SQL PROJECT

MUSIC STORE ANALYSIS

– PIYUSH SAINI

OBJECTIVE

- THE PROJECT AIMS TO ANALYZE A DIGITAL MUSIC STORE DATABASE USING SQL, PROVIDING STAKEHOLDERS WITH VALUABLE INSIGHTS FOR DECISION-MAKING.

- THROUGH SQL QUERIES, IT ADDRESSES QUESTIONS REGARDING GEOGRAPHICAL GROWTH, PURCHASE POWER, REVENUE, GENRE PERFORMANCE, AND MUSIC BAND POPULARITY.

- THE ANALYSIS OFFERS ACTIONABLE RECOMMENDATIONS TO DRIVE BUSINESS GROWTH AND OPTIMIZE PERFORMANCE.

LEVEL OF QUERIES

EASY

INCLUDES:
SELECT,GROUP BY,
ORDER BY,LIMIT,
DESC/ASC

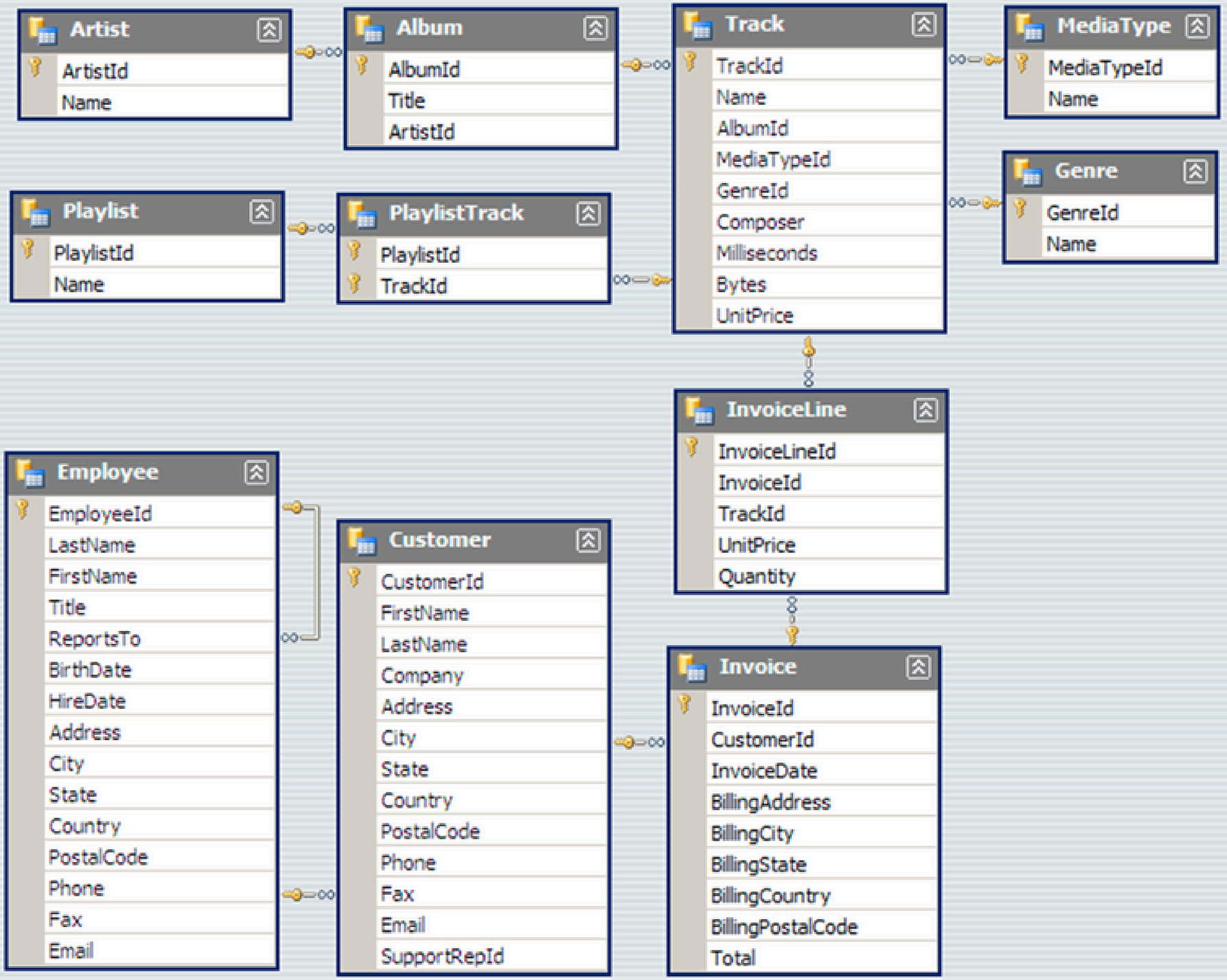
MODERATE

INCLUDES:
JOINS,ORDER BY,
GROUP BY,LIMITS

ADVANCE

INCLUDES:
CTE(COMMON
TABLE
EXPRESSION)

DATABASE SCHEMA



EASY

- QUESTION :1 WHO IS THE SENIOR MOST EMPLOYEE BASED ON JOB TITLE?

