**(ALL the code files are also 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.

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

select

t.name Top\_selling\_track,

a.name Top\_artist,

g.name Top\_genre

from

track t

left join invoice\_line il on t.track\_id = il.track\_id

left join invoice i on i.invoice\_id = il.invoice\_id

left join album al on al.album\_id = t.album\_id

left join artist a on a.artist\_id = al.artist\_id

left join genre g on g.genre\_id = t.genre\_id

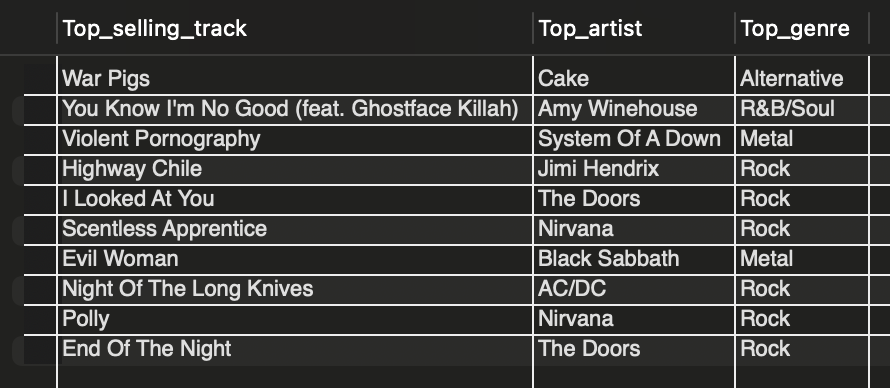
where

billing\_country = "usa"

group by 1,2,3

order by sum(quantity) desc

limit 10;



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

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

select

country,

count(customer\_id)

from customer

group by 1

order by 1;

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)

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

select

billing\_country,

billing\_state,

billing\_city,

count(invoice\_id) num\_of\_invoices,

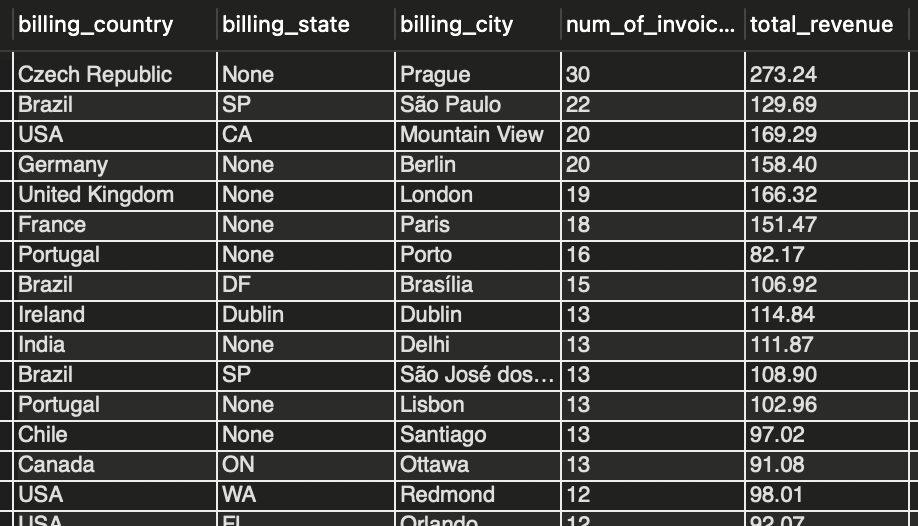
sum(total) total\_revenue\_city

from invoice

group by 1,2,3

order by count(invoice\_id) desc, sum(total) desc

;



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 .

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

with temp1 as

(

select

first\_name,

last\_name,

country,

sum(t.unit\_price \* il.quantity) total\_revenue

from customer c

left join invoice i on i.customer\_id = c.customer\_id

left join invoice\_line il on il.invoice\_id = i.invoice\_id

left join track t on t.track\_id = il.track\_id

group by 1,2,3

order by country

),

temp2 as

(

select

country,

first\_name,

last\_name,

row\_number() over(partition by country order by total\_revenue desc) n

from temp1

)

select

concat(first\_name," ",last\_name) name,

country

from temp2

where n <= 5;



