SQL – PROJECT

BUSINESS – CASE STUDY

Done & Submitted by,

KASI MUTHUVEERAPPAN

target

Table of Contents

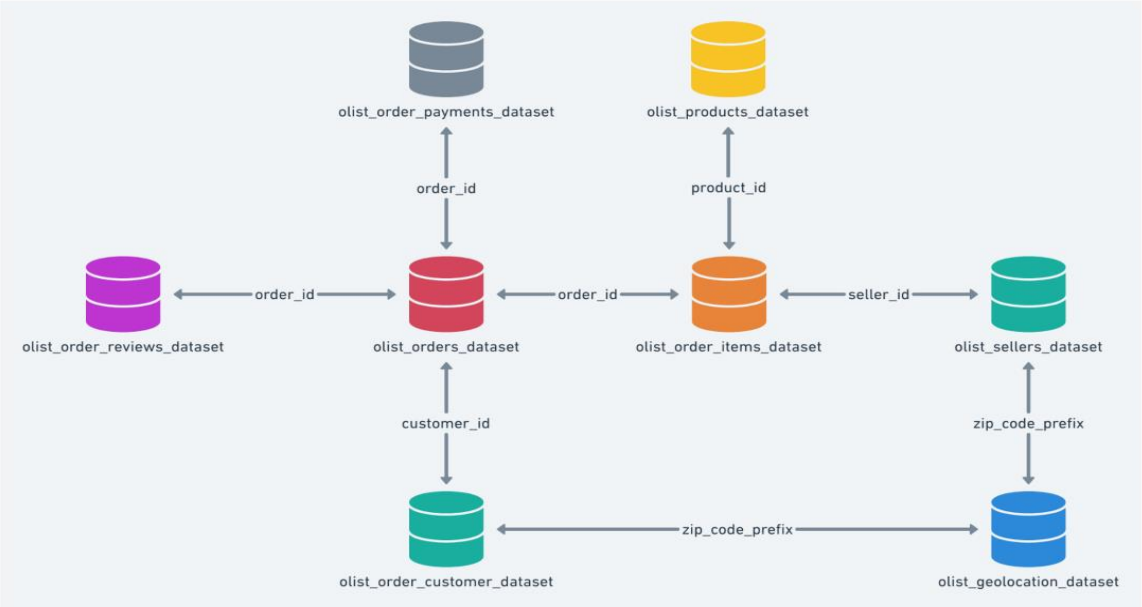
TARGET – BUSINESS CASE STUDY

Company Description

- ✓ Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation, and an exceptional guest experience that no other retailer can deliver.
- ✓ This business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between Sep-2016 and Aug-2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.
- ✓ By analysing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

Dataset:

Dataset schema:



The dataset consist of 8 tables such as ,

Customers Table, Geolocation Table, Order_items Table, Payments Table ,

Reviews Table, Orders Table, Products Table, Sellers Table …

Questions :-

1. Exploratory analysis:

a) Data type of all columns in the “customers” table.

QUERY –

```
select column_name,data_type
from `target`.INFORMATION_SCHEMA.COLUMNS
where table_name= 'customers'
```

OUTPUT –

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH

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

INSIGHTS –

- We have queried to fetch specific information like the datatype of each column present in the 'customers' table. This helps for ensuring data integrity and can be helpful for understanding the table's structure, designing queries, or performing data analysis
- “Customer ID” is typically a unique identifier assigned to each individual customer within a company's database system. It helps the company keep track of customer information, purchase history and other relevant data and this Id may be used for the tracking of order shipment and other requirements when the credentials of the customers are necessary.
- “Customer unique ID" refers to a more specific and individualized identifier assigned to a customer. This unique ID distinguishes one customer from the other in a more granular way. It may be used internally within Target's infrastructure for operational purposes, data analysis, or as a reference in their databases.

---

b) Get the time range between which the orders were placed

QUERY –

```
select
min(order_purchase_timestamp) as mintime,
max(order_purchase_timestamp) as maxtime
from `target.orders`;
```

OUTPUT –

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH

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

INSIGHTS –

- We have queried to find the date and time of the First ever order placed and last order placed. Time period ranges approximately 2 Years for the given data.

- The time range is calculated to oversee the sales performance to identify trends, peak periods, optimize seasonal management, marketing strategies, promotions, analysing customer patterns, purchasing behaviour, order fulfilment rates, feedbacks, sales growth, customer reviews or customer support response times, annual events, can help identify seasonal trends and preferences. This information can guide and help in predicting future demand to improve overall customer satisfaction and to achieve operational Efficiency.

---

c) Count the number of Cities and States in our dataset

QUERY –

```
select
count(distinct geolocation_city) as NoOfCities,
count(distinct geolocation_state) as NoOfStates
from `target.geolocation`;
```

OUTPUT –

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	NoOfCities	NoOfStates
1	8011	27

Another possibility:

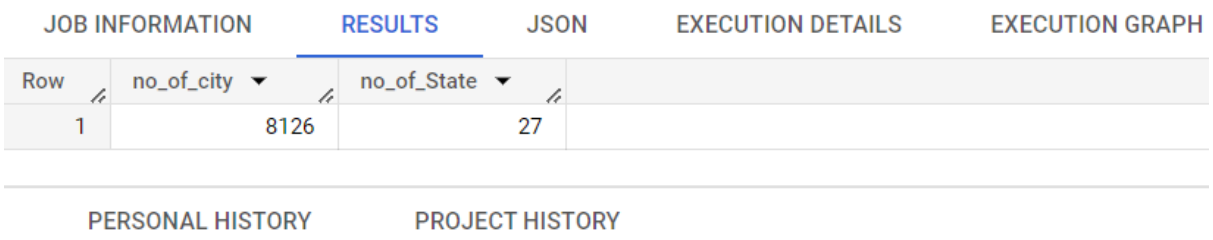
Since asked from the whole dataset :-

QUERY –

```
with cte as
(
select
geolocation_city as city,
geolocation_state as state
from `target.geolocation`
group by geolocation_city,geolocation_state

union all

select
customer_city as city,
customer_state as state
from `target.customers`

union all

select
seller_city as city,
seller_state as state
from `target.sellers`
)

select
```

```
count(distinct city) as no_of_city,
count(distinct state) as no_of_State
from cte;
```

OUTPUT –



JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	no_of_city	no_of_State
1	8126	27

PERSONAL HISTORY | PROJECT HISTORY

INSIGHTS –

- Target has established a significant presence across various regions in Brazil, indicating a robust market position and a widespread customer base spanning the entire country. This highlights their extensive reach and strong foothold in the Brazilian market.
- We can infer that the count of unique no.of.cities and unique no.of.states in the geolocation has missed around 115 cities that are mentioned in union of whole dataset.
- The count of cities and states offers valuable information for conducting geographic analysis, identifying potential opportunities, customers based on locations and their behavioural pattern, evaluating regional performance, and facilitating decision-making processes to understand market potential and make informed decisions regarding resource allocation and strategic planning.

