TARGET BUSINESS CASE STUDY

Context:

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 particular business case focuses on the operations of **Target in Brazil** and provides insightful information about 100,000 orders placed between 2016 and 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 analyzing 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.

Problem Statement:

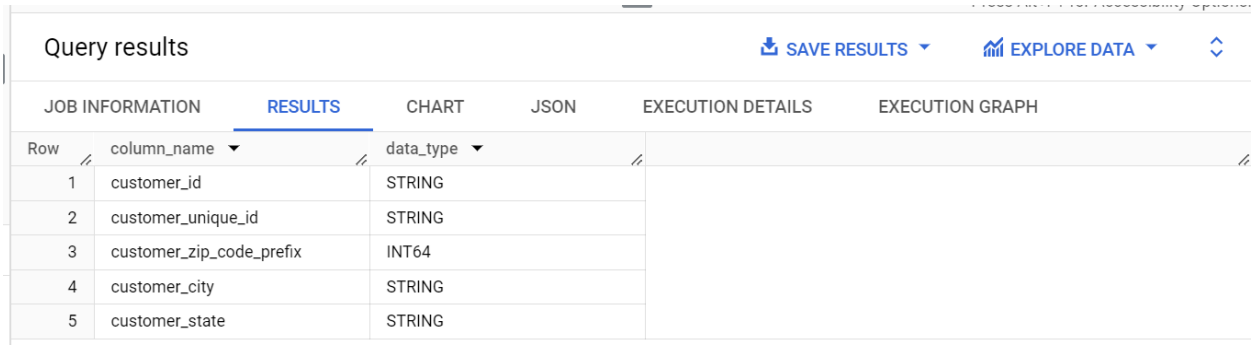
Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

What does 'good' look like?

1. **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.

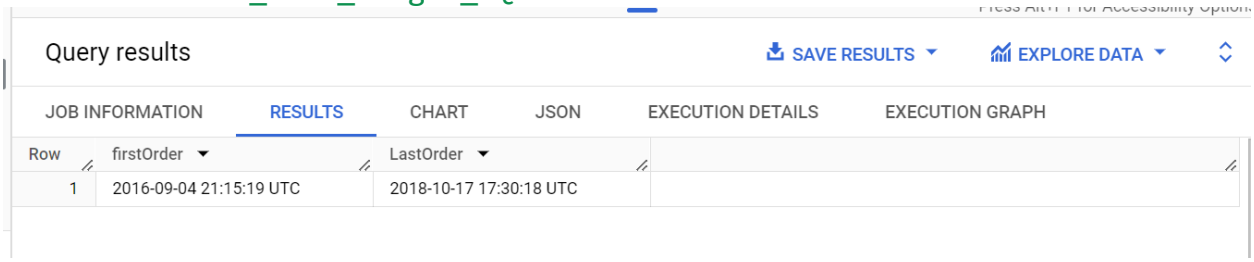
```
SELECT column_name, data_type
FROM `Business_Case_Target_SQL.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers'
```



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

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

```
SELECT min(order_purchase_timestamp) as first_order,
       max(order_purchase_timestamp) as Last_order
FROM `Business_Case_Target_SQL.orders`
```



Insights:-

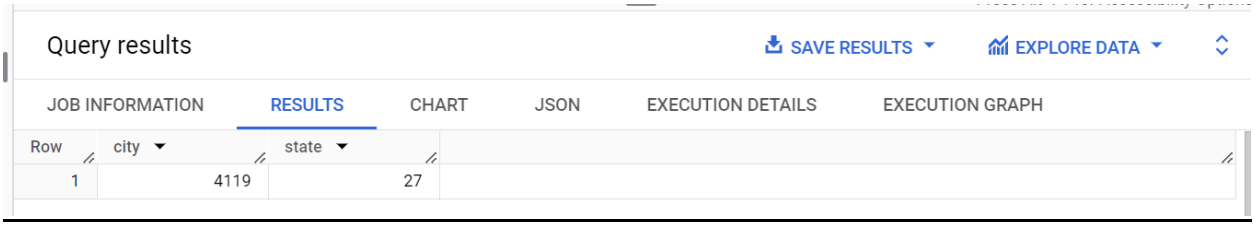
Earliest Order (first_order): The earliest order in the table was placed on September 4, 2016, at 21:15:19 UTC. This is the starting point of the time range.

