

# Music Store Sales Analysis

## 1. Executive Summary

This project analyzes transactional sales data from a digital music store using **PostgreSQL** to derive actionable business insights. The analysis focuses on **revenue trends, customer behavior, geographic performance, genre popularity, and artist contribution**.

Key findings indicate that revenue is concentrated in specific countries and cities, customer retention is exceptionally strong, Rock music dominates global sales, and revenue dependency on top artists exists but remains balanced. These insights provide a strong foundation for **data-driven marketing, catalog optimization, and customer engagement strategies**.

---

## 2. Business Problem

The music store operates across multiple countries with a diverse catalog of artists, albums, and genres. While sales data is available, the business lacks clarity on:

- Whether revenue is growing or seasonal
- Which markets and customers generate the most value
- What genres and artists drive profitability
- Whether revenue is overly dependent on a small subset of artists
- How music preferences differ across countries

Without structured analysis, decisions related to **marketing investment, promotions, catalog expansion, and pricing** remain inefficient and risk-prone.

---

## 3. Objectives

The primary objectives of this analysis are to:

1. Evaluate overall sales performance and trends
  2. Identify top-performing markets, cities, and customers
  3. Understand customer retention and value
  4. Analyze genre and artist performance
  5. Assess revenue concentration and business risk
  6. Provide actionable recommendations for growth and optimization
- 

#### 4. Dataset Overview

The dataset is a relational database consisting of the following core tables:

- **Customer** – customer demographics and location
- **Invoice / Invoice\_Line** – purchase transactions
- **Track / Album / Artist** – music catalog structure
- **Genre / Media\_Type** – music classification
- **Employee** – internal sales support structure
- **Playlist** – curated collections

The schema enforces **primary and foreign key constraints** to ensure data integrity.

---

#### 5. Tools & Technologies

- **Database:** PostgreSQL
- **Language:** SQL
- **Techniques Used:**
  - Joins (INNER, LEFT)
  - Aggregations

- Window Functions (RANK)
  - Common Table Expressions (CTEs)
  - Date-based analysis (DATE\_TRUNC)
- 

## 6. Methodology

1. Designed and validated a relational database schema
  2. Imported CSV data using staging tables and data cleaning techniques
  3. Wrote business-focused SQL queries
  4. Prioritized insights by business relevance
  5. Interpreted results from a stakeholder perspective.
- 

## 7. Analysis & SQL Queries

### Q1. Are sales increasing or decreasing over time?

SELECT

DATE\_TRUNC('month', invoice\_date) AS month,

SUM(total) AS revenue

FROM invoice

GROUP BY month

ORDER BY month;

	month timestamp without time zone	revenue double precision
1	2017-01-01 00:00:00	126.71999999999998
2	2017-02-01 00:00:00	141.57000000000002
3	2017-03-01 00:00:00	103.95
4	2017-04-01 00:00:00	142.55999999999997
5	2017-05-01 00:00:00	104.94
6	2017-06-01 00:00:00	75.24000000000001
7	2017-07-01 00:00:00	108.89999999999999
8	2017-08-01 00:00:00	88.11000000000001
9	2017-09-01 00:00:00	107.91
10	2017-10-01 00:00:00	79.2
11	2017-11-01 00:00:00	94.05
12	2017-12-01 00:00:00	28.70999999999997
13	2018-01-01 00:00:00	183.14999999999998
Total rows: 48		Query complete 00:00:00.088

### Insight:

Revenue shows **clear seasonality** with recurring peaks and dips rather than steady growth. Early-year and late-year months consistently perform better, while mid-year periods tend to underperform.

This pattern suggests strong potential for **time-based promotions** to stabilize and increase overall sales.

### Q2. Which countries generate the most revenue?

```
SELECT billing_country, SUM(total) AS revenue
FROM invoice
GROUP BY billing_country
ORDER BY revenue DESC;
```

	<b>billing_country</b> character varying (30) 	<b>revenue</b> double precision 
1	USA	1040.489999999998
2	Canada	535.5900000000001
3	Brazil	427.6800000000006
4	France	389.0699999999999
5	Germany	334.62
6	Czech Republic	273.24000000000007
7	United Kingdom	245.52
8	Portugal	185.13000000000002
9	India	183.1499999999998

### Insight:

Revenue is **heavily concentrated in the USA**, which generates nearly double the revenue of the next largest market, Canada.

North America and select European countries dominate sales, while many other countries contribute relatively small but steady revenue.

Emerging markets like **Brazil and India** show strong potential for growth with targeted pricing and localized marketing strategies.