INPUT

```
SELECT title, first_name, last_name, employee_id, email
FROM employee
ORDER BY levels desc
LIMIT 1;
```

OUTPUT

	title character varying (50)	first_name character	last_name character	employee_id [PK] character varying (50)	email character varying (30)
1	Senior General Manager	Mohan	Madan	9	madan.mohan@chinookcorp.com

EASY

- QUESTION :2 WHICH COUNTRIES HAVE THE MOST INVOICES?

INPUT

```
SELECT count (*) AS c, billing_country
FROM invoice
Group by billing_country
ORDER BY c DESC;
```

OUTPUT

c bigint	billing_country character varying (30)
131	USA
76	Canada
61	Brazil
50	France
41	Germany
30	Czech Republic
29	Portugal
28	United Kingdom
21	India

EASY

- QUESTION :3 WHAT ARE TOP 3 VALUES OF TOTAL INVOICE?

INPUT

```
SELECT invoice_id, customer_id, invoice_date, total
FROM invoice
ORDER BY total DESC
LIMIT 3;
```

OUTPUT

	invoice_id [PK] integer	customer_id integer	invoice_date timestamp without time zone	total double precision
	183	42	2018-02-09 00:00:00	23.759999999999998
	92	32	2017-07-02 00:00:00	19.8
	31	3	2017-02-21 00:00:00	19.8

EASY

● QUESTION :4 WHICH CITY HAS THE BEST CUSTOMERS? WE WOULD LIKE TO THROW A PROMOTIONAL MUSIC FESTIVAL IN THE CITY WE MADE THE MOST MONEY.
WRITE A QUERY THAT RETURNS ONE CITY THAT HAS THE HIGHEST SUM OF INVOICE TOTALS. RETURN BOTH THE CITY NAME & SUM OF ALL INVOICE TOTALS.

INPUT

```
SELECT SUM (total) AS invoice_total, billing_city
FROM invoice
GROUP BY billing_city
ORDER BY invoice_total DESC;
```

OUTPUT

	invoice_total double precision	billing_city character varying (30)
1	273.24000000000007	Prague
2	169.29	Mountain View
3	166.32	London
4	158.4	Berlin
5	151.47	Paris
6	129.69	São Paulo
7	114.83999999999997	Dublin
8	111.86999999999999	Delhi
9	108.89999999999998	São José dos Campos

EASY

● QUESTION :5 WHO IS THE BEST CUSTOMER? THE CUSTOMER WHO HAS SPENT THE MOST MONEY WILL BE DECLARED THE BEST CUSTOMER.
WRITE A QUERY THAT RETURNS THE PERSON WHO HAS SPENT THE MOST MONEY.

INPUT

```sql
SELECT customer.customer_id, first_name, last_name, city, phone, SUM(total) AS total_spending
FROM customer
JOIN invoice ON customer.customer_id = invoice.customer_id
GROUP BY customer.customer_id
ORDER BY total_spending DESC
LIMIT 1;
```

OUTPUT

	customer_id [PK] integer	first_name character	last_name character	city character varying (50)	phone character varying (50)	total double precision
1	5	R ...	Madhav	Prague	+420 2 4172 5555	144.54000000000002

MODERATE

● QUESTION :1 WRITE QUERY TO RETURN THE EMAIL, FIRST NAME, LAST NAME, & GENRE OF ALL ROCK MUSIC LISTENERS.
RETURN YOUR LIST ORDERED ALPHABETICALLY BY EMAIL STARTING WITH A.

