

I. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.

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

Query:

```
select
  column_name,
  data_type
from
  scaler-dsml-sql-402115.target.INFORMATION_SCHEMA.COLUMNS
where table_name = 'customers';
```

Result:

Query results				
JOB INFORMATION		RESULTS	CHART	PREVIEW
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: Most of the data in Customer table is being stored in string format as the columns in customer table are of STRING data type.

Recommendation: NA

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

Query:

```
select
  min(order_purchase_timestamp) as First_Order_Timestamp,
  max(order_purchase_timestamp) as Last_Order_Timestamp
from target.orders;
```

Result:

Query results			
JOB INFORMATION		RESULTS	CHART
Row	First_Order_Timestamp	Last_Order_Timestamp	PREVIEW
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC	JSON

Insights: The first order in the dataset was placed on 4th of September 2016 and Last order was placed on 17th October 2018.

Recommendations: N/A

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

Query:

```
select
  count(distinct c.customer_state) as States_Count,
  count(distinct c.customer_city) as Cities_Count
from target.customers c join target.orders o on c.customer_id=o.customer_id;
```

Result:

Query results

JOB INFORMATION		RESULTS	CHART	PREVIEW
Row	States_Count	Cities_Count		
1	27	4119		

Insights: The customers that have ordered are spread over 4119 different cities from 27 States.

Recommendations: N/A

II. In-depth Exploration:

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

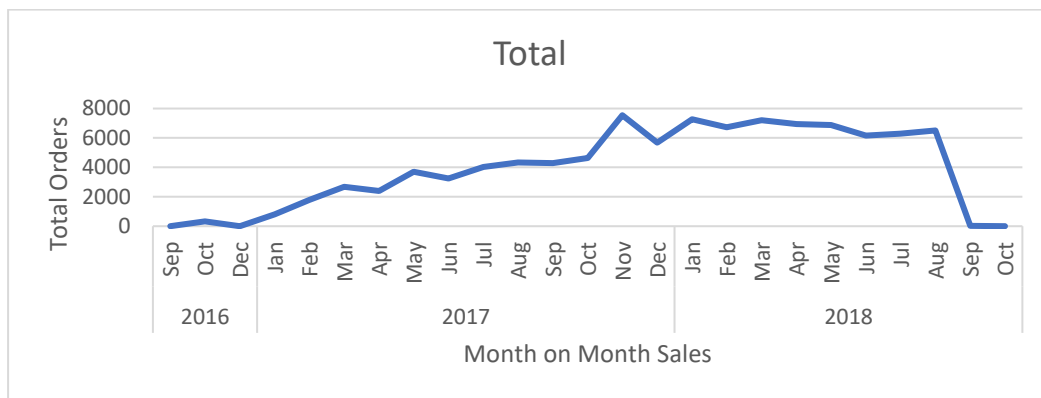
Query:

```
select
  FORMAT_DATE('%Y-%m',order_purchase_timestamp) YearMonth,
  count(order_id) as Total_Orders
from target.orders
group by 1
order by 1;
```

Result:

Row	YearMonth	Total_Orders
1	2016-09	4
2	2016-10	324
3	2016-12	1
4	2017-01	800
5	2017-02	1780
6	2017-03	2682
7	2017-04	2404
8	2017-05	3700
9	2017-06	3245
10	2017-07	4026

Insights:



- There is an overall steady increase in orders with a peak in November 2017.
- With few dips in orders next to peaks like Nov 2018 and January 2018 might be because of festive season sales were high and again back to its normal course.
- There was slight decrease trend in orders from March 2018 with peak in August.

Recommendations:

- Offer seasonal discounts on the products specific to a particular season to boost the sales.
- Implement new marketing strategies to attract untapped customer base.

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

Query:

```
select
  FORMAT_DATE('%Y-%m', order_purchase_timestamp) YearMonth,
  count(order_id) as Total_Orders
from target.orders
group by 1
order by 2 desc;
```

Result:

Row	YearMonth	Total_Orders
1	2017-11	7544
2	2018-01	7269
3	2018-03	7211
4	2018-04	6939
5	2018-05	6873
6	2018-02	6728
7	2018-08	6512
8	2018-07	6292
9	2018-06	6167
10	2017-12	5673

Insights:

- In 2017, there is a steady increase in orders from January to December, with a peak in November.
- The trend continues in 2018, with a peak in orders during January, followed by a decline in the subsequent months. There is also a small spike in August.
- These patterns suggest that there may be seasonality factors like holidays, promotional sales, or other seasonal trends.

