Exercise Week 2 (Mentoring 1)

SQL & Relational Database

By : Oktavio Reza Putra

1. Determine which countries have the most number of invoices (top 10). Order them by the number of invoices in descending order and if there are the same number of invoices, sort them by country name in ascending order. Show Country Name and total number of invoices.

   a. **SQL Query Syntax:**

```sql
-- 1. Top 10 Negara dengan Jumlah Invoice Terbanyak
select
    country,
    sum(jumlah_invoice) total_invoice
from
    (
        select
            customerid,
            country
        from
            exercise_pacmann.customer
    ) as c
inner join
    (
        select
            customerid,
            count(invoiceid) as jumlah_invoice
        from
            exercise_pacmann.invoice
        group by
            customerid
    ) as i
on
    c.customerid = i.customerid
group by
    country
order by
    total_invoice desc,
    country
limit
    10;
-- Jawaban : USA menjadi negara dengan Jumlah Unique Invoice Terbanyak yaitu sebanyak 91 Invoice
```

   b. **Query Result:**

customer 1

```
select country, sum(jumlah_invoice) to
```

	country	total_invoice
1	USA	91
2	Canada	56
3	Brazil	35
4	France	35
5	Germany	28
6	United Kingdom	21
7	Czech Republic	14
8	Portugal	14
9	India	13
10	Argentina	7

   c. **Description of Query Result:**

   The United States (USA) is the country with the highest number of unique invoices, which is 91 invoices.

2. The top 10 genres by total sales in the database. The total sales are obtained by multiplying the quantity of items sold by their respective prices. Shows Genre Name and Total Sales

a. SQL Query Syntax:

```
-- 2. Mencari Top 10 Genres berdasarkan Total Sales
-- Total Sales = Quantity * Price
with
    track_genre
as
(
    select
        trackid,
        name as genre_name
    from
        exercise_pacmann.genre g
    inner join
        (
            select
                genreid,
                trackid
            from
                exercise_pacmann.track
        ) as t
    on
        g.genreid = t.genreid
)
select
    genre_name,
    sum(quantity * unitprice) as total_sales
from
    track_genre tg
inner join
    (
        select
            trackid,
            unitprice,
            quantity
        from
            exercise_pacmann.invoiceline
    ) as il
on
    tg.trackid = il.trackid
group by
    genre_name
order by
    total_sales desc
limit
    10;
-- Jawaban: Genre Rock adalah genre dengan Penjualan Tertinggi yaitu sebanyak 826.65
```

b. Query Result:

Grid | Text

	genre_name	total_sales
1	Rock	826.65
2	Latin	382.14
3	Metal	261.36
4	Alternative & Punk	241.56
5	TV Shows	93.53
6	Jazz	79.2
7	Blues	60.39
8	Drama	57.71
9	Classical	40.59
10	R&B/Soul	40.59

c. Description of Query Result:

The Rock genre is the Highest Selling genre at 826.65

3. Who are the top 10 customers by their total spending? Shows Customer Name (consist of first name and last name), Email, and Total Spending

a. SQL Query Syntax:

```sql
-- 3. Top 10 Customer Spender
with
    cust_invoice
as
(
    select
        *
    from
        (
            select
                customerid,
                concat(firstname,' ',lastname) as customername,
                email
            from
                exercise_pacmann.customer
        ) as c
    join
        (
            select
                customerid,
                invoiceid
            from
                exercise_pacmann.invoice
        ) as i
    on
        c.customerid = i.customerid
)
select
    customername,
    email,
    sum(quantity * unitprice) as total_spending
from
    cust_invoice ci
inner join
    (
        select
            invoiceid,
            quantity,
            unitprice
        from
            exercise_pacmann.invoiceline
    ) as i
on
    ci.invoiceid = i.invoiceid
group by
    customername,
    email
order by
    total_spending desc
limit
    10;
-- Jawaban: Customer atas nama Helena Hol adalah Top Spender tertinggi dengan total spending 49.62
```

b. Query Result:

