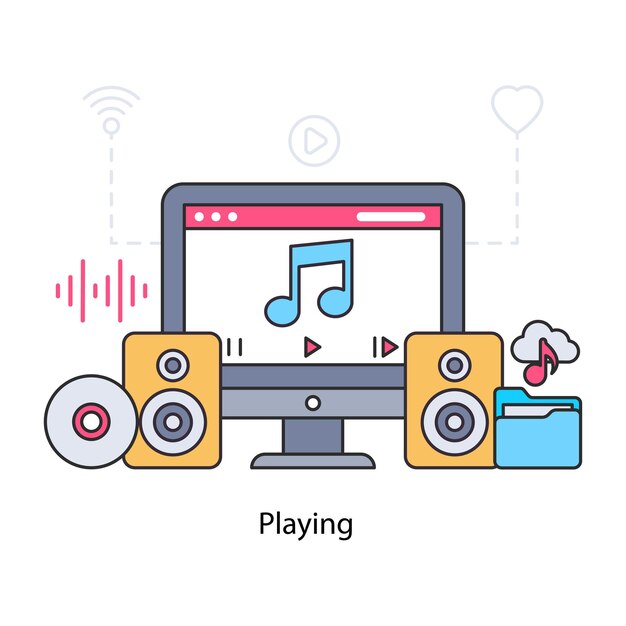
  Mentoring Week 2 - SQL Query

SQL and Relational Database - Job Preparation Program - Pacmann AI

****

source: [freepik](https://img.freepik.com/premium-vector/premium-download-illustration-playing-music_362714-95.jpg)

Task Description

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

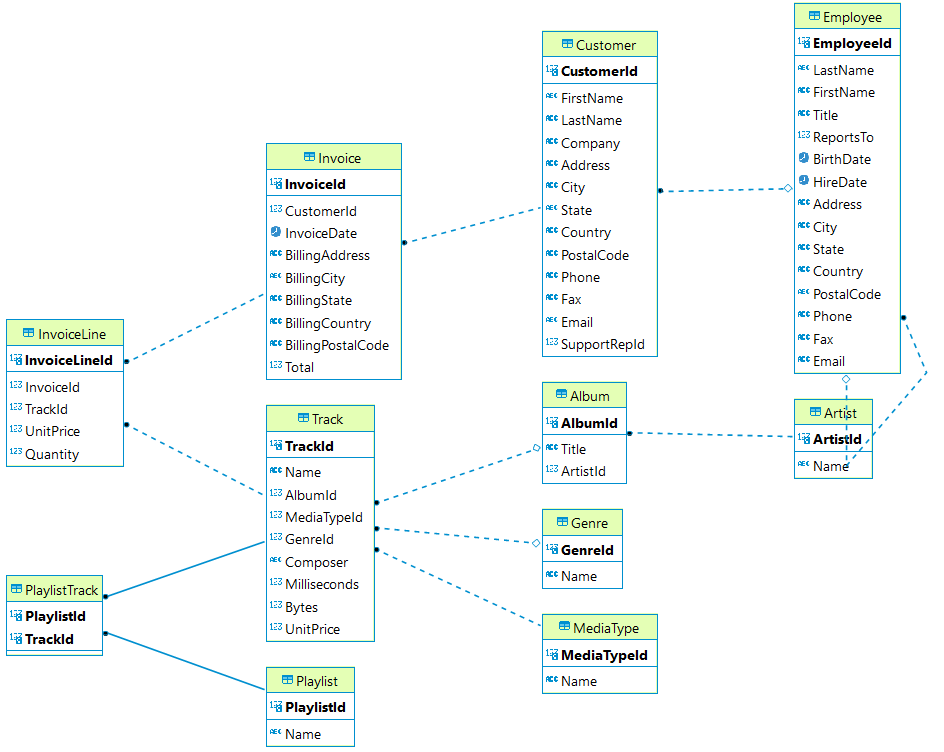
* Give some **insights of current transaction histories**.
* Give information for the Product team or marketing team related to music genres, artist, track, ect

The Dataset

* Download Dataset: [week 2](https://drive.google.com/drive/u/0/folders/13D_vkH_vtqh8LUPj1cIgJNA0IUMWvYI1) (*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("BillingCountry") as number\_of\_invoice**

