",
    COUNT("InvoiceId") AS "TotalNumberInvoices"
FROM "Invoice"
WHERE "BillingCountry" IN (
    SELECT
        "BillingCountry"
    FROM
        "Invoice"
    GROUP BY
        1
    ORDER BY
        COUNT("InvoiceId") DESC,
        "BillingCountry" ASC
    LIMIT 10)
GROUP BY 1,2
ORDER BY COUNT("InvoiceId") DESC
```

- **Screenshot of Query Results:**

	CountryName character varying (40) 🔒	CityName character varying (40) 🔒	TotalNumberInvoices bigint 🔒
1	France	Paris	14
2	Czech Republic	Prague	14
3	Germany	Berlin	14
4	USA	Mountain View	14
5	United Kingdom	London	14
6	Brazil	São Paulo	14
7	Canada	Winnipeg	7
8	USA	Cupertino	7
9	France	Bordeaux	7
10	USA	Chicago	7
11	Brazil	Rio de Janeiro	7
12	France	Dijon	7
13	USA	Fort Worth	7
14	Canada	Halifax	7
15	Portugal	Porto	7
16	USA	New York	7
17	Canada	Montréal	7
18	Brazil	Brasília	7

○ **Description of Query Results:**

This query lists cities from the top 10 countries by invoice count, showing each city's total invoices.

Top Cities (14 invoices): Paris, Prague, Berlin, Mountain View, London, São Paulo

**Next, we can find deeper information to help Product Team**

5. (10 point ) The product team is looking to add some tracks from new artists to the store and market them in the United Kingdom. Due to budget constraints for marketing, the product team needs to select 4 out of 6 songs to include in the store. The product team assumes that they should choose songs with genres that are popular in the United Kingdom. Here are the tracks and their respective genres that **will be added** to the store:

- "Lalaland": R&B/Soul
- "Soul Sister": Pop
- "Good to See You": Rock
- "Nothing On You": Jazz
- "Get Ya Before Sunrise": Reggae
- "Before The Coffee Gets Cold": Hip Hop/Rap

Assist the product team in selecting the songs to be included in the store.



*(Hint: Find the genres that are popular in the United Kingdom. Popularity is determined by the number of purchases of tracks (quantity) in that genre.)*

Answer here

○ **SQL Query Syntax**

```
SELECT
    "Genre"."Name",
    SUM("InvoiceLine"."Quantity") AS number_of_purchase
FROM "Genre"
JOIN "Track" USING("GenreId")
JOIN "InvoiceLine" USING("TrackId")
JOIN "Invoice" Using ("InvoiceId")
JOIN "Customer" USING("CustomerId")
WHERE "Customer"."Country" = 'United Kingdom'
GROUP BY 1
ORDER BY 2 DESC, 1 ASC
```

○ **Screenshot of Query Results:**

	Name character varying (120) 🔒	number_of_purchase bigint 🔒
1	Rock	37
2	Latin	31
3	Metal	20
4	Alternative & Punk	9
5	Reggae	5
6	Jazz	4
7	Hip Hop/Rap	3
8	Pop	2
9	R&B/Soul	2
10	World	1

○ **Description of Query Results:**

The choosen songs is :

- "Good to See You": Rock
- "Get Ya Before Sunrise": Reggae
- "Nothing On You": Jazz
- "Before The Coffee Gets Cold": Hip Hop/Rap

6. (10 point ) The Product Team wants to market albums that are popular in the USA to be marketed in other countries. Help the product team by searching for the 10 most popular albums in the USA based on album units sold

Answer here

- **SQL Query Syntax**

