Objective Questions

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

|  |  |
| --- | --- |
| **Tables With NULL Values** | **Approach Used** |
| **Track** – **Composer** Column | Replaced with a placeholder “Unknown” |
| **Customer** – **company, postal\_code, state,phone,fax** columns | Replaced with placeholder “Unknown” |

**Query:**

|  |
| --- |
| **update customer**  **set company='Unknown'**  **where company is NULL;**  **update customer**  **set state='Unknown'**  **where state is NULL;**  **update customer**  **set postal\_code='Unknown'**  **where postal\_code is NULL;**  **update customer**  **set phone='Unknown'**  **where phone is NULL;**  **update customer**  **set fax='Unknown'**  **where fax is NULL;**  **update track**  **set composer='Unknown'**  **where composer is NULL;**  **select \* from customer;** |

**Reference:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **customer\_id** | **company** | **postal\_code** | **state** | **phone** | **fax** |
| **44** | **Unknown** | **530** | **Unknown** | **+358 09 870 2000** | **Unknown** |
| **45** | **Unknown** | **H-1073** | **Unknown** | **Unknown** | **Unknown** |
| **46** | **Unknown** | **Unknown** | **Dublin** | **+353 01 6792424** | **Unknown** |
| **47** | **Unknown** | **192** | **RM** | **+39 06 39733434** | **Unknown** |
| **48** | **Unknown** | **1016** | **VV** | **+31 020 6223130** | **Unknown** |

**Insights:**

* The name of composers for tracks is missing from the data.
* Some customers have their company, postal\_code, state, phone or fax number missing.
* We have to replace the data with a placeholder for now since missing values will make the data inconsistent.

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

**Query:**

|  |
| --- |
| **-- top selling track**  **with topSellingTrack**  **as**  **(select track\_id,count(invoice\_line\_id) as sales\_count from invoice i**  **left join invoice\_line i1 on i.invoice\_id=i1.invoice\_id**  **where billing\_country='USA'**  **group by track\_id**  **order by sales\_count desc**  **limit 10)**  **select t.name,sales\_count from topSellingTrack ts**  **join track t on ts.track\_id=t.track\_id;**  **-- top selling artist**  **with topAlbum**  **as**  **(select album\_id,count(invoice\_line\_id) as albumSales from invoice\_line i**  **join track t on i.track\_id=t.track\_id**  **join invoice i1 on i.invoice\_id=i1.invoice\_id**  **where billing\_country='USA'**  **group by album\_id)**  **select aa.name,sum(albumSales) as artistSale from topAlbum ta**  **join album a on ta.album\_id=a.album\_id**  **join artist aa on a.artist\_id=aa.artist\_id**  **group by aa.name**  **order by artistSale desc**  **limit 10;**  **-- their most famous genre**  **with topAlbum**  **as**  **(select album\_id,count(invoice\_line\_id) as albumSales from invoice\_line i**  **join track t on i.track\_id=t.track\_id**  **join invoice i1 on i.invoice\_id=i1.invoice\_id**  **where billing\_country='USA'**  **group by album\_id),**  **top10artist**  **as**  **(select aa.artist\_id,aa.name,sum(albumSales) as artistSale from topAlbum ta**  **join album a on ta.album\_id=a.album\_id**  **join artist aa on a.artist\_id=aa.artist\_id**  **group by aa.artist\_id,aa.name**  **order by artistSale desc**  **limit 10),**  **cte1**  **as**  **(select t.artist\_id,t.name,genre\_id,count(tt.track\_id) as salecount from top10artist t**  **join album a on t.artist\_id=a.artist\_id**  **join track tt on a.album\_id=tt.album\_id**  **join invoice\_line il on tt.track\_id=il.track\_id**  **group by 1,2,3),**  **bestgenre**  **as**  **(select \*,rank() over (partition by artist\_id order by salecount desc) as ranking from cte1)**  **select b.name,g.name from bestgenre b**  **join genre g on b.genre\_id=g.genre\_id**  **where ranking=1;** |

