

Target: Business case study

1. Import the dataset and do usual exploratory analysis step like checking the structure and characteristics of dataset:

1) Data type of all columns in "Customertable"

```
select column_name, data_type
from ecommerce.INFORMATION_SCHEMA.COLUMNS
where table_name="customers"
```

Row	column_name	data_type
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

2) Get the time range between which order were placed.

```
select min(order_purchase_timestamp) as min, max(order_purchase_timestamp) as max
from ecommerce.orders
```

Row	min	max
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

Insights: Time range between which order were placed (minimum time =2016-09-04 21:15:19 UTC) and maximum(2018-10-17 17:30:18 UTC)

3) Count the cities and states of customer who ordered during given period.

```
select count(distinct customer_city) as city, count(distinct customer_state) as state
from ecommerce.customers as c join ecommerce.orders as o
on c.customer_id=o.customer_id
```

Row	city	state
1	4119	27

Insights: In given period 2016 to 2018 there are 4119 cities and 27 state who ordered the item.

2) In- depth exploration

1. Is there a growing trend in the number of order placed over the past years?

```
select extract(year from order_purchase_timestamp) as year,
count(*) as no_of_order
from ecommerce.orders
group by year
order by year
```

Row	year	no_of_order
1	2016	329
2	2017	45101
3	2018	54011

Insights:yes there is growing trend in the number of order.In every year number of order increases so that its growing .

Recommendation:we have to follow the same strategy

2.can we see some kind of monthly seasonality in terms of the number of orders being placed.

```
select extract(year from order_purchase_timestamp) as year,
extract(month from order_purchase_timestamp) as month,
count(*) as no_of_orders
from ecommerce.orders
group by year,month
order by year,month
```

Row	year ▼	month ▼	no_of_orders ▼
4	2017	1	800
5	2017	2	1780
6	2017	3	2682
7	2017	4	2404
8	2017	5	3700
9	2017	6	3245
10	2017	7	4026
11	2017	8	4331
12	2017	9	4285
13	2017	10	4631
14	2017	11	7544

Insights: In every month there are no. of order increases and little bit decreases but in November month number of orders increases very fast because in Brazilian people ordered more product in this month. December month is full festive season. Christmas is main festival so that they ordered more product.

Recommendation: During festive season we kept the product in stock the and rest of month we provided some offer like buy 1 get 1 free, cashback, Wednesday in morning time give 20% discount on certain item.

3) During what time of the day, do the Brazilian customers mostly place their order?

```
select case
when extract(hour from o.order_purchase_timestamp)<=6 then "dawn"
when extract(hour from o.order_purchase_timestamp) between 7 and 12 then "morning"
when extract(hour from o.order_purchase_timestamp) between 13 and 18 then
"afternoon"
when extract(hour from o.order_purchase_timestamp) between 19 and 23 then "night"
end as hour,
count(o.order_id) as ordercount
from ecommerce.orders as o join ecommerce.customers as c
on o.customer_id=c.customer_id
group by hour
order by ordercount desc;
```

Row	hour ▼	ordercount ▼
1	afternoon	38135
2	night	28331
3	morning	27733
4	dawn	5242

Insights: In afternoon Brazilian people mostly ordered the item. In Dawn time people ordered less item.

Recommendation: In Dawn time people order very item so that we have to make strategy that during 0-6 hours there is 10-20% discount in product, buy 1 get 1 free, combo offer, cashback if order 5000 get cashback

3) Evolution of E-commerce orders in Brazil region:

1. Get the month to month no. of orders placed in each state.

```
select extract(year from order_purchase_timestamp) as year,
extract(month from order_purchase_timestamp) as month,
customer_state, count(*) as num_orders
from ecommerce.orders as o inner join ecommerce.customers as c
on o.customer_id=c.customer_id
group by year, month, customer_state
order by num_orders desc
```

Row	year ▼	month ▼	customer_state ▼	num_orders ▼
1	2018	8	SP	3253
2	2018	5	SP	3207
3	2018	4	SP	3059
4	2018	1	SP	3052
5	2018	3	SP	3037
6	2017	11	SP	3012
7	2018	7	SP	2777
8	2018	6	SP	2773
9	2018	2	SP	2703
10	2017	12	SP	2357
11	2017	10	SP	1793

Insights:each state month on month number of orders .In 2018 august month state SP has 3253 no. of order

2)How are the customers distributed across all the states.

```
select customer_state,count(customer_id)as no_of_customer
from ecommerce.customers
group by customer_state
order by no_of_customer desc
```

Row	customer_state	no_of_customer
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020

Insights:Highest number of customers 41746 are distributed in customerstate SP .