Recommendations:

- There were peaks in November 2017 and January 2018 might be related to festive seasons, business can club low selling products with higher sales products along with discounts to be lucrative.
- Operations can plan for inventory management and logistics accordingly to fulfil the orders in timely manner without any delays or cancellations.

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:

```
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'
    when extract(HOUR from order_purchase_timestamp) between 19 and 23 then
      'Night'
    end as Duration_Of_Day ,
    count(order_id) as Total_Orders
from target.orders
group by 1
order by 2 desc;
```

Result:

JOB INFORMATION		RESULTS
Row	Duration_Of_Day	Total_Orders
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

Insights:

- Most of the orders from Brazilian customers are placing the orders during Afternoon.
- During Morning & Night almost same number of orders being placed.
- Least number of orders are being placed during Dawn.

Recommendations: Since most of the orders are being placed during afternoon, business can run more discounted sales or promotional offers during this time to increase the number of sales.

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

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

Query:

```
select
  c.customer_state,
  FORMAT_DATE('%Y-%m', o.order_purchase_timestamp) YearMonth,
  count(order_id) as Total_Orders_Per_Month
from target.orders o join target.customers c on o.customer_id=c.customer_id
group by c.customer_state, YearMonth
order by YearMonth, c.customer_state;
```

Result:

Row	customer_state	YearMonth	Total_Orders_Per_Month
1	RR	2016-09	1
2	RS	2016-09	1
3	SP	2016-09	2
4	AL	2016-10	2
5	BA	2016-10	4
6	CE	2016-10	8
7	DF	2016-10	6
8	ES	2016-10	4
9	GO	2016-10	9
10	MA	2016-10	4

Insights:

- The state SP consistently stands out being significant contributor to the total order count each month overall.
- The states MG, RJ, BA shows steady growth in terms of number orders.
- RR, AP, AC, AM & RO states does not show any growth.

Recommendations: Customer outreach program might help with boost in sales in the states with less growth.

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

Query:

```
select
  customer_state,
  count(distinct customer_id) as Customers_Statewise
from target.customers
group by customer_state
order by 2 desc;
```

Result:

Row	customer_state	Customers_Statewise
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:

- The states SP, RJ, MG, RS & PR are the top 5 states combined having almost 77% of the total customers.
- RR, AP, AC, AM & RO are the states with least number of customers

Recommendations:

The states with least number of customers are having low order volume and high freight charges and takes more delivery time. By reducing the freight charges and efficient shipping plans to deliver orders in time along with better marketing might increase the customer base.

IV. Impact on Economy: Analyze 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:

```
with cte as
(
    select
        extract (year from order_purchase_timestamp) as year,
        round(sum(p.payment_value),3) as Total_cost
    from target.payments p
        join target.orders o on p.order_id=o.order_id
    where extract (date from order_purchase_timestamp) between '2017-01-01'
and '2018-08-31'
/*Conditon to include only month from Jan to Aug for 2017 & 2018 as
mentioned in Question*/

group by 1
)

select
    c1.Total_cost as Cost_for_2017,
    c1.Year as Year_2017,
    c2.Total_cost as Cost_for_2018,
    c2.Year as Year_2018,
    round(((c2.Total_cost - c1.Total_cost)/c1.Total_cost)*100,3)
as Percentage_Increase
from cte c1 join cte c2 on c1.year=2017 and c2.year = 2018;
```

Result:

JOB INFORMATION		RESULTS	CHART	PREVIEW	JSON	EXECUTION DETA
low	Cost_for_2017	Year_2017	Cost_for_2018	Year_2018	Percentage_Increase	
1	7249746.73	2017	8694733.84	2018	19.932	

Insights: During the given period the cost of orders from 2017 to 2018 has increased by almost 20%.

Recommendations: With better marketing we can reach to new customers along with loyalty programs we increase sales with existing customers might increase cost of order sales.

