SQL PROJECT

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

1.a.Data type of columns in a table

```sql
select *
from `Target.orders`
```

1.b.Time period for which the data is given

```sql
Select

min(order_purchase_timestamp) as first_order_date,max(order_purchase_timestamp)
as last_order_date

from `Target.orders`
```

Row	first_order_date ▼	last_order_date ▼
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

1.c.Cities and States of customers ordered during the given period

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

Row	city ▼	state ▼
1	4119	27

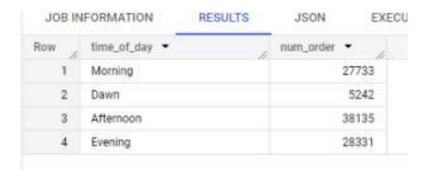## 2.In-depth Exploration:

1. Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?
2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

```sql
select
  extract(year from order_purchase_timestamp) as year,
  extract(month from order_purchase_timestamp) as month,
  count (*) as num_orders
from `Target.orders`
group by month,year
```

	A	B	C
1	year	month	num_orders
2	2017	11	7544
3	2017	12	5673
4	2018	2	6728
5	2017	4	2404
6	2017	7	4026
7	2018	5	6873
8	2017	10	4631
9	2018	1	7269
10	2017	6	3245

```sql
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 24 then
'Evening'
  end as time_of_day,
  count (order_id) as num_order
from `Target.orders`
group by time_of_day
```


Row	time_of_day ▼	num_order ▼
1	Morning	27733
2	Dawn	5242
3	Afternoon	38135
4	Evening	28331

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

1. Get month on month orders by states
2. Distribution of customers across the states in Brazil

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

	year	month	num_orders
1			
2	2017	11	7544
3	2017	12	5673
4	2018	2	6728
5	2017	4	2404
6	2017	7	4026
7	2018	5	6873
8	2017	10	4631
9	2018	1	7269
10	2017	6	3245

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

1. Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use "payment_value" column in payments table
2. Mean & Sum of price and freight value by customer state

```sql
with cte as
(select
sum(case
when extract(year from order_purchase_timestamp)=2017 and
extract(month from order_purchase_timestamp)between 1 and 8
then payment_value
else 0
end)as cost_of_2017,
sum(case
when extract(year from order_purchase_timestamp)=2018 and
extract(month from order_purchase_timestamp)between 1 and 8
then payment_value
else 0
end)as cost_of_2018
from `Target.orders` o
join `Target.payments` p
on o.order_id=p.order_id)

select
 ((cost_of_2018-cost_of_2017)/cost_of_2017)*100 as
 perc_of_inc from cte
```

Row	perc_of_inc ▼
1	136.9768716466...

5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery 2. Find time_to_delivery & diff_estimated_delivery. Formula for the same given
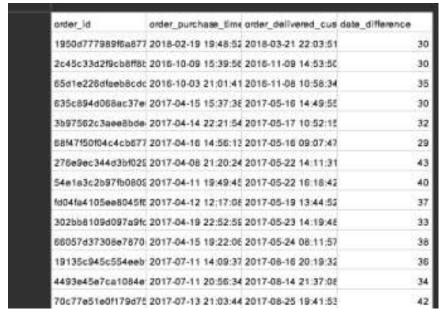
below:

- ○ time_to_delivery = order_delivered_customer_date order_purchase_timestamp

    diff_estimated_delivery = order_estimated_delivery_date order_delivered_customer_date
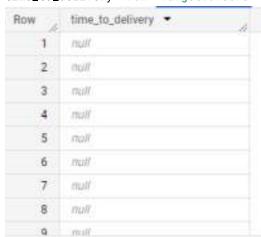
```sql
select
o.order_id,order_purchase_timestamp,order_delivered_customer_date,date_diff(orde r_delivered_customer_date,order_purchase_timestamp,day) as date_difference
from `Target.orders` o ;
```
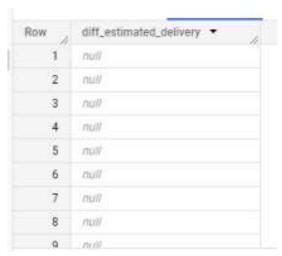