---

### Q3.Which cities have the best customers?

```
SELECT billing_city, SUM(total) AS total_revenue
FROM invoice
GROUP BY billing_city
ORDER BY total_revenue DESC
LIMIT 5;
```

	<b>billing_city</b> character varying (30) 	<b>total_revenue</b> double precision 
1	Prague	273.24000000000007
2	Mountain View	169.29
3	London	166.32
4	Berlin	158.4
5	Paris	151.47

### **Insight:**

Prague generates the **highest total revenue**, indicating a highly valuable and engaged customer base in that city.

Mountain View, London, Berlin, and Paris also show strong spending, highlighting them as **key urban markets** for targeted promotions.

These cities are ideal candidates for **localized marketing campaigns, events, or premium offerings** to maximize revenue impact.

---

### **Q4. Who are the best customers overall?**

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

	<b>customer_id</b> [PK] character varying (30)	<b>first_name</b> character varying (30)	<b>last_name</b> character varying (30)	<b>total_spent</b> double precision
1	5	František	Wichterlová	144.54000000000002
2	6	Helena	Holý	128.7
3	46	Hugh	O'Reilly	114.8399999999997
4	58	Manoj	Pareek	111.86999999999999
5	1	Luís	Gonçalves	108.89999999999998

### **Insight:**

František Wichterlová is the **highest-spending customer**, followed closely by Helena Holý and Hugh O'Reilly, indicating a small group of **high-value customers** at the top.

Customer spending declines gradually rather than sharply, suggesting a **healthy spread of mid-value customers** rather than reliance on a single buyer.

Targeted loyalty programs for the top spenders could significantly increase retention and overall lifetime value.

---

### **Q5. Repeat vs one-time customers**

```
SELECT
```

```

COUNT(*) FILTER (WHERE cnt = 1) AS one_time_customers,
COUNT(*) FILTER (WHERE cnt > 1) AS repeat_customers
FROM (
  SELECT customer_id, COUNT(*) AS cnt
  FROM invoice
  GROUP BY customer_id
) sub;

```

	one_time_customers	repeat_customers
	bigint	bigint
1	0	59

### Insight:

All customers in the dataset are **repeat customers**, with none making only a single purchase. This indicates **exceptionally strong customer retention** and consistent engagement with the store.

The business should focus on **upselling and loyalty rewards** rather than basic retention campaigns.

---

### Q6.Top customer per country

```

WITH customer_spending AS (
  SELECT
    c.customer_id,
    c.first_name,
    c.last_name,
    i.billing_country,
    SUM(i.total) AS total_spent,
    RANK() OVER (
      PARTITION BY i.billing_country
      ORDER BY SUM(i.total) DESC
    ) AS rank_num
)

```

```

FROM invoice i
JOIN customer c ON c.customer_id = i.customer_id
GROUP BY c.customer_id, c.first_name, c.last_name, i.billing_country
)
SELECT *
FROM customer_spending
WHERE rank_num = 1;

```

	<b>customer_id</b> character varying (30) 	<b>first_name</b> character varying (30) 	<b>last_name</b> character varying (30) 	<b>billing_country</b> character varying (30) 	<b>total_spent</b> double precision 	<b>rank_num</b> bigint 
1	56	Diego	Gutiérrez	Argentina	39.6	1
2	55	Mark	Taylor	Australia	81.18	1
3	7	Astrid	Gruber	Austria	69.3	1
4	8	Daan	Peeters	Belgium	60.38999999999999	1
5	1	Luís	Gonçalves	Brazil	108.8999999999998	1
6	3	François	Tremblay	Canada	99.99	1
7	57	Luis	Rojas	Chile	97.0200000000001	1
8	5	František	Wichterlová	Czech Republic	144.5400000000002	1
9	9	Kara	Nielsen	Denmark	37.61999999999999	1
10	44	Terhi	Hämäläinen	Finland	79.2	1

### Insight:

Each country has a clear top-spending customer, with no ties at the highest level of spending. Top customers in countries like the Czech Republic, Ireland, India, and Brazil spend significantly more than others, highlighting strong individual contributors within those markets. Identifying and rewarding these country-level VIP customers presents an opportunity for highly targeted, personalized retention and upselling strategies.

---

### Q7.Revenue by genre

```

SELECT g.name, SUM(il.unit_price * il.quantity) AS revenue
FROM invoice_line il
JOIN track t ON t.track_id = il.track_id
JOIN genre g ON g.genre_id = t.genre_id
GROUP BY g.name

```

ORDER BY revenue DESC;

	name character varying (50) 	revenue numeric 
1	Rock	2608.65
2	Metal	612.81
3	Alternative & Punk	487.08
4	Latin	165.33
5	R&B/Soul	157.41
6	Blues	122.76
7	Jazz	119.79
8	Alternative	115.83
9	Easy Listening	73.26
10	Pop	62.37
11	Electronica/Dance	54.45
12	Classical	46.53
13	Reggae	34.65
14	Hip Hop/Rap	32.67
15	Heavy Metal	7.92
16	Soundtrack	4.95
17	TV Shows	1.98
18	Drama	0.99

#### Insight:

Rock overwhelmingly dominates revenue, generating **more than four times** the revenue of the next highest genre, Metal.

