

# Chinook Database

---

Prepared By:  
Sama Ahmed

# 1. Introduction

This project analyzes the **Chinook Music Store dataset** (Kaggle), using queries in SQLite. The objectives are to:

- Explore sales, customers, and genres
- Identify revenue drivers by product, region, and time
- Apply advanced SQL (JOINS, window functions)
- Deliver actionable insights for business decisions

---

## 2. Step 1 – Sample Invoices

### SQL Query:

```
SELECT InvoiceId, CustomerId, InvoiceDate, BillingAddress,  
       BillingCity, BillingState, BillingCountry, BillingPostalCode, Total  
FROM Invoice  
LIMIT 5;
```

### Output (sample):

InvoiceId	CustomerId	InvoiceDate	City	Country	Total
1	2	2009-01-01	Stuttgart	Germany	1.98
2	4	2009-01-02	Oslo	Norway	3.96
3	8	2009-01-03	Brussels	Belgium	5.94
4	14	2009-01-06	Edmonton	Canada	8.91
5	23	2009-01-11	Boston	USA	13.86

**Observation:** Invoices record sales, customers, and regions. Small basket sizes (\$2–14) suggest track-based purchases.

---

### 3. Step 2 – Top-Selling Products (Revenue) SQL Query:

```
SELECT t.TrackId, t.Name AS TrackName, a.Title AS Album,  
       ar.Name AS Artist, SUM(il.UnitPrice * il.Quantity) AS Revenue,  
       SUM(il.Quantity) AS Quantity  
FROM InvoiceLine il  
JOIN Track t ON il.TrackId = t.TrackId  
JOIN Album a ON t.AlbumId = a.AlbumId  
JOIN Artist ar ON a.ArtistId = ar.ArtistId  
GROUP BY t.TrackId  
ORDER BY Revenue DESC  
LIMIT 10;
```

**Output (sample):**

TrackId	TrackName	Artist	Album	Revenue	Qty
2832	The Woman King	Battlestar Galactica	Battlestar Galactica, S3	3.98	2
2850	The Fix	Heroes	Heroes, Season 1	3.98	2
2868	Walkabout	Lost	Lost, Season 1	3.98	2
3177	Hot Girl	The Office	The Office, Season 1	3.98	2

**Observation:** Premium TV show tracks (\$1.99) dominate revenue. They're usually bought twice, generating \$3.98 per title.

---

## 4. Step 3 – Top-Selling Products (Quantity)

Output (sample):

TrackId	TrackName	Qty	Revenue
2	Balls to the Wall	2	1.98
8	Inject The Venom	2	1.98
20	Overdose	2	1.98
32	Deuces Are Wild	2	1.98
84	Welcome Home	2	1.98

**Observation:** Classic Rock/Metal tracks (AC/DC, Aerosmith, Metallica) dominate sales by quantity. Price is lower (\$0.99), so revenue per track is smaller than TV shows.

---

## 5. Step 4 – Revenue per Region

Output (top regions):

Country	Invoices	Revenue
USA	91	523.06
Canada	56	303.96
France	35	195.10

**Observation:** USA (31%) and Canada (18%) = nearly 50% of revenue. France, Brazil, Germany, UK are secondary markets

---

## 6. Step 5 – Monthly Performance

Output (sample):

Month	Invoices	Revenue
2009-01	6	35.64
2009-02	7	37.62
2010-01	7	52.62
2011-04	7	51.62
2011-11	6	23.7

Observation:

- Baseline: ~\$37/month, 7 invoices
  - Spikes: Jan 2010, Apr 2011, Jun 2011, Nov 2013
  - Dip: Nov 2011 (\$23.76)
- 

## 7. Step 6 – Revenue per Region per Month

Observation:

- USA & Canada provide stable revenue every month.
- Other regions show spikes (e.g., Czech Republic \$25.86 in Nov 2013, Brazil \$22.77 in Aug 2013).
- Smaller countries = occasional large invoice

## 8. Step 7 – Genre Performance (with Ranking Bonus) **SQL**

### Query with RANK:

```
SELECT g.Name AS Genre,  
       COUNT(il.InvoiceLineId) AS UnitsSold,  
       SUM(il.UnitPrice * il.Quantity) AS Revenue,  
       RANK() OVER (ORDER BY SUM(il.UnitPrice * il.Quantity) DESC) AS Rank  
FROM InvoiceLine il  
JOIN Track t ON il.TrackId = t.TrackId  
JOIN Genre g ON t.GenreId = g.GenreId  
GROUP BY g.Name  
ORDER BY Revenue DESC;
```

### Output (top genres):

Rank	Genre	Units	Revenue
1	Rock	835	826.65
2	Latin	386	382.14
3	Metal	264	261.36
4	Alternative & Punk	244	241.56
5	TV Shows	47	93.53

### Observation:

- Rock = 48% of all revenue, foundation genre.
- Latin, Metal, Alt/Punk follow.
- TV Shows small in volume, but high per-unit pricing.

## 9. Step 8 – Customer Segmentation (Bonus)

Output (top customers):

CustomerId	Name	Country	Invoices	Revenue
6	Helena Holý	Czech Republic	7	49.62
26	Richard Cunningham	USA	7	47.62
57	Luis Rojas	Chile	7	46.62
45	Ladislav Kovács	Hungary	7	45.62

**Observation:**

- All top 10 customers spend ~\$42–50, with 7 invoices each.
  - High-value customers are geographically diverse (USA, Europe, South America).
  - Loyalty = steady repeat purchasing, not one-off big orders.
-



## 10. Final Insights & Recommendations

### Insights

1. **Revenue vs. Volume mismatch:** Rock drives volume, but premium TV/Drama tracks bring higher margin.
2. **Geography:** USA + Canada = 49% of revenue. But top individual spenders are spread globally.
3. **Monthly pattern:** Very stable baseline, occasional spikes from single customers.
4. **Genres:** Rock = foundation, TV/Drama = strategic niche.
5. **Customers:** Loyal repeat buyers exist, but no extreme “whales”.

### Recommendations

- **Double down on Rock** for volume-driven campaigns (bundles, playlists).
- **Upsell premium TV/Drama tracks** — high margin with consistent niche demand.
- **Focus on North America** for a reliable baseline, but nurture high-value individuals abroad.
- **Personalize offers** to repeat buyers (7 invoices = clear loyalty signal).
- Explore **regional promotions** in Brazil, Czech Republic, Ireland to capture emerging high-value markets.