```
SELECT
    "Album"."Title",
    SUM("InvoiceLine"."Quantity") AS number_of_purchase
FROM "Album"
JOIN "Track" USING("AlbumId")
JOIN "InvoiceLine" USING("TrackId")
JOIN "Invoice" Using ("InvoiceId")
WHERE "Invoice"."BillingCountry" = 'USA'
GROUP BY 1
ORDER BY 2 DESC, 1 ASC;
```

- **Screenshot of Query Results:**

	Title character varying (160)	number_of_purchase bigint
1	The Office, Season 3	14
2	Prenda Minha	11
3	Unplugged	11
4	Chill: Brazil (Disc 2)	10
5	Back to Black	9
6	International Superhits	8
7	A-Sides	7
8	B-Sides 1980-1990	7
9	Vinícius De Moraes - Sem Limite	7
10	Black Album	6
11	Serie Sem Limite (Disc 1)	6
12	Serie Sem Limite (Disc 2)	6

- **Description of Query Results:**

This query shows the top 10 albums sold in the USA, with The Office, Season 3 being the most popular at 14 units. Brazilian music albums like Prenda Minha and Chill: Brazil also performed well, suggesting strong niche appeal. The product team should focus on marketing these highest-selling albums first when expanding to other countries

7. (10 point ) Provide a table that aggregates purchase data by country. In cases where a country has only one customer, group these countries as 'Other.' The results should be sorted by total sales in descending order.

Information to calculate:

- Total Number of Customers: Calculate the count of unique customers within each country.
- Total Value of Sales: Sum the total sales value for each country.
- Average Value of Sales per Customer: Divide the total sales value by the number of unique customers in each country
- Average Order Value: Divide the total sales value by the number of orders (invoices) placed in each country to calculate the average order value.

Answer here

○ **SQL Query Syntax**

```
WITH CountrySales AS (  
    SELECT  
        CASE  
            WHEN COUNT(DISTINCT c."CustomerId") = 1 THEN 'Other'  
            ELSE c."Country"  
        END AS Country,  
        COUNT(DISTINCT c."CustomerId") AS TotalCustomers,  
        SUM(i."Total") AS TotalSales,  
        COUNT(DISTINCT i."InvoiceId") AS TotalOrders  
    FROM "Customer" c  
    JOIN "Invoice" i ON c."CustomerId" = i."CustomerId"  
    GROUP BY c."Country"  
)
```

```
SELECT  
    Country,  
    SUM(TotalCustomers) AS "Total Number of Customers",  
    ROUND(SUM(TotalSales), 2) AS "Total Value of Sales",  
    ROUND(SUM(TotalSales) / SUM(TotalCustomers), 2) AS "Average Value of Sales  
per Customer",  
    ROUND(SUM(TotalSales) / SUM(TotalOrders), 2) AS "Average Order Value"  
FROM CountrySales  
GROUP BY Country  
ORDER BY  
    2 DESC;
```

○ **Screenshot of Query Results:**

	country character varying 🔒	Total Number of Customers numeric 🔒	Total Value of Sales numeric 🔒	Average Value of Sales per Customer numeric 🔒	Average Order Value numeric 🔒
1	Other	15	604.30	40.29	5.76
2	USA	13	523.06	40.24	5.75
3	Canada	8	303.96	38.00	5.43
4	France	5	195.10	39.02	5.57
5	Brazil	5	190.10	38.02	5.43
6	Germany	4	156.48	39.12	5.59
7	United Kingdom	3	112.86	37.62	5.37
8	Portugal	2	77.24	38.62	5.52
9	Czech Republic	2	90.24	45.12	6.45
10	India	2	75.26	37.63	5.79

○ **Description of Query Results:**

**Top Performing Countries:**

1. **Other** (grouped small markets) leads with \$604.30 total sales from 15 customers
2. **USA** follows closely with \$523.06 from 13 customers
3. **Canada** ranks third with \$303.96 from 8 customers

8. (10 point ) Some **genres** have low sales, the product team wants to analyze which genres need to be boosted by carrying out additional promotion or other strategies. Because each country has different behavior, the product team started by analyzing sales in USA

( The total sales are obtained by multiplying the quantity of items sold by their respective prices)

*Answer here*

○ **SQL Query Syntax**

