# Music Store case study

Hi everyone! I'm excited to present a SQL case study where I dive into analysing data from a music store. Through a series of complex SQL queries and a little bit of pandas, I uncovered valuable insights that showcase my analytical skills and problem-solving abilities. Join me as I explore the data and reveal interesting findings!

To start with this is the Schema diagram.

There are 11 tables with different data related to music tracks, playlists, artists, employees, customers and invoices.
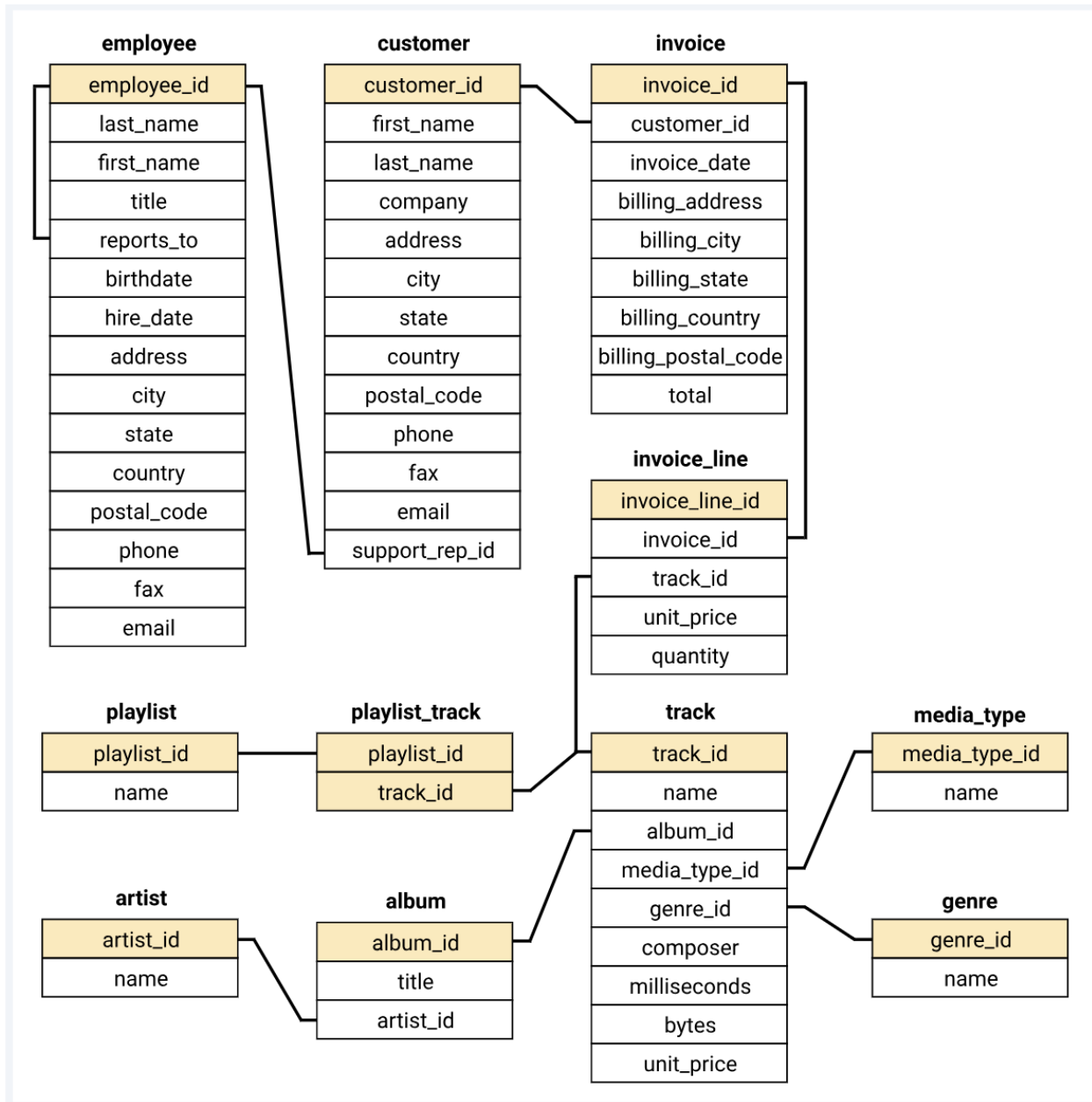
These are the problem statements I tackled, ranging from basic to advanced, to derive meaningful insights from the tables:

Basic:

1. **Customer Purchase Analysis** - Finding the total number of purchases made by each customer.