**Approach Used:** Common Table Expressions, window functions, aggregation functions.

**References:**

|  |  |
| --- | --- |
| **Song Name** | **Sales Count** |
| **War Pigs** | **6** |
| **You Know I'm No Good (feat. Ghostface Killah)** | **5** |
| **I Looked At You** | **4** |
| **Violent Pornography** | **4** |
| **Night Of The Long Knives** | **4** |
| **End Of The Night** | **4** |
| **Highway Chile** | **4** |
| **Scentless Apprentice** | **4** |
| **Evil Woman** | **4** |
| **You're What's Happening (In The World Today)** | **3** |

|  |  |
| --- | --- |
| **Artist** | **Sales** |
| **Van Halen** | **43** |
| **R.E.M.** | **38** |
| **The Rolling Stones** | **37** |
| **Nirvana** | **35** |
| **Eric Clapton** | **34** |
| **Foo Fighters** | **34** |
| **Guns N' Roses** | **32** |
| **Green Day** | **32** |
| **Pearl Jam** | **31** |
| **Amy Winehouse** | **30** |

|  |  |
| --- | --- |
| **Artist** | **Genre** |
| **Foo Fighters** | **Rock** |
| **Guns N' Roses** | **Rock** |
| **Nirvana** | **Rock** |
| **Pearl Jam** | **Rock** |
| **The Rolling Stones** | **Rock** |
| **Van Halen** | **Rock** |
| **Green Day** | **Alternative & Punk** |
| **R.E.M.** | **Alternative & Punk** |
| **Eric Clapton** | **Blues** |
| **Amy Winehouse** | **R&B/Soul** |

**Insights:**

* Rock leads the market in terms of sales.
* Van Halen is the most popular artist in the US.

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

**Gender and age data is not available to us**.

**Query:**

|  |
| --- |
| **select country,count(customer\_id) CustomerStrength from customer**  **group by country**  **order by CustomerStrength desc;** |

**Approach Used:** Aggregation Function

**Reference:**

|  |  |
| --- | --- |
| **Country** | **CustomerStrength** |
| **USA** | **13** |
| **Canada** | **8** |
| **Brazil** | **5** |
| **France** | **5** |
| **Germany** | **4** |
| **United Kingdom** | **3** |
| **Czech Republic** | **2** |
| **Portugal** | **2** |
| **India** | **2** |
| **Norway** | **1** |
| **Austria** | **1** |
| **Belgium** | **1** |
| **Denmark** | **1** |
| **Finland** | **1** |
| **Hungary** | **1** |
| **Ireland** | **1** |
| **Italy** | **1** |
| **Netherlands** | **1** |
| **Poland** | **1** |
| **Spain** | **1** |
| **Sweden** | **1** |
| **Australia** | **1** |
| **Argentina** | **1** |
| **Chile** | **1** |

**Insights:**

* Customer table has the personal information of all the customers including their location.
* We can group customers by their country and measure our customer demographic.
* Age and gender data for customers is missing.

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

**Query:**

|  |
| --- |
| **select billing\_country as country,sum(total) as total\_revenue,count(invoice\_id) invoice\_count from invoice**  **group by country**  **order by total\_revenue desc,invoice\_count desc;**  **select billing\_state as state,sum(total) as total\_revenue,count(invoice\_id) invoice\_count from invoice**  **group by state**  **order by total\_revenue desc,invoice\_count desc;**  **select billing\_city as city,sum(total) as total\_revenue,count(invoice\_id) invoice\_count from invoice**  **group by city**  **order by total\_revenue desc,invoice\_count desc;** |

**Approach Used:** Aggregation Function, CTE

**Reference:**







**Insights:**

* US is the global leader in terms of music sales in our organisation.
* Due to inconsistent data, we do not have the data for the State which is highest in sales.