---

2. <u>In-depth Exploration:</u>

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

QUERY –

Month wise

```
with cte as
(
select
concat(year,'-',mnth)as period,
sum(no_of_orders)as total_orders,
lag(sum(no_of_orders)) over(order by year,mnth)as lagg
from
(
select
extract(year from order_purchase_timestamp)as year,
extract(month from order_purchase_timestamp)as mnth,
count(*)as no_of_orders
from `target.orders`
where order_status not in ('canceled','unavailable')
and extract(year from order_purchase_timestamp) >= 2017
group by order_purchase_timestamp,year,mnth
order by year,mnth
)
group by year,mnth
order by year,mnth
)

select
period,
total_orders,
Growth_trend
from
(
select * ,
concat(round(((total_orders-lagg)/lagg)*100,2),'  %') as Growth_trend
from cte
);
```

Year wise:-

- Data is insufficient to calculate based on year , missing of 2016 november data , also data for 2017 and 2018 are partially available to do a comparative analysis.

```
with cte as
(
(
  select extract(year from order_purchase_timestamp) as year,
  count(*) as no_of_orders
  from `target.orders`
```

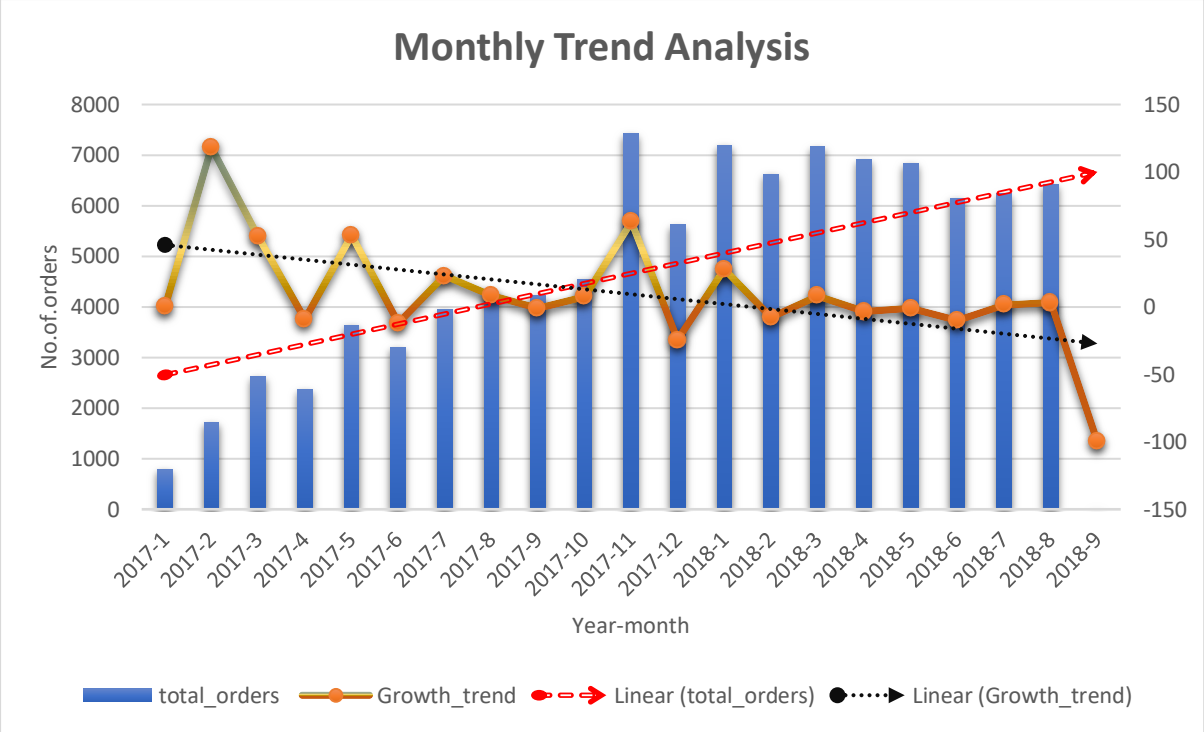
```
  where order_status <>'canceled'
  group by 1
  order by 1
)
)

select *,
concat(round(((No_of_Orders-lagg)/lagg)*100,2),'  %') as growth_trend
from
(
  select *,
  lag(no_of_orders) over(order by year)as lagg
  from cte
  order by 1
);
```

OUTPUT –

Row	period	total_orders	Growth_trend
2	2017-2	1718	118.3 %
3	2017-3	2617	52.33 %
4	2017-4	2377	-9.17 %
5	2017-5	3640	53.13 %
6	2017-6	3205	-11.95 %
7	2017-7	3946	23.12 %
8	2017-8	4272	8.26 %
9	2017-9	4227	-1.05 %
10	2017-10	4547	7.57 %
11	2017-11	7423	63.25 %
12	2017-12	5620	-24.29 %

Row	year	no_of_orders	lagg	growth_trend
1	2016	303	*null*	*null*
2	2017	44836	303	14697.36 %
3	2018	53677	44836	19.72 %

ANALYSIS –



INSIGHTS –

- By query analysis, the number of orders placed year-wise and month-wise excluding cancelled and unavailable orders, and considering the "Growth_trend" column indicating order growth rate compared to the previous year, it is evident that there was substantial order growth between 2016 and 2017, followed by a decrease in growth rate in 2018. This analysis helps identify patterns and trends in order placements over time.

- Overall, it demonstrates a consistent and significant upward trend in the number of orders in terms of orders placed whereas based on the revenue, there is a leap trend over the Brazil region, indicating positive business performance and increasing customer demand for Target's products in Brazil but must work on the revenue based to gain an uptrend is what we can from the chart analysis.

---

b) Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

QUERY –

```
select
concat(period,' -- is ',mth)as monthly_season,
Max_orders
from
(
select
concat(year,' - ',mnth)as period,
sum(no_of_orders)as Max_orders,
case
when mnth=1 then 'January'
when mnth=2 then 'Feburary'
when mnth=3 then 'March'
when mnth=4 then 'April'
when mnth=5 then 'May'
when mnth=6 then 'June'
when mnth=7 then 'July'
when mnth=8 then 'August'
when mnth=9 then 'September'
when mnth=10 then 'October'
when mnth=11 then 'November'
else 'December'
end as mth
from
(
select
extract(year from order_purchase_timestamp)as year,
extract(month from order_purchase_timestamp)as mnth,
count(*) as no_of_orders,
from `target.orders`
where order_status <>'canceled'
group by order_purchase_timestamp,year,mnth
order by year,mnth
)
group by year,mnth
)
order by  Max_orders desc;
```

OUTPUT –

Row	monthly_season	Max_orders
1	2017 - 11 -- is November	7507
2	2018 - 1 -- is January	7235
3	2018 - 3 -- is March	7185
4	2018 - 4 -- is April	6924
5	2018 - 5 -- is May	6849
6	2018 - 2 -- is Feburary	6655
7	2018 - 8 -- is August	6428
8	2018 - 7 -- is July	6251
9	2018 - 6 -- is June	6149
10	2017 - 12 -- is December	5662
11	2017 - 10 -- is October	4605

ANALYSIS-

INSIGHTS –

- The objective of the query was to calculate the total count orders for each month over the years to get an understanding and managing the business's order patterns.
- There is a clear monthly seasonal pattern in the number of orders, with increases and decreases throughout the year, perhaps influenced by holidays or specific shopping periods.
- We can infer that Peak is at November 2017 shows the highest number of orders, while September 2016 has the lowest number of orders. we can simply understand the uptrend in the orders place using chart analysis.
- Comparing with the world events and Brazilian culture, Having welcome the new year is great even , on the occasion they materialize the home appliances and needs, bath tubs, beds and hence the order were peaked during year end & November2017 is the month of **Black Friday** & they are more inclined towards soccer game and **FIFA world cup 2018** happened , during that time orders peaked were placed. Therefore, we see the Monthly seasonality hotspots.
- Comparing the number of orders across different years reveals growth and trough in number of orders. There is significant growth from 2016 to 2017, but a decrease in average orders in 2018 compared to the previous year.