Latest Order (Last_order): The most recent order in the table was placed on October 17, 2018, at 17:30:18 UTC. This represents the endpoint of the time range.

These insights are valuable for understanding the temporal distribution of orders within dataset. We can use this information to answer questions about the overall duration of the dataset, track changes in ordering behavior over time, or filter and analyze orders within specific date ranges.

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

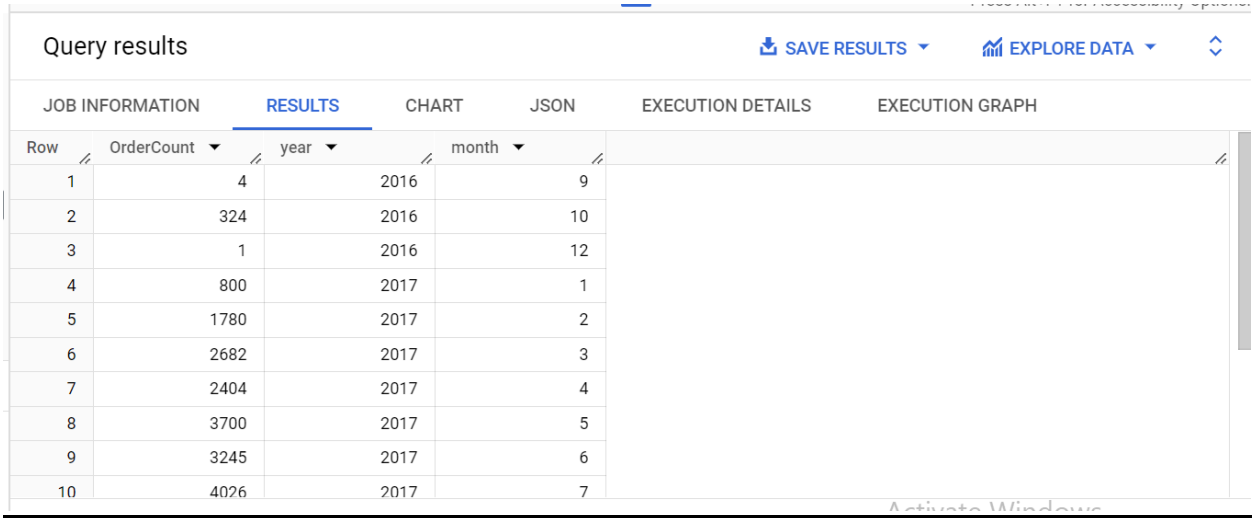
```
select count(distinct (C.customer_city)) as city,
       count(distinct(C.customer_state)) as state
from `Business_Case_Target_SQL.customers` C
inner join `Business_Case_Target_SQL.orders` O
on C.customer_id = O.customer_id
```