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

**Query:**

|  |
| --- |
| **with**  **cust**  **as**  **(select billing\_country, customer\_id from (select \*, rank() over(partition by billing\_country order by rev) as ranking from (select customer\_id,billing\_country,sum(total) as rev from invoice**  **group by customer\_id,billing\_country**  **order by billing\_country) a**  **order by billing\_country) aa**  **where ranking<=5)**  **select billing\_country as Country,concat(first\_name,' ',last\_name) as Customer\_Name from cust c**  **join customer cc on c.customer\_id=cc.customer\_id;** |

**Approach Used:** Aggregation Function, CTE

**Reference:**



**Insights:**

* **Most countries have total customers less than 5.**
* **USA has 13 customers which is the most for a country.**

1. Identify the top-selling track for each customer

**Query:**

|  |
| --- |
| **with**  **cte1**  **as**  **(select c.customer\_id,track\_id,count(track\_id) as sale from customer c**  **join invoice i on c.customer\_id=i.customer\_id**  **join invoice\_line i1 on i.invoice\_id=i1.invoice\_id**  **group by 1,2**  **order by sale desc),**  **rownum**  **as**  **(select \*,row\_number() over(partition by customer\_id order by sale desc) as rownum from cte1)**  **select customer\_id,track\_id from rownum**  **where rownum=1;** |

**Approach Used:** Aggregation function, row\_number() function

**Reference:**



**Insights:**

* For each customer, maximum sales for one track are 2, shared by a lot of tracks.
* Top track for each customer is given by sales count and order of purchase date.

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

**Query:**

|  |
| --- |
| **with**  **frequency**  **as**  **(select customer\_id,round(avg(purchase\_frequency),0) as avg\_days\_bw\_purchase from (select customer\_id,invoice\_date,next\_date,datediff(next\_date,invoice\_date) as purchase\_frequency**  **from (select \*,lead(invoice\_date) over(partition by customer\_id) as next\_date from invoice**  **order by customer\_id) a) b**  **group by customer\_id),**  **average\_order**  **as**  **(select customer\_id,round(avg(total),2) as average\_order\_value from invoice**  **group by customer\_id**  **order by customer\_id)**  **select f.customer\_id,avg\_days\_bw\_purchase,average\_order\_value from frequency f**  **join average\_order a on f.customer\_id=a.customer\_id;** |

There is no data on payment methods.

**Reference:**

|  |  |  |
| --- | --- | --- |
| **Customer ID** | **Average days between Purchases** | **Average Order Value** |
| **1** | **106** | **8.38** |
| **2** | **129** | **7.47** |
| **3** | **148** | **11.11** |
| **4** | **136** | **8.03** |
| **5** | **74** | **8.03** |
| **6** | **104** | **10.73** |
| **7** | **134** | **7.7** |
| **8** | **132** | **8.63** |
| **9** | **121** | **3.76** |
| **10** | **112** | **5.03** |
| **11** | **108** | **6.93** |
| **12** | **138** | **7.47** |
| **13** | **92** | **7.13** |
| **14** | **128** | **2.97** |
| **15** | **167** | **7.37** |
| **16** | **186** | **9.28** |
| **17** | **104** | **8.17** |
| **18** | **165** | **9.9** |
| **19** | **140** | **6.05** |
| **20** | **130** | **7.92** |
| **21** | **136** | **8.28** |
| **22** | **120** | **7.67** |
| **23** | **124** | **6.63** |
| **24** | **194** | **8.91** |
| **25** | **152** | **7.62** |
| **26** | **122** | **7.18** |
| **27** | **172** | **9.35** |
| **28** | **126** | **7.23** |
| **29** | **392** | **10.15** |
| **30** | **116** | **7.01** |
| **31** | **141** | **5.67** |
| **32** | **173** | **8.79** |
| **33** | **128** | **6.27** |
| **34** | **111** | **7.92** |
| **35** | **95** | **5.14** |
| **36** | **111** | **7.74** |
| **37** | **144** | **9.41** |
| **38** | **149** | **8.14** |
| **39** | **144** | **8.8** |
| **40** | **174** | **8.03** |
| **41** | **130** | **7.15** |
| **42** | **141** | **9.09** |
| **43** | **107** | **6.11** |
| **44** | **141** | **7.2** |
| **45** | **149** | **7.82** |
| **46** | **116** | **8.83** |
| **47** | **157** | **5.61** |
| **48** | **116** | **6.53** |
| **49** | **160** | **7.62** |
| **50** | **118** | **8.91** |
| **51** | **151** | **7.52** |
| **52** | **192** | **8.54** |
| **53** | **142** | **8.91** |
| **54** | **119** | **8.8** |
| **55** | **156** | **8.12** |
| **56** | **202** | **7.92** |
| **57** | **98** | **7.46** |
| **58** | **103** | **8.61** |
| **59** | **205** | **8.91** |

