SQL project - Ecommerce Sales Analysis

Your client, a Brazilian E-commerce store, wants you to help them understand how their sales trend across different states over the years and why?.

For this analysis, you are given the data for the time period 2016-2018

a. **Create the different metrics like Sales, customer acquisitions, total no. of orders for each Year across the different states they serve.**

Does all the metrices show similar trends or is there any disparity amongst each of them?

```
select year(o.order_purchase_timestamp) as years ,
 c.customer_state as state, count(o.order_id) as number_of_order
from orders_dataset as o inner join customers as c on
c.customer_id=o.customer_id
group by year(o.order_purchase_timestamp),c.customer_state order by
year(o.order_purchase_timestamp)
```

```
Sales
select * into sales from(
select year(o.order_purchase_timestamp) as years ,oi.order_id as
order_id,oi.product_id as product_id,
 c.customer_state as state ,c.customer_id as customer_id,
p.product_category_name
from orders_dataset as o inner join customers as c on
c.customer_id=o.customer_id
inner join order_items as oi on o.order_id=oi.order_id  inner join
olist_products_dataset$ as p on oi.product_id=p.product_id
group by
year(o.order_purchase_timestamp),p.product_category_name,c.customer_state,
c.customer_id
 ,o.order_id,oi.order_id ,oi.product_id
)c
```

```
select years,product_category_name , count(order_id) as
number_of_order,state
from sales
where product_category_name <> 'NULL'  group by
years,product_category_name,state
```

sales aquesition

```
select * ,CASE WHEN order_purchase_timestamp = first_purchase_date THEN 1 ELSE 0
END AS 'isNewCustomer'
from (SELECT DISTINCT
customer_id, order_purchase_timestamp, MIN(order_purchase_timestamp) OVER
(PARTITION BY customer_id) AS
first_purchase_date
FROM orders_dataset)c
```

b. Using the above metrics, identify the top 2 States which show
i. Declining trend over the years

```
select distinct years,state, COUNT(order_id) over
(partition by years, state)
as number_of_orders from sales where order_status = 'canceled';
```

2.Increasing trend over the years

```
select * from(
select  distinct years,state,product_category_name ,count(order_id)  over
(partition by product_category_name order by years)
as number_of_orders from sales where product_category_name != 'null' )c
order by number_of_orders desc
```

c. For the States identified above, do the Root Cause analysis for their performance across a variety of metrics.
You can utilize the following metrics and explore a few yourself as well by analyzing the data.
Category level Sales and orders placed, post-order reviews, Seller performance in terms of deliveries, product-level sales & orders placed,
% of orders delivered earlier than the expected date, % of orders delivered later than the expected date, etc.*/

```
select * into pos from(
select order_status,status * 100.0/ SUM(status) OVER()
'Percentage(%) of status' from (
select distinct order_status, count(order_status)  over (partition
by order_status)
as status from orders_dataset )c
group by order_status,status) c

select * from pos

select * into payment_type from(
select payment_type,no_of_payment * 100.0/ SUM(no_of_payment) OVER()
'Percentage(%)' from (
select distinct payment_type, count(payment_type)  over (partition by
payment_type) as no_of_payment from order_payments )c
```

```
group by payment_type,no_of_payment)c

select distinct
s.seller_id,o.order_purchase_timestamp,o.order_delivered_customer_date,
datediff(day,o.order_purchase_timestamp,o.order_delivered_customer_date)
as
'no of day to take delivery',avg(r.review_score) as 'reviews to seller'
from
sellers_dataset as s
inner join order_items as oi on s.seller_id=oi.seller_id inner join
orders_dataset as o
on oi.order_id=o.order_id inner join order_reviews as r on
o.order_id=r.order_id
where o.order_status = 'delivered'
group by
s.seller_id,o.order_purchase_timestamp,o.order_delivered_customer_date
```

d. Do the above analysis for the top 2 cities which are causing the trend for each of the states identified in point (b)

```
select top 2 *  from (
select  distinct s.years,s.state,s.product_category_name ,sd.seller_city,
count(order_id)
over (partition by product_category_name order by years)
as number_of_orders from sales as s inner join sellers_dataset as sd on
s.state=sd.seller_state
where product_category_name != 'null' )c order by number_of_orders desc
```

e. After doing the Root cause analysis, help the client by suggesting ways to
improve the performance of the States and the cities

- Sellers should improve delivery time.
- Sellers having less than 3 rating, needs to improve their product quality.
- Sellers should improve delivery time.
- Sellers having less than 3 rating, needs to improve their product quality.
- Some of state has very less sale so, companies needs to work on it .
- If they concentrate on these state they can improve their business
- And they earn more.