

Chinook Database project

Исследуйте данные музыкального магазина с помощью SQL!

Обзор проекта

В этом проекте я выполнил запросы для базы данных Chinook. Chinook содержит информацию о музыкальном магазине.

Мои задачи:

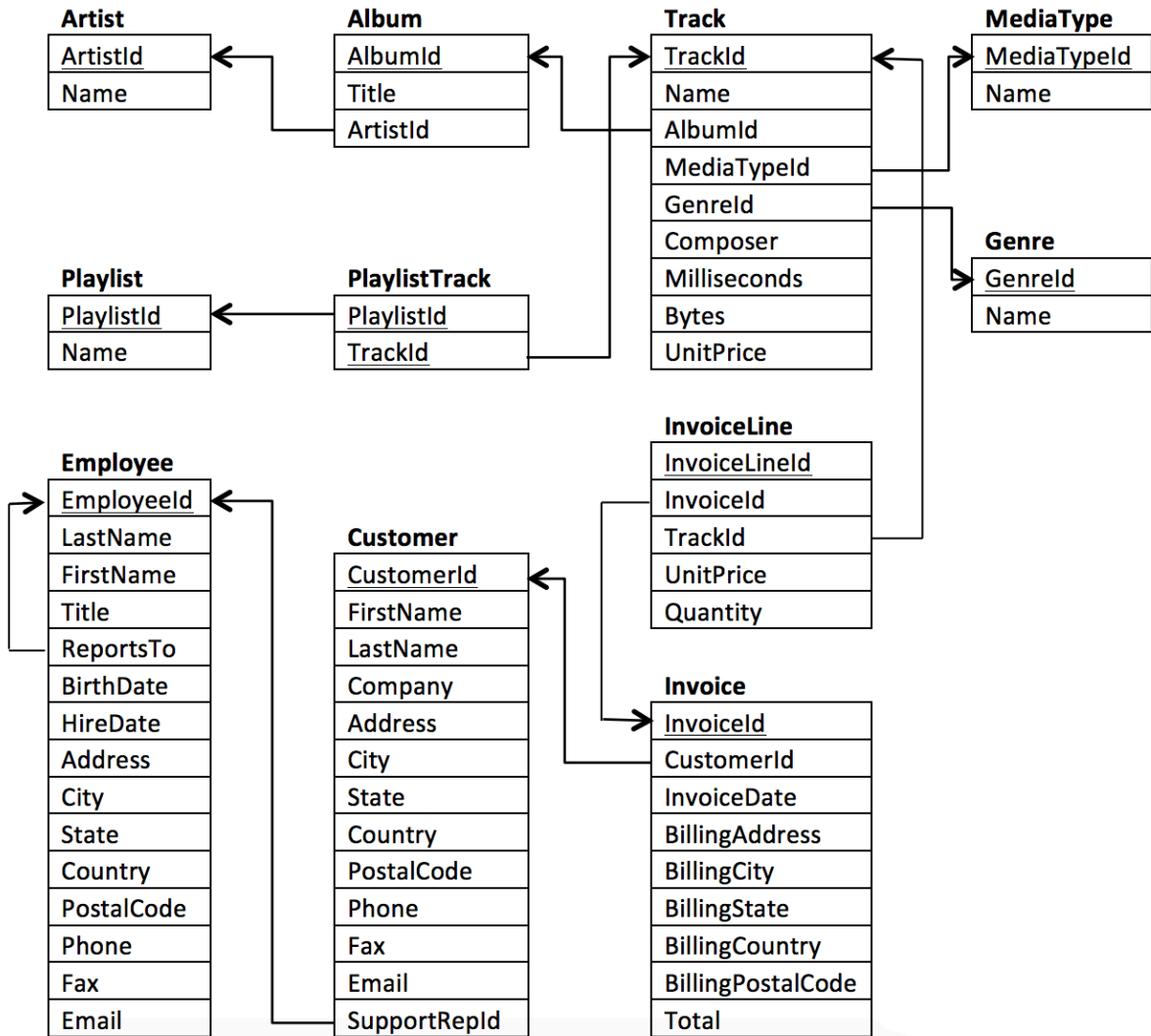
- использовать основные функционалы SQL
- отвечать на вопросы с помощью правильных запросов
- найти интересные инсайты в данных используя JOIN-ы и подзапросы
- переносить данные с одной платформы в другую
- получить удовольствие от процесса

Моя роль

В рамках этого проекта я помог команде Chinook разобраться в товарах в их магазине, их покупателях и сотрудниках, а также в информации об их счетах.