**Insights:**

* Most customers take 120 days between purchases.

1. What is the customer churn rate?

**Query:**

|  |
| --- |
| **with last\_purchase**  **as**  **(select customer\_id, max(date(invoice\_date)) as last\_purchase\_date from invoice**  **group by customer\_id**  **order by customer\_id),**  **active\_or\_inactive**  **as**  **(select customer\_id,last\_purchase\_date, case when last\_purchase\_date>date\_sub('2020-12-31',interval 6 month) then 'Active Customer'**  **when last\_purchase\_date<date\_sub('2020-12-31',interval 6 month) then 'Churned Customer' end as status from last\_purchase),**  **churned\_count**  **as**  **(select status,count(customer\_id) number\_of\_customers from active\_or\_inactive**  **group by status)**  **select round((sum(case when status='Churned Customer' then number\_of\_customers else 0 end)/sum(number\_of\_customers))\*100,2) as churn\_rate from churned\_count;** |

**Churn Rate-** 27.12%

**Approach Used:** Aggregation Function, CTE

**Reference:**



**Insights:**

* The customer churn is high.
* The churn is calculated for people who haven’t purchased anything in the last 6 months.

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

**Query:**

|  |
| --- |
| **with cte1**  **as**  **(select i1.track\_id as trackid,i1.unit\_price,count(invoice\_line\_id) as sales\_count from invoice i**  **left join invoice\_line i1 on i.invoice\_id=i1.invoice\_id**  **where billing\_country='USA'**  **group by track\_id,unit\_price**  **order by sales\_count desc),**  **cte2**  **as**  **(select c.trackid,sales\_count,t.genre\_id,t.album\_id,aa.artist\_id,aa.name,c.unit\_price,g.name as genre\_name from cte1 c**  **join track t on c.trackid=t.track\_id**  **join genre g on t.genre\_id=g.genre\_id**  **join album a on t.album\_id=a.album\_id**  **join artist aa on a.artist\_id=aa.artist\_id),**  **cte3**  **as**  **(select genre\_name,sum(sales\_count\*unit\_price) as rev from cte2**  **group by genre\_name**  **order by rev desc)**  **select genre\_name,round((rev/(select sum(rev) from cte3)\*100),2) as percentage\_sales from cte3;**  **-- artists**  **with cte1**  **as**  **(select i1.track\_id as trackid,i1.unit\_price,count(invoice\_line\_id) as sales\_count from invoice i**  **left join invoice\_line i1 on i.invoice\_id=i1.invoice\_id**  **where billing\_country='USA'**  **group by track\_id,unit\_price**  **order by sales\_count desc),**  **cte2**  **as**  **(select c.trackid,sales\_count,t.genre\_id,t.album\_id,aa.artist\_id,aa.name as artist\_name,c.unit\_price,g.name as genre\_name from cte1 c**  **join track t on c.trackid=t.track\_id**  **join genre g on t.genre\_id=g.genre\_id**  **join album a on t.album\_id=a.album\_id**  **join artist aa on a.artist\_id=aa.artist\_id),**  **cte3**  **as**  **(select artist\_name,sum(sales\_count\*unit\_price) as rev from cte2**  **group by artist\_name**  **order by rev desc)**  **select artist\_name,round((rev/(select sum(rev) from cte3)\*100),2) as percentage\_sales from cte3;** |

