TARGET BUSINESS ANALYSIS
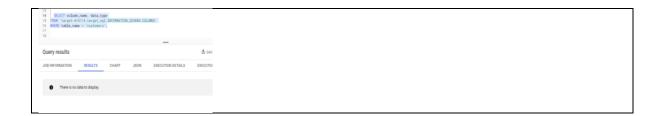
A. Data type of all columns in the "customers" table.

SELECT column_name, data_type

FROM `target-414114.target_sql.INFORMATION_SCHEMA.COLUMNS`

WHERE table_name = 'customers';

While running the above code i am not getting any result and this is prompting on the result window (There is no data to display.)

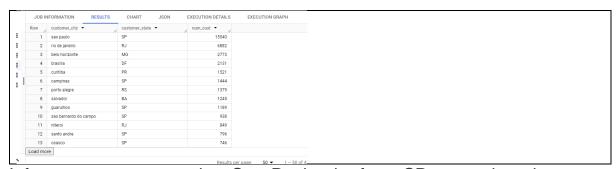## B. Get the time range between which the orders were placed.

select min(order_purchase_timestamp) as start_time,

max(order_purchase_timestamp) as end_time

from `target-414114.target_sql.orders `

Inference-The start and end dates of the data i.e. from 4th September 2016 to 17th October 2018. , providing a context for our analysis.

C. Count the Cities & States of customers who ordered during the given period.

Inference-we can see that Sao Paulo city from SP state alone has more orders than the following 5 cities combined

II. In-depth Exploration:

A. Is there a growing trend in the no. of orders placed over the past years?

```sql
select
extract(year from order_purchase_timestamp) as year ,
extract(month from order_purchase_timestamp) as month,
count(distinct o.order_id) as order_count
from `target-414114.target_sql.orders ` o
join `target-414114.target_sql.customers ` c
on o.customer_id=c.customer_id
group by year, month
order by year ,month
```

Inference-it can be observed that there is a growing trend in e-commerce in Brazil. The count of purchases has shown an overall upward trend, with some fluctuations but to gain accurate understanding we should consider revenue growth as well.

B. Can we see some kind of monthly seasonality in terms of the no. of orders being Placed?

Inference-Here we can see the similarity in the count of order being placed on monthly bases as it is showing a growing trend with some fluctuation comparing January order count with December it shows decline in order count and highest order count is in the month of august and the lowest order count is in the month of September.

C. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

● 0-6 hrs : Dawn

● 7-12 hrs : Mornings

● 13-18 hrs : Afternoon

● 19-23 hrs : Night

```
select
case when extract(hour from order_purchase_timestamp) between 0 and 6 then 'Dawn'
```

```sql
when extract(hour from order_purchase_timestamp) between 7 and 12 then 'Mornings'

when extract (hour from order_purchase_timestamp) between 13 and 18 then 'Afternoon'

when extract(hour from order_purchase_timestamp) between 19 and 23 then 'Night'

end as hour ,

count(distinct o.order_id) as order_count

from `target-414114.target_sql.orders ` o

join `target-414114.target_sql.customers ` c

on o.customer_id=c.customer_id

group by hour

order by order_count desc
```

JOB INFORMATION	RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH

Row	hour	order_count
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Inference-Based on the analysis the Brazilian customer tends to place the orders in the afternoon and night by analysis the buying pattern of the customer with the peak hour of buying the e-commerce company target can focus there add campaign on the peak hour of buying of products and maximise the order placed by   focusing and increasing the number of resources in the specific time .

III. Evolution of E-commerce orders in the Brazil region:

A. Get the month-on-month no. of orders placed in each state.

```sql
select c.customer_state,

extract(month from order_purchase_timestamp ) as month ,

count(o.order_id) as order_count

from `target-414114.target_sql.orders ` o

join `target-414114.target_sql.customers ` c

on o.customer_id=c.customer_id

group by 1,month

order by 1,order_count desc
```
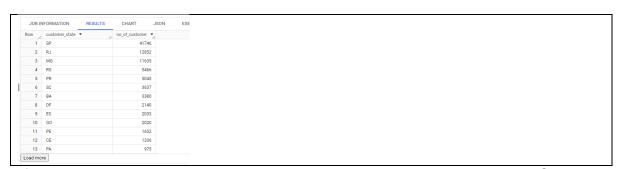
Row	customer_state	month	order_count
1	AC	5	10
2	AC	4	9
3	AC	7	9
4	AC	1	8
5	AC	8	7
6	AC	6	7
7	AC	10	6
8	AC	2	6
9	AC	11	5
10	AC	12	5
11	AC	9	5
12	AC	3	4
13	AL	4	51

