Objective Questions

1. Does any table have missing values or duplicates? If yes how would you handle it ?

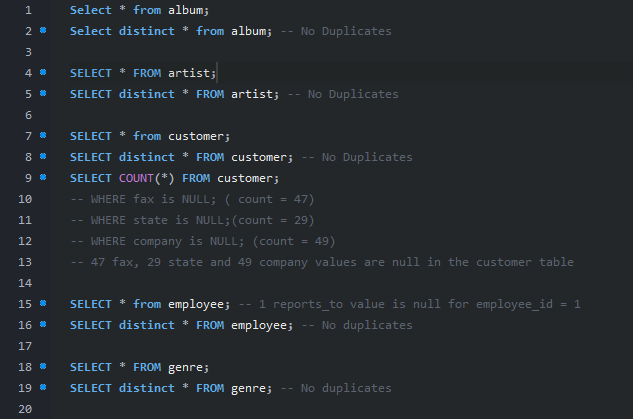
The data provided possess 0 duplicates although there are missing values in 3 tables.

**Customer table**: 47 fax, 29 state and 49 company values are null in the customer table

**Employee table**: 1 reports to value is null for employee\_id = 1

**Track table**: 978 composers are not assigned any value/are null in the track table

These Null values could be handled easily by using the Coalesce function in SQL.



A screenshot of a computer

Description automatically generated

1. Find the top-selling tracks and top artist in the USA and identify their most famous genres.

Top-Selling Tracks in the USA

The Top selling track in USA is ‘War Pigs’ with a total sales of 5.94 .

A screenshot of a computer

Description automatically generated

Top Artist in the USA

The Top Artist in USA is ‘Van Halen’ with sales of 42.57 .

A screenshot of a computer program

Description automatically generated

Most Famous Genre of the Top Artist

The most Famous of top artist ‘Van Helen’ is rock

A screenshot of a computer program

Description automatically generated A screenshot of a computer program

Description automatically generated



1. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?

Chinook music store customer base is spread across 24 countries. USA is the country with most number of customers .

A screenshot of a computer

Description automatically generated

1. Calculate the total revenue and number of invoices for each country, state, and city:

Czech republic is the highest in terms of revenue with 273.24

and Denmark lands the lowest.

A screenshot of a computer

Description automatically generated

1. Find the top 5 customers by total revenue in each country

The Top 5 Customers are

‘Diego’

‘Mark’

‘Astrid’

‘Daan’

‘Luis’

A screenshot of a computer

Description automatically generated

1. Identify the top-selling track for each customer

The top selling track for customer is ‘Put on your Lights’(track name).

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a music album

Description automatically generated

1. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?

The average order a customer takes is around 8-9 with the least one is 4

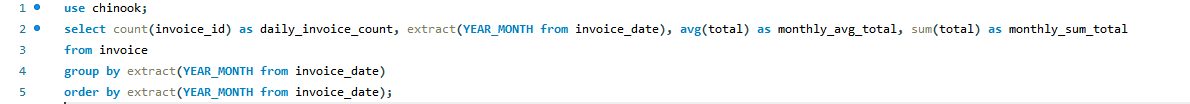
A computer code with black text

Description automatically generated

A table with numbers and text

Description automatically generated

The monthly avg total and monthly sum is likely same around from January to November but there’s a Drop in December values.



A screenshot of a data

Description automatically generated

1. What is the customer churn rate?

**Overall Churn Rate :**

Formula used:

Churn Rate = Total Churned Customers\*100​/Average Customer Base

Total Churned Customers = Churned in Year 2018 + Churned in Year 2019 + Churned in Year 2020 = 2 + 4 + 5 = 11

A close-up of a website

Description automatically generated A screenshot of a computer

Description automatically generated A close-up of a website

Description automatically generated

Average Customer Base =

(Customers at Start of Year 2018 + Customers at Start of Year 2019 + Customers at Start of Year 2020)/3

= (57+55+54)/3 = 55.33

A black text on a white background

Description automatically generated A close-up of a computer code

Description automatically generated A close-up of a computer screen

Description automatically generated

**Churn Rate** = 11 \*100/55.33 = 19.88%

1. Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.

The best selling genre in ‘USA’ is ROCK and best artist is ‘Van Helen’ (ref. Q2)

A screenshot of a computer code

Description automatically generated

A screenshot of a music list

Description automatically generated

1. Find customers who have purchased tracks from at least 3 different genres

Leonie Kohler being the top buyer and there are total 59 customers who have bought atleast 3 different genre.

A screen shot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Rank genres based on their sales performance in the USA

Rock genre is being in the top of the list

And followed by rock, Punk and metal are in 2nd and 3rd position

A screenshot of a computer code

Description automatically generated

A screenshot of a music list

Description automatically generated

1. Identify customers who have not made a purchase in the last 3 months

There are total of 24 customer who have not purchased anything from last 3 months.

A screenshot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

Subjective Questions

1. Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.