2. In-depth Exploration:

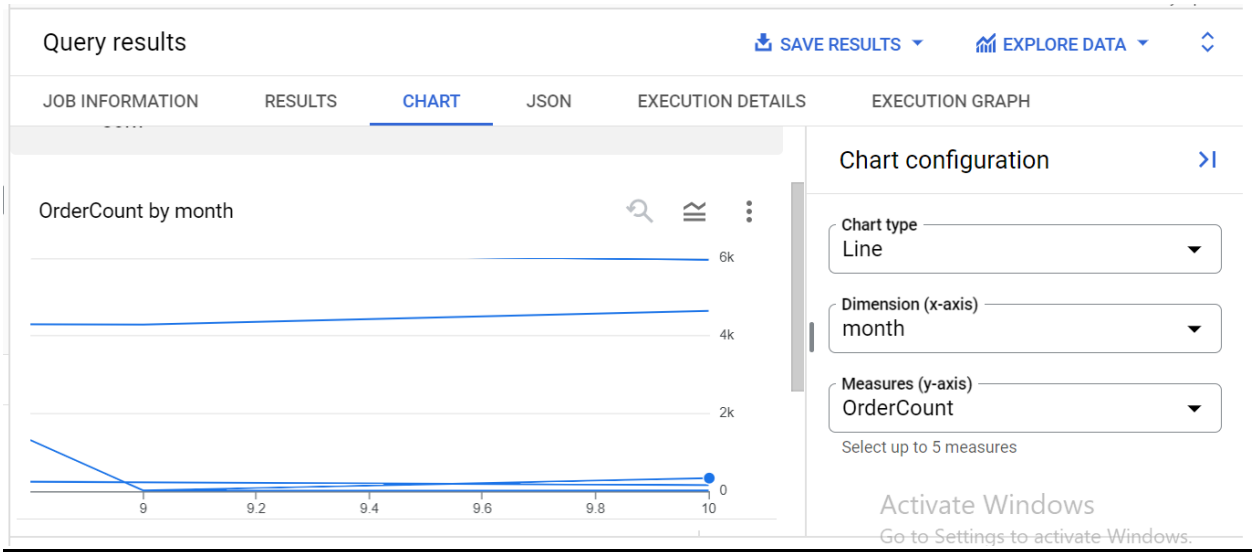
A. Is there a growing trend in the no. of orders placed over the past years?

```
select count(distinct(O.order_id)) as OrderCount,
extract(year from O.order_purchase_timestamp) as year,
extract(month from O.order_purchase_timestamp) as month
from `Business_Case_Target_SQL.orders` O
inner join `Business_Case_Target_SQL.customers` C
on O.customer_id = C.customer_id group by 2, 3 order by 2, 3 asc
```



Row	OrderCount	year	month
1	4	2016	9
2	324	2016	10
3	1	2016	12
4	800	2017	1
5	1780	2017	2
6	2682	2017	3
7	2404	2017	4
8	3700	2017	5
9	3245	2017	6
10	4026	2017	7

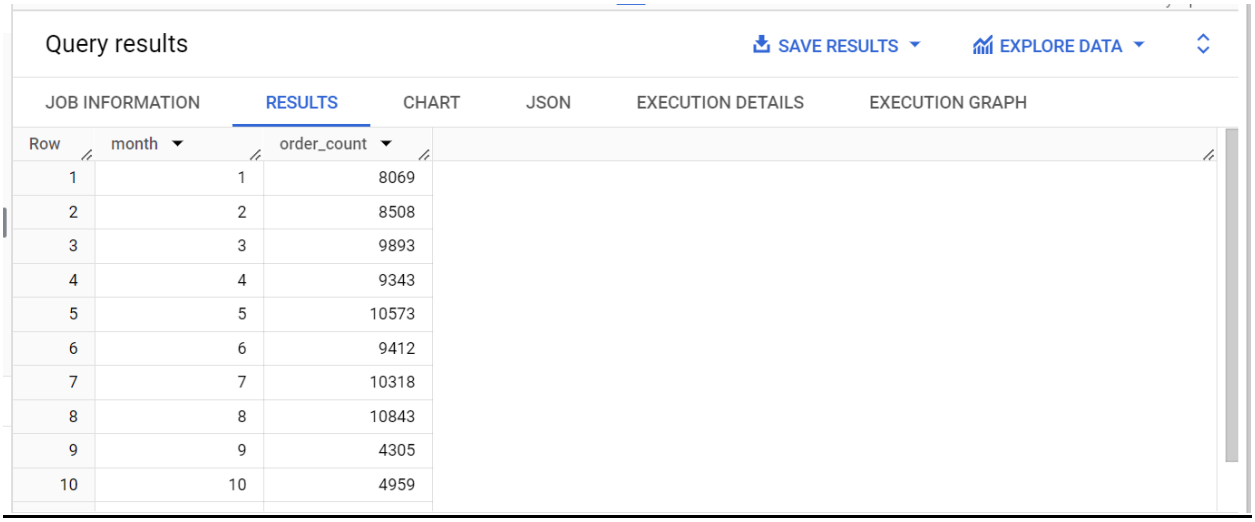
Order Count by Month Graph



Based on the analysis, there is a growing trend of number of orders over the past years.

