(ALL the code files are present in the SQL file along)

Objective Questions

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

   Yes there are 2 tables which have missing values .
   In the employee table there is general manager who has NULL value in it's reports_to columns which could be given 0 as there will never be a employee who has employee_id as 0.

   Also we have customer table where :-
   - 47 NULL entries in fax column
   - 29 NULL entries in state
   - 49 NULL entries in company

   In theses we could drop the Fax and company column all together using #ALTER TABLE *_table name_* DROP COLUMN *_column name_*; as we wont be needing those columns , to handle the NULL values in state column I'll be using the COALESCE function.

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

With the above result set we can conclude that the the top genre in USA is Rock.

Top_selling_track	Top_artist	Top_genre
War Pigs	Cake	Alternative
You Know I'm No Good (feat. Ghostface Killah)	Amy Winehouse	R&B/Soul
Violent Pornography	System Of A Down	Metal
Highway Chile	Jimi Hendrix	Rock
I Looked At You	The Doors	Rock
Scentless Apprentice	Nirvana	Rock
Evil Woman	Black Sabbath	Metal
Night Of The Long Knives	AC/DC	Rock
Polly	Nirvana	Rock
End Of The Night	The Doors	Rock

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

country	COUNT(customer_i...
Argentina	1
Australia	1
Austria	1
Belgium	1
Brazil	5
Canada	8
Chile	1
Czech Republic	2
Denmark	1
Finland	1
France	5
Germany	4
Hungary	1
India	2
Ireland	1
Italy	1
Netherlands	1
Norway	1
Poland	1
Portugal	2
Spain	1
Sweden	1
United Kingdom	3
USA	13

1) The customer Demographic background is very diverse and is from 24 different countries.
2) The maximum number of customers are from USA
3) There are many countries which have their count of customers at 1.
(As the dataset does not have any age/gender column we can-not get a demographic breakdown on those parameters)

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

billing_country	billing_state	billing_city	num_of_invoic...	total_revenue
Czech Republic	None	Prague	30	273.24
Brazil	SP	São Paulo	22	129.69
USA	CA	Mountain View	20	169.29
Germany	None	Berlin	20	158.40
United Kingdom	None	London	19	166.32
France	None	Paris	18	151.47
Portugal	None	Porto	16	82.17
Brazil	DF	Brasília	15	106.92
Ireland	Dublin	Dublin	13	114.84
India	None	Delhi	13	111.87
Brazil	SP	São José dos...	13	108.90
Portugal	None	Lisbon	13	102.96
Chile	None	Santiago	13	97.02
Canada	ON	Ottawa	13	91.08
USA	WA	Redmond	12	98.01
USA	FL	Orlando	12	92.07

The most amount spend in any city combined is in Prague in Czech Republic at $273.24 whereas the minimum spent is in Edmonton Canada at $29.70 .

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

Name	country
Diego Gutiérrez	Argentina
Mark Taylor	Australia
Astrid Gruber	Austria
Daan Peeters	Belgium
Luís Gonçalves	Brazil
Fernanda Ramos	Brazil
Roberto Almeida	Brazil
Alexandre Rocha	Brazil
Eduardo Martins	Brazil
François Tremblay	Canada
Edward Francis	Canada
Ellie Sullivan	Canada
Aaron Mitchell	Canada
Jennifer Peterson	Canada

6. Identify the top-selling track for each customer.

Name	Track_name
Patrick Gray	War Pigs
Roberto Almeida	Love And Marriage
Stanisław Wójcik	Faceless
Victor Stevens	Untitled
Wyatt Girard	Changes
Eduardo Martins	Like A Bird
Fernanda Ramos	24 Caprices, Op. 1, No. 24, for Solo Violin, in A…
François Tremblay	Sting Me
Fynn Zimmermann	Radio/Video
Hannah Schneider	I Can't Explain
Hugh O'Reilly	Drain You
Isabelle Mercier	Tease Me Please Me

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

customer_id	num_invoices	avg_sales
5	18	8.03
35	16	5.14
13	15	7.13
46	13	8.83
58	13	8.61
1	13	8.38
34	13	7.92
57	13	7.46
30	13	7.01
6	12	10.73
17	12	8.17
20	12	7.92
22	12	7.67
26	12	7.18
33	12	6.27

With the above dataset there was no co-relation or trend between customers and frequency/average Order value.

8. What is the customer churn rate?

Churn rate :

Churn rate is a metric that measures the percentage of customers or subscribers who stop doing business with a company over a specific period of time.

churn_rate
4.08

Here to find out churn rate i have used Number of customers who made a purchase in first 3 months of dataset and the number of number of customers in last 3 months.

