***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

1. 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 acquisition**

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';

**ii. 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.