	customername	email	total_spending
1	Helena Hol	hholy@gmail.com	49.62
2	Richard Cunningham	ricunningham@hotmail.com	47.62
3	Luis Rojas	luisrojas@yahoo.cl	46.62
4	Ladislav Kov cs	ladislav_kovacs@apple.hu	45.62
5	Hugh O'Reilly	hughoreilly@apple.ie	45.62
6	Fynn Zimmermann	fzimmermann@yahoo.de	43.62
7	Frank Ralston	fralston@gmail.com	43.62
8	Julia Barnett	jubarnett@gmail.com	43.62
9	Astrid Gruber	astrid.gruber@apple.at	42.62
10	Victor Stevens	vstevens@yahoo.com	42.62

c. Description of Query Result:

Customer by the name of Helena Holy is the highest Top Spender with a total spending of 49.62

4. In the results list of countries in number 1, which city has the most number of invoices? Show Country Name, City Name and total number of invoices.

Answer Number 1:

customer 1 ×

select country, sum(jumlah_invoice) to

	country	total_invoice
1	USA	91
2	Canada	56
3	Brazil	35
4	France	35
5	Germany	28
6	United Kingdom	21
7	Czech Republic	14
8	Portugal	14
9	India	13
10	Argentina	7

a. SQL Query Syntax:

```
-- 4. Berikan Data Negara dengan Total Invoice Terbanyak berdasarkan Jawaban Nomor 1
-- Tampilkan Country, City dan Total Invoice
create view
    top10_country_invoices
as
    select
        country,
        sum(jumlah_invoice) total_invoice
    from
        (
            select
                customerid,
                country
            from
                exercise_pacmann.customer
        ) as c
    inner join
        (
            select
                customerid,
                count(invoiceid) as jumlah_invoice
            from
                exercise_pacmann.invoice
            group by
                customerid
        ) as i
    on
        c.customerid = i.customerid
    group by
        country
    order by
        total_invoice desc,
        country
    limit
        10;

select
    country,
    city,
    sum(jumlah_invoice) as total_invoice
    from
    (
        select
            c.customerid,
            country,
            city,
            count(invoiceid) as jumlah_invoice
        from
            (
                select
                    customerid,
                    country,
                    city
                from
                    exercise_pacmann.customer
            ) as c
        inner join
            (
                select
                    customerid,
                    invoiceid
                from
                    exercise_pacmann.invoice
            ) as i
        on
            c.customerid = i.customerid
        group by
            c.customerid,
            country,
            city
    ) as ti
where
    country in (
                    select
                        country
                    from
                        exercise_pacmann.top10_country_invoices tci
                )
group by
    country,
    city
order by
    total_invoice desc;
-- Jawaban: Kota London, Mountain View, Paris, Berlin, Prague, Sao Pulo adalah Kota-Kota yang invoice nya terbanyak sebanyak 14 Invoice
```

b. Query Result:

	country	city	total_invoice
1	United Kingdom	London	14
2	USA	Mountain View	14
3	France	Paris	14
4	Germany	Berlin	14
5	Czech Republic	Prague	14
6	Brazil	S o Paulo	14
7	Canada	Montr al	7
8	Canada	Ottawa	7
9	Canada	Toronto	7
10	Canada	Vancouver	7
11	Canada	Winnipeg	7
12	Canada	Yellowknife	7
13	France	Bordeaux	7
14	France	Dijon	7
15	France	Lyon	7
16	Germany	Frankfurt	7
17	Germany	Stuttgart	7
18	India	Delhi	7
19	Portugal	Lisbon	7
20	Portugal	Porto	7
21	United Kingdom	Edinburgh	7
22	USA	Boston	7
23	USA	Chicago	7

c. Description of Query Result:

London, Mountain View, Paris, Berlin, Prague, Sao Pulo are the cities that have the most invoices with a total of 14 invoices in each of these cities.

