Music Store SQL Analysis Report

Author: Rahul Reddy
GitHub: https://github.com/RahuReddy/Music_Store_Sales_SQL_Analysis

Executive Summary

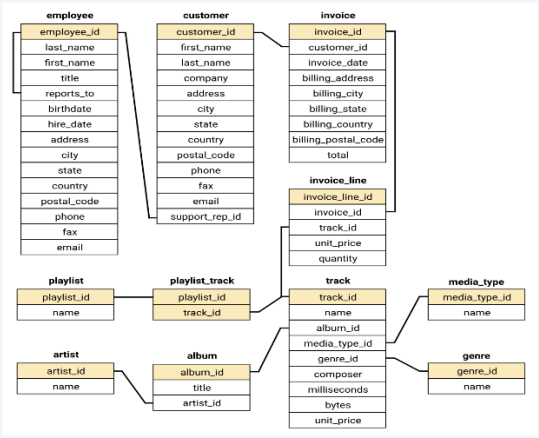
This report analyzes a music store's sales and operations using SQL queries. The objective was to translate raw transactional data into actionable insights for marketing, sales, and management. Key findings highlight top-performing genres, customers, media formats, and employees. Recommendations are provided to improve marketing focus, content acquisition, and customer retention.

1. Introduction

This project leverages SQL to analyze sales and customer data from a music store. The aim is to identify trends in revenue, customer behavior, product performance, and sales operations. Insights generated from this analysis help shape data-driven business strategies.

2. Dataset & Schema

The dataset consists of multiple tables including Customer, Employee, Invoice, Invoice Line, Playlist, Playlist Track, Album, Artist, Track, Genre, and Media Type. These tables are linked through primary and foreign keys.



3. Methodology

The analysis followed these steps:

1. Data Exploration of all tables
2. Problem Formulation based on schema
3. SQL Querying using `JOIN`, `GROUP BY`, `SUM`, `COUNT`, and window functions like `LAG`
4. Insight Generation
5. Recommendation Development

4. Key Business Questions

- Which top 10 countries are generating the most revenue, and what are their key sales metrics?
- Which are the top 10 best performing cities based on revenue?
- What are the top 5 most popular genres and artists by sales volume?
- Which media formats are most popular and profitable?
- Who are the top 10 most valuable customers?
- Which sales support agents are driving the most revenue?
- What does sales performance look like over time (yearly and monthly)?
- Which top 10 albums and tracks are selling well?
- What is the purchasing behavior of repeat customers vs. less frequent customers?
- What is the total revenue generated by less repeated customers vs more repeated customers?
- What is the average customer lifetime value for each sales support agent?

5. Analysis & Results

This section presents the SQL queries generated from the dataset. Each subsection focuses on sales & customer insights, product & content insights, and employee & operational insights.

Sales & Customer Insights

- The United States is the top revenue-generating country, with a total revenue of **$1,049.49** and **131 sales**, making it the most significant market.
- The top-performing city by revenue is **Prague** (**$273.24**), followed by Mountain View, London, and Berlin.
- The most valuable customers are **František Wichterlová** and **Helena Holy**, with lifetime revenues of **$144.54** and **$128.70** respectively.

- Repeat customers drive more revenue than new ones: **$3,268.98 (69.41%)** of total revenue comes from customers with more than 10 purchases.

Content & Product Insights

- **Rock** is the most popular genre, generating **$2,608.65** in revenue (**64.71%** of total), followed by Latin and Metal.
- Top-selling artists: **Queen**, **Jimi Hendrix**, and **Nirvana**.
- The **MPEG audio format** is the most profitable, generating **$4,216.41**.
- Best-selling album: *Cake: B-Sides and Rarities*.
- Best-selling track: *War Pigs*.

Employee & Operational Insights

- The top-performing sales agent is **Jane Peacock**, generating **$1,731.51** in revenue and achieving the highest average customer lifetime value.
- The year **2019** was the most profitable (**$1,221.66**), with **March** as the best-performing month.

6. Recommendations

Strategic Marketing & Geographic Focus

- Prioritize the **US market** and top-performing cities (Prague, Mountain View, London).
- Explore new urban markets such as Berlin, Paris, São Paulo, Dublin, and Delhi.