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

Query:

```
select
customer_state,
round(sum(Order_Price),2) as Total_Order_Price,
round(avg(Order_Price),2) as Average_Order_Price
from (
select
c.customer_state,
o.order_id,
sum(oi.price) as Order_Price
from target.customers c
join target.orders o on c.customer_id=o.customer_id
join target.order_items oi on o.order_id =oi.order_id
group by c.customer_state,o.order_id
)t
group by customer_state
order by 3 desc;
```

Result:

Row	customer_state	Total_Order_Price	Average_Order_Price
1	PB	115268.08	216.67
2	AP	13474.3	198.15
3	AC	15982.95	197.32
4	AL	80314.81	195.41
5	RO	46140.64	186.8
6	PA	178947.81	184.48
7	TO	49621.74	177.86
8	PI	86914.08	176.3
9	MT	156453.53	173.26
10	RN	83034.98	172.27

Insights:

- The state SP significantly has the highest total revenue compared to any other states however has the lowest average order price.
- There is disparity in both total revenue and average order value across states SP, RJ & MG contributes more to the total revenue compared to other states.
- States like PB, AC & AP have high average order cost. While this indicates a higher amount spent per order, it also suggest that the total number of customers is smaller.
- Smaller revenue states with high average order cost suggest challenges in attracting a larger volume of orders.

Recommendations: Strategies like better customer support, marketing, or promotional sales to be explored in the states with less total revenue to boost sales and attract new customers to increase order volume.

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

Query:

```
select
  customer_state,
  round(sum(Freight_Price),2) as Total_Freight_Price,
  round(avg(Freight_Price),2) as Average_Freight_Price
from (
  select
    c.customer_state,
    o.order_id,
    sum(oi.freight_value) as Freight_Price
  from target.customers c
       join target.orders o on c.customer_id=o.customer_id
       join target.order_items oi on o.order_id =oi.order_id
  group by c.customer_state,o.order_id
)t
group by customer_state
order by 3 desc;
```

Result:

Row	customer_state	Total_Freight_Price	Average_Freight_Price
1	RR	2235.19	48.59
2	PB	25719.73	48.35
3	RO	11417.38	46.22
4	AC	3686.75	45.52
5	PI	21218.2	43.04
6	MA	31523.77	42.6
7	TO	11732.68	42.05
8	AP	2788.5	41.01
9	SE	14111.47	40.9
10	PA	38699.3	39.9

Insights:

- All the states with highest average Freight Price are also states with highest average order price indicating that these were remote with respect to warehouses or lack of better shipping routes leading to increased shipping costs
- States with lower average freight prices may have more efficient logistics or near to the warehouses or distribution centres resulting in lower shipping costs per order

Recommendations:

- Businesses should explore bulk shipping or partnering with logistics companies that have frequent shipping to the states with higher shipping costs to achieve efficiency and reduce average freight prices.
- Resulting in lower freight charges it might also increase the customer base.

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.

Query:

```
select
  order_id,
  order_purchase_timestamp,
  order_estimated_delivery_date,
  order_delivered_customer_date,
  date_diff(order_delivered_customer_date, order_purchase_timestamp, day) as
time_to_deliver,
  date_diff(order_estimated_delivery_date, order_delivered_customer_date, day)
as diff_estimated_delivery
from target.orders
where order_status='delivered'
order by 2;
```

Result:

Row	order_id	order_purchase_timestamp	order_estimated_delivery_date	order_delivered_customer_date	time_to_deliver	diff_estimated_delivery
1	bfb0f9bdef84302105ad712db648a6c	2016-09-15 12:16:38 UTC	2016-10-04 00:00:00 UTC	2016-11-09 07:47:38 UTC	54	-36
2	3b697a20d9e427646d92567910af6d57	2016-10-03 09:44:50 UTC	2016-10-27 00:00:00 UTC	2016-10-26 14:02:13 UTC	23	0
3	be5bc2f0da14d8071e2d45451ad119d9	2016-10-03 16:56:50 UTC	2016-11-07 00:00:00 UTC	2016-10-27 18:19:38 UTC	24	10
4	a41c8759fbc7aab36ea07e038b2d4465	2016-10-03 21:13:36 UTC	2016-11-29 00:00:00 UTC	2016-11-03 10:58:07 UTC	30	25
5	d207cc272675637bfed0062edff0818	2016-10-03 22:06:03 UTC	2016-11-23 00:00:00 UTC	2016-10-31 11:07:42 UTC	27	22
6	cd3b8574c82b42fc8129f6d502690c3e	2016-10-03 22:31:31 UTC	2016-11-23 00:00:00 UTC	2016-10-14 16:08:00 UTC	10	39
7	ae8a60e4b03c5a4ba9ca0672c164b181	2016-10-03 22:44:10 UTC	2016-12-01 00:00:00 UTC	2016-11-03 14:04:50 UTC	30	27
8	ef1b29b591d31d57c0d7337460dd83c9	2016-10-03 22:51:30 UTC	2016-11-25 00:00:00 UTC	2016-11-01 15:14:45 UTC	28	23
9	0a0837a5eee9e7a9ce2b1fa831944d27	2016-10-04 09:06:10 UTC	2016-11-24 00:00:00 UTC	2016-10-22 14:51:18 UTC	18	32
10	1ff217aa612f6cd7c4255c9bfe931c8b	2016-10-04 09:16:33 UTC	2016-11-24 00:00:00 UTC	2016-10-24 16:33:45 UTC	20	30