5. The product team is looking to add some tracks from new artists to the store and market them in the United Kingdom. Due to budget constraints for marketing, the product team needs to select 4 out of 6 songs to include in the store. The product team assumes that they should choose songs with genres that are popular in the United Kingdom.

a. SQL Query Syntax:

```sql
-- 5. Memilih 4 dari 6 Lagu untuk dimasukkan dalam Toko (Store)
with
    customer_invoice_track
as
(
    select
        customerid,
        country,
        il.invoiceid,
        trackid,
        quantity
    from
    (
        select
            c.customerid,
            country,
            invoiceid
        from
            (
                select
                    customerid,
                    country
                from
                    exercise_pacmann.customer
                where
                    country = 'United Kingdom'
            ) as c
        inner join
            (
                select
                    customerid,
                    invoiceid
                from
                    exercise_pacmann.invoice
            ) as i
        on
            c.customerid = i.customerid
    ) as cust_country_invoice
    inner join
        (
            select
                invoiceid,
                trackid,
                quantity
            from
                exercise_pacmann.invoiceline
        ) as il
    on
        cust_country_invoice.invoiceid = il.invoiceid
)

select
    country,
    genre,
    sum(quantity) as purchase_total
from
    customer_invoice_track as cit
join
    (
        select
            trackid,
            t.genreid,
            t.name as song_name,
            g.name as genre
        from
            exercise_pacmann.track t
        join
            exercise_pacmann.genre g
        on
            t.genreid = g.genreid
    ) as t_join
on
    cit.trackid = t_join.trackid
group by
    country,
    genre
order by
    purchase_total desc;
```

b. Query Result:

	country	genre	purchase_total
1	United Kingdom	Rock	37
2	United Kingdom	Latin	31
3	United Kingdom	Metal	20
4	United Kingdom	Alternative & Punk	9
5	United Kingdom	Reggae	5
6	United Kingdom	Jazz	4
7	United Kingdom	Hip Hop/Rap	3
8	United Kingdom	R&B/Soul	2
9	United Kingdom	Pop	2
10	United Kingdom	World	1

c. Description of Query Result:

Based on the query results, it is found that the most purchased Rock Genre Songs, so there are 4 songs that can be added are Rock, Reggae, Jazz, Hip Hop/Rap genres.

So the songs that can be added are :

- "Good to See You": Rock,
- "Got Ya Before Sunrise":Reggae,
- "Nothing On You":Jazz,
- "Before The Coffee Gets Cold":Hip Hop/Rap.

6. The Product Team wants to market albums that are popular in the USA to be marketed in other countries. Help the product team by searching for the 10 most popular albums in the USA based on album units sold

a. SQL Query Syntax:

```sql
-- 6. Album that Popular in USA
with
    invoice_track_qty
as
(
    select
        invoiceid,
        trackid,
        quantity
    from
        (
            select
                customerid,
                country
            from
                exercise_pacmann.customer
            where
                country = 'USA'
        ) as cust_country
    inner join
        (
            select
                customerid,
                i.invoiceid,
                trackid,
                quantity
            from
            (
                select
                    customerid,
                    invoiceid
                from
                    exercise_pacmann.invoice
            ) as i
            inner join
                (
                    select
                        invoiceid,
                        trackid,
                        quantity
                    from
                        exercise_pacmann.invoiceline
                ) as il
            on
                i.invoiceid = il.invoiceid
        ) as invoice_track
    on
        cust_country.customerid = invoice_track.customerid
)
select
    album_title,
    sum(quantity) as total_purchase
from
    invoice_track_qty itq
inner join
    (
        select
            trackid,
            title as album_title
        from
            (
                select
                    trackid,
                    albumid
                from
                    exercise_pacmann.track
            ) as t
        inner join
            (
                select
                    albumid,
                    title
                from
                    exercise_pacmann.album
            ) a
        on
            t.albumid = a.albumid
    ) a
on
    itq.trackid = a.trackid
group by
    album_title
order by
    total_purchase desc
limit
    10;
```

b. Query Result:

album_title	total_purchase
The Office, Season 3	14
Prenda Minha	11
Unplugged	11
Chill: Brazil (Disc 2)	10
Back to Black	9
International Superhits	8
Vin cius De Moraes - Sem Limite	7
A-Sides	7
B-Sides 1980-1990	7
Serie Sem Limite (Disc 1)	6

c. Description of Query Result:

The Office, Season 3 became the highest of Top 10 most popular Albums in USA with 14 Orders. Then followed by Prenda Minha, Unplugged, Chill: Brazil (Disc 2), Back to Black, International Superhits, etc.

