

Data Analysis Project: Exploring the Chinook Database

Objective

The objective of this project is to perform exploratory data analysis (EDA) on the Chinook database to showcase proficiency in SQL by utilizing aggregate functions, joins, subqueries, and window functions. This project will uncover meaningful insights related to revenue, customer behavior, artist and track performance, and salesperson efficiency.

Dataset Overview

The Chinook database is a sample database used for educational purposes. It contains data related to a digital media store. Below is a brief overview of the main tables used:

- **artist:** Contains information about artists (*ArtistId, Name*).
- **album:** Represents albums with their respective artists (*AlbumId, ArtistId, Title*).
- **track:** Tracks with metadata such as album, genre, and pricing (*TrackId, Name, AlbumId, GenreId, UnitPrice*).
- **genre:** Different genres of tracks (*GenreId, Name*).
- **customer:** Information about customers (*CustomerId, Name, Country, Email, etc.*).
- **invoice:** Records of customer purchases (*InvoiceId, CustomerId, InvoiceDate, Total*).
- **invoiceline:** Line items within invoices, detailing tracks sold (*InvoiceLineId, InvoiceId, TrackId, Quantity*).
- **employee:** Employee information, including sales support roles (*EmployeeId, FirstName, LastName, Title*).

SQL Analysis

1. Record Counts in Each Table

To understand the database size:

```
1 SELECT 'artist' AS table_name, COUNT(*) AS record_count FROM artist
2 UNION ALL
3 SELECT 'album', COUNT(*) FROM album
4 UNION ALL
5 SELECT 'customer', COUNT(*) FROM customer
6 UNION ALL
7 SELECT 'track', COUNT(*) FROM track
8 UNION ALL
9 SELECT 'invoice', COUNT(*) FROM invoice;
```

Insight: Displays the number of records in each table for context.

table_name	record_count
artist	275
album	347
customer	59
track	3503
invoice	412

2. Total Revenue From All Sales

```
1 SELECT SUM(total) AS total_revenue FROM invoice;
```

Insight: The total revenue generated by all invoices is \$2328.60.

3. Average Purchase Amount Per Invoice

```
1 SELECT AVG(total) AS avg_invoice_amount FROM invoice;
```

Insight: The average revenue per customer invoice is \$5.651942.

4. Monthly Revenue Trends

```
1 SELECT
2     DATE_FORMAT(InvoiceDate, '%Y-%m') AS invoice_month,
3     SUM(Total) AS monthly_revenue
4 FROM invoice
5 GROUP BY invoice_month
6 ORDER BY invoice_month DESC;
```

Insight: Identifies revenue trends by month, highlighting peak and off-peak months.

invoice_month	monthly_revenue
2025-12	38.62
2025-11	49.62
2025-10	37.62
2025-09	37.62
2025-08	37.62
2025-07	37.62
2025-06	37.62
2025-05	37.62
2025-04	33.66
2025-03	37.62
2025-02	27.72
2025-01	37.62
2024-12	37.62
2024-11	37.62
2024-10	42.62
2024-09	46.71
2024-08	47.62
2024-07	39.62
2024-06	37.62
2024-05	37.62
2024-04	37.62
2024-03	37.62
2024-02	37.62
2024-01	37.62
2023-12	37.62
2023-11	23.76
2023-10	37.62
2023-09	37.62
2023-08	37.62
2023-07	37.62

Table 1: Monthly Revenue Trends 1

invoice_month	monthly_revenue
2023-06	50.62
2023-05	42.62
2023-04	51.62
2023-03	37.62
2023-02	37.62
2023-01	37.62
2022-12	37.62
2022-11	37.62
2022-10	37.62
2022-09	36.63
2022-08	37.62
2022-07	37.62
2022-06	37.62
2022-05	37.62
2022-04	37.62
2022-03	44.62
2022-02	46.62
2022-01	52.62
2021-12	37.62
2021-11	37.62
2021-10	37.62
2021-09	37.62
2021-08	37.62
2021-07	37.62
2021-06	37.62
2021-05	37.62
2021-04	37.62
2021-03	37.62
2021-02	37.62
2021-01	35.64

Table 2: Monthly Revenue Trends 2

5. Top 5 Customers by Spending

```

1 SELECT
2     c.CustomerId,
3     CONCAT(c.FirstName, ' ', c.LastName) AS customer_name,
4     SUM(i.Total) AS total_spent
5 FROM customer c
6 JOIN invoice i ON c.CustomerId = i.CustomerId
7 GROUP BY c.CustomerId
8 ORDER BY total_spent DESC
9 LIMIT 5;

```

Insight: Lists the highest-spending customers, which can guide targeted marketing strategies.