```
SELECT
    "Genre"."Name" AS Genre_name,
    "InvoiceLine"."Quantity",
    SUM("InvoiceLine"."UnitPrice" * "InvoiceLine"."Quantity") AS Total_Sales,
    "Invoice"."BillingCountry" AS Country
FROM "Genre"
JOIN "Track" USING("GenreId")
JOIN "InvoiceLine" USING("TrackId")
JOIN "Invoice" USING("InvoiceId")
WHERE "Invoice"."BillingCountry" = 'USA'
GROUP BY 1,4,2
ORDER BY 3 ASC
```

○ **Screenshot of Query Results:**

	genre_name character varying (120) 🔒	Quantity integer 🔒	total_sales numeric 🔒	country character varying (40) 🔒
1	Science Fiction	1	1.99	USA
2	Rock And Roll	1	2.97	USA
3	Easy Listening	1	2.97	USA
4	Soundtrack	1	3.96	USA
5	Heavy Metal	1	3.96	USA
6	Hip Hop/Rap	1	3.96	USA
7	Pop	1	4.95	USA
8	Alternative	1	4.95	USA
9	Reggae	1	5.94	USA
10	Bossa Nova	1	6.93	USA
11	Classical	1	7.92	USA
12	Sci Fi & Fantasy	1	9.95	USA
13	R&B/Soul	1	11.88	USA
14	Drama	1	11.94	USA
15	Blues	1	14.85	USA
16	Comedy	1	15.92	USA
17	Jazz	1	21.78	USA
18	TV Shows	1	27.86	USA

○ **Description of Query Results:**

**Low-Performing Genres (Need Promotion):**

- Science Fiction (\$1.99) - Lowest sales
- Rock And Roll (\$2.97)
- Easy Listening (\$2.97)

These genres sold just 1 unit each with minimal revenue, indicating weak demand or visibility.

**Now, let's deep dive into the behavior of our customers**

9. (10 point ) We want to advertise songs to the customer based on how much each customers spent per genre. Help Marketing Team to find Top genre for each customers with the most spent

Answer here

○ **SQL Query Syntax**

**WITH CustomerGenreSpending AS**

**(SELECT**

**"Customer"."CustomerId",**

**CONCAT("Customer"."FirstName", ' ', "Customer"."LastName") AS FullName,**

```

    "Genre"."Name" AS Genre,
    SUM("InvoiceLine"."UnitPrice" * "InvoiceLine"."Quantity") AS Total_spending
FROM "Customer"
JOIN "Invoice" USING("CustomerId")
JOIN "InvoiceLine" USING("InvoiceId")
JOIN "Track" USING("TrackId")
JOIN "Genre" USING("GenreId")
GROUP BY 1,2,3
ORDER BY 1 ),
RankedGenres AS (
    SELECT
        "CustomerId",
        FullName,
        Genre,
        Total_spending,
        RANK() OVER (PARTITION BY "CustomerId" ORDER BY Total_spending DESC)
    AS GenreRank
    FROM CustomerGenreSpending)
SELECT
    "CustomerId",
    FullName,
    Genre AS TopGenre,
    Total_spending AS AmountSpent
FROM RankedGenres
WHERE GenreRank = 1
ORDER BY 1

```

- Screenshot of Query Results:

	CustomerId integer	fullname text	topgenre character varying (120)	amountspent numeric
1	1	Luís Gonçalves	Rock	13.86
2	2	Leonie Köhler	Rock	16.83
3	3	François Tremblay	Metal	9.90
4	4	Bjørn Hansen	Rock	16.83
5	5	František Wichterlová	Rock	14.85
6	6	Helena Holý	TV Shows	11.94
7	7	Astrid Gruber	Rock	14.85
8	8	Daan Peeters	Rock	20.79
9	9	Kara Nielsen	Rock	20.79
10	10	Eduardo Martins	Rock	28.71
11	11	Alexandre Rocha	Latin	15.84
12	12	Roberto Almeida	Rock	15.84
13	12	Roberto Almeida	Latin	15.84
14	13	Fernanda Ramos	Rock	10.89
15	14	Mark Philips	Rock	12.87
16	15	Jennifer Peterson	Rock	11.88
17	16	Frank Harris	Metal	11.88
18	17	Jack Smith	Rock	7.92

- **Description of Query Results:**

Focus on Rock music promotions since most customers prefer it. For the few who like Latin or Metal, send them targeted offers.

10. (10 point ) The Marketing team wants to increase advertising in countries with customers who have spent the most money. Help the Marketing team find the top 10 countries with the highest-spending customers.

*Answer here*

- **SQL Query Syntax**
- **Screenshot of Query Results:**

	country character varying (40) 🔒	total_spend numeric 🔒
1	USA	523.06
2	Canada	303.96
3	France	195.10
4	Brazil	190.10
5	Germany	156.48
6	United Kingdom	112.86
7	Czech Republic	90.24
8	Portugal	77.24
9	India	75.26
10	Chile	46.62

- **Description of Query Results:**  
Focus advertising budget first on the USA (highest spending), then Canada and France. These top 3 countries account for over 50% of the total spending shown.