7. Provide a table that aggregates purchase data by country. In cases where a country has only one customer, group these countries as 'Other.' The results should be sorted by total sales in descending order. Here are Informations to calculate:

- **Total Number of Customers**: Calculate the count of unique customers within each country.
- **Total Value of Sales**: Sum the total sales value for each country.
- **Average Value of Sales per Customer**: Divide the total sales value by the number of unique customers in each country
- **Average Order Value**: Divide the total sales value by the number of orders (invoices) placed in each country to calculate the average order value

a. SQL Query Syntax:

```sql
-- 7. Buat Tabel Aggregate Purchase Data by Country
with
    sales_summary
as
(
    with
        invoice_cust
    as
    (
        select
            c.customerid,
            country,
            invoiceid
        from
            (
                select
                    customerid,
                    country
                from
                    exercise_pacmann.customer
            ) c
        inner join
            (
                select
                    customerid,
                    invoiceid
                from
                    exercise_pacmann.invoice
            ) i
        on
            c.customerid = i.customerid
    )
    select
        case
            when
                count(distinct customerid) = 1
            then
                'Other'
            else
                country
        end as country,
        count(distinct customerid) as unique_customer_cnt,
        sum(quantity * unitprice) as total_value_of_sales,
        count(distinct ic.invoiceid) as total_order
    from
        invoice_cust ic
    inner join
        (
            select
                invoiceid,
                quantity,
                unitprice
            from
                exercise_pacmann.invoiceline
        ) il
    on
        ic.invoiceid = il.invoiceid
    group by
        country
    order by
        country
)
select
    country,
    sum(unique_customer_cnt) as total_number_of_customers,
    sum(total_value_of_sales) as total_value_of_sales,
    sum(total_value_of_sales) / sum(unique_customer_cnt) as avg_value_of_sales_per_cust,
    sum(total_value_of_sales) / sum(total_order) as avg_order_value
from
    sales_summary
group by
    country
order by
    country;
```

b. Query Result:

	country	total_number_of_customers	total_value_of_sales	avg_value_of_sales_per_cust	avg_order_value
1	Brazil	5	190.1	38.02	5.4314285714
2	Canada	8	303.96	37.995	5.4278571429
3	Czech Republic	2	90.24	45.12	6.4457142857
4	France	5	195.1	39.02	5.5742857143
5	Germany	4	156.48	39.12	5.5885714286
6	India	2	75.26	37.63	5.7892307692
7	Other	15	604.3	40.2866666667	5.7552380952
8	Portugal	2	77.24	38.62	5.5171428571
9	United Kingdon	3	112.86	37.62	5.3742857143
10	USA	13	523.06	40.2353846154	5.7479120879

c. Description of Query Result:

- Highest Total Customer is USA with 13 customers (Others are not counted because they consist of many countries)
- Highest Total Value of Sales is USA with 523.06
- The highest Average Value of Sales per Customer is Czech Republic
- The highest Average Order Value (AOV) is Czech Republic with 6.45