База данных

Схему базы данных можно увидеть ниже:



Часть 1

Вопрос 1: В каких странах больше всего счетов (Invoices)?

Используйте таблицу **Invoice**, чтобы определить страны с наибольшим количеством счетов. Предоставьте таблицу со столбцами **BillingCountry** и **Invoices**, упорядоченную по количеству счетов для каждой страны. Страна с наибольшим количеством счетов должна отображаться первой.

Мой запрос:

```
SELECT BillingCountry, COUNT(InvoiceId) AS Invoices
FROM `da-nfactorial.chinook.invoice`
GROUP BY BillingCountry
ORDER BY Invoices DESC;
```

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	BillingCountry ▼	Invoices ▼				
1	USA	91				
2	Canada	56				
3	Brazil	35				
4	France	35				
5	Germany	28				

Results per page: 50 ▼ 1 – 24 of 24

Вопрос 2: В каком городе самые лучшие клиенты?

Мы хотели бы устроить рекламный музыкальный фестиваль в городе, в котором мы заработали больше всего денег. Напишите запрос, который возвращает город с наибольшей суммой итоговых счетов (invoice totals). Предоставьте название города и общую сумму итоговых счетов.

Мой запрос:

```
SELECT BillingCity, SUM(Total) AS Total_invoices FROM `chinook.invoice`  
  
GROUP BY BillingCity  
  
ORDER BY Total_invoices DESC  
  
LIMIT 1;
```

Row	BillingCity ▼	Total_invoices ▼
1	Prague	90.24

Вопрос 3: Кто является лучшим клиентом?

Клиент, потративший больше всего денег, будет объявлен лучшим клиентом. Напишите запрос, который возвращает человека, который потратил больше всего денег. Мы нашли решение, связав следующие три таблицы: **Invoice**, **InvoiceLine** и **Customer**, чтобы получить эту информацию, но вы, вероятно, можете сделать это с меньшим количеством таблиц.

Мой запрос:

```
SELECT c.FirstName, c.LastName, SUM(i.Total) AS Total_Paid FROM `chinook.customer` c
INNER JOIN `chinook.invoice` i
ON c.CustomerId = i.CustomerId
GROUP BY c.FirstName, c.LastName
ORDER BY Total_Paid DESC
LIMIT 1;
```

Row	FirstName	LastName	Total_Paid
1	Helena	Holý	49.620000000000...

Вопрос 4: Кто слушает рок-музыку?

Напишите запрос, чтобы вернуть адрес электронной почты, имя, фамилию и жанр всех слушателей рок-музыки. Верните свой список в алфавитном порядке по адресу электронной почты.

Мой запрос:

```
SELECT DISTINCT c.Email, c.FirstName, c.LastName, g.Name FROM `chinook.customer` c
INNER JOIN `chinook.invoice` i
ON c.CustomerId = i.CustomerId
INNER JOIN `chinook.invoiceline` il
ON i.InvoiceId = il.InvoiceId
INNER JOIN `chinook.track` t
ON il.TrackId = t.TrackId
INNER JOIN `chinook.genre` g
ON t.GenreId = g.GenreId
WHERE g.Name = 'Rock'
ORDER BY c.Email;
```

Row	Email	FirstName	LastName	Name
1	aaronmitchell@yahoo.ca	Aaron	Mitchell	Rock
2	alero@uol.com.br	Alexandre	Rocha	Rock
3	astrid.gruber@apple.at	Astrid	Gruber	Rock
4	bjorn.hansen@yahoo.no	Bjørn	Hansen	Rock
5	camille.bernard@yahoo.fr	Camille	Bernard	Rock

Results per page: 50 ▼ 1 – 50 of 59

Вопрос 5: Кто пишет рок-музыку?

Теперь, когда мы знаем, что наши клиенты любят рок-музыку, мы можем решить, каких музыкантов пригласить сыграть на концерте.

Давайте пригласим артистов, которые написали больше всего рок-музыки в нашем наборе данных. Напишите запрос, который возвращает имя исполнителя и общее количество треков топ 10 рок-групп по количеству треков.