CustomerId	customer_name	total_spent
6	Helena Holý	49.62
26	Richard Cunningham	47.62
57	Luis Rojas	46.62
45	Ladislav Kovács	45.62
46	Hugh O'Reilly	45.62

6. Revenue by Country

```

1 SELECT
2     c.Country,
3     SUM(i.Total) AS total_revenue
4 FROM customer c
5 JOIN invoice i ON c.CustomerId = i.CustomerId
6 GROUP BY c.Country
7 ORDER BY total_revenue DESC;

```

Country	total_revenue
USA	523.06
Canada	303.96
France	195.10
Brazil	190.10
Germany	156.48
United Kingdom	112.86
Czech Republic	90.24
Portugal	77.24
India	75.26
Chile	46.62
Hungary	45.62
Ireland	45.62
Austria	42.62
Finland	41.62
Netherlands	40.62
Norway	39.62
Sweden	38.62
Belgium	37.62
Denmark	37.62
Italy	37.62
Poland	37.62
Spain	37.62
Australia	37.62
Argentina	37.62

Insight: Identifies which countries contribute the most revenue.

7. Most Popular Genre

```
1 SELECT
2     g.name AS genre,
3     COUNT(il.quantity) AS tracks_sold
4 FROM genre g
5 JOIN track t ON g.GenreId = t.GenreId
6 JOIN invoice il ON t.TrackId = il.TrackId
7 GROUP BY g.GenreId
8 ORDER BY tracks_sold DESC;
```

genre	tracks_sold
Rock	835
Latin	386
Metal	264
Alternative & Punk	244
Jazz	80
Blues	61
TV Shows	47
R&B/Soul	41
Classical	41
Reggae	30
Drama	29
Pop	28
Soundtrack	20
Sci Fi & Fantasy	20
Hip Hop/Rap	17
Bossa Nova	15
Alternative	14
World	13
Heavy Metal	12
Electronica/Dance	12
Easy Listening	10
Comedy	9
Rock And Roll	6
Science Fiction	6

Insight: Highlights the genres that drive the most sales, useful for inventory planning.

8. Top 5 Most Sold Tracks

```
1 SELECT
2     t.Name AS track_name,
3     SUM(il.Quantity) AS total_sales
4 FROM track t
5 JOIN invoice il ON t.TrackId = il.TrackId
6 GROUP BY t.TrackId
7 ORDER BY total_sales DESC
8 LIMIT 5;
```

track_name	total_sales
Balls to the Wall	2
Deuces Are Wild	2
Inject The Venom	2
Not The Doctor	2
Overdose	2

Insight: Shows the most popular tracks by sales volume.

9. Top 5 Artists by Number of Tracks

```
1 SELECT
2     ar.Name AS artist_name,
3     COUNT(t.TrackId) AS total_tracks
4 FROM artist ar
5 JOIN album al ON ar.ArtistId = al.ArtistId
6 JOIN track t ON al.AlbumId = t.AlbumId
7 GROUP BY ar.ArtistId
8 ORDER BY total_tracks DESC
9 LIMIT 5;
```

artist_name	total_tracks
Iron Maiden	213
U2	135
Led Zeppelin	114
Metallica	112
Lost	92

Insight: Identifies the most prolific artists in the store.

10. Artist With the Highest Album Sales

```
1 SELECT
2     a.Name AS artist_name,
3     SUM(il.quantity) AS total_tracks_sold
4 FROM artist a
5 JOIN album al ON a.ArtistId = al.ArtistId
6 JOIN track t ON al.AlbumId = t.AlbumId
7 JOIN invoiceline il ON t.TrackId = il.TrackId
8 GROUP BY a.ArtistId
9 ORDER BY total_tracks_sold DESC
10 LIMIT 1;
```

artist_name	total_tracks_sold
Iron Maiden	140

Insight: The artist Iron Maiden has generated the most sales of 140 tracks in total.