4) Impact on Economy: Analyze the money movement by e-commerce by looking at order prices,freight and others

1.Get the % increase in the cost of order from year 2017 to 2018 (include months between jan to aug only)

```
with costoforder2017 as(
select extract(year from o.order_purchase_timestamp)as year,
```

```

sum(p.payment_value)as total_cost
from ecommerce.payments as p join ecommerce.orders as o
on p.order_id=o.order_id
where extract(year from o.order_purchase_timestamp)=2017 and extract(month
from o.order_purchase_timestamp) between 1 and 8
group by year),

```

```

costoforder2018 as(
    select extract(year from o.order_purchase_timestamp)as
year,sum(p.payment_value)as total_cost
    from ecommerce.payments as p join ecommerce.orders as o
    on p.order_id=o.order_id
    where extract(year from o.order_purchase_timestamp)=2018 and
extract(month from o.order_purchase_timestamp) between 1 and 8
    group by year)
select (t.total_cost)total_cost2018,(p.total_cost)as total_cost2017,(t.total_cost-
p.total_cost)/p.total_cost*100 as prcentage_increase
from costoforder2018 as t,costoforder2017 as p

```

Row	total_cost2018	total_cost2017	prcentage_increase
1	8694733.839999...	3669022.120000...	136.9768716466...

Insights:There is 136% increase in the revenue from the period of jan 2017 to aug 2017ana jan2018-aug2018.

Recommendation:we observed there is tremendous increase in sales in 2018we need to follow same business strategy,customer services,promotional offersin upcoming year to improve the sales.

2.calculate the total and average value of order price for each state

```

select customer_state,ROUND(sum(price),2)as total_price,
ROUND(avg(price),2) as avg_price
from ecommerce.orders inner join ecommerce.order_items USING(order_id)
inner join ecommerce.customers USING(customer_id)
group by customer_state
order by total_price

```

Row	customer_state ▼	total_price ▼	avg_price ▼
1	RR	7829.43	150.57
2	AP	13474.3	164.32
3	AC	15982.95	173.73
4	AM	22356.84	135.5
5	RO	46140.64	165.97
6	TO	49621.74	157.53
7	SE	58920.85	153.04
8	AL	80314.81	180.89
9	RN	83034.98	156.97
10	PI	86914.08	160.36
11	PB	115268.08	191.48

Insights: There is total average value of order price for each state.

3. calculate the total and average value of order freight for each state.

```
select (c.customer_state) as state, round(sum(oi.freight_value), 2) as totalfreight, round(avg(oi.freight_value), 2) as avgfreight
from ecommerce.orders as o join ecommerce.customers as c
on o.customer_id=c.customer_id
join ecommerce.order_items as oi
on oi.order_id=o.order_id
group by state
order by totalfreight
```

low	state ▼	totalfreight ▼	avgfreight ▼
1	RR	2235.19	42.98
2	AP	2788.5	34.01
3	AC	3686.75	40.07
4	AM	5478.89	33.21
5	RO	11417.38	41.07
6	TO	11732.68	37.25
7	SE	14111.47	36.65
8	AL	15914.59	35.84

read more

Insights: Total price rate at which a product is delivered from one point to another in each state .and avg price rate at which a product is delivered from one point to another.

Recommendation: Don't increase the price rate too much because if increase the price rate then number of customer decreases.

5) Analysis based on sales, freight and delivery time

1. Find the no. of days taken to deliver each order from orders purchase date as delivery time also, calculate the difference (in days) between the estimated and actual delivery date of on order .

```
select order_id,
       date_diff(order_delivered_customer_date, order_purchase_timestamp, day) as
time_to_deliver,
       date_diff(order_delivered_customer_date, order_estimated_delivery_date, day) as
diff_estimated_delivery
from ecommerce.orders
where order_status="delivered"
```

Row	order_id	time_to_deliver	diff_estimated_delivery
1	635c894d068ac37e6e03dc54e...	30	-1
2	3b97562c3aee8bdedcb5c2e45...	32	0
3	68f47f50f04c4cb6774570cfde...	29	-1
4	276e9ec344d3bf029ff83a161c...	43	4
5	54e1a3c2b97fb0809da548a59...	40	4
6	fd04fa4105ee8045f6a0139ca5...	37	1
7	302bb8109d097a9fc6e9cefc5...	33	5
8	66057d37308e787052a32828...	38	6
9	19135c945c554eebfd7576c73...	36	2
10	4493e45e7ca1084efcd38ddeb...	34	0

Insights: There is no. days to deliver the each order from delivery time is actual date. means time taken by the vendor to deliver the product to the customer.

Estimated delivery - no. of days between the estimated delivery date and actual delivery date. difference between actual and estimated delivery is how much late delivered the product.

Recommendation: provide good customer service .delivered the product before the estimated delivery time.