Inference –Month 4 and 5 has the highest numbers of order count

B. How are the customers distributed across all the states?

select customer_state,

count(distinct customer_id) as no_of_customer

from `target-414114.target_sql.customers `

group by 1

order by 2 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
11	PE	1652
12	CE	1336
13	PA	975

Inference-By analysing the data we can clearly come to and conclusion that Sao Paulo has the most number of customer and PA is the state with least customer by this data Target as an e-commerce company can alter or change its marketing strategy and allocate maximum resources to the states where they are getting more customers wo maximise their profit and increase the revenue numbers and gain new customers too

IV. Impact on Economy: Analyse the money movement by e-commerce by looking at order prices, freight and others.

A. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
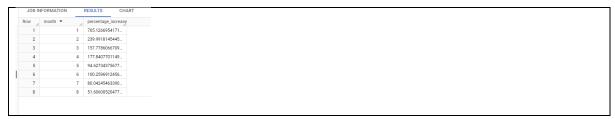
WITH orders_2017_2018 AS (

  SELECT

    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS year,

    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,

    p.payment_value

```
    FROM

      `target-414114.target_sql.orders ` o

    JOIN

      `target-414114.target_sql.payments` p

    ON

      o.order_id = p.order_id

    WHERE

      EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017, 2018) AND

      EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8

)

SELECT

    month,

    ((SUM(CASE WHEN year = 2018 THEN payment_value ELSE 0 END) - SUM(CASE WHEN year = 2017 THEN
    payment_value ELSE 0 END)) / NULLIF(SUM(CASE WHEN year = 2017 THEN payment_value ELSE 0 END), 0)) * 100
    AS percentage_increase

FROM

    orders_2017_2018

GROUP BY

    month

ORDER BY

    month;
```

Row	month ▾	percentage_increase
1	1	705.1266954171...
2	2	239.9918145445...
3	3	157.7786066709...
4	4	177.8407701149...
5	5	94.62734375677...
6	6	100.2596912456...
7	7	80.04245463390...
8	8	51.60600520477...

Inference-The overall percentage increase in the cost of orders from 2017 to 2018, including only the months from January to August, is 138.53%. Upon examining the month-wise increase, January shows the highest percentage increase, followed by February and Apri

B. Calculate the Total & Average value of order price for each state.

```
select customer_state ,

       sum(price + freight_value) as total ,

       avg(price + freight_value) as average
```

```sql
from `target-414114.target_sql.customers ` c

join `target-414114.target_sql.orders ` o

  on c.customer_id = o.customer_id

join `target-414114.target_sql.order_items` oi

  on o.order_id = oi.order_id

group by customer_state
```


Row	customer_state	total	average
1	RN	101895.0800000...	192.6182986767...
2	CE	275606.3000000...	186.4724627875...
3	RS	885826.7599999...	142.0732574178...
4	SC	610213.5999999...	146.1239463601...
5	SP	5921678.119999...	124.8009045501...
6	MG	1856161.490000...	141.3787409551...
7	BA	611506.6700000...	160.9651671492...
8	RJ	2129681.980000...	146.0787420262...
9	GO	347706.9300000...	149.0385469352...
10	MA	151171.9900000...	183.4611529126...
11	PE	322237.6899999...	178.4261849390...
12	PB	140987.8099999...	234.1990199335...
13	ES	324801.9100000...	143.9724778368...

Inference-PB has the highest average order price value

C. Calculate the Total & Average value of order freight for each state.

```sql
select c.customer_state, sum (ot.freight_value) as total_frieght_value,

avg(ot.freight_value) as avg_frieght_value

from `target-414114.target_sql.customers ` c

join `target-414114.target_sql.orders ` o

on c.customer_id=o.customer_id

join `target-414114.target_sql.order_items` ot

on o.order_id=ot.order_id

group by 1
```


Row	customer_state	total_frieght_value	avg_frieght_value
1	RN	18860.09999999...	35.65236294896...
2	CE	48351.58999999...	32.71420162381...
3	RS	135522.7400000...	21.73580433039...
4	SC	89660.26000000...	21.47036877394...
5	SP	718723.0699999...	15.14727539041...
6	MG	270853.4600000...	20.63016680630...
7	BA	100156.6799999...	26.36395893656...
8	RJ	305589.3100000...	20.96092393168...
9	GO	53114.97999999...	22.76681525932...
10	MA	31523.77000000...	38.25700242718...
11	PE	59449.65999999...	32.91786267995...
12	PB	25719.72999999...	42.72380398671...
13	ES	49764.59999999...	22.05877659574...