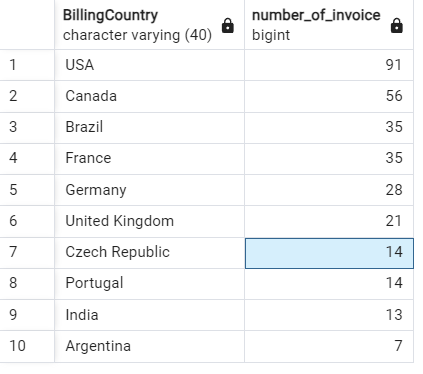
**from "Invoice"**

**Group by "BillingCountry"**

**Order By number\_of\_invoice desc, "BillingCountry" asc**

**Limit 10;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, saya mendapatkan 10 negara dengan jumlah invoice terbanyak**

1. (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**

**Gnr."Name" as Genres ,**

**sum(Inv."UnitPrice" \*Inv."Quantity") as Total\_Sales**

**from "InvoiceLine" Inv**

**Join "Track" using("TrackId")**

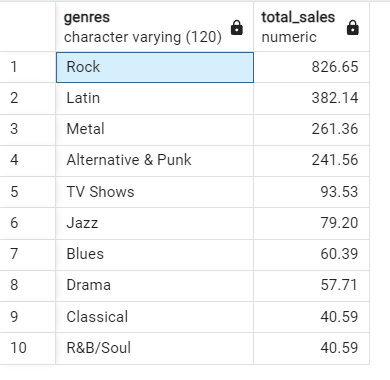
**Join "Genre" Gnr using("GenreId")**

**Group by 1**

**order by 2 desc**

**limit 10;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, saya mendapatkan genre terlaris berdasarkan total penjualan**

1. (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(Cust."FirstName",' ',Cust."LastName") as Customer\_Name,**

**"Email" as Email,**

**Inv."Total" as Total\_Spending**

**from "Customer" Cust**

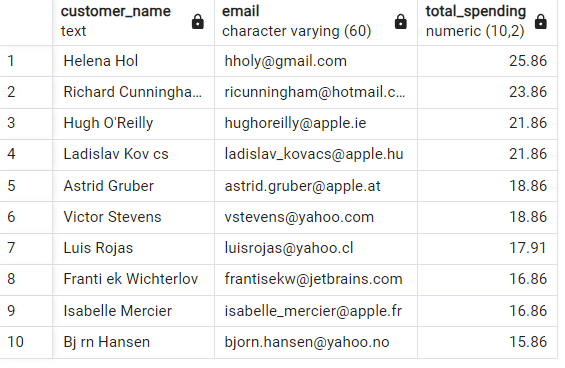
**Join "Invoice" Inv using ("CustomerId")**

**Group by 1, 2,3**

**order by 3 Desc**

**limit 10;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, saya mendapatkan data top 10 data pelanggan(nama, email dan total pembelian)**

1. (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 Country\_Name,**

**"BillingCity" as City\_Name,**

**Count("InvoiceId") as Total\_Number\_of\_Invoice**

**from "Invoice"**

**where "BillingCountry" IN (**

**select "BillingCountry"**

**from "Invoice"**

**Group by "BillingCountry"**

**Order By count("InvoiceId") desc, "BillingCountry" asc**

**Limit 10**

**)**

**group by 1,2**

**order by Total\_Number\_of\_Invoice desc,**

**City\_Name asc;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, saya mendapatkan top negara beserta kota pembelian terbanyak berdasarkan total invoice**

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

1. (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**

**Gnr."Name" as Genre,**

**"BillingCountry" as Country,**

**sum("InvoiceLine"."Quantity") as Track\_Purchase**

**from "Invoice"**

**Join "InvoiceLine" using ("InvoiceId")**

**Join "Track" Trc using ("TrackId")**

**Join "Genre" Gnr using ("GenreId")**

**where "BillingCountry" = 'United Kingdom'**

**Group by 1,2**

**order by 3 desc;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, kita dapat merekomendasikan 4 lagu yang terpopuler di UK berdasarkan genre:**