A steep drop exists after the top three genres, indicating that **most revenue is driven by a small subset of genres**.

Lower-revenue genres represent niche interests and could benefit from **bundling or targeted promotions** rather than broad marketing.

---

#### Q8.Most popular genre by country

WITH genre\_purchases AS (

SELECT

```

c.country,
g.name AS genre,
COUNT(il.quantity) AS purchases,
RANK() OVER (
    PARTITION BY c.country
    ORDER BY COUNT(il.quantity) DESC
) AS rank_num

FROM invoice_line il
JOIN invoice i ON i.invoice_id = il.invoice_id
JOIN customer c ON c.customer_id = i.customer_id
JOIN track t ON t.track_id = il.track_id
JOIN genre g ON g.genre_id = t.genre_id
GROUP BY c.country, g.name
)

SELECT country, genre, purchases
FROM genre_purchases
WHERE rank_num = 1;

```

	country character varying (30) 	genre character varying (50) 	purchases bigint 
1	Argentina	Alternative & Punk	17
2	Australia	Rock	34
3	Austria	Rock	40
4	Belgium	Rock	26
5	Brazil	Rock	205
6	Canada	Rock	333
7	Chile	Rock	61
8	Czech Republic	Rock	143
9	Denmark	Rock	24
10	Finland	Rock	46

### **Insight:**

Rock is the **most popular genre in nearly every country**, often by a wide margin, confirming it as the store's **core global revenue driver**.

Argentina is the only exception, where **Alternative & Punk** slightly outperforms Rock, highlighting a unique regional preference.

This consistency suggests global Rock-focused promotions will be effective, while **localized genre campaigns** can further optimize sales in select markets.

---

### **Q9.Genre diversity by country**

```
SELECT billing_country, COUNT(DISTINCT genre_id) AS genre_diversity  
FROM invoice  
JOIN invoice_line USING (invoice_id)  
JOIN track USING (track_id)  
GROUP BY billing_country  
ORDER BY genre_diversity DESC;
```

	<b>billing_country</b> character varying (30)	<b>genre_diversity</b> bigint
1	USA	17
2	Canada	15
3	France	14
4	Germany	14
5	Czech Republic	14
6	Portugal	14
7	Finland	13
8	Brazil	13
9	Ireland	12
10	United Kingdom	12

### **Insight:**

The USA shows the **highest genre diversity**, indicating a mature market with broad and varied music preferences.

Countries like Canada, France, Germany, and the Czech Republic also display high diversity, suggesting openness to experimenting with different genres.

Markets with lower genre diversity (e.g., Denmark, Austria) may respond better to **focused, genre-specific strategies** rather than wide catalog expansion.

---

### Q10. Most popular artists

```
SELECT ar.name AS artist_name, COUNT(il.quantity) AS purchases
FROM invoice_line il
JOIN track t ON t.track_id = il.track_id
JOIN album al ON al.album_id = t.album_id
JOIN artist ar ON ar.artist_id = al.artist_id
GROUP BY ar.name
ORDER BY purchases DESC
LIMIT 10;
```

	artist_name character varying (150)	locked	purchases bigint	locked
1	Queen		192	
2	Jimi Hendrix		187	
3	Nirvana		130	
4	Red Hot Chili Peppers		130	
5	Pearl Jam		129	
6	Guns N' Roses		124	
7	AC/DC		124	
8	Foo Fighters		121	
9	The Rolling Stones		117	
10	Metallica		106	

#### Insight:

A small group of **iconic rock artists** dominates overall purchases, with Queen and Jimi Hendrix leading by a clear margin.

The top 10 artists are closely clustered in purchase counts, indicating **strong but competitive**

**demand** among leading bands.

This highlights both the importance of maintaining these high-performing artists and the opportunity to **diversify promotions** to avoid over-reliance on a few names.

---

### **Q11.Revenue concentration (Top 10 artists)**

WITH artist\_revenue AS (

```
SELECT ar.name, SUM(il.unit_price * il.quantity) AS revenue
```

```
FROM invoice_line il
```

```
JOIN track t ON t.track_id = il.track_id
```

```
JOIN album al ON al.album_id = t.album_id
```

```
JOIN artist ar ON ar.artist_id = al.artist_id
```

```
GROUP BY ar.name
```

```
)
```

```
SELECT
```

```
SUM(revenue) FILTER (WHERE rank <= 10) * 100.0 / SUM(revenue) AS top10_percentage
```

```
FROM (
```

```
SELECT *, RANK() OVER (ORDER BY revenue DESC) AS rank
```

```
FROM artist_revenue
```

```
) r;
```

	top10_percentage numeric
1	28.5894471305444608

#### **Insight:**

The **top 10 artists contribute ~28.6% of total revenue**, indicating noticeable but not extreme revenue concentration.