- Approach: Basic counting of purchases with straightforward joins:
- Use-case:
  - Finding the total number of purchases made by each customer is significant in various industries, particularly in retail, e-commerce, and subscription services.
  - This helps in identifying high-volume purchasers, businesses can classify customers into different segments (e.g., VIP customers, frequent buyers, occasional buyers).

2. **Playlist Popularity** - Counting how often songs appear in playlists.
   - Approach: Using the `playlist_track` and `track` tables to count how many times each track appears in playlists.
   - Use-case:
     - Identifying popular tracks across playlists can help build more personalised recommendations for users.
     - Curators creating themed playlists can reference the most frequently included tracks to see which songs are popular and suitable for inclusion.

3. **Employee Supervision** - Identify which employees report to specific supervisors.
   - Approach: Using the `employee` table and its `reports_to` column to list the employees reporting to each manager.
   - Use-case:
     - Understanding the reporting hierarchy within a company helps clarify the structure of teams and departments. This can be essential for organisational charts, restructuring efforts, or workflow optimisation.
     - Identifying which employees report to certain supervisors can help allocate resources more effectively, ensuring that supervisors aren't managing too many direct reports.

Moderate:

1. **Top-Selling Tracks** - Find the top 10 tracks with the highest sales.

   - Approach: Summing sales as sales summary and joining multiple tables. We can also use CTE if the sales summary as it can be reused with a different query.

   - Use case:

     - Evaluate artist performance: To assess which artists are driving sales and reward them accordingly.

     - Make informed marketing decisions: To focus marketing efforts on the most popular tracks and artists.

2. **Media Type Sales** - Compare sales of different media types (e.g., Audio vs video).

   - Approach - Summing sales by media type using a few joins.

   - Use case:

     - Companies can then tailor their approach to the platform that yields the best results. This data-driven approach ensures promotions align with consumer behaviour and maximise the return on investment (ROI)

3. **Genre Popularity** - Determine which genre has the highest number of tracks sold.

   - Approach: Similar to top-selling tracks, but with additional grouping by genre using a subquery or a CTE.

   - Use case:

     - Identifying popular genres

     - Evaluating artists performance

     - Personalise recommendations

4. **Invoice Trends** - Analyse monthly or yearly sales trends based on invoices.

   - Approach: Grouping data by time periods and summing up sales over month and year and just the year.

   - Use case:

- By understanding sales trends over time, businesses can make more accurate revenue forecasts and create realistic budget plans.

- Identifying renewal and churn patterns can lead to targeted customer retention strategies.

5. **Employee Sales Performance** - Identifying which employees (support representatives) are responsible for the highest total sales.

  - Approach: Joining multiple tables and calculating total sales for each employee.

  - Use case:

    - Performance evaluation and incentives.

    - Resource Allocation

Advanced:

1. **Customer Segmentation by Region** - Analyse sales by geographic region (city, state, country) to identify the regions with the highest customer activity.

  - Approach: Involves more detailed analysis by region and geographic grouping.

  - Use Case:

    - Targeted Marketing and Advertising.

    - Budget and Resource Allocation

2. **Customer Lifetime Value (CLV)** - Calculate the lifetime value of each customer based on their total purchases.

  - Approach: Aggregating over time to calculate total revenue per customer.

  - Use Case:

    - Calculating Customer Lifetime Value (CLV) helps businesses understand the long-term revenue each customer brings, guiding strategies for customer retention and targeted marketing efforts.

    - By analysing total purchases, companies can prioritise high-value customers and optimise resource allocation.

3. **Album Sales Insights :** Which albums are generating the most revenue?

   - Approach: Join the `album` , `track` , `invoice_line` , and `invoice` tables to calculate total sales. If we want to know album wise, First we should calculate the individual track sales using the previous CTE

   - Use case:

     - Identifying which albums generate the most revenue allows businesses to focus their marketing and production efforts on top-performing products.

     - This insight helps optimise inventory, promotional strategies, and investment in similar successful content.

4. **Track Duration and Sales Correlation** -

   - Approach: After exporting the required data from SQL using pandas to get the correlation.

   - Use case:

     - Analysing the correlation between track duration and sales helps determine if longer or shorter tracks are more profitable.

     - This insight can inform decisions on content production and marketing strategies to maximise revenue.