* 1. **"Good to See You": Rock**
  2. **"Get Ya Before Sunrise": Reggae**
  3. **"Nothing On You": Jazz**
  4. **"Before The Coffee Gets Cold": Hip Hop/Rap**

1. (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" as Title,**

**sum("InvoiceLine"."Quantity") as Quantity\_sold,**

**"Invoice"."BillingCountry" as Country**

**from "Invoice"**

**join "InvoiceLine" using ("InvoiceId")**

**join "Track" using ("TrackId")**

**join "Album" using ("AlbumId")**

**where "Invoice"."BillingCountry" = 'USA'**

**group by 1,3**

**order by 2 DESC, 1 ASC**

**limit 10;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, kita mendapatkan top ten album terlaris di USA**

1. (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 "Total Sales" as (**

**select**

**case**

**when "Customer"."Country" in (**

**select**

**"Country"**

**from "Customer"**

**group by 1**

**having count(distinct "CustomerId") = 1)**

**then 'Other'**

**else "Customer"."Country"**

**end as "Edited Country",**

**count(distinct "CustomerId") as "Total Customers",**

**sum("Invoice"."Total") as "Total Value of Sales",**

**sum("InvoiceLine"."Quantity") as "Total Quantity"**

**from "Invoice"**

**join "Customer" using ("CustomerId")**

**join "InvoiceLine" using ("InvoiceId")**

**group by 1**

**)**

**select**

**"Edited Country",**

**"Total Customers",**

**"Total Value of Sales",**

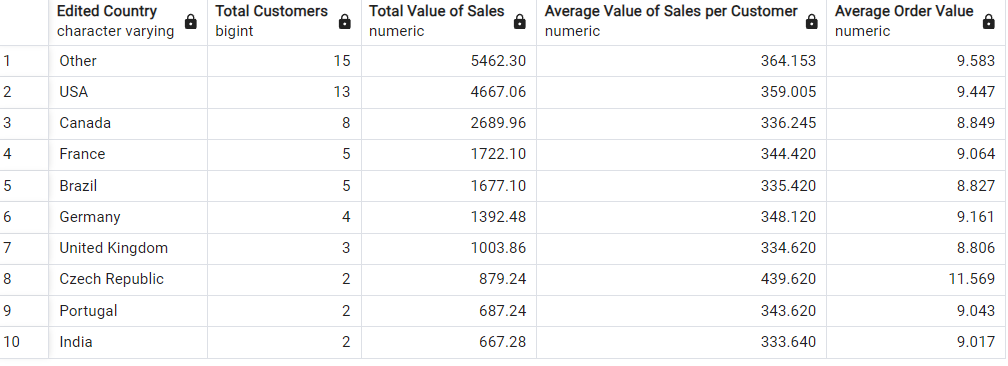
**round("Total Value of Sales"/"Total Customers",3) as "Average Value of Sales per Customer",**

**round("Total Value of Sales"/"Total Quantity",3) as "Average Order Value"**

**from "Total Sales"**

**order by 3 DESC;**

* **Screenshot of Query Results:**



* **Description of Query Results:**

**Pada query ini, kita mendapatkan total 9 negara dengan jumlah customer lebih dari satu. Dan juga table total customer per country, total penjualan, rata2 pembelian per customer dan rata2 nilai pembelian.**