While these artists are important drivers, the majority of revenue is still generated by the **long tail of other artists**, reducing dependency risk.

This balanced distribution allows the business to **leverage top artists for promotions** while safely investing in catalog diversification.

---

## Q12. Most popular songs

SELECT

```
ar.name AS artist_name,  
t.name AS song_name,  
al.title AS album_name,  
COUNT(il.quantity) AS purchases  
  
FROM invoice_line il  
  
JOIN track t ON t.track_id = il.track_id  
  
JOIN album al ON al.album_id = t.album_id  
  
JOIN artist ar ON ar.artist_id = al.artist_id  
  
GROUP BY ar.name, t.name, al.title  
  
ORDER BY purchases DESC  
  
LIMIT 3;
```

	artist_name character varying (150) 	song_name character varying (250) 	album_name character varying (150) 	purchases bigint 
1	Cake	War Pigs	Cake: B-Sides and Rariti...	31
2	Jimi Hendrix	Are You Experienced?	Are You Experienced?	14
3	Jimi Hendrix	Highway Chile	Are You Experienced?	14

### Insight:

“War Pigs” by **Cake** is the most popular track by a wide margin, indicating a **clear standout song** that drives disproportionate engagement.

Jimi Hendrix appears twice in the top results, showing **strong album-level appeal**, where multiple tracks from the same album attract listeners.

This suggests opportunities to promote **high-performing songs individually** while also leveraging **successful albums and artists** for bundled or curated recommendations.

---

## 8. Key Findings

### 1. Revenue shows strong seasonality rather than steady growth

Sales peak during early and late-year months, while mid-year periods underperform, indicating predictable seasonal buying behavior.

### 2. Revenue is geographically concentrated

The USA dominates total revenue, followed by Canada and select European countries. Emerging markets such as Brazil and India contribute meaningful revenue and show growth potential.

### 3. Customer retention is exceptionally strong

All customers in the dataset are repeat buyers, reflecting consistent engagement and a loyal customer base.

### 4. A small group of customers and cities drive disproportionate revenue

High-value customers and cities like Prague, London, and Mountain View generate significantly higher spending than others.

### 5. Rock music is the dominant genre globally

Rock generates the highest revenue and is the most popular genre in nearly every country, with only minor regional exceptions.

### 6. Markets differ in genre maturity

Countries such as the USA and Canada show high genre diversity, while others prefer a narrower range of genres.

### 7. Artist revenue concentration is moderate

The top 10 artists account for ~28.6% of revenue, indicating influence without excessive dependency.

### 8. A few tracks and albums drive high engagement

Songs like “War Pigs” and albums by Jimi Hendrix attract disproportionate attention and sales.

---

## 9. Business Recommendations

### 1. Implement time-based promotional strategies

Align marketing campaigns with historically high-performing months and introduce discounts or bundles during low-revenue periods to stabilize sales.

## **2. Protect core markets while scaling emerging ones**

Maintain strong investment in North America and top European markets, while applying localized pricing and promotions in Brazil and India to accelerate growth.

## **3. Target high-value cities with localized campaigns**

Launch city-specific promotions, premium offerings, or digital events in top-revenue cities to maximize return on marketing spend.

## **4. Shift focus from retention to upselling and loyalty**

Introduce tiered loyalty programs and personalized recommendations to increase average order value among already-retained customers.

## **5. Establish country-level VIP programs**

Reward top-spending customers in each country with exclusive benefits, early access, or personalized offers to strengthen long-term relationships.

## **6. Maintain Rock as the flagship genre while monetizing niches**

Continue global Rock-focused promotions while bundling or targeting niche genres to smaller, specific audiences.

## **7. Apply market-specific catalog strategies**

Expand genre offerings in high-diversity markets and concentrate on dominant genres in low-diversity markets to improve conversion rates.

## **8. Diversify artist exposure to reduce long-term risk**

Promote emerging and mid-tier artists through playlists and recommendations to balance reliance on top performers.

## **9. Leverage top tracks and albums for engagement**

Feature best-selling songs and successful albums prominently in playlists and promotions to drive incremental sales.

---

## Conclusion

This project demonstrates how **SQL-driven analysis** can convert raw transactional data into **actionable business insights**.

The findings highlight strong customer loyalty, clear revenue drivers, and distinct geographic and genre-based patterns. While Rock music and a handful of artists dominate sales, revenue remains sufficiently diversified to support sustainable growth.

By adopting **data-driven, localized, and customer-centric strategies**, the music store can improve revenue consistency, reduce dependency risk, and enhance customer lifetime

value.

Overall, this analysis provides a solid foundation for **strategic decision-making in marketing, catalog management, and market expansion.**