8. Some genres have low sales, the product team wants to analyze which genres need to be boosted by carrying out additional promotion or other strategies. Because each country has different behavior, the product team started by analyzing sales in USA *(The total sales are obtained by multiplying the quantity of items sold by their respective prices)*

a. SQL Query Syntax:

```
with
    sales_summary
as
(
with
    customer_invoice_track
as
(
    select
        customerid,
        country,
        il.invoiceid,
        trackid,
        quantity,
        unitprice
    from
    (
        select
            c.customerid,
            country,
            invoiceid
        from
            (
                select
                    customerid,
                    country
                from
                    exercise_pacmann.customer
                where
                    country = 'USA'
            ) as c
        inner join
            (
                select
                    customerid,
                    invoiceid
                from
                    exercise_pacmann.invoice
            ) as i
        on
            c.customerid = i.customerid
    ) as cust_country_invoice
    inner join
        (
            select
                invoiceid,
                trackid,
                quantity,
                unitprice
            from
                exercise_pacmann.invoiceline
        ) as il
    on
        cust_country_invoice.invoiceid = il.invoiceid
)
    select
        country,
        genre,
        sum(quantity) as qty,
        sum(quantity * unitprice) as total_sales
    from
        customer_invoice_track as cit
    join
        (
            select
                trackid,
                t.genreid,
                t.name as song_name,
                g.name as genre
            from
                exercise_pacmann.track t
            join
                exercise_pacmann.genre g
            on
                t.genreid = g.genreid
        ) as t_join
    on
        cit.trackid = t_join.trackid
    group by
        country,
        genre
    order by
        total_sales
)
select
    *,
    case
        when
            total_sales < avg(total_sales) over()
        then
            'Total Sales Below Sales Average'
        else
            'Total Sales Above Sales Average'
    end as sales_category
from
    sales_summary
;
```

b. Query Result:

	country	genre	qty	total_sales	sales_category
1	USA	Science Fiction	1	1.99	Total Sales Below Sales Average
2	USA	Easy Listening	3	2.97	Total Sales Below Sales Average
3	USA	Rock And Roll	3	2.97	Total Sales Below Sales Average
4	USA	Heavy Metal	4	3.96	Total Sales Below Sales Average
5	USA	Hip Hop/Rap	4	3.96	Total Sales Below Sales Average
6	USA	Soundtrack	4	3.96	Total Sales Below Sales Average
7	USA	Alternative	5	4.95	Total Sales Below Sales Average
8	USA	Pop	5	4.95	Total Sales Below Sales Average
9	USA	Reggae	6	5.94	Total Sales Below Sales Average
10	USA	Bossa Nova	7	6.93	Total Sales Below Sales Average
11	USA	Classical	8	7.92	Total Sales Below Sales Average
12	USA	Sci Fi & Fantasy	5	9.95	Total Sales Below Sales Average
13	USA	R&B/Soul	12	11.88	Total Sales Below Sales Average
14	USA	Drama	6	11.94	Total Sales Below Sales Average
15	USA	Blues	15	14.85	Total Sales Below Sales Average
16	USA	Comedy	8	15.92	Total Sales Below Sales Average
17	USA	Jazz	22	21.78	Total Sales Below Sales Average
18	USA	TV Shows	14	27.86	Total Sales Above Sales Average
19	USA	Alternative & Pun	50	49.5	Total Sales Above Sales Average
20	USA	Metal	64	63.36	Total Sales Above Sales Average
21	USA	Latin	91	90.09	Total Sales Above Sales Average
22	USA	Rock	157	155.43	Total Sales Above Sales Average

c. Description of Query Result:

Genres other than Rock, Latin, Metal, Alternative & Punk and TV Shows are below Average Sales. So these genres need to be boosted in sales, especially the Science Fiction genre which has the lowest total sales among other genres in the USA.

9. We want to advertise songs to the customer based on how much each customers spent per genre. Help Marketing Team to find Top genre for each customers with the most spent