1. (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**

**"BillingCountry" as "Country",**

**"Genre"."Name" as "Genre",**

**sum(IL."UnitPrice" \* IL."Quantity") as "Total Sales",**

**sum(IL."Quantity") as "Total Quantity"**

**from "InvoiceLine" IL**

**join "Invoice" Inv using ("InvoiceId")**

**join "Track" using ("TrackId")**

**join "Genre" using ("GenreId")**

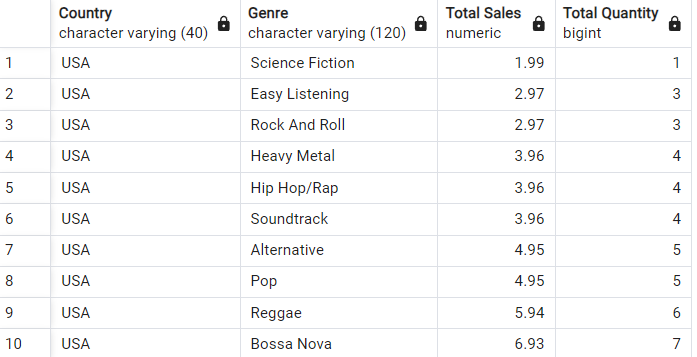
**where Inv."BillingCountry" = 'USA'**

**group by 1,2**

**order by 3 ASC, 2 ASC**

**limit 10;**

* **Screenshot of Query Results:**

****

* **Description of Query Results:**

**Pada query ini, kita mendapatkan 10 genre penjualan terendah di USA.**

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

1. (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 "Top Genre per Customer" as(**

**select**

**Concat("FirstName",' ',"LastName") as "Name Customer",**

**"Genre"."Name" as "Genre",**

**sum("Invoice"."Total") as "Total Spend",**

**dense\_rank() over(partition by Concat("FirstName",' ',"LastName")**

**order by sum("Invoice"."Total") DESC ) as "Rank"**

**from "Customer"**

**join "Invoice" using ("CustomerId")**

**join "InvoiceLine" using ("InvoiceId")**

**join "Track" using ("TrackId")**

**join "Genre" using ("GenreId")**

**group by 1,2**

**order by 1,3 DESC**

**)**

**select**

**"Name Customer",**

**"Genre",**

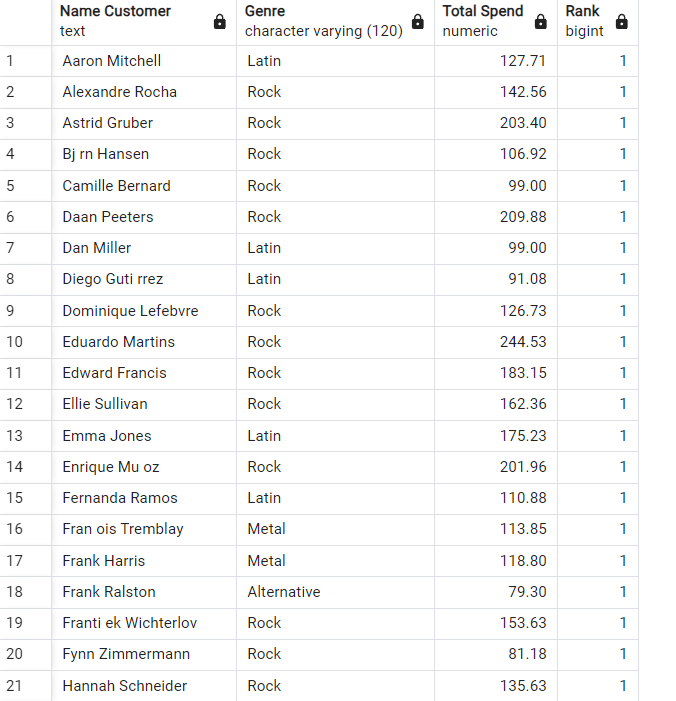
**"Total Spend",**

**"Rank"**

**from "Top Genre per Customer"**

**where "Rank" = 1;**

* **Screenshot of Query Results:**

****

* **Description of Query Results:**

**Pada query ini, kita mendapatkan top genre per customer berdasarkan total pembelian.**

1. (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**

**select**

**"BillingCountry" as "Country",**

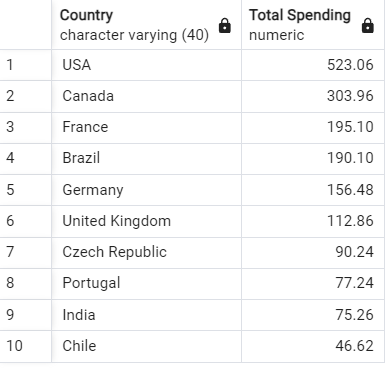
**sum("Total") as "Total Spending"**

**from "Invoice"**

**group by 1**

**order by 2 DESC**

**limit 10;**

* **Screenshot of Query Results:**
* 
* **Description of Query Results:**

**Pada query ini, kita mendapatkan top 10 negara berdasarkan penjulan tertinggi.**