INPUT

```
SELECT DISTINCT email, first_name, last_name
FROM customer
JOIN invoice ON customer.customer_id = invoice.customer_id
JOIN invoice_line ON invoice.invoice_id = invoice_line.invoice_id
WHERE track_id IN(
SELECT track_id FROM track
JOIN genre ON track.genre_id = genre. genre_id
WHERE genre.name LIKE 'Rock'
)
ORDER BY email;
```

OUTPUT

	email character varying (50)	first_name character	last_name character
1	aaronmitchell@yahoo.ca	Aaron	Mitchell
2	alero@uol.com.br	Alexandre	Rocha
3	astrid.gruber@apple.at	Astrid	Gruber
4	bjorn.hansen@yahoo.no	Bjørn	Hansen
5	camille.bernard@yahoo.fr	Camille	Bernard
6	daan_peeters@apple.be	Daan	Peeters
7	diego.gutierrez@yahoo.ar	Diego	Gutiérrez
8	dmiller@comcast.com	Dan	Miller
9	dominiquelefebvre@gmail.c...	Dominique	Lefebvre

MODERATE

● QUESTION :2 LET'S INVITE THE ARTISTS WHO HAVE WRITTEN THE MOST ROCK MUSIC IN OUR DATASET.
WRITE A QUERY THAT RETURNS THE ARTIST NAME AND TOTAL TRACK COUNT OF THE TOP 10 ROCK BANDS.

INPUT

```
SELECT artist.artist_id, artist.name,COUNT(artist.artist_id) AS number_of_songs
FROM track
JOIN album ON album.album_id = track.album_id
JOIN artist ON artist.artist_id = album.artist_id
JOIN genre ON genre.genre_id = track.genre_id
GROUP BY artist.artist_id
ORDER BY number_of_songs DESC
LIMIT 10;
```

OUTPUT

	artist_id [PK] character varying (50)	name character varying (120)	number_of_songs bigint
1	90	Iron Maiden	213
2	150	U2	135
3	22	Led Zeppelin	114
4	50	Metallica	112
5	149	Lost	92
6	58	Deep Purple	92
7	118	Pearl Jam	67
8	100	Lenny Kravitz	57
9	21	Various Artists	56
10	156	The Office	53

MODERATE

● QUESTION :3 RETURN ALL THE TRACK NAMES THAT HAVE A SONG LENGTH LONGER THAN THE AVERAGE SONG LENGTH.
RETURN THE NAME AND MILLISECONDS FOR EACH TRACK. ORDER BY THE SONG LENGTH WITH THE LONGEST SONGS LISTED FIRST.

INPUT

```sql
SELECT name,miliseconds
FROM track
WHERE miliseconds > (
    SELECT AVG(miliseconds) AS avg_track_length
    FROM track )
ORDER BY miliseconds DESC;
```

OUTPUT

	name character varying (150)	miliseconds integer
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
6	Battlestar Galactica, Pt. 1	2952702
7	Murder On the Rising Star	2935894
8	Battlestar Galactica, Pt. 3	2927802
9	Take the Celestra	2927677
10	Fire In Space	2926593
11	The Long Patrol	2925008
12	The Magnificent Warriors	2924716

ADVANCE

● QUESTION :1 FIND HOW MUCH AMOUNT SPENT BY EACH CUSTOMER ON ARTISTS? WRITE A QUERY TO RETURN CUSTOMER NAME, ARTIST NAME AND TOTAL SPENT.

INPUT

```sql
WITH best_selling_artist AS(
SELECT artist.artist_id AS artist_id, artist.name AS artist_name,
SUM(invoice_line.unit_price * invoice_line.quantity) AS total_sales
FROM invoice_line
JOIN track ON track.track_id = invoice_line.track_id
JOIN album ON album.album_id = track.album_id
JOIN artist ON artist.artist_id = album.artist_id
GROUP BY 1
ORDER BY 3 DESC
LIMIT 1
)
SELECT c.customer_id, c.first_name, c.last_name, bsa.artist_name,
SUM(il.unit_price*il.quantity) AS amount_spent
FROM invoice i
JOIN customer c ON c.customer_id = i.customer_id
JOIN invoice_line il ON il.invoice_id = i.invoice_id
JOIN track t ON t.track_id = il.track_id
JOIN album alb ON alb.album_id = t.album_id
JOIN best_selling_artist bsa ON bsa.artist_id = alb.artist_id
GROUP BY 1,2,3,4
ORDER BY 5 DESC;
```