order_id	order_purchase_time	order_delivered_cus	date_difference
1950d777989f6a877	2018-02-19 19:48:52	2018-03-21 22:03:51	30
2c45c33d2f9cb8ff8b	2016-10-09 15:39:58	2016-11-09 14:53:50	30
65d1e228dfeeb8cdc	2016-10-03 21:01:41	2016-11-08 10:58:34	35
635c894d068ac37e	2017-04-15 15:37:38	2017-05-16 14:49:55	30
3b97562c3aee8bde	2017-04-14 22:21:54	2017-05-17 10:52:15	32
68f47150f04c4cb677	2017-04-16 14:56:12	2017-05-16 09:07:47	29
278e9ec344d3bf025	2017-04-08 21:20:24	2017-05-22 14:11:31	43
54e1a3c2b97fb080S	2017-04-11 19:49:45	2017-05-22 16:18:42	40
fd04fe4105ee6045f0	2017-04-12 12:17:08	2017-05-19 13:44:52	37
302bb8109d097a9fc	2017-04-19 22:52:59	2017-05-23 14:19:48	33
68057d37308e7870	2017-04-15 19:22:06	2017-05-24 08:11:57	38
19135c945c554eeb	2017-07-11 14:09:37	2017-08-16 20:19:32	36
4493e45e7ca1084e	2017-07-11 20:56:34	2017-08-14 21:37:08	34
70c77e51e0f179d7S	2017-07-13 21:03:44	2017-08-25 19:41:52	42

```sql
select
(order_delivered_customer_date-order_purchase_timestamp) as
time_to_delivery from `Target.orders`
```

Row	time_to_delivery ▼
1	null
2	null
3	null
4	null
5	null
6	null
7	null
8	null
9	null

```sql
select
(order_estimated_delivery_date-order_delivered_customer_date)
as diff_estimated_delivery
from `Target.orders`
```

Row	diff_estimated_delivery ▼
1	null
2	null
3	null
4	null
5	null
6	null
7	null
8	null
9	null

6. Payment type analysis:

6.a.Month over Month count of orders for different payment types

```sql
SELECT
    extract(year from o.order_purchase_timestamp) AS year,
    extract(month from o.order_purchase_timestamp) AS
    month, p.payment_type,
    COUNT(*) AS order_count
FROM
    `Target.orders` o
JOIN
    `Target.payments` p ON o.order_id = p.order_id
GROUP BY
    year,month, p.payment_type
ORDER BY
    year,month, p.payment_type;
```

	year	month	payment_type	order_count
1	year	month	payment_type	order_count
2	2016	9	credit_card	3
3	2016	10	UPI	63
4	2016	10	credit_card	254
5	2016	10	debit_card	2
6	2016	10	voucher	23
7	2016	12	credit_card	1
8	2017	1	UPI	197
9	2017	1	credit_card	583
10	2017	1	debit_card	9

6.b.Count of orders based on the no. of payment installments

```sql
SELECT
    p.payment_installments,
    COUNT(*) AS order_count
```

```
FROM
    `Target.payments` p
GROUP BY
    p.payment_installments
ORDER BY
    p.payment_installments;
```

	payment_installmen	order_count
1		
2	0	2
3	1	52546
4	2	12413
5	3	10461
6	4	7098
7	5	5239
8	6	3920
9	7	1626
10	8	4266

Based on the analysis of the dataset and the insights derived, here are some actionable insights and recommendations:

Growing trend of e-commerce in Brazil:

The analysis of month-on-month orders indicates a growing trend in e-commerce in Brazil. If the order volume is consistently increasing over time, it suggests a positive outlook for the e-commerce industry. To leverage this trend, businesses can focus on expanding their online presence, improving customer experience, and investing in marketing strategies to capture a larger share of the growing e-commerce market.
Seasonality with peaks at specific months:

If there are noticeable peaks in order volume at specific months, it suggests the presence of seasonality in e-commerce. Businesses can capitalize on these peak periods by planning targeted marketing campaigns, offering special promotions or discounts, and optimizing inventory and logistics to meet the increased demand. Understanding the seasonality patterns can help in resource allocation and planning for effective sales and operations management.
Time preferences of Brazilian customers:
Analysing the time preferences of Brazilian customers for making purchases can help businesses optimize their operations, customer support, and marketing efforts. By understanding the preferred time periods (dawn, morning, afternoon, or night) when customers tend to buy, businesses can align their staffing and customer service availability accordingly. Additionally, targeted promotions or personalized offers during specific time periods can enhance customer engagement and drive sales.
Money movement and cost analysis:

Analysing the money movement in e-commerce, including order prices, freight values, and cost increases, provides valuable insights for businesses. Based on the findings, businesses can optimize pricing strategies, identify opportunities for cost savings in logistics and operations, and evaluate the impact of pricing changes on customer behaviour and overall revenue. It also helps in understanding the overall economic impact of e-commerce on the local and national economy.

Mean and sum of price and freight value by customer state:

The analysis of mean and sum of price and freight value by customer state provides insights into the spending patterns and preferences of customers across different states. This information can be used for targeted marketing and sales strategies, such as tailoring product offerings, pricing, and promotional campaigns based on regional preferences. It can also help in optimizing logistics and supply chain management by identifying high-value regions and addressing any disparities in pricing or freight costs.

Geographic expansion and customer acquisition:

Understanding the distribution of customers across states in Brazil can inform business expansion plans and customer acquisition strategies. Businesses can prioritize regions with a high concentration of customers and focus on targeted marketing and localized initiatives to penetrate new markets. Additionally, analysing customer demographics, preferences, and behaviour within each state can help tailor marketing messages and offerings to effectively engage and convert potential customers.

Competitive analysis and benchmarking:

Benchmarking against industry peers and competitors can provide valuable insights into market trends, pricing strategies, customer satisfaction levels, and service quality. By analysing the dataset in comparison to industry benchmarks, businesses can identify areas for improvement, innovate in product offerings or customer experience, and stay competitive in the dynamic e-commerce landscape.

These actionable insights and recommendations can guide businesses in optimizing their strategies, enhancing customer experiences, and driving growth in the e-commerce sector in Brazil. However, it's important to further analyse the specific business context and combine these insights with additional market research for more targeted and customized recommendations.

Average Freight Value by State:

The top 5 states with the highest average freight value should be targeted for potential pricing optimizations or negotiation with logistics partners to reduce shipping costs.
The top 5 states with the lowest average freight value can be analysed to identify potential areas for cost optimization or efficiency improvements in the shipping process. Average Time to Delivery by State:

The top 5 states with the highest average time to delivery should be closely monitored to

identify any bottlenecks or inefficiencies in the order fulfilment and delivery process. Efforts can be made to improve logistics operations, optimize delivery routes, or collaborate with carriers to expedite deliveries.

The top 5 states with the lowest average time to delivery can be analysed to identify best practices and strategies that contribute to faster delivery times. These insights can be applied to other regions to improve overall delivery performance.

Delivery Performance compared to Estimated Date:

The top 5 states where delivery is exceptionally fast compared to the estimated date can be promoted to customers, highlighting the efficient delivery service.

The top 5 states where delivery is not as fast compared to the estimated date should be examined to identify potential issues causing delays. Strategies such as better inventory management, proactive communication with customers, or improved coordination with logistics partners can be implemented to address the delays and enhance customer satisfaction.

Payment Type Analysis:

Analysing the count of orders for different payment types month over month can help identify trends and preferences among customers. This information can be used to tailor marketing campaigns, offers, and payment options to align with customer preferences. Understanding the count of orders based on the number of payment instalments can provide insights into customers' financing preferences and purchase behaviour. It can be useful for optimizing payment options, partnering with financial institutions for instalment plans, or designing targeted promotional strategies.

Overall, the actionable insights and recommendations aim to optimize freight costs, improve delivery performance, align payment options with customer preferences, and enhance the overall customer experience. Regular monitoring and analysis of these metrics will help identify opportunities for improvement and drive business growth.