---

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

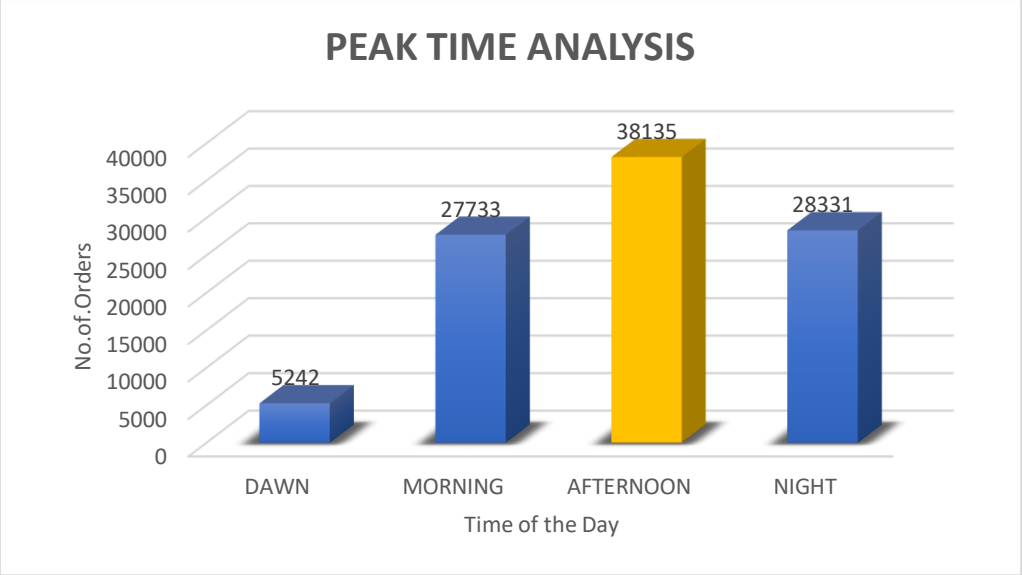
QUERY –

```
with cte as
(
select
case
when extract(hour from order_purchase_timestamp)
between 0 and 6 then 'DAWN'
when extract(hour from order_purchase_timestamp)
between 7 and 12 then 'MORNING'
when extract(hour from order_purchase_timestamp)
between 13 and 18 then 'AFTERNOON'
else 'NIGHT'
end as TIME_OF_DAY,
count(*)as Orders_placed
from `target.orders`
group by TIME_OF_DAY
)

select * from cte
order by 2 desc;
```

OUTPUT –

Row	TIME_OF_DAY	Orders_placed
1	AFTERNOON	38135
2	NIGHT	28331
3	MORNING	27733
4	DAWN	5242

ANALYSIS-



INSIGHTS –

- We have queried to find the apt time as when the most of the orders are being placed and we have inferred the same using chart analysis.
- Based on the output, Brazilian customers tend to place the highest number of orders during the afternoon hours, followed by the night hours.
- Morning hours also show a substantial volume of orders, while dawn hours have fewer orders placed.
- This suggests that customers in Brazil are actively engaged in online shopping during the afternoon and evening, possibly during leisure time, with a significant number of orders placed in the morning as well.

---

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

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

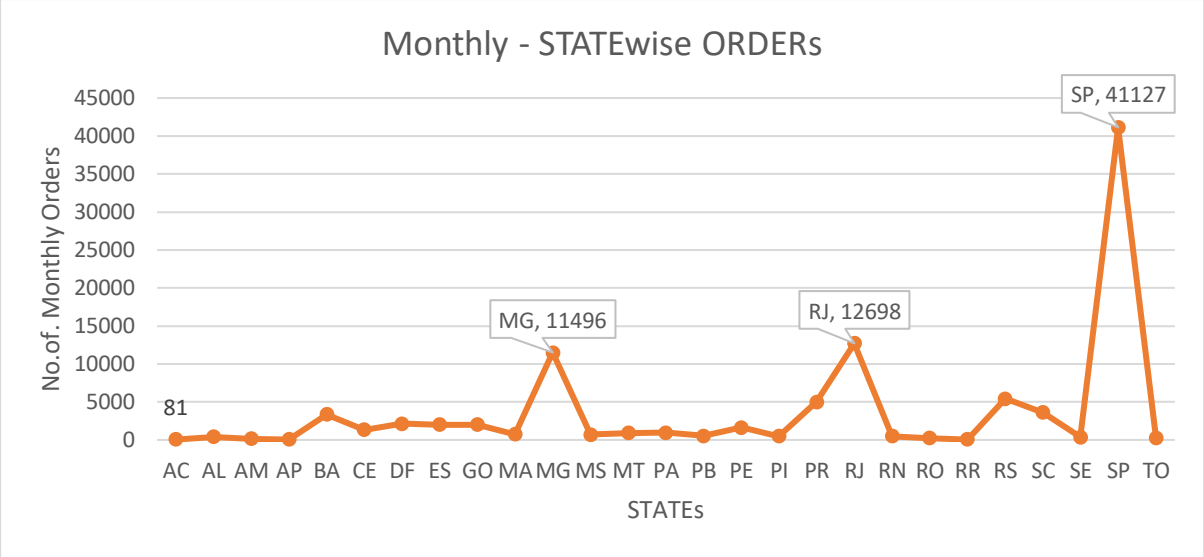
QUERY –

```
select yr_mnth,state,
min(monthly_orders) over(partition by yr_mnth)as Min_orders_recd,
max(monthly_orders) over(partition by yr_mnth) as Max_orders_recd,
round(avg(monthly_orders) over(partition by state order by yr_mnth),3)as Avg_monthly_order,
monthly_orders as Total_ordersPerMonth,
sum(monthly_orders) over(partition by state)as Monthly_state_order
from
(
select customer_state as state,
format_timestamp("%Y-%m",order_purchase_timestamp) as yr_mnth,
count(*)as monthly_orders
from `target.customers` c
join `target.orders` o
on c.customer_id=o.customer_id
where order_status not in ('canceled','unavailable')
group by customer_state,yr_mnth
order by customer_state,yr_mnth
)
group by monthly_orders,yr_mnth,State
order by state ;
```

OUTPUT –

Row	yr_mnth	state	Min_orders_recd	Max_orders_recd	Avg_monthly_order	Total_ordersPerMont	Monthly_state_order
1	2017-01	AC	1	294	2.0	2	81
2	2017-02	AC	2	630	2.5	3	81
3	2017-03	AC	2	991	2.333	2	81
4	2017-04	AC	2	895	3.0	5	81
5	2017-05	AC	2	1396	4.0	8	81
6	2017-06	AC	1	1308	4.0	4	81
7	2017-07	AC	1	1570	4.143	5	81
8	2017-08	AC	3	1698	4.125	4	81
9	2017-09	AC	1	1605	4.222	5	81
10	2017-10	AC	3	1754	4.4	6	81
11	2017-11	AC	2	2949	4.455	5	81

ANALYSIS-



INSIGHTS –

- The query states the number of orders placed, minimum and maximum order values, and state wise running average of orders placed. On a broader perspective, comparing order counts with minimum and maximum values helps identify states that consistently Peaks or troughs. This information allows prioritization of geolocations and targeted strategies to enhance orders in specific states.