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

genre_id	genre_name	artist_name	percentage_contribut...	ranking
1	Rock	The Posies	53.38	1
1	Rock	Scorpions	53.38	1
1	Rock	Ozzy Osbourne	53.38	1
1	Rock	Dread Zeppelin	53.38	1
1	Rock	Velvet Revolver	53.38	1
1	Rock	Van Halen	53.38	1
1	Rock	U2	53.38	1
1	Rock	The Who	53.38	1
1	Rock	The Rolling Stones	53.38	1
1	Rock	The Police	53.38	1
1	Rock	The Doors	53.38	1
1	Rock	The Cult	53.38	1
1	Rock	Terry Bozzio, Ton…	53.38	1
1	Rock	Stone Temple Pilots	53.38	1

Based on sales in USA, Rock is the top most contributing genre in USA sales which has artists like The posies, Scorpions, Ozzy Osbourne …

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

Name_of_customer	No_of_genres
Leonie Köhler	14
Terhi Hämäläinen	13
Madalena Sampaio	13
Edward Francis	13
Heather Leacock	13
František Wichterlová	13
Wyatt Girard	12
Niklas Schröder	12
Michelle Brooks	12
Marc Dubois	12
Julia Barnett	12
John Gordon	12
Jack Smith	12
Hugh O'Reilly	12

This is a result set of all the customers who have purchased tracks from at least 3 different genres where 'Leonie Köhler' has purchased from 14 different genres.

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

Name	Sale_performan...	Rank
Rock	555.39	1
Alternative & Punk	128.70	2
Metal	122.76	3
R&B/Soul	52.47	4
Blues	35.64	5
Alternative	34.65	6
Latin	21.78	7
Pop	21.78	7
Hip Hop/Rap	19.80	8
Jazz	13.86	9
Easy Listening	12.87	10
Reggae	5.94	11
Electronica/Dance	4.95	12

The result set clearly shows us that Rock genre in USA has the most sales among other genres.

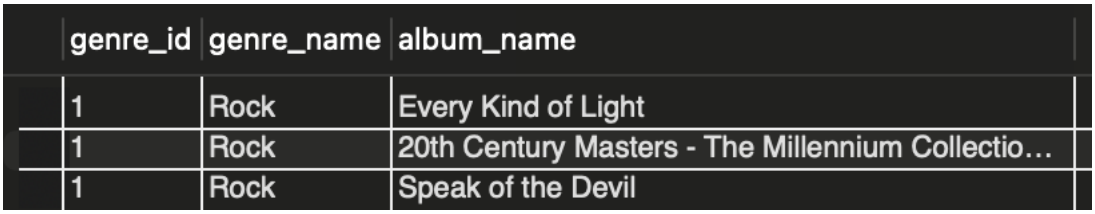
12. Identify customers who have not made a purchase in the last 3 months

first_name	last_name
Luís	Gonçalves
François	Tremblay
Bjørn	Hansen
Astrid	Gruber
Daan	Peeters
Kara	Nielsen
Eduardo	Martins
Alexandre	Rocha
Jack	Smith
Michelle	Brooks
Tim	Goyer
Hannah	Schneider
Fynn	Zimmermann
Niklas	Schröder
Camille	Bernard
Isabelle	Mercier
Johannes	Van der Berg
Enrique	Muñoz
Steve	Murray
Diego	Gutiérrez
Luis	Rojas
Manoj	Pareek

The Following are the customers who have not made a purchase in the last 3 months .

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.



genre_id	genre_name	album_name
1	Rock	Every Kind of Light
1	Rock	20th Century Masters - The Millennium Collectio…
1	Rock	Speak of the Devil

As rock Genre is at performing in USA we sort the albums from rock genre and the top performing albums within those are :-

- Every Kind of Light
- 20th Century Masters - The Millennium Collection: The Best of Scorpions
- Speak of the Devil

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

genre_id	name	total_revenue_for_ge...
1	Rock	2053.26
3	Metal	490.05
4	Alternative & Punk	358.38
7	Latin	143.55
2	Jazz	105.93
14	R&B/Soul	104.94
6	Blues	87.12
23	Alternative	81.18
12	Easy Listening	60.39
15	Electronica/Dance	49.50
24	Classical	42.57
9	Pop	40.59
8	Reggae	28.71
17	Hip Hop/Rap	12.87
13	Heavy Metal	4.95
10	Soundtrack	2.97
21	Drama	1.99
19	TV Shows	1.99

Comparing this result set with USA sales there are a lot of commonalities as the rock Genre is the most revenue generating genre of all followed by Metal . jazz where is was 10^{th} in USA is now ranked 5^{th} in the whole world excluding USA.