1. Identify the top-selling track for each customer

with temp1 as(

select

first\_name,

last\_name,

t.name track\_name,

sum(quantity) total\_quantity,

row\_number()over(partition by first\_name,last\_name order by sum(quantity) desc) rn

from customer c

left join invoice i on i.customer\_id = c.customer\_id

left join invoice\_line il on il.invoice\_id = i.invoice\_id

left join track t on t.track\_id = il.track\_id

group by 1,2,3

order by sum(quantity) desc

)

select

concat(first\_name," ",last\_name) name,

track\_name

from temp1

where rn=1;



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

select

customer\_id,

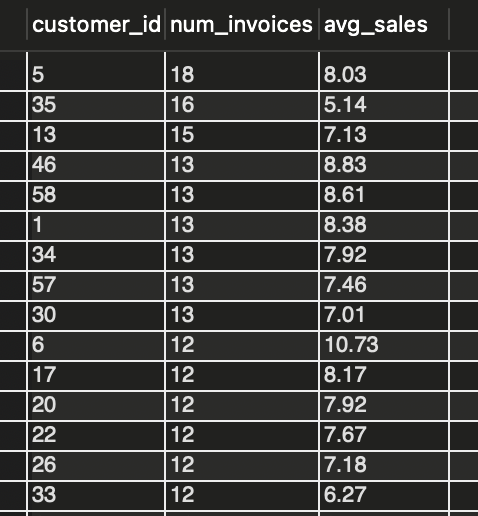
count(invoice\_id) num\_invoices,

round(avg(total),2) avg\_sales

from invoice

group by 1

order by count(invoice\_id) desc, avg(total) desc;



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

1. 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.

--calculating churn rate on first 3 and last 3 months

with cust\_in\_1st\_3m as

(

select

count(customer\_id) ttl

from invoice

where

invoice\_date between '2017-01-01' and '2017-03-31'

),

cust\_in\_last\_3m as

(

select

count(customer\_id) l\_num

from invoice

where

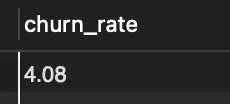
invoice\_date between '2020-10-01' and '2020-12-31'

)

select

round(((select ttl from cust\_in\_1st\_3m)-(select l\_num from cust\_in\_last\_3m))/(select ttl from cust\_in\_1st\_3m) \* 100,2) as churn\_rate

;



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.

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

with cte as

(

select

sum(total) total\_revenue\_for\_usa

from invoice

where billing\_country = 'usa'

),

genre\_sales as

(

select

g.genre\_id,

g.name,

sum(t.unit\_price \* il.quantity) total\_revenue\_for\_genre

from track t

left join genre g on g.genre\_id = t.genre\_id

left join invoice\_line il on il.track\_id = t.track\_id

left join invoice i on i.invoice\_id = il.invoice\_id

where billing\_country = 'usa'

group by 1,2

order by total\_revenue\_for\_genre desc

),

ranking as

(

select

genre\_id,

name,

round(total\_revenue\_for\_genre/(select total\_revenue\_for\_usa from cte) \* 100,2) percentage\_contribution,

dense\_rank() over(order by round(total\_revenue\_for\_genre/(select total\_revenue\_for\_usa from cte) \* 100,2) desc) rk

from genre\_sales

)

select

ranking.genre\_id,

ranking.name genre\_name,

-- a.name artist\_name,

percentage\_contribution,

rk as `ranking`

from ranking

left join track t on t.genre\_id = ranking.genre\_id

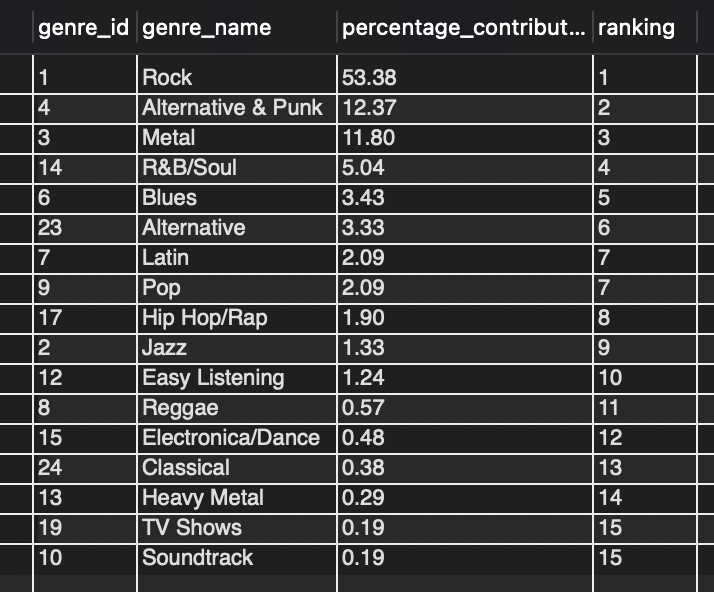
left join album al on al.album\_id = t.album\_id