Insights:

- Of all the orders delivered in the given dataset around 94 % of the orders were delivered well in advance to the estimated delivery date.
- Remaining 6% orders took more than estimated delivery date for delivering the order to customer.
- Around 6354 of orders that were delayed, 52% of the orders belongs to the customers from only two states **SP (28.6%) & RJ (23.5%)**
- Below are the top 8 states that has majority of order delays

Customer_State	Total_Orders_Delayed	% Orders Delayed
SP	1820	28.64
RJ	1495	23.52
MG	519	8.16
BA	396	6.26
RS	325	5.11
SC	291	4.57
ES	214	3.36
PR	199	3.13

Recommendations:

- Additional logistics must be routed to states SP and RJ to decrease the delay in delivering order to customers.
- If there are any operational constraints on delivering the order to these states, while preparing the order itself we can update the estimated delivery time as per the logistical planning to make customer better informed and less dissatisfied on delay.
- Logistics need to be planned properly for the states where orders are being delayed.

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

Query:

```
with freight as
(
select
    c.customer_state,
    o.order_id,
    sum(oi.freight_value) as Freight_Price
from target.customers c
    join target.orders o on c.customer_id=o.customer_id
    join target.order_items oi on o.order_id =oi.order_id
group by c.customer_state,o.order_id
),
avg_freight as(
select
    customer_state,
    round(sum(Freight_Price),2) as Total_Freight_Price,
    round(avg(Freight_Price),2) as Average_Freight_Price
from freight
group by customer_state
)
```

#Top 5 states with least Average Freight Price

```
select
    customer_state,
    Average_Freight_Price
from avg_freight
order by 2
limit 5;
```

#Top 5 states with Highest Average Freight Price

```
select
    customer_state,
    Average_Freight_Price
from avg_freight
order by 2 desc
limit 5;
```

Result:

Highest Average Freight

Row	customer_state	Average_Freight_Price
1	RR	48.59
2	PB	48.35
3	RO	46.22
4	AC	45.52
5	PI	43.04

Lowest Average Freight

Row	customer_state	Average_Freight_Price
1	SP	17.37
2	MG	23.46
3	PR	23.58
4	DF	23.82
5	RJ	23.95

Insights:

- All the top 5 states with highest average freight cost are in the top 10 of states having least number of orders, since there were a smaller number of orders it could have been that operations cannot plan bulk shipping which resulted in high freight costs.
- Similarly, all 5 states with least average freight cost are in the top 8 states with the greatest number of orders which resulted in less freight cost because of frequent bulk ship based on order volume.

Recommendations:

- Partnering with logistics companies that have frequent shipping to the states with higher shipping costs to achieve efficiency and reduce average freight prices.
- Also attracting new customers from these states also reduce the freight cost as there will be more orders

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

Query:

```
with delivery as(
select
    distinct o.order_id,
    c.customer_state,
    o.order_purchase_timestamp,
    o.order_estimated_delivery_date,
    o.order_delivered_customer_date,
    date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day)
as time_to_deliver,
    date_diff(o.order_estimated_delivery_date,o.order_delivered_customer_date,
day) as diff_estimated_delivery
from target.customers c
    join target.orders o on c.customer_id=o.customer_id
    join target.order_items oi on o.order_id =oi.order_id
where order_status='delivered'
),
avg_delivery as(
select
    customer_state,
    round(avg(time_to_deliver),2) as Avg_Delivery_Time,
    round(avg(diff_estimated_delivery),2) as Avg_Estimated_Delivery_Time
from delivery
group by customer_state
)
```

#Top 5 states with Lowest Average Delivery Time

```
select
    customer_state,
    Avg_Delivery_Time
from avg_delivery
order by 2
limit 5;
```