- The state-wise average order counts indicate overall order count trends, highlighting states with steady or increasing order placed. This insight identifies growth opportunities and effective marketing strategies in different regions

- In summary _ by analyzing this data, this enables data-driven decision-making, aiding in production planning, considering seasonality, state-wise performance, and growth trends, to optimize marketing, operations, and resource allocation strategies across different states throughout the Brazil country.

---

b) How are the customers distributed across all the states?

QUERY –

```
select
customer_state,
count(customer_unique_id) AS No_of_unique_customers
from `target.customers`
group by customer_state
order by customer_state;
```

Another Approach:

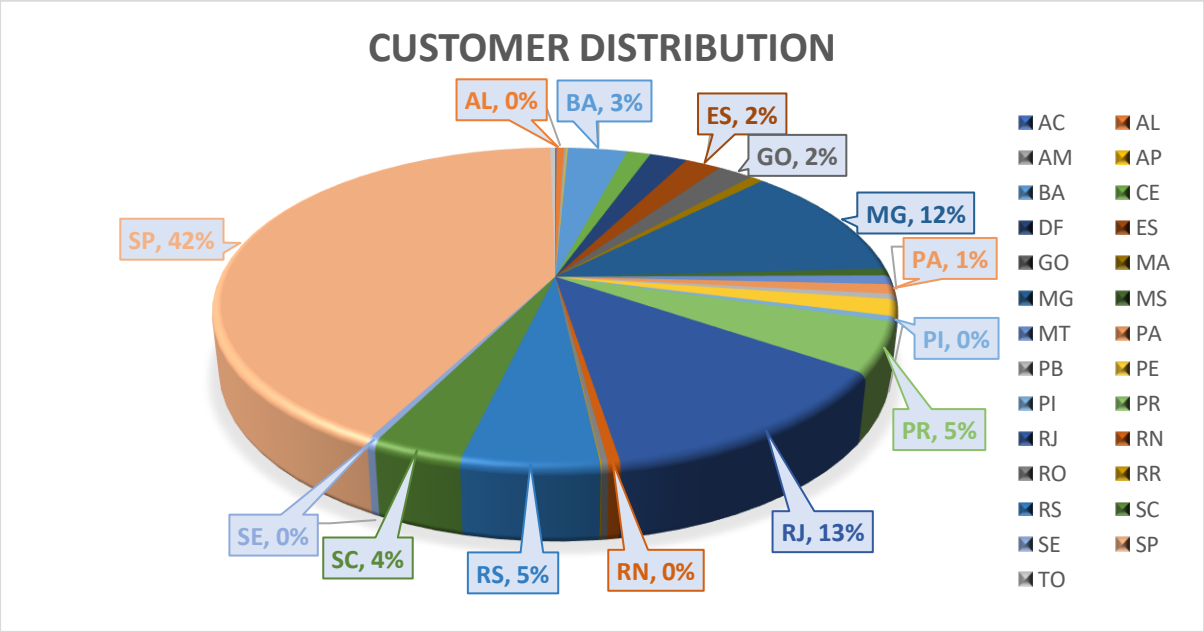
```
select *,
(no_of_customers-no_of_unique_customers)as customers_purchased_more
from
(
select customer_state,
count(distinct customer_id)as no_of_customers,
count(distinct customer_unique_id)as no_of_unique_customers
from `target.customers`
group by customer_state
order by 1
)as nt;
```

OUTPUT –

Row	customer_state	No_of_unique_customers
1	AC	81
2	AL	413
3	AM	148
4	AP	68
5	BA	3380
6	CE	1336
7	DF	2140
8	ES	2033
9	GO	2020
10	MA	747
11	MG	11635

Row	customer_state	no_of_customers	no_of_unique_customers	customers_purchased_more
1	AC	81	77	4
2	AL	413	401	12
3	AM	148	143	5
4	AP	68	67	1
5	BA	3380	3277	103
6	CE	1336	1313	23
7	DF	2140	2075	65
8	ES	2033	1964	69
9	GO	2020	1952	68
10	MA	747	726	21
11	MG	11635	11259	376

ANALYSIS -



INSIGHTS –

- The query captures the total number of customers, number of customers whom made more than 1 purchase and total orders placed state wise helps us to do the Geographical Analysis.

- By examining customer distribution across states, regional trends and patterns can be identified. States like Sao Paulo (SP) and Rio de Janeiro (RJ) with a high count of customers indicate potential market hotspots. State

with lower customer counts such as Roraima (RR) and Tocantins (TO) with only 45 customers, present potential hotspots for market opportunities for expansion and growth opportunities. States with a higher number of regular customers suggests reflects higher customer likability's or satisfaction.

- This information helps in decision making of where to concentrate marketing campaigns, expansion into new regions, improvement strategies for retaining customers, and identification of areas for growth and investment.

---

4. 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).

QUERY –

Monthwise :-

```
with cte as (
select month,
round(sum(if(year=2018,amt,0)),2) as cost_2018,
round(sum(if(year=2017,amt,0)),2) as cost_2017
from
(
select
extract(year from order_purchase_timestamp)as year,
extract(month from order_purchase_timestamp)as mn,
format_timestamp("%b",order_purchase_timestamp) as month,
sum(p.payment_value)as amt
from `target.payments` p
join `target.orders` o
on p.order_id=o.order_id
where order_status not in ('canceled','unavailable')
group by mn,order_purchase_timestamp
order by mn
)as n
where n.mn < 9  and year <> 2016
group by mn,month
order by mn
```

```
)

select month,
concat(cost_2017,'  -  REAIs')as Amt_2017,
concat(cost_2018,'  -  REAIs')as Amt_2018,
concat(round((((cost_2018-cost_2017)/cost_2017)*100),2),'  %') as percent_increased
from cte;
```

Yearwise :-

```
with cte as (
select
round(sum(if(year=2018,amt,0)),2) as cost_2018,
round(sum(if(year=2017,amt,0)),2) as cost_2017
from
(
select
extract(year from order_purchase_timestamp)as year,
extract(month from order_purchase_timestamp)as mn,
sum(p.payment_value)as amt
from `target.payments` p
join `target.orders` o
on p.order_id=o.order_id
group by order_purchase_timestamp
order by year
)as n
where n.mn < 9  and year <> 2016
)

select
concat(cost_2017,'  -  REAIs')as Amt_2017,
concat(cost_2018,'  -  REAIs')as Amt_2018,
concat(round((((cost_2018-cost_2017)/cost_2017)*100),2),'  %') as percent_increased
from cte;
```