**Approach Used:** Aggregation Function, CTE

**Reference:**

|  |  |
| --- | --- |
| **Genre** | **Percentage of Sales in USA** |
| **Rock** | **53.38** |
| **Alternative & Punk** | **12.37** |
| **Metal** | **11.8** |
| **R&B/Soul** | **5.04** |
| **Blues** | **3.43** |
| **Alternative** | **3.33** |
| **Latin** | **2.09** |
| **Pop** | **2.09** |
| **Hip Hop/Rap** | **1.9** |
| **Jazz** | **1.33** |
| **Easy Listening** | **1.24** |
| **Reggae** | **0.57** |
| **Electronica/Dance** | **0.48** |
| **Classical** | **0.38** |
| **Heavy Metal** | **0.29** |
| **Soundtrack** | **0.19** |
| **TV Shows** | **0.1** |

|  |  |
| --- | --- |
| **Artist** | **Percentage of Sales** |
| **Van Halen** | **4.09** |
| **R.E.M.** | **3.62** |
| **The Rolling Stones** | **3.52** |
| **Nirvana** | **3.33** |
| **Foo Fighters** | **3.24** |
| **Eric Clapton** | **3.24** |
| **Green Day** | **3.04** |
| **Guns N' Roses** | **3.04** |
| **Pearl Jam** | **2.95** |
| **Amy Winehouse** | **2.85** |

**Insights:**

* Rock has the highest sales in terms of genre.
* Van Halen is the most popular artist in USA in terms of sales.

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

**Query:**

|  |
| --- |
| **select customer\_id,count(distinct genre\_id) as genre\_count from invoice i**  **join invoice\_line il on i.invoice\_id=il.invoice\_id**  **join track t on il.track\_id=t.track\_id**  **group by customer\_id**  **having count(distinct genre\_id)>=3;** |

All the customers have purchased tracks belonging to more than 3 genres.

**Approach Used:** Aggregation Function, CTE

**Reference:**

|  |  |
| --- | --- |
| **Customer ID** | **Number of Genres Purchased** |
| **1** | **10** |
| **2** | **14** |
| **3** | **8** |
| **4** | **9** |
| **5** | **13** |
| **6** | **11** |
| **7** | **8** |
| **8** | **10** |
| **9** | **6** |
| **10** | **9** |
| **11** | **8** |
| **12** | **8** |
| **13** | **12** |
| **14** | **9** |
| **15** | **6** |
| **16** | **8** |
| **17** | **12** |
| **18** | **12** |
| **19** | **10** |
| **20** | **9** |
| **21** | **10** |
| **22** | **13** |
| **23** | **12** |
| **24** | **8** |
| **25** | **6** |
| **26** | **10** |
| **27** | **9** |
| **28** | **12** |
| **29** | **5** |
| **30** | **13** |
| **31** | **8** |
| **32** | **9** |
| **33** | **11** |
| **34** | **10** |
| **35** | **13** |
| **36** | **10** |
| **37** | **6** |
| **38** | **12** |
| **39** | **11** |
| **40** | **8** |
| **41** | **12** |
| **42** | **12** |
| **43** | **10** |
| **44** | **13** |
| **45** | **11** |
| **46** | **12** |
| **47** | **10** |
| **48** | **11** |
| **49** | **9** |
| **50** | **11** |
| **51** | **9** |
| **52** | **7** |
| **53** | **8** |
| **54** | **11** |
| **55** | **11** |
| **56** | **10** |
| **57** | **11** |
| **58** | **10** |
| **59** | **8** |

**Insights:**

* Most customers have purchased tracks belonging to more than 3 genres.
* Customers usually prefer multiple genres in their purchasing pattern.