Customer Loyalty & Retention

- Launch a **tiered loyalty program** for high-value customers (e.g., František Wichterlová, Helena Holy).
- Offer **discounts for first-time customers** to encourage repeat purchases.

Content & Inventory Management

- Acquire more **Rock, Metal, and Alternative/Punk** content.
- Focus on top artists (Queen, Jimi Hendrix) with exclusive content or promotions.
- Ensure new acquisitions are available in the **MPEG format**.

Sales & Operational Efficiency

- Recognize and reward top-performing agents like Jane Peacock.
- Implement a **training program** to spread best practices from top performers.
- Use **customer lifetime value (CLV) per agent** as a performance metric.

7. Conclusion

This project demonstrates my ability to apply SQL in a complete, end-to-end data analysis workflow from problem formulation to actionable recommendations. The analysis provided tangible insights into the music store's market dynamics, customer behavior, and operational efficiency, highlighting my technical skills and business acumen. These insights could further be used to build interactive dashboards or predictive models.

8. Limitations

- The dataset is static and may not reflect current or future market conditions.
- Results may vary in a real-world setting where additional external factors (marketing campaigns, economic shifts) influence sales.

9. Future Work

1. **Interactive Dashboards**

    - Develop real-time dashboards in Power BI so stakeholders can filter revenue, customers, and products by country, genre, or sales agent.
    - Business Value: Enables managers to spot opportunities and issues quickly without waiting for static reports.

2. **Cohort & Retention Tracking**

    - Build monthly or quarterly cohort analysis to track how customer groups behave over time.
    - Business Value: Provides insight into retention trends and the effectiveness of promotional activities.

Appendix

- Full SQL scripts used in the analysis.

- Top 10 countries are generating the most revenue.

```sql
-- 1.Which top 10 countries are generating the most revenue,
-- and what are their key sales metrics?

SELECT TOp 10 c.country,
       ROUND(SUM(i.total),2) AS total_revenue,
       COUNT(i.invoice_id) AS total_sales
FROM invoice i
JOIN customer c
ON i.customer_id = c.customer_id
GROUP BY c.country
ORDER BY 2 DESC, 3 DESC;
```

No issues found

Results | Messages

	country	total_revenue	total_sales
1	USA	1040.49	131
2	Canada	535.59	76
3	Brazil	427.68	61
4	France	389.07	50
5	Germany	334.62	41
6	Czech Republic	273.24	30
7	United Kingdom	245.52	28
8	Portugal	185.13	29
9	India	183.15	21
10	Ireland	114.84	13

- Top 10 best performing cities based on revenue.

```sql
-- 2. Find out top 10 best performing cities based on Revenue

SELECT TOP 10 c.country, c.city,
       ROUND(SUM(i.total),2) AS total_revenue
FROM invoice i
JOIN customer c ON i.customer_id = c.customer_id
GROUP BY c.country, c.city
ORDER BY total_revenue DESC;
```

No issues found

Results | Messages

	country	city	total_revenue
1	Czech Republic	Prague	273.24
2	USA	Mountain View	169.29
3	United Kingdom	London	166.32
4	Germany	Berlin	158.4
5	France	Paris	151.47
6	Brazil	São Paulo	129.69
7	Ireland	Dublin	114.84
8	India	Delhi	111.87
9	Brazil	São José dos Campos	108.9
10	Brazil	Brasília	106.92

- Top 5 most popular genres and artists by sales volume.

```sql
-- 3. What are the top 5 most popular genres and artists by sales volume?
-- Top 5 Most Popular genres by sales volume
WITH CTE AS (
SELECT TOP 5 g.name AS Genre_name,
       ROUND(SUM(il.unit_price),2) AS total_revenue
FROM invoice_line il
JOIN track t ON il.track_id = t.track_id
JOIN genre g ON t.genre_id = g.genre_id
JOIN album a ON t.album_id = a.album_id
JOIN artist ar ON a.artist_id = ar.artist_id
GROUP BY g.name
ORDER BY total_revenue DESC)

SELECT *,
CONCAT(ROUND((total_revenue *100.0 / SUM(total_revenue) OVER()),2),'%') AS percentage0ftotal
FROM CTE
;
```