The top 3 Rock albums are "From The Muddy Banks Of The Wishkah [live]", "Are You Experienced?", and "The Best Of Van Halen, Vol. I".

A screenshot of a computer screen

Description automatically generated

1. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

These are the top selling genres excluding the country ’USA’

A screenshot of a computer

Description automatically generated

1. Customer Purchasing Behavior Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?

Calculated the average of customer period of association with chinook

Divided the customers based on their period of association:If any customer is associated term > avg of all customers, then he/she is a long term customer. Otherwise, short term.

Calculated the other required metrics for analysis between short term and long term customers like, total expenditure, items bought, frequency of purchase.

A screenshot of a computer

Description automatically generated



1. Product Affinity Analysis: Which music genres, artists, or albums are frequently purchased together by customers? How can this information guide product recommendations and cross-selling initiatives?

Different genres purchased over an invoice

A screenshot of a computer

Description automatically generated

Different albums purchased over an invoice

A screenshot of a music list

Description automatically generated

Different artists prefered in a single invoice

A screenshot of a computer

Description automatically generated

1. Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?

From the below data we can see that countries like the USA, Canada, Germany, UK, Australia, Brazil have either very low churn rate or negative churn rate. As these nations come under developed economies with high GDP, this could be a factor supporting the trend in the data.

On the other hand, developing nations or less developed nations like Norway, Netherlands, Chile, Denmark, India,etc. Have comparatively higher churn rate depicting decreasing scope of business.



1. Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?

Inactivity: Customers with a higher gap of purchase period between two purchases are more likely to churn.

Low Purchase Frequency: Customers with fewer total purchases might be less engaged.

Low Total Spending: Customers with lower total spending could be at risk of reducing spending further.

Demographics: Certain age groups or gender segments in specific countries might display higher churn or lower spending compared to others. The effect of the economy on churn rate has been analysed in the previous question.

A screenshot of a computer

Description automatically generated A computer screen shot of a computer

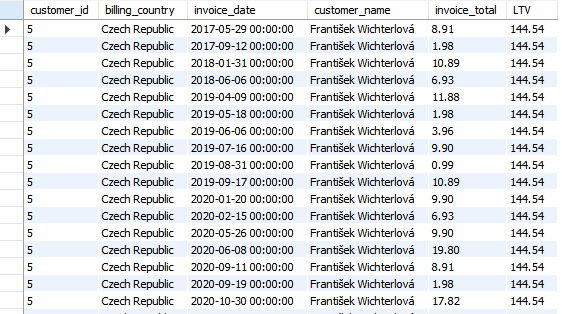
Description automatically generated

1. Customer Lifetime Value Modeling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns among customers who have stopped purchasing?

Customer’s data such as purchase history, tenure, engagement,etc can be utilised to analyse traits such as the loyalty, high value/low value, frequent buyer or not ,etc

Recent customers with high purchase value can be future loyal customers and therefore, targeted marketing or promotional campaigns can be deployed for such potential buyers.

Customers can be segmented based on their purchase history and therefore, targeted marketing campaign can be deployed for such customers.



1. If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?

Identifying the tracks that have not been sold in last 6 months(i.e. Between 01 July 2020 and 31 Dec 2020)

A screenshot of a computer

Description automatically generated

3006 such songs/tracks were Identified out of 3503 total songs/tracks

Identifying customers that haven’t made any purchase in last 6 months.

A screenshot of a computer

Description automatically generated

1. How would you approach this problem, if the objective and subjective questions weren't given?

Here's how I'd approach the Chinook music sales data analysis if I wasn't given specific objectives:

1. Data Understanding

Schema Familiarisation: Evaluating the schema to determine tables, column, data type and relationship among tables.

Data Exploration: Use SQL queries to:

Detect for missing data and redundance

2. Ask questions pertaining to the data

Based on the initial exploration, formulate broad questions about the data, such as:

Sales Performance:

A screenshot of a report

Description automatically generated

Customer Behaviour:

A screenshot of a spreadsheet

Description automatically generated

3. Analyze and Illustration

Deep Dive with SQL: Generate more particular SELECT statements to the questions created in step 2. This may include aggregation, denormalisation, sub-query, join, window functions and many others.

Visualization: Go ahead and analyze the raw data using analytical tools such as Tableau, Power BI or even excel to see the outcome of the analyzed patterns of the population. Create charts like:

The type of product graphs of sales dynamics for the identified period

Line graphs used in analyzing comparisons between sales, by genre, artist or by regions.

1. How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?

Alter table Albums add ReleaseYear int;



For Doing alteration in the following table

Steps to be followed:

GO to Edit > Preferences> Safe UPDATES (Tick)>OK

1. Chinook is interested in understanding the purchasing behavior of customers based on their geographical location. They want to know the average total amount spent by customers from each country, along with the number of customers and the average number of tracks purchased per customer. Write an SQL query to provide this information.

The average total amount spent per customer seems to be in the range of 8 to 12 approximately with an outlier of 5.158 in Denmark.

The number of customers are very few (single digits) in most of the countries.

USA is the country with highest number of customers.