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

**Query:**

|  |
| --- |
| **with cte1**  **as**  **(select track\_id,count(invoice\_line\_id) as sales\_count from invoice i**  **left join invoice\_line i1 on i.invoice\_id=i1.invoice\_id**  **where billing\_country='USA'**  **group by track\_id**  **order by sales\_count desc),**  **genrerank**  **as**  **(select genre\_id,salesum,rank() over(order by salesum desc) as genre\_sales\_rank from (select genre\_id,sum(sales\_count) as salesum from cte1 c**  **join track t on c.track\_id=t.track\_id**  **group by genre\_id**  **order by salesum desc) a)**  **select gg.name,salesum,genre\_sales\_rank from genrerank g**  **join genre gg on g.genre\_id=gg.genre\_id**  **order by genre\_sales\_rank;** |

**Approach Used:** Window Functions, aggregation functions

**Reference:**

|  |  |  |
| --- | --- | --- |
| **Genre** | **Sales** | **Genre Sales Rank** |
| **Rock** | **561** | **1** |
| **Alternative & Punk** | **130** | **2** |
| **Metal** | **124** | **3** |
| **R&B/Soul** | **53** | **4** |
| **Blues** | **36** | **5** |
| **Alternative** | **35** | **6** |
| **Latin** | **22** | **7** |
| **Pop** | **22** | **7** |
| **Hip Hop/Rap** | **20** | **9** |
| **Jazz** | **14** | **10** |
| **Easy Listening** | **13** | **11** |
| **Reggae** | **6** | **12** |
| **Electronica/Dance** | **5** | **13** |
| **Classical** | **4** | **14** |
| **Heavy Metal** | **3** | **15** |
| **Soundtrack** | **2** | **16** |
| **TV Shows** | **1** | **17** |

**Insights:**

* Rock, Alternative & Punk and Metal occupy the top 3 places in terms of sales in USA.
* TV Shows genre is at the bottom of the pile.

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

**Query:**

|  |
| --- |
| **with cte1**  **as**  **(select customer\_id from invoice**  **where date(invoice\_date) between '2020-10-01' and '2020-12-30')**  **select distinct customer\_id,max(date(invoice\_date)) as last\_purchase\_date from invoice**  **where customer\_id not in (select \* from cte1)**  **group by customer\_id**  **order by customer\_id;** |

**Approach Used:** Aggregation Function, CTE

**Reference:**



**Insights:**

* 27% of the customers have not made a purchase in the last 3 months.
* The churn rate shows that a considerable number of customers have stopped purchasing.

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.

**Based on Genre analysis, the following 3 albums should be prioritised for advertising and promotion since these 3 albums show the most promise in terms of sales.**

**Reference:**



**Insights:**

* Due to the huge popularity of these albums, advertising and promoting these will attract new customers as well as drive up the sales among existing customers.
* A Focused marketing campaign in areas that have not been explored before will ensure more revenue.

**Recommendations:**

* In order to capitalise on the popularity of these 3 albums, more promotional campaigns should be launched in favour of these.
* Similar albums can be discounted to ensure customer engagement.

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

**Reference:**



**Insights:**

* Rock is the best-selling genre in USA as well as outside of it.
* The lowest selling genres have a lot of genres in common i.e. TV Shows, Heavy Metal, Soundtrack etc
* Metal is more popular than Alternative & Punk outside USA.
* There are no sales for Drama genre in USA.

**Recommendations:**

* Genres showing good sales except rock can be promoted in order to extract more revenue and enhance customer engagement.
* Low performing genres can be supplemented to include more tracks.

1. Customer Purchasing Behaviour 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?

**Reference:**



**Insights:**

* New customers tend to have more days between their purchases than old customers.
* The average order size of both new and old customers is similar.
* Long Term customers spend more on average on their orders.

**Recommendations:**

* To reward loyalty of the old customers, promotions should be aimed with taking into account previous purchases.
* In order to retain the long-term customers, they should be given occasional discounts and rebates on purchases.

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?