#Top 5 states with Highest Average Delivery Time

```
select
    customer_state,
    Avg_Delivery_Time
from avg_delivery
order by 2 desc
limit 5;
```

Result:

Highest Average Delivery Time

Row	customer_state	Avg_Delivery_Time
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

Lowest Average Delivery Time

Row	customer_state	Avg_Delivery_Time
1	SP	8.3
2	DF	12.51
3	MG	11.54
4	PR	11.53
5	ES	15.33

Insights:

- All the top 5 states with highest average delivery time are the states having least number of orders, because of which it could have been possible that there was no frequent shipping route resulted in highest delivery time.
- Similarly, all 5 states with least average freight cost are in the top 10 states with the large number of orders so frequent shipping would result in low delivery time.

Recommendations:

- Partnering with logistics companies that have frequent shipping to the states with higher delivery time might reduce average delivery time.
- Also attracting new customers from these states means increase in volume orders results in frequent shipping.

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:

```
with delivery as(
select
  distinct o.order_id,
  c.customer_state,
  o.order_purchase_timestamp,
  o.order_estimated_delivery_date,
  o.order_delivered_customer_date,
  date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp, day)
as time_to_deliver,
  date_diff(o.order_estimated_delivery_date,o.order_delivered_customer_date,
day) as diff_estimated_delivery
from target.customers c
      join target.orders o on c.customer_id=o.customer_id
      join target.order_items oi on o.order_id =oi.order_id
where order_status='delivered'
)
select
  customer_state,
  round(avg(time_to_deliver),2) as Avg_Delivery_Time,
  round(avg(diff_estimated_delivery),2) as Avg_Estimated_Delivery_diff_Time
from delivery
group by customer_state
order by 3 desc
limit 5
```

Result:

Row	customer_state	Avg_Delivery_Time	Avg_Estimated_Delivery_diff_Time
1	AC	20.64	19.76
2	RO	18.91	19.13
3	AP	26.73	18.73
4	AM	25.99	18.61
5	RR	28.98	16.41

Insights:

- The top 5 states with fastest delivery compared to estimated delivery time are the states with least number of orders. Also, all these states have highest estimated delivery time probably because of not frequent shipping due to less volume but still all these orders are being delivered way in advance than the Estimated delivery.
- States like SP, PR, and MG shows efficiency in both average delivery time and estimation accuracy.

Recommendations: Delivery estimation mechanism should be updated to consider all the factors affecting delivery times, such as location of delivery, logistics planning. This can improve the accuracy of delivery time estimates.

VI. Analysis based on the payments:

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

Query:

```
select
  p.payment_type,
  FORMAT_DATE('%Y-%m', o.order_purchase_timestamp) as Month_Year,
  count(distinct o.order_id) as Total_Orders
from target.payments p join target.orders o
  on o.order_id=p.order_id
group by 1,2
order by 2,3,1;
```

Result:

Row	payment_type	Month_Year	Total_Orders
1	credit_card	2016-09	3
2	debit_card	2016-10	2
3	voucher	2016-10	11
4	UPI	2016-10	63
5	credit_card	2016-10	253
6	credit_card	2016-12	1
7	debit_card	2017-01	9
8	voucher	2017-01	33
9	UPI	2017-01	197
10	credit_card	2017-01	582

Insights:

- Payments with credit card has the highest number of total orders across all months. This suggests that credit cards are the preferred payment method for majority of customers.
- UPI Payments has shown a growing trend overall.
- There were some spikes in Voucher payment type could have been associated with discounts on this type of payment.

Recommendations:

- Credit card payments have the highest number of total orders, considering this business can implement discounts, cashback offers or reward points exclusively for credit card payments.
- UPI shows a growing trend in total orders, ensure a seamless UPI integration to make payments hassle free.

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

Query:

```
select
    payment_installments,
    count(distinct order_id) as Total_Orders
from target.payments
where payment_installments>=1
group by 1
order by 2 desc;
```

Result:

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

Insights:

- Almost 50% of the orders have used single installment payment.
- There is a decreasing trend in the number of orders as the installment count increases. It could be comprehended suggesting longer installment plans are less used by customers.
- The number of orders with 10 or more instalments is not significant as it has very less orders.

Recommendations:

- Business should promote the benefits of single payment options to attract customers who prefers immediate payment.
- Should offer installment plans with 2 to 4 instalments, as these seem to be used by customers after immediate payment.