SQL PROJECT ON E-COMMERCE SALES ANALYSIS

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 metrics show similar trends or is there any disparity amongst each of them?

Ans. Here I simply count customer, in each state for customer acquisition for each year

```
---qun a, customer aquasition for year 2016
select customer_state,count(customer_id) as cuntomer_count
from dbo.olist_customers_dataset
where customer_id in(select customer_id from dbo.olist_orders_dataset where datepart(year,order_delivered_customer_date) = 2016)
group by customer_state;

---qun a, for year 2017
select customer_state,count(customer_id) as cuntomer_count
from dbo.olist_customers_dataset
where customer_id in(select customer_id from dbo.olist_orders_dataset where datepart(year,order_delivered_customer_date) = 2017)
group by customer_state;

---qun a, for year 2018
select customer_state,count(customer_id) as cuntomer_count
from dbo.olist_customers_dataset
where customer_id in(select customer_id from dbo.olist_orders_dataset where datepart(year,order_delivered_customer_date) = 2018)
group by customer_state;
```

The out put is as given

customer_state	cuntomer_count
AL	1
BA	3
CE	6
DF	6
ES	3
GO	7
MA	4
MG	35
MT	1
PA	4
PB	1
PE	6
PI	1
PR	20
RJ	43
RN	4
RR	1
RS	17
SC	10
SE	3
SP	96

2016

customer_state	cuntomer_count
AL	188
AM	63
AP	23
BA	1382
CE	574
DF	827
ES	879
GO	838
MA	320
MG	4999
MS	273
MT	374
PA	435
PB	215
PE	664
PI	199
PR	2072
RJ	5442
RN	209
RO	127
RR	17
RS	2405
SC	1536
SE	166
SP	16533

2017

customer_state	cuntomer_count
AC	32
AL	208
AM	82
AP	44
BA	1871
CE	699
DF	1247
ES	1113
GO	1112
MA	393
MG	6321
MS	428
MT	511
PA	507
PB	301
PE	923
PI	276
PR	2831
RJ	6868
RN	261
RO	116
RR	23
RS	2922
SC	2001
SE	166
SP	23866
TO	152

2018

Here is total number of order

```
select year(o.order_purchase_timestamp) as year,
c.customer_state as state,count(o.order_id) as count_of_order
from dbo.olist_orders_dataset as o inner join dbo.olist_customers_dataset 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);
```

Ther output is

year	state	count_of_order
2016	MT	3
2016	BA	4
2016	AL	2
2016	RR	2
2016	SP	115
2016	PA	4
2016	CE	8
2016	DF	6
2016	MA	4
2016	GO	9
2016	PB	1
2016	PE	7
2016	PI	1
2016	ES	4
2016	PR	20
2016	RS	25
2016	RN	4
2016	MG	40
2016	SC	11
2016	RJ	56
2016	SE	3

year	state	count_of_order
2017	DF	921
2017	MA	387
2017	AC	54
2017	GO	955
2017	PA	504
2017	ES	968
2017	MG	5414
2017	CE	660
2017	AL	207
2017	PE	772
2017	RO	141
2017	PB	260
2017	AP	29
2017	MS	301
2017	BA	1592
2017	RN	234
2017	RJ	6225
2017	RR	18
2017	RS	2661
2017	PI	226
2017	MT	420
2017	SC	1720
2017	TO	134
2017	PR	2270
2017	SP	17760
2017	SE	193

2018	RR	26
2018	PB	275
2018	DF	1213
2018	SE	154
2018	PR	2755
2018	BA	1784
2018	AP	39
2018	PA	467
2018	RS	2780
2018	AC	27
2018	RN	247
2018	ES	1061
2018	SP	23871
2018	MS	414
2018	MT	484
2018	CE	668
2018	MA	356
2018	PI	268
2018	MG	6181
2018	RO	112
2018	PE	873
2018	SC	1906
2018	TO	146

Total sales

```
---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 dbo.olist_orders_dataset as o inner join dbo.olist_customers_dataset as c on
c.customer_id=o.customer_id
inner join dbo.olist_order_items_dataset as oi on o.order_id=oi.order_id inner join
dbo.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
```

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 count_of_orders from sales
where product_category_name='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.

Percentage of comparing delivery report

```
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 dbo.olist_orders_dataset )c
group by order_status,status) c
select * from pos
```

output is

order_status	Percentage(%) of status
canceled	0.628513389849
delivered	97.020343721402
approved	0.002011242847
invoiced	0.315765127060
processing	0.302692048551
created	0.005028107118
shipped	1.113222916101
unavailable	0.612423447069

Percentage of payment methods

```
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 dbo.olist_order_payments_dataset )c
group by payment_type,no_of_payment)c
select * from payment_type
```

column no_of_payment(int, null)

Output is

	payment_type	Percentage(%)
1	boleto	19.043952024334
2	debit_card	1.471805633097
3	voucher	5.558978110621
4	credit_card	73.922376451109
5	not_defined	0.002887780836

Number of day takes for delivery and day takes for delivery

```sql
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
dbo.olist_sellers_dataset as s
inner join dbo.olist_order_items_dataset as oi on s.seller_id=oi.seller_id inner join
dbo.olist_orders_dataset as o
on oi.order_id=o.order_id inner join dbo.olist_order_reviews_dataset 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)

```sql
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 dbo.olist_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

1. Most of state have very less sales hence the have to improve that.
2. Then delivery time is needs to improve
3. The product quality is low,where customer rating is low
4. If they concentrate on these state they can improve their business and they earn more.
5. Credit card user is more,we can earn from there by increasing tax on credit card.