Мой запрос:

```
SELECT ar.Name, COUNT(t.TrackId) AS Total_tracks FROM `chinook.artist` ar
INNER JOIN `chinook.album` al
ON ar.ArtistId = al.ArtistId
INNER JOIN `chinook.track` t
ON al.AlbumId = t.AlbumId
INNER JOIN `chinook.genre` g
ON t.GenreId = g.GenreId
WHERE g.Name = 'Rock'
GROUP BY ar.Name
ORDER BY Total_tracks DESC
LIMIT 10;
```

Row	Name	Total_tracks
1	Led Zeppelin	114
2	U2	112
3	Deep Purple	92
4	Iron Maiden	81
5	Pearl Jam	54

Results per page: 50 1 – 10 of 10

Вопрос 6: Кто зарабатывает и кто тратит больше всего?

Во-первых, найдите, какой артист заработал больше всего по данным из **InvoiceLines**. Теперь используйте этого исполнителя, чтобы определить, какой клиент потратил больше всего на этого исполнителя.

Мой запрос:

```
SELECT c.FirstName, c.LastName, SUM(i.Total) AS total_spent FROM `chinook.customer` c

JOIN `chinook.invoice` i

ON c.CustomerId = i.CustomerId

JOIN `chinook.invoice` il

ON i.InvoiceId = il.InvoiceId

JOIN `chinook.track` t

ON il.TrackId = t.TrackId

JOIN `chinook.album` al

ON t.AlbumId = al.AlbumId

JOIN `chinook.artist` ar

ON al.ArtistId = ar.ArtistId

WHERE ar.name =

(

SELECT ar.Name FROM `chinook.artist` ar

JOIN `chinook.album` al

ON ar.ArtistId = al.ArtistId

JOIN `chinook.track` t

ON al.AlbumId = t.AlbumId
```



```

JOIN `chinook.invoice` il
ON t.TrackId = il.TrackId

GROUP BY ar.Name

ORDER BY SUM(il.Quantity * il.UnitPrice) DESC

LIMIT 1
)

GROUP BY c.FirstName, c.LastName

ORDER BY total_spent DESC

LIMIT 1;

```

Row	FirstName ▼	LastName ▼	total_spent ▼
1	Mark	Taylor	209.8800000000...

Вопрос 7: Какой жанр самый популярный?

Мы хотим найти самый популярный музыкальный жанр для каждой страны. Мы определяем самый популярный жанр как жанр с наибольшим количеством покупок. Напишите запрос, который возвращает каждую страну вместе с топ 1 жанром. Для стран, где несколько жанров с максимальное количество покупок, верните все жанры соответствующие критериям.

Мой запрос:

WITH

main_table AS

(

SELECT i.BillingCountry AS Country, g.Name AS Genre, SUM(il.Quantity) AS
total_quantity FROM `chinook.invoice` i

JOIN `chinook.invoiceline` il

ON i.InvoiceId = il.InvoiceId

JOIN `chinook.track` t

ON il.TrackId = t.TrackId

JOIN `chinook.genre` g

ON t.GenreId = g.GenreId

GROUP BY Country, Genre

)

SELECT t1.Country, t1.Genre, t1.total_quantity

FROM main_table t1

WHERE t1.total_quantity =

(

SELECT MAX(t2.total_quantity)

```
FROM main_table t2

WHERE t2.Country = t1.Country

);
```

Row	Country ▼	Genre ▼	total_quantity ▼	
1	USA	Rock	157	
2	Chile	Rock	9	
3	India	Rock	25	
4	Italy	Rock	18	
5	Spain	Rock	22	

Results per page: 50 ▼ 1 – 25 of 25

Вопрос 8: Какая песня самая долгая?

Предоставьте названия всех песен, длина которых превышает среднюю длину песни в наборе данных.

Предоставьте столбцы **Name** и **Milliseconds** для каждой песни. Упорядочьте по длине песни, самые длинные песни укажите первыми.