OUTPUT –

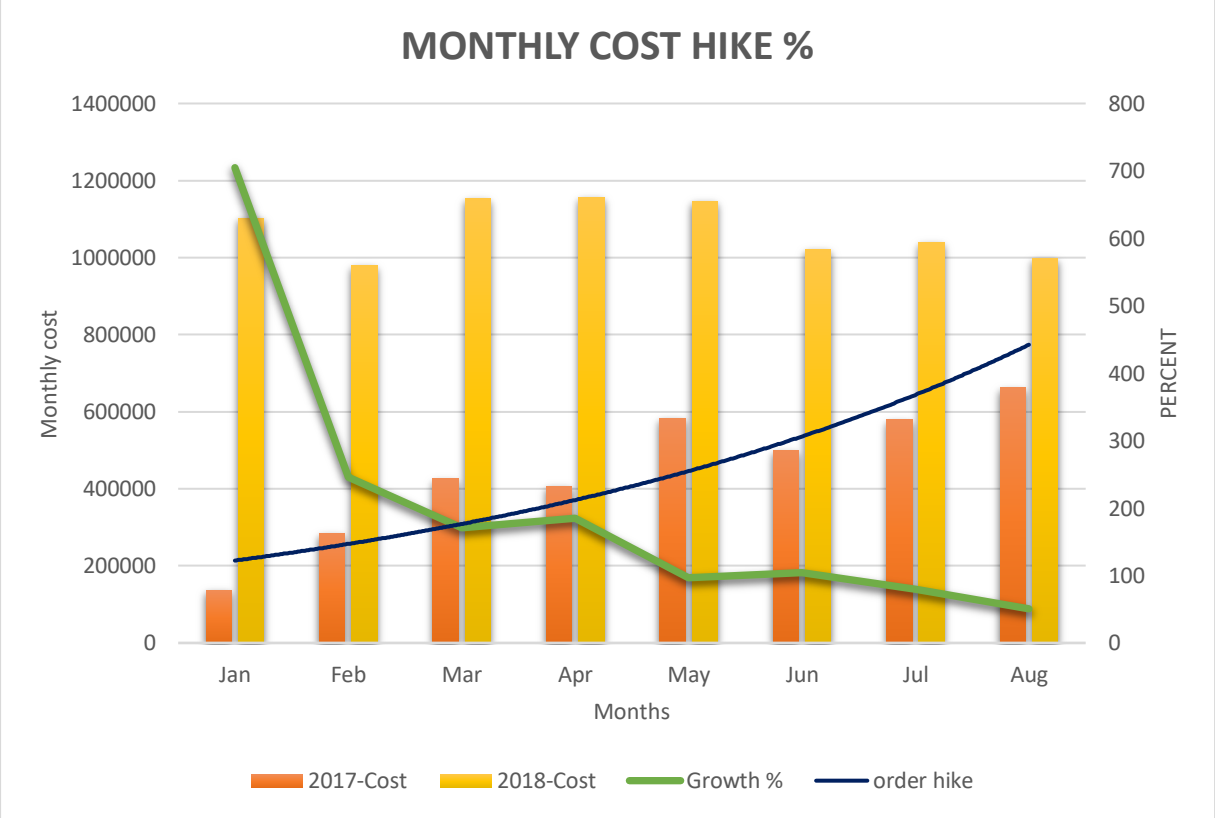
Yearwise:-

Row	Amt_2017	Amt_2018	percent_increased
1	3669022.12 - REAIs	8694733.84 - REAIs	136.98 %

Monthwise:-

Row	month	Amt_2017	Amt_2018	percent_increased
1	Jan	137006.76 - REAIs	1102639.41 - REAIs	704.81 %
2	Feb	283621.94 - REAIs	979966.23 - REAIs	245.52 %
3	Mar	425656.4 - REAIs	1152736.74 - REAIs	170.81 %
4	Apr	405988.38 - REAIs	1156303.91 - REAIs	184.81 %
5	May	582926.16 - REAIs	1145748.63 - REAIs	96.55 %
6	Jun	499827.47 - REAIs	1020494.29 - REAIs	104.17 %
7	Jul	578858.58 - REAIs	1039880.16 - REAIs	79.64 %
8	Aug	662071.77 - REAIs	996896.15 - REAIs	50.57 %

ANALYSIS –



INSIGHTS –

- The query shows the percentage increase in the total cost month wise and year wise which signifies the increase in order costs of products suggests a pricing adjustment or changes in product offerings, influenced by factors like procurement, raw material costs, and demand-supply dynamics.

- The data shows a positive trend of increasing order amounts from 2017 to 2018, a substantial decrease in cost of the products makes us understand the economic growth of the country indicating potential growth and success in the business across Brazilian states.
- Understanding these market dynamics is to be considered to make the change and replicate successful strategies in the future. These insights help understand seasonal patterns and growth trends, supporting strategic planning and decision-making.

---

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

QUERY –

-using payment_value column from payments table

```
select customer_state ,
concat(round(sum(payment_value),2),'  REAIs') as Total_amount,
concat(round(avg(payment_value),2),'  REAIs') as Avg_amt
from `target.customers` c
join `target.orders` o
on c.customer_id=o.customer_id
join`target.payments` p
on p.order_id = o.order_id
group by customer_state
order by customer_state;
```

Another Approach:

-using order_items price column

```
select customer_state ,
concat(round(sum(oi.price),2),'  REAIs') as Total_amount,
concat(round(avg(oi.price),2),'  REAIs') as Avg_amt
from `target.customers` c
join `target.orders` o
on c.customer_id=o.customer_id
join`target.order_items`oi
on oi.order_id = o.order_id
group by customer_state
order by customer_state;
```

OUTPUT –

Row	customer_state	Total_amount	Avg_amt
1	AC	19680.62 REAIs	234.29 REAIs
2	AL	96962.06 REAIs	227.08 REAIs
3	AM	27966.93 REAIs	181.6 REAIs
4	AP	16262.8 REAIs	232.33 REAIs
5	BA	616645.82 REAIs	170.82 REAIs
6	CE	279464.03 REAIs	199.9 REAIs
7	DF	355141.08 REAIs	161.13 REAIs
8	ES	325967.55 REAIs	154.71 REAIs
9	GO	350092.31 REAIs	165.76 REAIs
10	MA	152523.02 REAIs	198.86 REAIs
11	MG	1872257.26 REAIs	154.71 REAIs

Row	customer_state	Total_amount	Avg_amt
1	AC	15982.95 REAIs	173.73 REAIs
2	AL	80314.81 REAIs	180.89 REAIs
3	AM	22356.84 REAIs	135.5 REAIs
4	AP	13474.3 REAIs	164.32 REAIs
5	BA	511349.99 REAIs	134.6 REAIs
6	CE	227254.71 REAIs	153.76 REAIs
7	DF	302603.94 REAIs	125.77 REAIs
8	ES	275037.31 REAIs	121.91 REAIs
9	GO	294591.95 REAIs	126.27 REAIs
10	MA	119648.22 REAIs	145.2 REAIs

INSIGHTS –

- We have queried to find total and average amount of orders. This can be inferred from either payments table payment_value column or order_item tables price column.

- The analysis reveals significant variation in total and average order prices across different states. São Paulo (SP) and Rio de Janeiro (RJ) have higher total prices, indicating larger revenue contributions, while Roraima (RR) and Amapá (AP) have relatively lower total prices.

- Amapá (AP) and Roraima (RR) have higher average order prices, suggesting potential higher-value customers or demand for premium-priced products, while states like Espírito Santo (ES) and Goiás (GO) have relatively lower average order prices.

- This information enables targeted marketing campaigns and provides insights into regional customer behavior and preferences.