3. 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?

category	total_spending	basket_size	frequency
Long-term Customer	28034.82	2762	32
Short-term Customer	19468.35	1995	27

When compared to the whole customer-set the basket is divided into parts (Long term and short term customers) where long term customers are almost 58% and contribute 59% of the total revenue generated .

It pushes us to think that the long-term customers are spending just as much as short-term customers which in turn means that the company could move in either direction but focusing in gathering new customers would be a smart move as they will be the one who'll generate the most revenue and also increase the customer-base.

4. 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?

   When the above query's output is plotted as a table in excel and a pivot table constructed with genres in rows, invoice id in columns and count of genres in values.

   The following observations were visible.

   - Genres bought together
     - Rock
     - Metal
     - Alternative
   - albums bought together
     - Mezmerize
     - The Doors
     - Dark Side Of The Moon
   - artists bought together
     - Green Day
     - Foo Fighters
     - U2

5. 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?

billing_country	billing_state	billing_city	num_invoices	avg_sales	churn_rate
Australia	NSW	Sidney	10	8.118000	-100.0000
Brazil	RJ	Rio de Janeiro	11	7.470000	-100.0000
Brazil	SP	São José dos…	13	8.376923	100.0000
Brazil	SP	São Paulo	22	5.895000	100.0000
Canada	NS	Halifax	11	5.670000	50.0000
Canada	NT	Yellowknife	12	6.270000	0.0000
Canada	ON	Ottawa	13	7.006154	50.0000
Canada	QC	Montréal	9	11.110000	100.0000
Chile	None	Santiago	13	7.463077	100.0000
Denmark	None	Copenhagen	10	3.762000	100.0000
Finland	None	Helsinki	11	7.200000	50.0000
France	None	Bordeaux	11	9.090000	0.0000
France	None	Dijon	12	6.105000	100.0000
France	None	Paris	18	8.415000	66.6667
Germany	None	Berlin	20	7.920000	100.0000
Germany	None	Frankfurt	10	9.405000	100.0000
Hungary	None	Budapest	10	7.821000	0.0000
India	None	Bangalore	8	8.910000	0.0000
India	None	Delhi	13	8.605385	100.0000
Ireland	Dublin	Dublin	13	8.833846	0.0000

With the above result set in view we could conclude the following conclusions:

- American countries have more varied churn rate
- Europe countries when compared with American countries have a higher churn rate.
- Asian countries (only 1 country INDIA) have a low churn rate when considered as a whole

6. 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?

With reference to the above question, (as there isn't adequate data for any more in-depth analysis for the same) I would probably go with

- AGE: divide the customers in to 3 profiles
  - Young
  - Adult
  - Old/senior's
- Gender :
  - Male
  - Female
- Location (already have this information)
- Purchase history :
  - High spenders- spends more than average per person
  - Low spenders- spends less than average per person

With all these information I could do the analysis and decide which categories are to churn more than the others

As we know young customers are more likely to try new services and explore which means that they are more susceptible to churn than adult or old age people as they prefer to stick with things which they are comfortable around .

Also with gender I would try to find out if gender to plays any role their customer base, or to find out which of the gender is more of a long term customer than the other

Purchase history would split the people into 2 different categories on the basis of how much amount they had spend .People who have spent more than average person spending would be kept in high spending bracket and the others in low spending.

With all these parameters in mind people from young age group from developing countries and who are low spenders would pose a threat/risk in long term.

7. 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?

   To analyse the customer database on the basis of Lifetime value modelling, we could use the following features in a dataset :

   - First I would distribute customers into 3 categories
     - Regular customer
     - Occasional customer
     - One-time customer
   - Then I would segregate them based on their total amount spent
     - Large purchase amount
     - Small purchase amount
   - After I segregate the customer database among these categories I would then try to run demographic analysis and find out what do they have common with their location (As in subjective question 6 there we found out that American countries have a higher churn rate as compared to)

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

   If data related to promotional campaigns was available in the dataset, the analysis on it would be as follows:-

   - Click through rate of customers.
   - Number of customers from each city, state, country .
   - Analysis of promotional campaigns with money spend and new customers
   - Sales brought in by promotional campaigns
   - Age group of all the new customers
   - Which customers are possibly to be long term customers (on the basis of previous promotional campaigns if possible)

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

With the above available dataset my steps would be as follows

- Average revenue from each customer .
- Sales from each country ,state ,city to analyse performance geographically.
- Sales generated by each employee to find hard working employees.
- Popular genres from each country, state, city for promotional activities.
- Popular artists and their albums with songs .
- Genre ranking based on countries.
- Churn rate of countries to find trends in customer retention.

10. 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

        Album

ADD

    ReleaseYear int

;
```