No issues found

Results | Messages

	Genre_name	total_revenue	percentage0ftotal
1	Rock	2608.65	64.71%
2	Metal	612.81	15.2%
3	Alternative & Punk	487.08	12.08%
4	Latin	165.33	4.1%
5	R&B/Soul	157.41	3.9%

```sql
-- Top 5 most popular Artists by total quantity and total revenue

SELECT TOP 5 ar.name,
       SUM(quantity) AS total_quantity,
       ROUND(SUM(il.unit_price),2) AS total_revenue
FROM invoice_line il
JOIN track t ON il.track_id = t.track_id
JOIN genre g ON t.genre_id = g.genre_id
JOIN album a ON t.album_id = a.album_id
JOIN artist ar ON a.artist_id = ar.artist_id
GROUP BY ar.name
ORDER BY total_quantity DESC, total_revenue DESC;
```

No issues found

Results | Messages

	name	total_quantity	total_revenue
1	Queen	192	190.08
2	Jimi Hendrix	187	185.13
3	Nirvana	130	128.7
4	Red Hot Chili Peppers	130	128.7
5	Pearl Jam	129	127.71

- These Media formats are most popular and profitable.

```sql
-- 4.Find out which media formats are most popular and profitable.

SELECT mt.name AS Media_Type,
       ROUND(SUM(il.unit_price*il.quantity),2) AS total_revenue,
       COUNT(il.quantity) AS total_quantity_sold
FROM invoice_line il
JOIN track t ON il.track_id = t.track_id
JOIN media_type mt ON t.media_type_id = mt.media_type_id
GROUP BY mt.name
ORDER BY total_revenue DESC;
```

100 % No issues found

Results Messages

	Media_Type	total_revenue	total_quantity_sold
1	MPEG audio file	4216.41	4259
2	Protected AAC audio file	434.61	439
3	Purchased AAC audio file	34.65	35
4	AAC audio file	20.79	21
5	Protected MPEG-4 video file	2.97	3

- Top 10 most valuable customers.

```sql
-- 5. Top 10 most valuable customers.

SELECT TOP 10 first_name, last_name, ROUND(SUM(total),2) AS total_revenue
FROM customer c
JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY first_name, last_name
ORDER BY total_revenue DESC;
```

100 % No issues found

Results Messages

	first_name	last_name	total_revenue
1	František	Wichterlová	144.54
2	Helena	Holý	128.7
3	Hugh	O'Reilly	114.84
4	Manoj	Pareek	111.87
5	Luís	Gonçalves	108.9
6	Fernanda	Ramos	106.92
7	João	Fernandes	102.96
8	Wyatt	Girard	99.99
9	François	Tremblay	99.99
10	Phil	Hughes	98.01

- These sales support agents are driving the most revenue.

```sql
-- 6. Identify which sales support agents are driving the most revenue.

SELECT e.first_name, e.last_name, ROUND(SUM(i.total),2) AS total_revenue
FROM customer c
JOIN employee e ON c.support_rep_id = e.employee_id
JOIN invoice i ON c.customer_id = i.customer_id
GROUP BY e.first_name, e.last_name
ORDER BY total_revenue DESC;
```

100 % No issues found

Results Messages

	first_name	last_name	total_revenue
1	Jane	Peacock	1731.51
2	Margaret	Park	1584
3	Steve	Johnson	1393.92

- Yearly and Monthly sales performance

```sql
-- 7. Identifying sales performance over time.

-- Yearly Revenue
SELECT YEAR(invoice_date) AS Year,
       ROUND(SUM(total),2) AS total_revenue
FROM invoice
GROUP BY YEAR(invoice_date)
ORDER BY total_revenue DESC;

-- Monthly Revenue

SELECT
       MONTH(invoice_date) AS Month,
       ROUND(SUM(total),2) AS total_revenue
FROM invoice
GROUP BY MONTH(invoice_date)
ORDER BY total_revenue DESC;
```

100 % — No issues found

Results | Messages

	Year	total_revenue
1	2019	1221.66
2	2017	1201.86
3	2018	1147.41
4	2020	1138.5

	Month	total_revenue
1	3	456.39
2	4	442.53
3	1	438.57
4	8	426.69
5	2	414.81
6	7	395.01
7	9	386.1
8	6	380.16
9	5	368.28
10	12	364.32
11	10	345.51
12	11	291.06