left join artist a on a.artist\_id = al.artist\_id

-- where rk=1

group by 1,2,3

;



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

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

select

concat(first\_name, ' ', last\_name) name\_of\_customer,

count(distinct g.name) no\_of\_genres

from customer c

left join invoice i on i.customer\_id = c.customer\_id

left join invoice\_line il on il.invoice\_id = i.invoice\_id

left join track t on t.track\_id = il.track\_id

left join genre g on g.genre\_id = t.genre\_id

group by 1

having count(distinct g.name) >= 3

order by count(distinct g.name) desc;

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.

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

with cte as

(

select

t.genre\_id,

g.name,

sum(t.unit\_price \* il.quantity) sale\_performance

from track t

left join genre g on g.genre\_id = t.genre\_id

left join invoice\_line il on il.track\_id = t.track\_id

left join invoice i on i.invoice\_id = il.invoice\_id

where billing\_country = 'usa'

group by 1, 2

)

select

name,

sale\_performance,

dense\_rank() over(order by sale\_performance desc) `rank`

from cte

;

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

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

select

first\_name,

last\_name

from customer

where customer\_id not in (

select

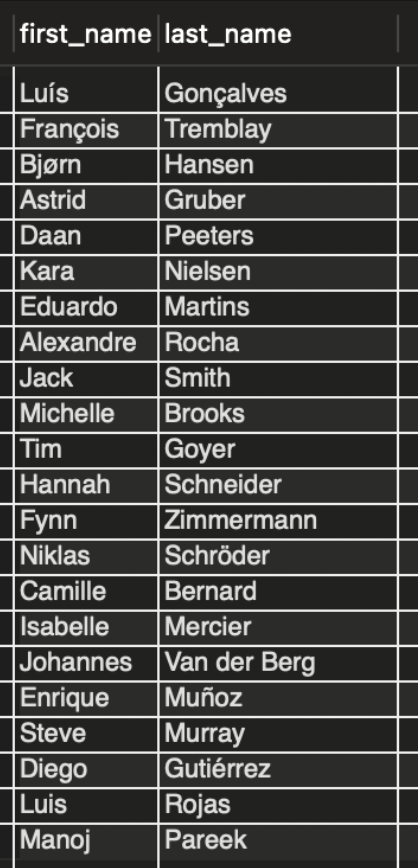
customer\_id

from invoice

where invoice\_date > (select max(invoice\_date) from invoice) - interval 3 month

)

;



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.

with genre\_sales as

(

select

g.genre\_id,

g.name,

sum(t.unit\_price \* il.quantity) total\_revenue\_for\_genre

from track t

left join genre g on g.genre\_id = t.genre\_id

left join invoice\_line il on il.track\_id = t.track\_id

left join invoice i on i.invoice\_id = il.invoice\_id

where billing\_country = 'usa'

group by 1,2

order by total\_revenue\_for\_genre desc

),

ranking as

(

select

genre\_id,

name,

total\_revenue\_for\_genre,

dense\_rank() over(order by total\_revenue\_for\_genre desc) rk

from genre\_sales

),

genre\_album as

(

select

ranking.genre\_id,

ranking.name genre\_name,

alb.title album\_name

from ranking

left join track t on t.genre\_id = ranking.genre\_id

left join album alb on alb.album\_id = t.album\_id

left join artist a on a.artist\_id = alb.artist\_id

where rk = 1

group by 1,2,3

),

best\_album as

(

select

alb.album\_id,

title,

sum(t.unit\_price \* il.quantity)

from album alb

left join track t on t.album\_id = alb.album\_id

left join invoice\_line il on il.track\_id = t.track\_id

group by 1,2

order by sum(t.unit\_price \* il.quantity) desc

)

select

genre\_id,

genre\_name,

album\_name

from genre\_album

inner join best\_album on best\_album.title = genre\_album.album\_name

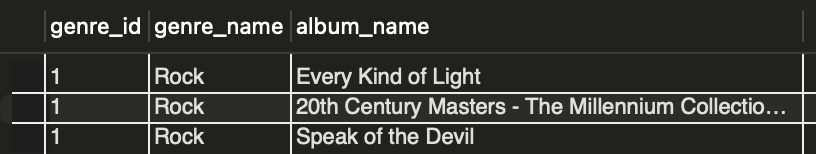
limit 3

;