---

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

QUERY –

```
select customer_state ,
concat(round(sum(freight_value),2),'  REAIs') as Total_freight_value,
concat(round(avg(freight_value),2),'  REAIs') as Avg_freight_value
from `target.customers` c
join `target.orders` o on c.customer_id=o.customer_id
join`target.order_items` oi on oi.order_id = o.order_id
group by customer_state
order by customer_state;
```

OUTPUT –

Row	customer_state	Total_freight_value	Avg_freight_value
1	AC	3686.75 REAIs	40.07 REAIs
2	AL	15914.59 REAIs	35.84 REAIs
3	AM	5478.89 REAIs	33.21 REAIs
4	AP	2788.5 REAIs	34.01 REAIs
5	BA	100156.68 REAIs	26.36 REAIs
6	CE	48351.59 REAIs	32.71 REAIs
7	DF	50625.5 REAIs	21.04 REAIs
8	ES	49764.6 REAIs	22.06 REAIs
9	GO	53114.98 REAIs	22.77 REAIs
10	MA	31523.77 REAIs	38.26 REAIs
11	MG	270853.46 REAIs	20.63 REAIs

INSIGHTS –

- The query showcases the cost of freight values for each states and we infer that states like São Paulo (SP) and Minas Gerais (MG) have higher total freight values, indicating higher overall shipping costs incurred. On the other hand, states like Roraima (RR) and Amapá (AP) have lower total freight values.
- Average freight values reflect the average shipping cost per order, with states like São Paulo (SP) and Minas Gerais (MG) having lower average freight values, suggesting more efficient logistics networks or economies of scale.
- States like Paraíba (PB) and Piauí (PI) have higher average freight values, indicating relatively higher shipping costs per order. These insights can assist in optimizing logistics operations and understanding regional shipping trends.

5.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.

QUERY –

```
select
order_id,
datetime_diff(order_delivered_customer_date,order_purchase_timestamp,day) as delivery_time,
datetime_diff(order_estimated_delivery_date,order_delivered_customer_date,day) as
diff_estimated_delivery
from
`target.orders`;
```

OUTPUT –

Row	order_id	delivery_time	diff_estimated_delive
1	1950d777989f6a877539f5379...	30	-12
2	2c45c33d2f9cb8ff8b1c86cc28...	30	28
3	65d1e226dfaeb8cdc42f66542...	35	16
4	635c894d068ac37e6e03dc54e...	30	1
5	3b97562c3aee8bdedcb5c2e45...	32	0
6	68f47f50f04c4cb6774570cfde...	29	1
7	276e9ec344d3bf029ff83a161c...	43	-4
8	54e1a3c2b97fb0809da548a59...	40	-4
9	fd04fa4105ee8045f6a0139ca5...	37	-1
10	302bb8109d097a9fc6e9cefc5...	33	-5

INSIGHTS –

- From Query Analysis, The "Time_to_deliver" column represents the number of days taken to deliver an order to the customer from the purchase date, while the "diff_estimated_delivery" column indicates the difference between the estimated delivery date and the actual delivery date.
- By analyzing this data, we can identify orders that took longer to deliver and compare each delivery time with the average delivery timeline to assess delivery efficiency. Negative values in the "diff_estimated_delivery" column indicate delayed deliveries, while positive values indicate early deliveries.
- Digging deeper, the reasons for these variances can help improve delivery timelines and reduce the difference between estimated and actual delivery dates, leading to enhanced logistics and delivery processes.

---

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

QUERY –

```
select state ,avg_freight_value from
(
(
select concat('HIGH #   ',customer_state) as state,
max(freight_value) as High_freight_value,
concat(round(avg(freight_value),2),' REAIs') as avg_freight_value
from `target.customers` c
join `target.orders` o on c.customer_id=o.customer_id
join `target.order_items` as p on o.order_id = p.order_id
group by customer_state
order by 3 desc
limit 5
)
union all
(
select concat('LOW #   ',customer_state) as state,
min(freight_value) as low_freight_value,
concat(round(avg(freight_value),2),' REAIs') as avg_freight_value
from `target.customers` c
join `target.orders` o on c.customer_id=o.customer_id
join `target.order_items` as p on o.order_id = p.order_id
group by customer_state
order by 3 asc
limit 5
)
)as t;
```

OUTPUT –

Row	state	avg_freight_value
1	HIGH # RR	42.98 REAIs
2	HIGH # PB	42.72 REAIs
3	HIGH # RO	41.07 REAIs
4	HIGH # AC	40.07 REAIs
5	HIGH # PI	39.15 REAIs
6	LOW # SP	15.15 REAIs
7	LOW # PR	20.53 REAIs
8	LOW # MG	20.63 REAIs
9	LOW # RJ	20.96 REAIs
10	LOW # DF	21.04 REAIs

Another Approach :-

```
with cte as
(
select State,'High'as val,avg_freight_value,
dense_rank() over (order by avg_freight_value desc) as avg_rank
from
(
select customer_state as state,round(avg(freight_value),2) as avg_freight_value
from `target.customers` as c
join `target.orders` as o on c.customer_id = o.customer_id
join `target.order_items` as p on o.order_id = p.order_id
group by customer_state
) nt1

union all

select state,"Low"as val,avg_freight_value,
dense_rank() over (order by avg_freight_value asc) as avg_rank
from
(
select customer_state as state,round(avg(freight_value),2) as avg_freight_value
from `target.customers` as c
join `target.orders` as o on c.customer_id = o.customer_id
join `target.order_items` as p on o.order_id = p.order_id
group by customer_state
) nt2
)

select concat(val," - ",avg_rank) as fv_order,state,concat(cte.avg_freight_value,'  REAIs')as
Avg_Freight_cost
from cte
where avg_rank<=5
order by fv_order;
```

OUTPUT –

Row	fv_order	state	Avg_Freight_cost
1	High - 1	RR	42.98 REAIs
2	High - 2	PB	42.72 REAIs
3	High - 3	RO	41.07 REAIs
4	High - 4	AC	40.07 REAIs
5	High - 5	PI	39.15 REAIs
6	Low - 1	SP	15.15 REAIs
7	Low - 2	PR	20.53 REAIs
8	Low - 3	MG	20.63 REAIs
9	Low - 4	RJ	20.96 REAIs
10	Low - 5	DF	21.04 REAIs

INSIGHTS

- Our query shows that Certain states exhibit higher average freight values, indicating specific characteristics or logistical challenges that result in increased freight costs.
- Certain states exhibit higher average freight values, indicating specific characteristics or logistical challenges that result in increased freight costs. Some states demonstrate lower average freight values, suggesting more favorable logistics infrastructure or other factors contributing to reduced freight costs. With average freight costs, we can that there is a notable difference (Significant Variation) in the average freight value among different states.
- These insights provide businesses with valuable information for optimizing their shipping and logistics operations. It enables them to make informed decisions regarding pricing strategies, supply chain optimization, and resource allocation.
- For states with the lowest average freight values, such as São Paulo (SP), Distrito Federal (DF), Rio de Janeiro (RJ), Minas Gerais (MG), and Paraná (PR), businesses can focus on further optimizing freight costs by negotiating better rates, consolidating shipments, and improving logistics routes.
- On the other hand, states with the highest average freight values, such as Piauí (PI), Acre (AC), Rondônia (RO), Roraima (RR), and Paraíba (PB), can explore opportunities to mitigate factors contributing to higher costs, including partnering with local logistics providers, optimizing

transportation routes, and leveraging technology for efficient freight management.