OUTPUT

	customer_id integer	first_name character	last_name character	artist_name character var	amount_spent double precision
1	46	Hugh	O'Reilly	Queen	27.719999999999985
2	38	Niklas	Schröder	Queen	18.81
3	3	François	Tremblay	Queen	17.82
4	34	João	Fernandes	Queen	16.830000000000002
5	53	Phil	Hughes	Queen	11.88
6	41	Marc	Dubois	Queen	11.88
7	47	Lucas	Mancini	Queen	10.89
8	33	Ellie	Sullivan	Queen	10.89
9	20	Dan	Miller	Queen	3.96
10	5	R	Madhav	Queen	3.96
11	23	John	Gordon	Queen	2.9699999999999998
12	54	Steve	Murray	Queen	2.9699999999999998
13	31	Martha	Silk	Queen	2.9699999999999998
14	16	Frank	Harris	Queen	1.98
15	17	Jack	Smith	Queen	1.98
16	24	Frank	Ralston	Queen	1.98
17	30	Edward	Francis	Queen	1.98
18	35	Madalena	Sampaio	Queen	1.98
19	36	Hannah	Schneider	Queen	1.98
20	11	Alexandre	Rocha	Queen	1.98
21	8	Daan	Peeters	Queen	1.98
22	42	Wyatt	Girard	Queen	1.98

ADVANCE

● QUESTION :2 WE WANT TO FIND OUT THE MOST POPULAR MUSIC GENRE FOR EACH COUNTRY. WE DETERMINE THE MOST POPULAR GENRE AS THE GENRE WITH THE HIGHEST AMOUNT OF PURCHASES. WRITE A QUERY THAT RETURNS EACH COUNTRY ALONG WITH THE TOP GENRE. FOR COUNTRIES WHERE THE MAXIMUM NUMBER OF PURCHASES IS SHARED RETURN ALL GENRES.

INPUT

```
WITH popular_genre AS
(
    SELECT COUNT(invoice_line.quantity) AS purchases, customer.country, genre.name, genre.genre_id,
    ROW_NUMBER() OVER(PARTITION BY customer.country ORDER BY COUNT(invoice_line.quantity) DESC) AS RowNo
    FROM invoice_line
    JOIN invoice ON invoice.invoice_id = invoice_line.invoice_id
    JOIN customer ON customer.customer_id = invoice.customer_id
    JOIN track ON track.track_id = invoice_line.track_id
    JOIN genre ON genre.genre_id = track.genre_id
    GROUP BY 2,3,4
    ORDER BY 2 ASC, 1 DESC
)
SELECT * FROM popular_genre WHERE RowNo <= 1
```

OUTPUT

	purchases bigint	country character varying (50)	name character varying (120)	genre_id character varying (50)	rowno bigint
1	17	Argentina	Alternative & Punk	4	1
2	34	Australia	Rock	1	1
3	40	Austria	Rock	1	1
4	26	Belgium	Rock	1	1
5	205	Brazil	Rock	1	1
6	333	Canada	Rock	1	1
7	61	Chile	Rock	1	1

ADVANCE

● QUESTION :3 WRITE A QUERY THAT DETERMINES THE CUSTOMER THAT HAS SPENT THE MOST ON MUSIC FOR EACH COUNTRY.
WRITE A QUERY THAT RETURNS THE COUNTRY ALONG WITH THE TOP CUSTOMER AND HOW MUCH THEY SPENT.
FOR COUNTRIES WHERE THE TOP AMOUNT SPENT IS SHARED, PROVIDE ALL CUSTOMERS WHO SPENT THIS AMOUNT.

INPUT

```sql
WITH Customter_with_country AS (
        SELECT customer.customer_id,first_name,last_name,billing_country,SUM(total) AS total_spending,
        ROW_NUMBER() OVER(PARTITION BY billing_country ORDER BY SUM(total) DESC) AS RowNo
        FROM invoice
        JOIN customer ON customer.customer_id = invoice.customer_id
        GROUP BY 1,2,3,4
        ORDER BY 4 ASC,5 DESC)
SELECT * FROM Customter_with_country WHERE RowNo <= 1
```

OUTPUT

	customer_id integer	first_name character	last_name character	billing_country character varying (30)	total_spending double precision	rowno bigint
1	56	Diego	Gutiérrez	Argentina	39.6	1
2	55	Mark	Taylor	Australia	81.18	1
3	7	Astrid	Gruber	Austria	69.3	1
4	8	Daan	Peeters	Belgium	60.38999999999999	1
5	1	Luís	Gonçalves	Brazil	108.89999999999998	1
6	3	François	Tremblay	Canada	99.99	1
7	57	Luis	Rojas	Chile	97.02000000000001	1
8	5	R	Madhav	Czech Republic	144.54000000000002	1
9	9	Kara	Nielsen	Denmark	37.61999999999999	1


AND THAT'S A WRAP.
THANK YOU!
– PIYUSH SAINI