a. SQL Query Syntax:

```sql
-- 9. Mencari Lagu yang bisa diiklankan berdasarkan Customer Spending per Genre
with
    cust_spend_genre
as
(
    with
        cust_invoice_track
    as
    (
        select
            customerid,
            lastname,
            firstname,
            il.invoiceid,
            trackid,
            quantity,
            unitprice
        from
            (
                select
                    c.customerid,
                    lastname,
                    firstname,
                    invoiceid
                from
                    exercise_pacmann.customer c
                inner join
                    exercise_pacmann.invoice i
                on
                    c.customerid = i.customerid
            ) cust_invoice
        inner join
            exercise_pacmann.invoiceline il
        on
            il.invoiceid = cust_invoice.invoiceid
    )
    select
        customerid,
        firstname,
        lastname,
        genre,
        sum(quantity * unitprice) as total_sales
    from
        cust_invoice_track cit
    inner join
        (
            select
                trackid,
                g.name as genre
            from
                exercise_pacmann.track t
            inner join
                exercise_pacmann.genre g
            on
                t.genreid = g.genreid
        ) t
    on
        cit.trackid = t.trackid
    group by
        customerid,
        firstname,
        lastname,
        genre
    order by
        customerid,
        total_sales desc
)
select
    customerid,
    firstname,
    lastname,
    genre,
    total_sales,
    dense_rank() over(partition by customerid order by total_sales desc) as genre_rank_per_customer
from
    cust_spend_genre
order by
    total_sales desc;
```

b. Query Result:

	customerid	firstname	lastname	genre	total_sales	genre_rank_per_customer
1	10	Eduardo	Martins	Rock	28.71	1
2	29	Robert	Brown	Rock	24.75	1
3	49	Stanislaw	W jcik	Rock	21.78	1
4	55	Mark	Taylor	Rock	21.78	1
5	50	Enrique	Mu oz	Rock	21.78	1
6	8	Daan	Peeters	Rock	20.79	1
7	9	Kara	Nielsen	Rock	20.79	1
8	38	Niklas	Schr der	Rock	20.79	1
9	33	Ellie	Sullivan	Rock	18.81	1
10	30	Edward	Francis	Rock	18.81	1
11	18	Michelle	Brooks	Rock	18.81	1
12	36	Hannah	Schneider	Metal	17.82	1
13	53	Phil	Hughes	Rock	17.82	1
14	48	Johannes	Van der Berg	Rock	17.82	1
15	47	Lucas	Mancini	Rock	17.82	1
16	44	Terhi	H m l inen	Rock	17.82	1
17	43	Isabelle	Mercier	Rock	17.82	1
18	4	Bj rn	Hansen	Rock	16.83	1
19	2	Leonie	K hler	Rock	16.83	1
20	28	Julia	Barnett	Rock	16.83	1
21	40	Dominique	Lefebvre	Rock	16.83	1
22	35	Madalena	Sampaio	Rock	15.84	1
23	12	Roberto	Almeida	Latin	15.84	1
24	12	Roberto	Almeida	Rock	15.84	1
25	27	Patrick	Gray	Rock	15.84	1
26	11	Alexandre	Rocha	Latin	15.84	1
27	19	Tim	Gover	Rock	14.85	1

c. **Description of Query Result:**
The Rock genre is the genre that every customer who transacts with a total of 43 customers and Eduardo Martins is the Rock enthusiast customer with the highest total sales.

10. The Marketing team wants to increase advertising in countries with customers who have spent the most money. Help the Marketing team find the top 10 countries with the highest-spending customers.

a. **SQL Query Syntax:**

```sql
-- 10. Negara yang Spending Customernya paling banyak
select
    country,
    sum(quantity * unitprice) as total_sales
from
    (
        select
            country,
            invoiceid
        from
            exercise_pacmann.customer c
        inner join
            exercise_pacmann.invoice i
        on
            c.customerid = i.customerid
    ) as t_cust
inner join
    exercise_pacmann.invoiceline il
on
    t_cust.invoiceid = il.invoiceid
group by
    country
order by
    total_sales desc
limit 10;
```

b. **Query Result:**

	country	total_sales
1	USA	523.06
2	Canada	303.96
3	France	195.1
4	Brazil	190.1
5	Germany	156.48
6	United Kingdom	112.86
7	Czech Republic	90.24
8	Portugal	77.24
9	India	75.26
10	Chile	46.62

c. **Description of Query Result:**
USA is the country with the highest Total Spending out of the Top 10 Countries