Мой запрос:

```
SELECT Name, Milliseconds FROM `chinook.track`

WHERE Milliseconds > (SELECT AVG(Milliseconds) FROM `chinook.track`)

GROUP BY Name, Milliseconds

ORDER BY Milliseconds DESC;
```

Row	Name ▼	Milliseconds ▼	
1	Occupation / Precipice	5286953	
2	Through a Looking Glass	5088838	
3	Greetings from Earth, Pt. 1	2960293	
4	The Man With Nine Lives	2956998	
5	Battlestar Galactica, Pt. 2	2956081	

Results per page: 50 ▼ 1 – 50 of 494

Вопрос 9:

Напишите запрос, определяющий **клиента, который потратил больше всего на музыку для каждой страны**. Напишите запрос, который возвращает страну вместе с топ 1 покупателем и суммой, которую он потратил. Для стран, в которых есть несколько покупателей потратившие максимальное количество денег, укажите всех клиентов, потративших эту сумму.

Мой запрос:

WITH

main_table_9 AS(

SELECT c.Country, c.FirstName, c.LastName, SUM(i.Total) AS total_spent FROM
`chinook.customer` c

JOIN `chinook.invoice` i

ON c.CustomerId = i.CustomerId

GROUP BY c.Country, c.FirstName, c.LastName

)

SELECT t1.Country, t1.FirstName, t1.LastName, total_spent FROM main_table_9 t1

WHERE t1.total_spent =

(

SELECT MAX(t2.total_spent)

FROM main_table_9 t2

WHERE t2.Country = t1.Country

);

Row	Country ▼	FirstName ▼	LastName ▼	total_spent ▼
1	USA	Richard	Cunningham	47.620000000000...
2	Chile	Luis	Rojas	46.62
3	India	Manoj	Pareek	38.620000000000...
4	Italy	Lucas	Mancini	37.620000000000...
5	Spain	Enrique	Muñoz	37.62

Results per page: 50 ▼ 1 – 25 of 25

Часть 2

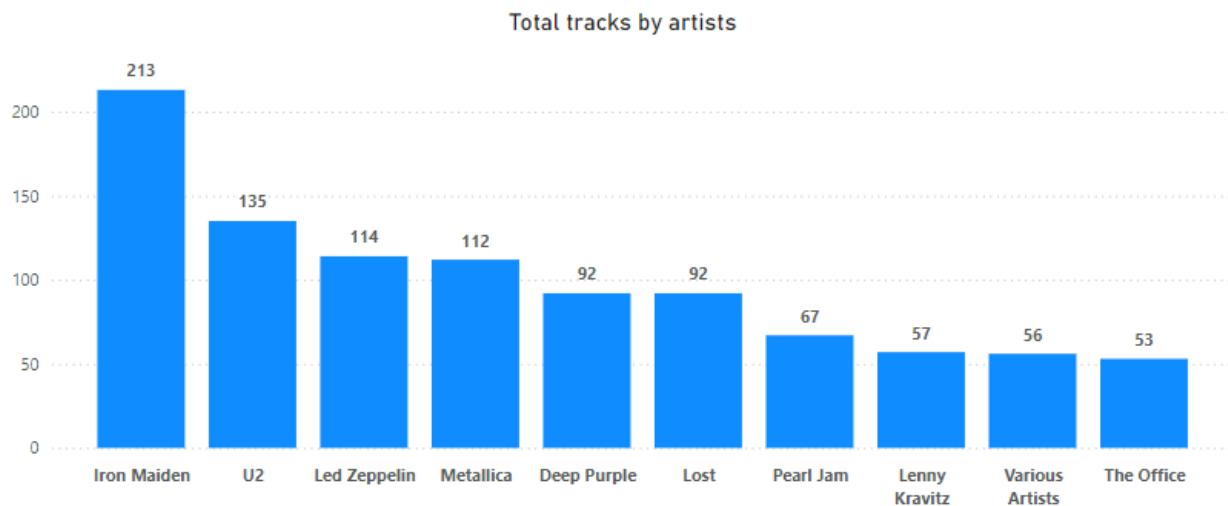
- Определите 4 вопроса о Chinook, на который вы хотите ответить на основе анализа данных.
- Затем напишете SQL-запросы, чтобы получить данные, необходимые для успешного ответа на Ваши вопросы.
- Визуализируйте полученные данные (используя гистограммы или другие графики), отвечающие на Ваш вопрос.
- Объясните ответ в 1-2 предложениях.
- Вопросы, которые вы задаете, зависят от вас, но все четыре запроса должны содержать JOIN и AGGREGATION.