Inferences- PB has the highest average freight value and SP has the highest total fieght value

V. Analysis based on sales, freight and delivery time.

A. Find the no. of days taken to deliver each order from the order's purchase date

as delivery time.

Also, calculate the difference (in days) between the estimated & actual delivery date of an order.

Do this in a single query.

```sql
SELECT order_id
 DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, day) AS delivery_time,
    DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date, day) AS diff_estimated_delivery
FROM `target-414114.target_sql.orders `
WHERE order_status = 'delivered';
```


Row	order_id	delivery_time	diff_estimated_deliv
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
11	70c77e51e0f179d75a64a6141...	42	11
12	d7918e406132d7c81f1b84527...	35	3
13	43f6604e77ce6433e7d68dd86...	32	7

B. Find out the top 5 states with the highest & lowest average freight value.

```sql
(select c.customer_state,
avg(ot.freight_value) as avg_freight_value ,'lowest' as type
from `target-414114.target_sql.customers `c
join `target-414114.target_sql.orders ` o
on c.customer_id=o.customer_id
join `target-414114.target_sql.order_items` ot
on o.order_id=ot.order_id
group by 1
order by 2
limit 5)
union all
(select c.customer_state,
avg(ot.freight_value) as avg_freight_value ,'highest' as type
from `target-414114.target_sql.customers `c
join `target-414114.target_sql.orders ` o
on c.customer_id=o.customer_id
join `target-414114.target_sql.order_items` ot
```
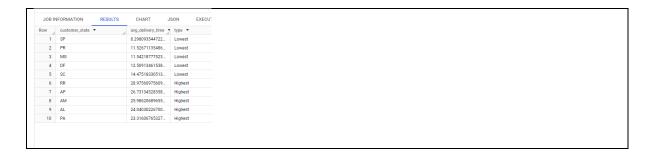
on o.order_id=ot.order_id

group by 1

order by 2 desc

limit 5)

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

C. Find out the top 5 states with the highest & lowest average delivery time.

(

SELECT

  customer_state,

  AVG(delivery_time) AS avg_delivery_time,

  'Lowest' AS type

FROM

(

SELECT

  customer_state,

  DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, day) AS delivery_time

FROM

  `target-414114.target_sql.orders ` o

JOIN

  `target-414114.target_sql.customers ` c

ON

  o.customer_id = c.customer_id

WHERE

  order_status = 'delivered'

  AND order_delivered_customer_date IS NOT NULL

) AS delivery_times

GROUP BY

  customer_state

```sql
    ORDER BY
      avg_delivery_time ASC
    LIMIT 5
)


UNION ALL


(
  SELECT
    customer_state,
    AVG(delivery_time) AS avg_delivery_time,
    'Highest' AS type
  FROM
  (
    SELECT
      customer_state,
      DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, day) AS delivery_time
    FROM
      `target-414114.target_sql.orders ` o
    JOIN
      `target-414114.target_sql.customers ` c
    ON
      o.customer_id = c.customer_id
    WHERE
      order_status = 'delivered'
      AND order_delivered_customer_date IS NOT NULL
  ) AS delivery_times
  GROUP BY
    customer_state
  ORDER BY
    avg_delivery_time DESC
  LIMIT 5
)
```

Row	customer_state	avg_delivery_time	type
1	SP	8.298093544722...	Lowest
2	PR	11.52671135486...	Lowest
3	MG	11.54218777523...	Lowest
4	DF	12.50913461538...	Lowest
5	SC	14.47518330513...	Lowest
6	RR	28.97560975609...	Highest
7	AP	26.73134328358...	Highest
8	AM	25.98620689655...	Highest
9	AL	24.04030226700...	Highest
10	PA	23.31606765327...	Highest

D. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

select customer_state ,

    avg(diff_estimated_delivery) as total

from

(select order_id , customer_state ,

date_diff(order_delivered_customer_date,order_estimated_delivery_date,

day) as diff_estimated_delivery

from `target-414114.target_sql.orders ` o

join `target-414114.target_sql.customers ` c

   on o.customer_id = c.customer_id

where order_status = 'delivered' and

    order_delivered_customer_date is not null)

group by customer_state

order by total

limit 5

Row	customer_state	total
1	AC	-19.7625000000...
2	RO	-19.1316872427...
3	AP	-18.7313432835...
4	AM	-18.6068965517...
5	RR	-16.4146341463...