11. Albums With the Highest Sales

```
1 SELECT
2     al.Title AS album_title,
3     ar.Name AS artist_name,
4     SUM(il.Quantity) AS total_sales
5 FROM album al
6 JOIN artist ar ON al.ArtistId = ar.ArtistId
7 JOIN track t ON al.AlbumId = t.AlbumId
8 JOIN invoiceline il ON t.TrackId = il.TrackId
9 GROUP BY al.AlbumId
10 ORDER BY total_sales DESC
11 LIMIT 5;
```

album_title	artist_name	total_sales
Minha Historia	Chico Buarque	27
Greatest Hits	Lenny Kravitz	26
Unplugged	Eric Clapton	25
Acústico	Titãs	22
Greatest Kiss	Kiss	20

Insight: Highlights the best-selling albums.

12. Customer Lifetime Value and Spending Tiers

```
1 WITH customer_spending AS (
2     SELECT
3         c.CustomerId,
4         CONCAT(c.FirstName, ' ', c.LastName) AS customer_name,
5         SUM(i.Total) AS total_spent
6     FROM customer c
7     JOIN invoice i ON c.CustomerId = i.CustomerId
8     GROUP BY c.CustomerId
```

```

9 )
10 SELECT
11     customer_name ,
12     total_spent ,
13     CASE
14         WHEN total_spent > 46 THEN 'High'
15         WHEN total_spent BETWEEN 39 AND 46 THEN 'Medium'
16         ELSE 'Low'
17     END AS spending_tier
18 FROM customer_spending
19 ORDER BY total_spent DESC;

```

Insight: Segments customers based on spending levels to inform retention strategies.

customer_name	total_spent	spending_tier
Helena Holý	49.62	High
Richard Cunningham	47.62	High
Luis Rojas	46.62	High
Ladislav Kovács	45.62	Medium
Hugh O'Reilly	45.62	Medium
Julia Barnett	43.62	Medium
Fynn Zimmermann	43.62	Medium
Frank Ralston	43.62	Medium
Astrid Gruber	42.62	Medium
Victor Stevens	42.62	Medium
Terhi Hämäläinen	41.62	Medium
František Wichterlová	40.62	Medium
Isabelle Mercier	40.62	Medium
Johannes Van der Berg	40.62	Medium
Luís Gonçalves	39.62	Medium
Jack Smith	39.62	Medium
François Tremblay	39.62	Medium
Bjørn Hansen	39.62	Medium
Dan Miller	39.62	Medium
Heather Leacock	39.62	Medium
João Fernandes	39.62	Medium
Wyatt Girard	39.62	Medium
Jennifer Peterson	38.62	Low
Tim Goyer	38.62	Low
Camille Bernard	38.62	Low
Dominique Lefebvre	38.62	Low
Joakim Johansson	38.62	Low
Manoj Pareek	38.62	Low
Robert Brown	37.62	Low
John Gordon	37.62	Low
Martha Silk	37.62	Low
Aaron Mitchell	37.62	Low
Ellie Sullivan	37.62	Low

Table 3: Customer Spending Tier Data 1

customer_name	total_spent	spending_tier
Leonie Köhler	37.62	Low
Madalena Sampaio	37.62	Low
Hannah Schneider	37.62	Low
Edward Francis	37.62	Low
Niklas Schröder	37.62	Low
Kara Nielsen	37.62	Low
Eduardo Martins	37.62	Low
Marc Dubois	37.62	Low
Alexandre Rocha	37.62	Low
Roberto Almeida	37.62	Low
Fernanda Ramos	37.62	Low
Mark Philips	37.62	Low
Daan Peeters	37.62	Low
Lucas Mancini	37.62	Low
Frank Harris	37.62	Low
Stanisław Wójcik	37.62	Low
Enrique Muñoz	37.62	Low
Michelle Brooks	37.62	Low
Emma Jones	37.62	Low
Phil Hughes	37.62	Low
Steve Murray	37.62	Low
Mark Taylor	37.62	Low
Diego Gutiérrez	37.62	Low
Kathy Chase	37.62	Low
Patrick Gray	37.62	Low
Puja Srivastava	36.64	Low

Table 4: Customer Spending Tier Data 2

13. Best-Performing Salesperson

```

1 SELECT
2     CONCAT(e.FirstName, ' ', e.LastName) AS employee_name,
3     COUNT(i.InvoiceId) AS total_sales,
4     SUM(i.total) AS total_revenue
5 FROM employee e
6 JOIN customer c ON e.EmployeeId = c.SupportRepId
7 JOIN invoice i ON c.CustomerId = i.CustomerId
8 GROUP BY e.EmployeeId
9 ORDER BY total_revenue DESC;

```

employee_name	total_sales	total_revenue
Jane Peacock	146	833.04
Margaret Park	140	775.40
Steve Johnson	126	720.16

Insight: Identifies the salesperson generating the most revenue and closing the most sales.