First we find out which genre is performing the best amongst all the genre’s , and then among that genre we find out the most bought artist from that album.

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



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

select

g.genre\_id,

g.name,

sum(t.unit\_price \* il.quantity) total\_revenue\_for\_genre

from track t

left join genre g on g.genre\_id = t.genre\_id

left join invoice\_line il on il.track\_id = t.track\_id

left join invoice i on i.invoice\_id = il.invoice\_id

where billing\_country <> 'usa'

group by 1,2

order by total\_revenue\_for\_genre desc

;

!USA USA

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 10th in USA is now ranked 5th in the whole world excluding USA.

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?

with cte as

(

select

i.customer\_id,

max(invoice\_date),

min(invoice\_date),

abs(timestampdiff(month, max(invoice\_date), min(invoice\_date))) time\_for\_each\_customer,

sum(total) sales,

sum(quantity) items,

count(invoice\_date) frequency

from invoice i

left join customer c on c.customer\_id = i.customer\_id

left join invoice\_line il on il.invoice\_id = i.invoice\_id

group by 1

order by time\_for\_each\_customer desc

),

average\_time as

(

select

avg(time\_for\_each\_customer) average

from cte

),-- 1244.3220 days or 40.36 months

categorization as

(

select

\*,

case

when time\_for\_each\_customer > (select average from average\_time) then "long-term customer"

else "short-term customer"

end category

from cte

)

select

category,

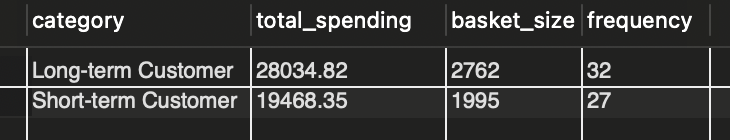
sum(sales) total\_spending,

sum(items) basket\_size,

count(frequency) frequency

from categorization

group by 1 ;



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.

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?

with cte as

(

select

invoice\_id,

count(distinct g.name) num

from invoice\_line il

left join track t on t.track\_id = il.track\_id

left join genre g on g.genre\_id = t.genre\_id

group by 1

having count(distinct g.name) > 1

)

select

cte.invoice\_id,

num,

g.name

from cte

left join invoice\_line il on il.invoice\_id = cte.invoice\_id

left join track t on t.track\_id = il.track\_id

left join genre g on g.genre\_id = t.genre\_id

group by 1,2,3;

with cte as

(

select

invoice\_id,

count(distinct al.title) num

from invoice\_line il

left join track t on t.track\_id = il.track\_id

left join album al on al.album\_id = t.album\_id

group by 1

having count(distinct al.title) > 1

)

select

cte.invoice\_id,

num,

al.title

from cte

left join invoice\_line il on il.invoice\_id = cte.invoice\_id

left join track t on t.track\_id = il.track\_id

left join album al on al.album\_id = t.album\_id

group by 1,2,3;

with cte as

(

select

invoice\_id,

count(distinct a.name) num

from invoice\_line il

left join track t on t.track\_id = il.track\_id

left join album al on al.album\_id = t.album\_id

left join artist a on a.artist\_id = al.artist\_id

group by 1

having count(distinct a.name) > 1

)

select

cte.invoice\_id,

num,

a.name

from cte

left join invoice\_line il on il.invoice\_id = cte.invoice\_id

left join track t on t.track\_id = il.track\_id

left join album al on al.album\_id = t.album\_id

left join artist a on a.artist\_id = al.artist\_id

group by 1,2,3;

When the above query's output is plotted as a table in excel and a pivot table was constructed . with genres/albums/artists with sum of quantity sold

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
  + U2
  + The Rolling Stones
  + Foo Fighters

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?