VI. Analysis based on the payments:

A. Find the month on month no. of orders placed using different payment types.

select

extract(month from order_purchase_timestamp) as order_month,

extract(year from order_purchase_timestamp) as order_year,

p.payment_type,count(*) as order_count

from `target-414114.target_sql.orders ` o

join `target-414114.target_sql.payments` p

on o.order_id=p.order_id

group by order_month,order_year,p.payment_type

order by order_year,order_month,p.payment_type

Row	order_month	order_year	payment_type	order_count
1	9	2016	credit_card	3
2	10	2016	UPI	63
3	10	2016	credit_card	254
4	10	2016	debit_card	2
5	10	2016	voucher	23
6	12	2016	credit_card	1
7	1	2017	UPI	197
8	1	2017	credit_card	583
9	1	2017	debit_card	9
10	1	2017	voucher	61
11	2	2017	UPI	398
12	2	2017	credit_card	1356
13	2	2017	debit_card	13

Load more

Results per page:

B. Find the no. of orders placed on the basis of the payment installments that have been paid.

SELECT

  p.payment_installments,

  COUNT(o.order_id) AS order_count

FROM

  `target-414114.target_sql.orders ` o

JOIN

  `target-414114.target_sql.payments` p

ON

  o.order_id = p.order_id

WHERE

  o.order_status != 'canceled'

GROUP BY

  1

ORDER BY

  2 DESC;

Actionable insights and Recommendations

There are following conclusions I would like to draw

As we can see that the count of order of Sao Paulo city from SP state is the highest order count of 15540 and so on. Here we can further analyse what all things people are looking forward to buy and what are the items which are less preferred. There are few things which we can understand about our customers like What customers are we targeting upon or what is the gender ratio or what is the best products amongst men or women in a particular category? We can recommend similar products so as to increase the traffic for that particular item. We can further give more discounts or some sales week off day so that customers can engage more into the online platform. The overall attempt here is to understand more about our customers so as to recommend them similar items based on their interests

However the least order placed is only count 1 for cities like MT and states like vila bela da santissima Trindade. There are many more states with such a less count of orders. We need to understand the reasons behind such a small count of orders. With this we should have some analysis about the reasons for such a low business. What all things can be done so as to improve business in those places like understand the actual market, Understand the gender ratio, Understand the competition with other retail stores. We should also have often likes and dislikes online surveys. With these we can understand our customers better and we can recommend them the

products they actual need. There also might be a possibility that these

people more often prefer to shop offline but given a good discounted price +

minimal transport cost can be a very good way to develop a healthy customers

relationships. We can also provide 50% off sale coupon for the first time

login users. These are some of the good ways we can implement to build a

good base to build more customers

After this we did observed a growing trend in number of orders which is

itself a positive things to look upon

However we have seen that the orders kept on increasing from 3rd quarter of

2016 to 1st quarter of 2018. But suddenly there has been a subsequent drop in

number of orders with quarter 4 of 2018 being the lowest. We can further

work on these factors so as to understand what is the reason behind a drop

in sales?

Then we have seen that the customers usually have a purchase in the afternoon time
and the night times. But during the dawn period the count is really low. The highest
customer count is from SP which is around 5620430 and lowest

around and lowest customer count is from RR around 2087

We have further seen that there is a subsequent growth in terms of percent

change of around 136.98 %. This is really positive in terms of business

growth

We have seen a subsequent total price of 43788 and average of 267 in

geolocation state of AC and total price of 1141328046.14 and average of

105.97

However the price of freight is SP state is around 198571257.84 and lowest

been around 5385.76

We can further try to reduce the delivery time so that the customers can

receive their orders really fast.

As per the analysis the top 5 states with highest averages include

CE,RO,PI,PB, AC and the top 5 states with lowest averages include RN, SP,

RJ, DF, PR

Further we have also seen that the delivery is much faster then the estimated delivery in some states like ES, RS, PR, MA, AL,SC, etc. This is really a positive thing so as to deliver the products as fast as possible so that our customers can further place a new order

We can also refer to different payments methods like credit_card, debit_card, Vouchers, UPI.

Finally we have seen that there are only two orders who have completed there payment   installation