Conclusions

1. Revenue Insights and Trends

- **Total Revenue:** The Chinook store has generated a total revenue of \$2,328.60.
- **Average Purchase Amount:** The average purchase amount per invoice is \$5.65, which offers insight into customer spending behavior.
- **Monthly Revenue Trends:** Monthly revenue exhibits fluctuations over time, with peaks observed in months like January 2022 (\$52.62) and April 2023 (\$51.62). Revenue remains relatively consistent across other months, indicating steady sales activity.
- **Overall Revenue Stability:** While revenue is stable, the slight variations suggest opportunities to capitalize on high-revenue months through targeted campaigns or seasonal promotions.

2. Customer Insights

- **Top Customers:** The top five customers by spending include Helena Holý (\$49.62), Richard Cunningham (\$47.62), and Luis Rojas (\$46.62). These customers are prime targets for loyalty programs and promotions.
- **Revenue by Country:** The USA (\$523.06) and Canada (\$303.96) contribute the highest revenues, while other countries like France, Brazil, and Germany also show significant contributions. This distribution highlights opportunities for focused marketing campaigns in high-revenue regions.
- **Spending Tiers:** Customers are segmented into three tiers—High, Medium, and Low spending:
 - High spenders like Helena Holý and Richard Cunningham contribute significantly, spending over \$45 each.
 - Medium spenders dominate the customer base, offering a large segment for upselling opportunities.
 - Low spenders provide an untapped segment where strategies to boost engagement could yield higher revenue.
- **Customer Lifetime Value (CLV):** Understanding CLV can inform retention strategies for high-value customers and engagement initiatives for low-tier customers.

3. Genre and Artist Performance

- **Most Popular Genre:** Rock dominates as the most popular genre with 835 tracks sold, followed by Latin (386 tracks) and Metal (264 tracks). These findings emphasize the importance of expanding offerings in these genres to meet customer demand.
- **Top Artists:** Iron Maiden leads with the highest number of tracks (213) and the highest album sales (140 tracks sold), showcasing their strong customer appeal. Other top artists include U2 and Led Zeppelin.

- **Top Albums:** The album *Minha Historia* by Chico Buarque ranks as the best-selling album with 27 tracks sold, followed by *Greatest Hits* by Lenny Kravitz (26 tracks sold) and *Unplugged* by Eric Clapton (25 tracks sold).
- **Top Artists and Albums:** Iron Maiden dominates in terms of tracks and album sales. Albums like *Minha Historia* by Chico Buarque and *Greatest Hits* by Lenny Kravitz also perform well, underlining the significance of artist and album appeal.
- **Top Tracks:** Tracks like "Balls to the Wall," "Deuces Are Wild," and "Inject The Venom" have achieved the highest sales, each with two units sold.

4. Sales and Employee Efficiency

- **Best-Performing Salespersons:** Jane Peacock leads as the top-performing salesperson, generating \$833.04 in revenue with 146 sales, followed by Margaret Park (\$775.40) and Steve Johnson (\$720.16). These results reflect exceptional customer engagement and sales efficiency.
- **Best Salesperson:** Jane Peacock leads in total sales and revenue, demonstrating exceptional performance. Margaret Park and Steve Johnson also show strong results. These findings suggest that sharing best practices among sales staff could enhance overall efficiency.

5. Strategic Recommendations

- **Enhance Employee Training:** Leverage top-performing salespeople to design training programs aimed at improving the effectiveness of other employees.
- **Targeting High-Revenue Regions:** Strengthen promotional strategies in the USA, Canada, and other high-revenue regions to capitalize on existing customer bases.
- **Boost Revenue in Low Months:** Identify factors contributing to lower revenue months and implement seasonal campaigns to improve sales consistency.
- **Expand Top-Performing Genres:** With Rock, Latin, and Metal leading sales, prioritize marketing and acquisitions in these categories.
- **Customer Retention Programs:** Engage top-spending customers with loyalty rewards and personalized offers to enhance retention.
- **Target High and Medium Spending Tiers:** Focus retention programs on high-tier customers while designing personalized upselling strategies for medium spenders.