2. Find out the top 5 states with highest and lowest average freight value

```
(select (c.customer_state) as state, avg(oi.freight_value) as
avgfreight,
from ecommerce.customers as c join ecommerce.orders as o
```



```

on c.customer_id=o.customer_id
join ecommerce.order_items as oi
on oi.order_id=o.order_id
group by state
order by avgfreight desc
limit 5)
union all

```

```

(select (c.customer_state)as state,avg(oi.freight_value)as
avgfreight,
from ecommerce.customers as c join ecommerce.orders as o
on c.customer_id=o.customer_id
join ecommerce.order_items as oi
on oi.order_id=o.order_id
group by state
order by avgfreight
limit 5)

```

Row	state	avgfreight
1	RR	42.98442307692...
2	PB	42.72380398671...
3	RO	41.06971223021...
4	AC	40.07336956521...
5	PI	39.14797047970...
6	SP	15.14727539041...
7	PR	20.53165156794...
8	MG	20.63016680630...
9	RJ	20.96092393168...
10	DF	21.04135494596...

Insights: Top 5 states with highest average freight value are: RR,PB,RO,AC,PI

Top 5 states with lowest average freight value are:SP,PR,MG,RJ,DF

3.Find out the top 5 states with highest and lowest average delivery time.

```

(select (c.customer_state)as
state,avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timest
amp,day))as avg
from
ecommerce.customers as c join ecommerce.orders as o
on c.customer_id=o.customer_id

```

```

group by state
order by avg desc
limit 5)

union all
(select (c.customer_state)as
state,avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timest
amp,day))as avg
from ecommerce.customers as c join ecommerce.orders as o
on c.customer_id=o.customer_id
group by state
order by avg
limit 5)

```

Row	state	avg
1	RR	28.97560975609...
2	AP	26.73134328358...
3	AM	25.98620689655...
4	AL	24.04030226700...
5	PA	23.31606765327...
6	SP	8.298061489072...
7	PR	11.52671135486...
8	MG	11.54381329810...
9	DF	12.50913461538...
10	SC	14.47956019171...

Insights: Top 5 states with highest average delivery time:RR,AP,AM,AL,PA

Top 5 states with lowest average delivery time:SP,PR,MG,DF,SC

Recommendation :reduce the delivery time

4.find out the 5 top state where the order delivery I really fast as compared to the estimated date of delivery.

```

select state,actualavg,estimatedavg,(actualavg-estimatedavg)as difference
from
(select (c.customer_state)as
state,avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day)
)as actualavg,
avg(date_diff(o.order_delivered_customer_date,o.order_estimated_delivery_date,day))
as estimatedavg
from ecommerce.orders as o join ecommerce.customers as c
on o.customer_id=c.customer_id
group by state
order by (actualavg-estimatedavg)
limit 5)

```

Row	state	actualavg	estimatedavg	difference
1	SP	8.298061489072...	-10.1353253488...	18.43338683788...
2	DF	12.50913461538...	-11.1187500000...	23.62788461538...
3	MG	11.54381329810...	-12.2969616908...	23.84077498899...
4	PR	11.52671135486...	-12.3642088157...	23.89092017062...
5	ES	15.33182957393...	-9.61854636591...	24.95037593984...

Insights:state SP has delivered fast as compared to other state .top 5 states which delivered fast SP,DF,MG,PR,ES

Recommendation:for increase the customer we have to decrease the delivery time.

6)Analysis based on payments:

1.Find the month on month no. of orders placed using different payment type

```
select extract(year from o.order_purchase_timestamp) as year,extract(month from
o.order_purchase_timestamp)as month,(p.payment_type)as paymenttype,
count(distinct o.order_id)as no_of_order
```

```
from ecommerce.orders as o join ecommerce.payments as p
on o.order_id=p.order_id
group by year,month,paymenttype
order by year,month
```

Row	year	month	paymenttype	no_of_order
1	2016	9	credit_card	3
2	2016	10	credit_card	253
3	2016	10	voucher	11
4	2016	10	debit_card	2
5	2016	10	UPI	63
6	2016	12	credit_card	1
7	2017	1	voucher	33
8	2017	1	UPI	197
9	2017	1	credit_card	582
10	2017	1	debit_card	9
11	2017	2	credit_card	1347
12	2017	2	voucher	69

Insights:month on month total number of orders using different payment type such as credit card,vocher,debit card ,UPI.in payment type no cash delivery available

Recommendation :no

2)find the no. of orders placed on the basis of the payment installments that have been paid.

```
select count(distinct order_id)as total_order,payment_installment
```

```
from ecommerce.payments as p
```

```
where p.payment_installments>=1
```

```
group by payment_installments
```

Row	total_order	payment_installment
1	49060	1
2	12389	2
3	10443	3
4	7088	4
5	5234	5
6	3916	6
7	1623	7
8	4253	8
9	644	9
10	5315	10
11	23	11

Insights: payment installments and the corresponding number of orders are essential.