---

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

QUERY –

```
with cte as
(
select state,'FAST'as val,avg(delivery_time) as avg_delivery_time,
dense_rank() over (order by avg(delivery_time) desc) as rnk
from
(
select customer_state as state,
datetime_diff(order_delivered_customer_date,order_purchase_timestamp,day) as delivery_time,
from `target.customers` as c
join `target.orders` as o on c.customer_id = o.customer_id
group by state,order_delivered_customer_date,order_purchase_timestamp,delivery_time
) nt1
group by state

union all

select state,'SLOW'as val,avg(delivery_time) as avg_delivery_time,
dense_rank() over (order by avg(delivery_time) asc) as rnk
from
(
select customer_state as state,
datetime_diff(order_delivered_customer_date,order_purchase_timestamp,day) as delivery_time,
from `target.customers` as c
join `target.orders` as o on c.customer_id = o.customer_id
group by state,order_delivered_customer_date,order_purchase_timestamp,delivery_time
) nt2
group by state
)

select concat(val," - ",rnk) as speed_of_delivery,state,
round(avg_delivery_time,2)as Avg_delivery_time
from cte
where rnk<=5
order by 1;
```

OUTPUT –

Row	speed_of_delivery	state	Avg_delivery_time
1	FAST - 1	RR	28.98
2	FAST - 2	AP	26.73
3	FAST - 3	AM	25.99
4	FAST - 4	AL	24.04
5	FAST - 5	PA	23.32
6	SLOW - 1	SP	8.3
7	SLOW - 2	PR	11.53
8	SLOW - 3	MG	11.54
9	SLOW - 4	DF	12.51
10	SLOW - 5	SC	14.48

INSIGHTS –

- From our queried output, we can say that states with the fastest average delivery times, such as Alagoas (AL), Pará (PA), Amazonas (AM), Amapá (AP), and Roraima (RR), the logistics are perfectly fine whereas the states with slow deliveries such as Santa Catarina(SC), Minas Gerais (MG), Distrito Federal (DF), (PR), Sao Paulo(SP)are to be considered.
- It is essential to identify and address the factors causing these prolonged delivery durations. Businesses should evaluate their logistics network, transportation routes, and last-mile delivery processes to optimize efficiency. Collaborating with local logistics providers or establishing strategic partnerships can also help improve delivery performance in these states. Implementing technology solutions like real-time tracking systems and efficient delivery scheduling tools can streamline the last-mile operations.
- Open communication and collaborative relationships with shipping carriers and logistics partners are crucial for monitoring and addressing any issues promptly. By optimizing the last-mile delivery process and maintaining efficient relationships with partners, businesses can reduce delivery times, enhance customer experience, and improve overall satisfaction.

d) Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

QUERY –

```
select customer_state as state,
round(avg(date_diff(o.order_estimated_delivery_date,o.order_delivered_customer_date,day)),2)as
avg_speed_delivery
from `target.customers` as c
join `target.orders` as o on c.customer_id = o.customer_id
group by state
order by avg_speed_delivery desc
limit 5
```

OUTPUT –

Row	state	avg_speed_delivery
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

INSIGHTS –

- From the Query Analysis, the top 5 states like Amazonas (AM), Amapá (AP), Roraima (RR), Acre (AC), Rondônia(RO) have average speed for order delivery compared to the estimated date, with an average difference of 16 to 19 days. This suggests efficient delivery processes, well-established transportation networks, optimized routing strategies, and proactive coordination with shipping carriers.

- Customers benefit from reliable and prompt delivery, increasing satisfaction. Businesses can leverage this information to highlight their effective logistics operations and gain a competitive advantage. It also helps identify areas for improving delivery efficiency by comparing with states that have longer delivery times.

6.Analysis based on the payments:

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

QUERY –

```
select ym,
payment_type,
sum(cnt)as Total_orders
from
(
select
format_timestamp("%Y-%m",order_purchase_timestamp) as ym,
payment_type,
count(p.order_id)as cnt
from `target.payments` p
join `target.orders` o
on p.order_id=o.order_id
group by order_purchase_timestamp,payment_type
)
where payment_type is not null
group by payment_type,ym
order by ym
```

Another Approach:

```
with cte as (
select
format_timestamp('%Y-%m', o.order_purchase_timestamp)as period,
count(case when payment_type='credit_card' then 1 end) as CREDIT_CARD,
count(case when payment_type='debit_card' then 1  end)as DEBIT_CARD,
count(case when payment_type='UPI' then 1 end) as UPI,
count(case when payment_type='voucher' then 1 end) as VOUCHER,
count(case when payment_type='not_defined' then 1 end)as NOT_DEFINED,
count(case when payment_type is null then 1 end)as NOT_KNOWN,
count(o.order_id)as total_orders
```

```
from `target.payments` p
join `target.orders` o
on p.order_id=o.order_id
where payment_type is not null and order_status not in ('canceled','unavailable')
group by period
order by period
)
```

OUTPUT –

Row	ym	payment_type	Total_orders
1	2016-09	credit_card	3
2	2016-10	credit_card	254
3	2016-10	voucher	23
4	2016-10	debit_card	2
5	2016-10	UPI	63
6	2016-12	credit_card	1
7	2017-01	voucher	61
8	2017-01	UPI	197
9	2017-01	credit_card	583
10	2017-01	debit_card	9
11	2017-02	credit_card	1356

Row	period	CREDIT_CARD	DEBIT_CARD	UPI	VOUCHER	NOT_DEFINED	NOT_KNOWN	total_orders
1	2016-09	1	0	0	0	0	0	1
2	2016-10	227	2	60	22	0	0	311
3	2016-12	1	0	0	0	0	0	1
4	2017-01	574	9	193	60	0	0	836
5	2017-02	1309	13	383	116	0	0	1821
6	2017-03	1959	30	582	200	0	0	2771
7	2017-04	1828	26	488	201	0	0	2543
8	2017-05	2807	30	758	288	0	0	3883
9	2017-06	2428	27	702	236	0	0	3393
10	2017-07	3024	22	829	349	0	0	4224
11	2017-08	3244	34	920	275	0	0	4473

INSIGHTS –

- The query provides insights into customer payment preferences by analyzing the monthly distribution of payment types. By analyzing the monthly distribution of payment types, businesses can gain insights into customer payment preferences, identify trends in payment methods, This information can be valuable for optimizing the checkout experience, expanding payment options and improving overall customer satisfaction.

- Credit card emerges as the preferred payment method, highlighting its convenience, security, and widespread acceptance. Businesses should ensure seamless credit card payment processing and maintain strong partnerships with payment service providers. Analyzing payment trends helps optimize the checkout experience, expand payment options, and enhance overall customer satisfaction.

---

b) Find the no. of orders placed on the basis of the payment installments that have been paid. (atleast 1 sucessful payment)

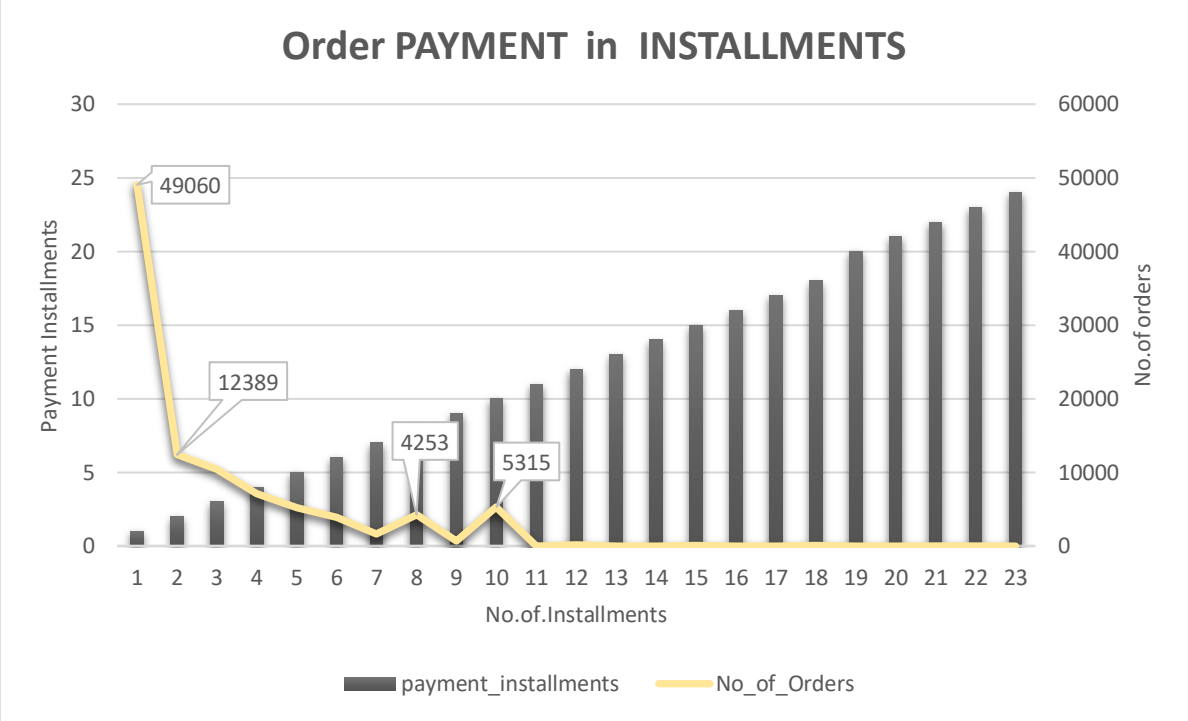
QUERY –

```
select payment_installments,
count(distinct order_id) as No_of_Orders
from `target.payments`
where payment_installments <> 0
group by 1
order by 1
```