- Best Selling top 10 albums and tracks.

```sql
-- 8. Which Top 10 albums and tracks are selling well
-- to ensure they are well-stocked and promoted.

SELECT TOP 10 a.title, t.name,
       COUNT(il.quantity) AS total_quantity
FROM invoice_line il
JOIN track t ON il.track_id = t.track_id
JOIN album a ON t.album_id = a.album_id
JOIN artist ar ON a.artist_id = ar.artist_id
GROUP BY a.title, t.name
ORDER BY total_quantity DESC;
```

100 % — No issues found

Results | Messages

	title	name	total_quantity
1	Cake: B-Sides and Rarities	War Pigs	31
2	Are You Experienced?	Are You Experienced?	14
3	Are You Experienced?	Highway Chile	14
4	Are You Experienced?	Hey Joe	13
5	For Those About To Rock We Salute You	Put The Finger On You	13
6	Are You Experienced?	Third Stone From The Sun	13
7	Are You Experienced?	Love Or Confusion	12
8	Mezmerize	Radio/Video	12
9	Are You Experienced?	51st Anniversary	11
10	Faceless	Dead And Broken	11

- Purchasing behavior of repeat customers vs. less frequent customers.

```sql
-- 9. Identifying purchasing behavior of more repeated customers versus less repeated customers.
-- Most Repeating Customers greater than 10 times

SELECT customer_id, COUNT(*) AS repeat_count
FROM invoice
GROUP BY customer_id
HAVING COUNT(*) > 10
ORDER BY repeat_count DESC;


-- Less Repeating Customers less than 10 times

SELECT customer_id, COUNT(*) AS repeat_count
FROM invoice
GROUP BY customer_id
HAVING COUNT(*) < 10
ORDER BY repeat_count;
```

100 % — No issues found

Results | Messages

	customer_id	repeat_count
1	5	18
2	35	16
3	13	15
4	1	13
5	30	13
6	34	13
7	46	13
8	57	13

	customer_id	repeat_count
1	29	4
2	56	5
3	8	7
4	16	8
5	18	8
6	24	8
7	59	8
8	52	8
9	32	8

- Total revenue generated by less repeated customers vs more repeated customers.

```sql
-- 10. Find out the total revenue generated by less repeated customers
-- versus more repeated customers.(less repeated like less than 10)

WITH CustomerPurchaseCount AS (
SELECT customer_id,
       COUNT(invoice_id) AS total_purchases
FROM invoice
GROUP BY customer_id)

SELECT
     CASE
         WHEN cpc.total_purchases < 10 THEN 'less_repeated_customers'
         ELSE 'more_repeated_customers'
     END AS customer_type,
     COUNT(cpc.customer_id) AS total_customers,
     ROUND(SUM(i.total),2) AS total_revenue
FROM invoice i
JOIN CustomerPurchaseCount cpc
    ON i.customer_id = cpc.customer_id
GROUP BY CASE
         WHEN cpc.total_purchases < 10 THEN 'less_repeated_customers'
         ELSE 'more_repeated_customers'
     END
ORDER BY total_revenue DESC;
```

100 % — No issues found

Results | Messages

	customer_type	total_customers	total_revenue
1	more_repeated_customers	442	3268.98
2	less_repeated_customers	172	1440.45

- Average customer lifetime value for each sales support agent.

```sql
-- 11.What is the average customer lifetime value for each sales support agent.

SELECT
e.first_name,
e.last_name,
ROUND(SUM(i.total),2) AS total_revenue_generated,
ROUND(AVG(i.total),2) AS avg_invoice_value,
COUNT(DISTINCT c.customer_id) AS total_customers_served
FROM invoice i
JOIN customer c ON i.customer_id = c.customer_id
JOIN employee e ON c.support_rep_id = e.employee_id
GROUP BY e.first_name,e.last_name
ORDER BY total_revenue_generated DESC;
```

100 % — No issues found

Results | Messages

	first_name	last_name	total_revenue_generated	avg_invoice_value	total_customers_served
1	Jane	Peacock	1731.51	8.17	21
2	Margaret	Park	1584	7.4	20
3	Steve	Johnson	1393.92	7.41	18

- Additional query results not included in the main body.
- The Complete set of SQL queries is available in the GitHub Repository under Music_Store_Sales_Queries.sql