with cust\_in\_first\_3M as

(

select

billing\_country,

billing\_state,

billing\_city,

count(customer\_id) ttl

from invoice

where invoice\_date between '2017-01-01' and '2017-03-31'

group by 1,2,3

),

cust\_in\_last\_3M as

(

select

billing\_country,

billing\_state,

billing\_city,

count(customer\_id) l\_num

from invoice

where invoice\_date between '2020-10-01' and '2020-12-31'

group by 1,2,3

),

Invoices as(

select

billing\_country,

billing\_state,

billing\_city,

count(invoice\_id) num\_invoices,

avg(total) avg\_sales

from invoice

group by 1,2,3

order by count(invoice\_id) desc, avg(total) desc

)

select

F.billing\_country,

F.billing\_state,

F.billing\_city,

num\_invoices,

avg\_sales,

ttl,

l\_num,

(ttl - coalesce(l\_num,0))/ttl \* 100 churn\_rate

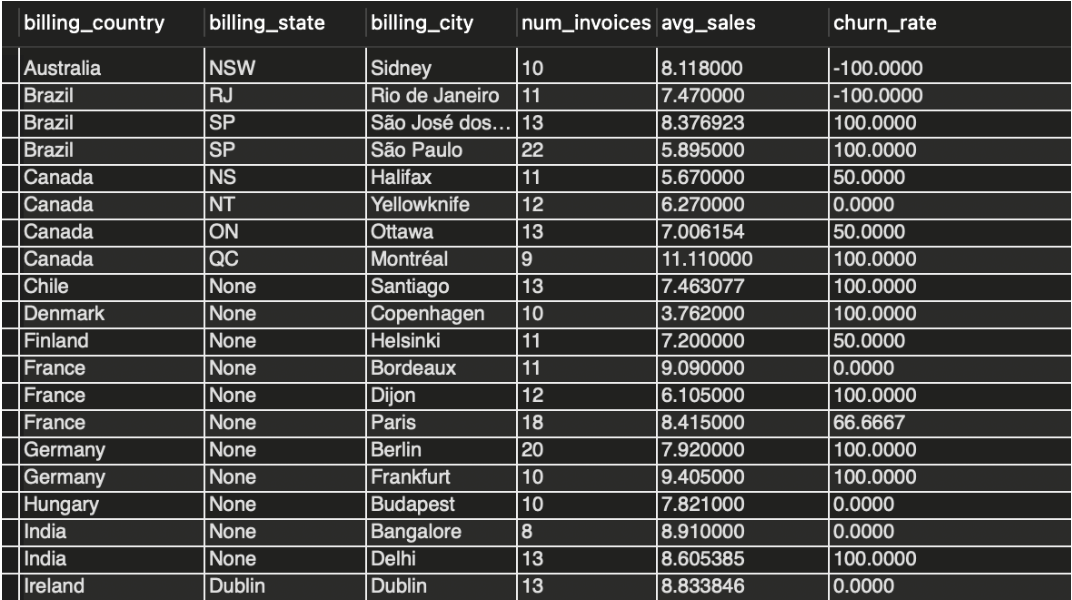
from cust\_in\_first\_3M F

left join cust\_in\_last\_3M L on F.billing\_city = L.billing\_city

join invoices i on F.billing\_city=i.billing\_city

order by 1,2,3

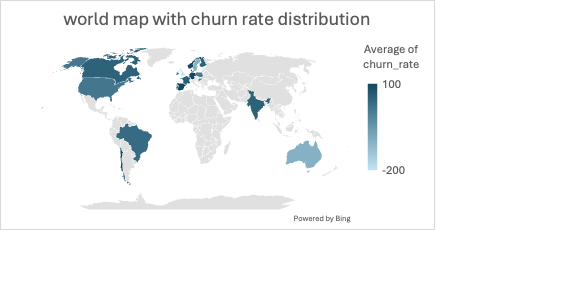
;



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



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?

select

i.customer\_id,

concat(first\_name, " ", last\_name) name,

billing\_country,

sum(total) total\_spending,

count(invoice\_id) num\_of\_orders

from invoice i

left join customer c on c.customer\_id = i.customer\_id

group by 1,2,3

order by name

;

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.

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?

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 (just like in subjective question 6 there we found out that American countries have a higher churn rate as compared to )

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?

* 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)

1. 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.

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

Album

ADD

ReleaseYear int