OUTPUT –

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

ANALYSIS-



INSIGHTS –

- Our Query Results shows the count of distinct orders for each payment instalment option. It indicates the distribution of customers who choose different instalment plans for their payments. Mostly preferred instalment is 1 or we can say that atleast 1 of many has been paid. The data indicates a notable customer preference for a particular focus on the 9-10 instalment range.
- To leverage this pattern and drive sales growth, businesses should consider providing customers with flexible payment options and instalment plans. By offering the ability to divide payments into multiple instalments, businesses can accommodate customer preferences, enhance affordability, and potentially expand their customer base. Incorporating instalment plans into the payment options can be an effective strategy to boost sales and improve overall customer satisfaction.

❖ 7. Additional Questions (DEFINED BY ME)

a) Explore how the customers are spread across BRAZILian cities :-

QUERY –

```
select
count(distinct customer_city) as city,
count(distinct customer_state) as state
from `target.customers`
```

OUTPUT –

Row	city	state
1	4119	27

INSIGHTS –

- We can infer that the count of unique no.of.cities and unique no.of.states in the customer table is 4119 and 27 where in geolocation table the cities count count to 8126 and 27 states , which mean the Target should target even more on expanding across the Brazilian country for a successful retailing and should be able give accessibility to the customers across Brazil.

b) Based on the order status , count the orders ?

QUERY –

```
select order_status,
count(order_status) as no_of_orders,
from `target.orders`
group by order_status
```

OUTPUT –

Row	order_status	no_of_orders
1	created	5
2	shipped	1107
3	approved	2
4	canceled	625
5	invoiced	314
6	delivered	96478
7	processing	301
8	unavailable	609

INSIGHTS –

- Based on the query , we can analyze the order status of every order and we infer that the cancelled orders and unavailable orders are at more than the nominal level than the usual .
- Knowing why the orders are being cancelled can help us in rectifying the error and huge concentration to be given on the backend team to analyze why the order status is unavailable would help much more gaining the percentage of successful order and attain the customers likeliness towards the firm
- It is important to understand these market dynamics is to be considered to make the change and replicate successful strategies in the future. The impact of payment value of the unavailable orders is huge enough and the retaining the data can help us identify the trend pattern and seasonality.

---

c) Count the total orders based on product category?

QUERY –

```
select product_category,
count(order_id) as no_of_orders,
from `target.products`p
join `target.order_items`o
on o.product_id=p.product_id
group by product_category
order by 2 desc
```

OUTPUT –

Row	product_category	no_of_orders
1	bed table bath	11115
2	HEALTH BEAUTY	9670
3	sport leisure	8641
4	Furniture Decoration	8334
5	computer accessories	7827
6	housewares	6964
7	Watches present	5991
8	telephony	4545
9	Garden tools	4347
10	automotive	4235
11	toys	4117

INSIGHTS –

- Here our Query finds the number of orders placed in each product category and this helps us identify the favourite category to be sold based on seasonality and to increase the sales more, we can campaign and market it accordingly to get the likeliness of the customers and thereby increasing the revenue.
- Most of the items are being order during the afternoon and night hours, so we can run time-limited promotions, offer personalized marketing promotions to capture the attention of users. we can also offer exclusive early bird discounts and ensure responsive customer support to capitalize on the high order volume in the afternoon.

---

d) calculate the average review score for each product category ?

QUERY –

```
select product_category, round(avg(review_score),2)as avg_rating
from `target.order_reviews`rw
join `target.order_items`oi on oi.order_id=rw.order_id
join `target.products`p on p.product_id=oi.product_id
group by 1
```

OUTPUT –

Row	product_category	avg_rating
1	Garden tools	4.04
2	sport leisure	4.11
3	Fashion Bags and Accessories	4.14
4	HEALTH BEAUTY	4.14
5	bed table bath	3.9
6	*null*	3.84
7	electronics	4.04
8	automotive	4.07
9	Watches present	4.02
10	musical instruments	4.15

INSIGHTS –

- Here from our Query analysis, we find that the review score of the product category. This helps us in inferring the customers goto product and their likability towards each category
- Hence the distribution of product in particular category, businesses can gain insights into customers category preferences, identify trends. This information can be valuable for optimizing the buying experience, expanding more products in favourite category and improving overall customer satisfaction.

RECOMMENDATIONS:-

- The count current customers are only from 4113 cities whereas Brazil has over 8126 and 27 states ,In order to give accessibility to the mass customers, **TARGET** should target even more on expanding across the Brazilian country.
- To address low-order months, businesses should conduct market research, collaborate with complementary businesses, offer promotions, and utilize targeted marketing. During different times of the day, strategies should be tailored to maximize sales. In states with higher customer bases, prioritize customer engagement, while in states with lower customer bases, target growth opportunities. Identify and leverage competitive advantages to stand out in the market

- Also inferring favourite product category, campaigning & marketing seasonal products related to the Brazilian culture like **NEWYEAR, BLACK FRIDAY, CARNIVALS, FIFA WORLD CUP & CAPOEIRA** helps us in understanding their which in turn gains more audience.
- To improve businesses, Optimize shipping operations, pricing strategies, and resource allocation. Focus on reducing the costs through negotiations and route improvements by having partnerships with local carriers and logistics partners are vital so that businesses can reduce delivery times and enhance logistics processes by using technology solutions which helps us improve performance by identifying the causing factors as last-mile delivery processes can streamline operations and reduce transit times. So that we can ensure customer satisfaction.
- Maintain proactive communication with customers about delivery expectations and provide timely updates on order status and potential delays.
- Ensure a secure payment infrastructure that supports various payment methods, adapting to evolving customer preferences and emerging technologies and also educate customers about alternative payment methods and offer incentives to encourage their usage. Promote the benefits of lower installment options and create value through targeted marketing campaigns.
- Collect customer feedback and analyze behaviour to understand preferences and to ensure customer satisfaction. With that , **Branding** can be gained.

Based on the data analysis conducted in this case study, businesses can gain valuable insights to optimize their operations, and recommendations improve customer satisfaction, and drive sales growth.

Data provided by: SCALER