**Reference:**



**Insights:**

* With this data, we can recommend the two albums, genres or artists as a combo and give discounts and promotions on them.
* We can also combine new albums on the basis of their genres together and put on offers and promotions.

**Recommendations:**

* The albums can be offered as complimentary or on discounted rates with the purchase of one of them.
* Post Purchase offers can be given to customers who have bought an album.

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?

**Reference:**



**Insights:**

* The purchase frequency is not varied too much among the different countries.
* USA is leading in terms of amount spending, this correlates with the customer engagement and higher average income.
* Due to a less global catalogue of music, there is a huge contrast among the different countries in terms of spending.
* The churn rate is also higher outside of the USA.

**Recommendations:**

* Discounts are recommended for countries where the average order value is less to increase customer engagement.
* Places with more days between purchases should be awarded discounts to pull them back in.

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?

**Reference:**



**Insights:**

* USA, Brazil, Germany and France customers tend to have the most churned customers along with customers who have reduced spending.
* Factors influencing this behaviour are reduced engagement, lack of new content and less promotional schemes.

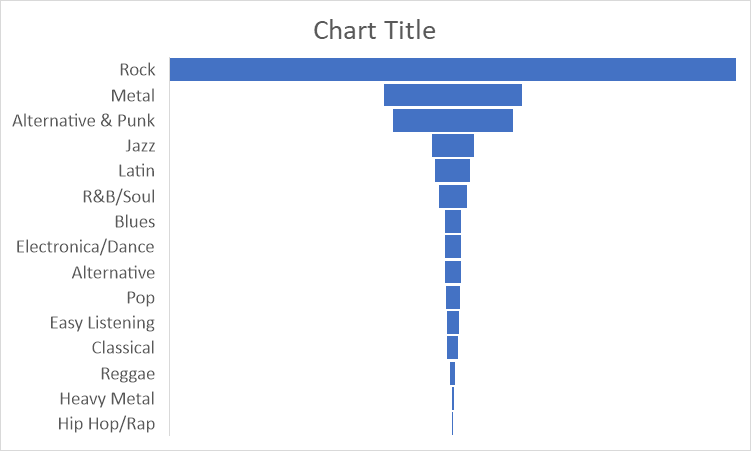
**Recommendations:**

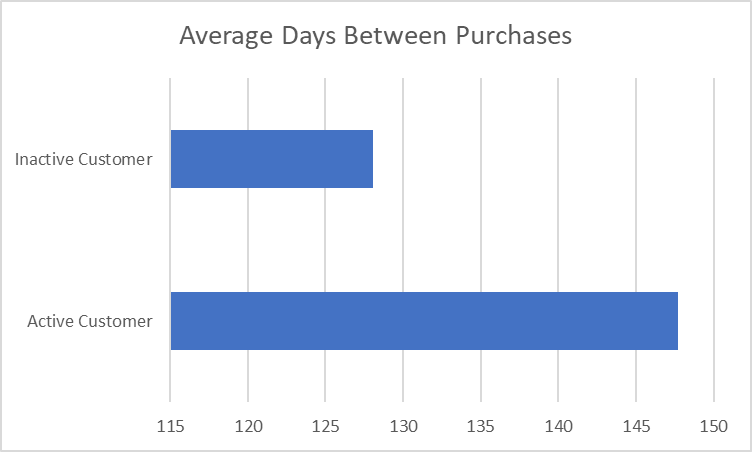
* To drive up the spending, regular promotional schemes are to be made available to the customers.
* Reduced spenders should be included in promotional marketing schemes to increase engagement.

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?

**Reference:**

Inactive customer genre preference:





**Insights:**

* Inactive customers have a huge preference to Rock with Metal and Alternative & Punk following.
* Inactive customers purchase products more frequently than the active customers.

**Recommendations:**

* To ensure better customer engagement, inactive customers are to be informed of new offers and deals as well as new release
* Inactive customers should be included in promotional emails more frequently.

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?