Вопрос 1: Топ 10 артистов у кого больше всего треков

Мой запрос и объяснение с графиком:

```
SELECT ar.Name, COUNT(t.TrackId) AS total_tracks
FROM `chinook.artist` ar
JOIN `chinook.album` al
ON ar.ArtistId = al.ArtistId
JOIN `chinook.track` t
ON al.AlbumId = t.AlbumId
GROUP BY ar.Name
ORDER BY total_tracks DESC
LIMIT 10;
```

Тут нужно было соединить таблицы Artist и Track, для этого заджойнил 3 таблицы Artist->Album->Track. Для кол-во использовал COUNT. Сгруппировал по артистам, отсортировал DESC по кол-ву треков и сделал лимит 10, чтобы вытащить топ 10.



Вопрос 2: Самый популярный трек по кол-ву продаж

Мой запрос и объяснение с графиком:

```
SELECT t.Name, COUNT(il.Quantity) AS total_profit FROM `chinook.track` t  
  
JOIN `chinook.invoice` il  
  
ON t.TrackId = il.TrackId  
  
GROUP BY t.Name  
  
ORDER BY total_profit DESC  
  
LIMIT 1;
```

Тут соединил таблицы Track->InvoiceLines, также использовал COUNT для подсчета кол-ва продаж трека. Сгруппировал по трекам и отсортировал DESC по кол-ву продаж с лимитом 1.

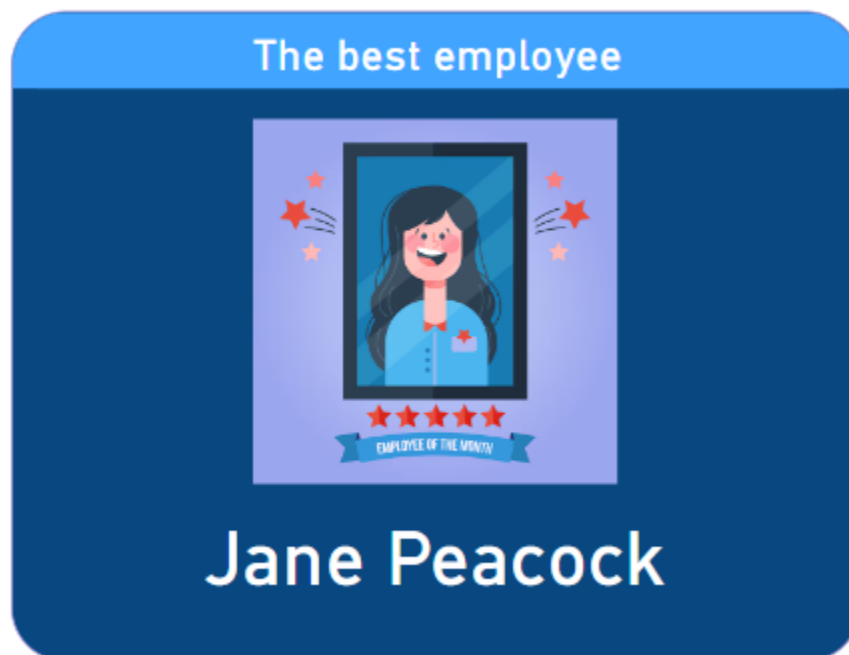


Вопрос 3: К какому сотруднику обращались больше всего?

Мой запрос и объяснение с графиком:

```
SELECT e.FirstName, e.LastName, COUNT(c.SupportRepId) AS total_reps
FROM `chinook.employee` e
JOIN `chinook.customer` c
ON e.EmployeeId = c.SupportRepId
GROUP BY e.FirstName, e.LastName
ORDER BY total_reps DESC
LIMIT 1;
```

Тут соединил таблицы Employee->Customer, также использовал COUNT для подсчета кол-ва обращения. Сгруппировал по имени и фамилии и отсортировал DESC по кол-ву обращения с лимитом 1.



Вопрос 4: В какой стране, средняя сумма продаж больше всего?

Мой запрос и объяснение с графиком:

```
SELECT i.BillingCountry, AVG(i.Total) AS avg_bill FROM `chinook.invoice` i  
  
JOIN `chinook.invoiceline` il  
  
ON i.InvoiceId = il.InvoiceId  
  
GROUP BY i.BillingCountry  
  
ORDER BY avg_bill DESC  
  
LIMIT 1
```

Тут соединил таблицы Invoice->InvoiceLines, использовал AVG для подсчета средней суммы продаж. Сгруппировал по странам и отсортировал DESC по средней сумме продаж с лимитом 1.