**Methods to measure impact of promotional campaigns**

* Email Conversion Rate – How many customers were acquired or purchased a product per email sent.
* Customer Acquisition – Measuring the number of customer before, during and after the campaign to have a metric of how successful the campaign was.
* Sales Growth – Measuring the sales difference between start and end of promotional campaign.
* Discount Usage – Measuring the number of customers actually purchasing a product because of discount.
* Customer Churn Rate – Measuring how many customers have stopped purchasing any products and calculating and improving customer retention.

**No Insights or recommendations required.**

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

* Understanding the problem statement
  + Determining the objective.
  + Recognising the scope of the problem
  + Defining a success criteria
* Analysing the data
  + Cleaning the data – Replacing NULL values with placeholders.
  + Identifying key tables - **Customer, Invoice, Track, Invoice\_line**.
  + Which tables have redundant or duplicate data
* Creating insights
  + Which genre is the most popular among the customers?
  + Which artist contribute most to the sales?
  + Establishing a trend among the customers based on their purchase history.
  + What was the impact of promotional campaigns on sales.
  + How many customers are inactive?
* Giving Recommendations
  + Which genre should the company focus on for marketing.
  + What is the best strategy to drive up sales.
  + How to acquire new customers.
  + Whether expanding the music catalogue will influence company performance.

**No insights or recommendations required.**

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?

**Query:**

|  |
| --- |
| **alter table album add ReleaseYear int;** |

**No insights or recommendations required**

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.

**Query:**

|  |
| --- |
| **with**  **totalamount**  **as**  **(select customer\_id,sum(total) as total\_amount from invoice**  **group by customer\_id),**  **tcount**  **as**  **(select customer\_id,count(track\_id) as track\_count from invoice i**  **join invoice\_line i1 on i.invoice\_id=i1.invoice\_id**  **group by customer\_id)**  **select billing\_country as Country ,count( distinct t.customer\_id) as Customer\_Strength,round(avg(total\_amount),2) avg\_sale\_per\_customer,round(avg(track\_count),2) as tracks\_per\_customer from invoice i**  **join totalamount t on i.customer\_id=t.customer\_id**  **join tcount t1 on i.customer\_id=t1.customer\_id**  **group by billing\_country**  **order by Customer\_Strength desc;** |

**Reference:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Customer\_Strength** | **avg\_sale\_per\_customer** | **tracks\_per\_customer** |
| **USA** | **13** | **81.3** | **82.12** |
| **Canada** | **8** | **69.63** | **70.33** |
| **Brazil** | **5** | **87.56** | **88.44** |
| **France** | **5** | **78.43** | **79.22** |
| **Germany** | **4** | **83.91** | **84.76** |
| **United Kingdom** | **3** | **83.48** | **84.32** |
| **Czech Republic** | **2** | **138.2** | **139.6** |
| **India** | **2** | **96.41** | **97.38** |
| **Portugal** | **2** | **91.49** | **92.41** |
| **Argentina** | **1** | **39.6** | **40** |
| **Australia** | **1** | **81.18** | **82** |
| **Austria** | **1** | **69.3** | **70** |
| **Belgium** | **1** | **60.39** | **61** |
| **Chile** | **1** | **97.02** | **98** |
| **Denmark** | **1** | **37.62** | **38** |
| **Finland** | **1** | **79.2** | **80** |
| **Hungary** | **1** | **78.21** | **79** |
| **Ireland** | **1** | **114.84** | **116** |
| **Italy** | **1** | **50.49** | **51** |
| **Netherlands** | **1** | **65.34** | **66** |
| **Norway** | **1** | **72.27** | **73** |
| **Poland** | **1** | **76.23** | **77** |
| **Spain** | **1** | **98.01** | **99** |
| **Sweden** | **1** | **75.24** | **76** |

**Insights:**

* USA is leading in terms of number of customers
* Average sale per customer and tracks per customer are highest in Czech Republic.

**Recommendations:**

* Customer acquisition must be done in other continents than the Americas.
* Catalogue expansion is needed to adhere to local tastes in music.