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

```
SELECT EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
      COUNT(DISTINCT order_id) AS order_count
FROM `Business_Case_Target_SQL.orders` GROUP BY month ORDER BY month
```



Row	month	order_count
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959

Insights:-

PEAK SEASON (MAY TO AUGUST):-

- MAY to AUGUST represent the peak season for orders, with counts consistently exceeding 9000 orders per month.
- The highest order count is in August, reaching over 10843 orders.

OFF-PEAK SEASON (SEPTEMBER TO APRIL):-

- September to April shows a decline in order counts, with September having the lowest count at 4,305 orders.
- December is the last month of the year and has a moderate order count compared to the peak season.

Action Plan:

Peak Season Strategies:

- During peak season, ensure that you have adequate inventory to meet the high demand. Work closely with suppliers and use demand forecasting to prevent stock outs.
- Allocate a significant portion of your marketing budget to these months. Launch special promotions, discounts, and loyalty programs to attract and retain customers.
- Increase customer support capacity to handle higher inquiries and ensure a seamless shopping experience.

Off-Peak Season Strategies:

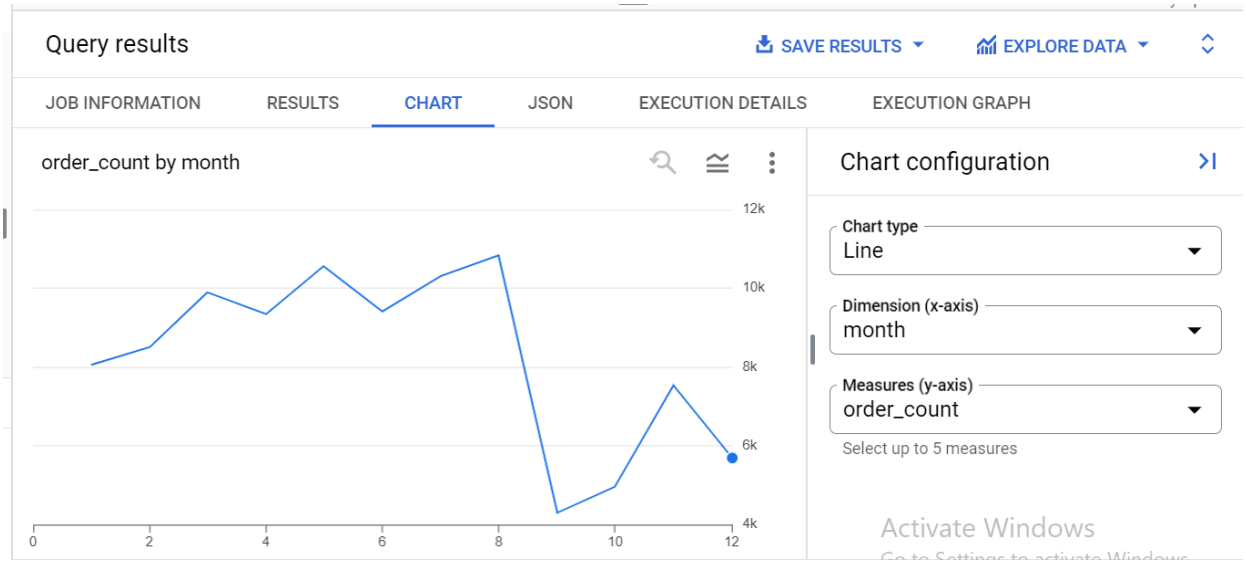
- During off-peak months, focus on optimizing inventory turnover and reducing carrying costs. Consider offering clearance sales to clear out old inventory.
- Run targeted marketing campaigns to incentivize purchases during off-peak months. Highlight unique selling points and create a sense of urgency.
- Implement customer engagement initiatives such as loyalty programs, referral bonuses, and personalized recommendations to maintain customer interest.

Year-End Strategies (December):

- Leverage the holiday season to boost sales with themed promotions and gift ideas.
- Plan early for the holiday season by expanding product offerings and ensuring timely delivery.

By implementing these strategies and actions, your company can better manage seasonality, optimize operations, and create a more balanced and sustainable revenue stream throughout the year.

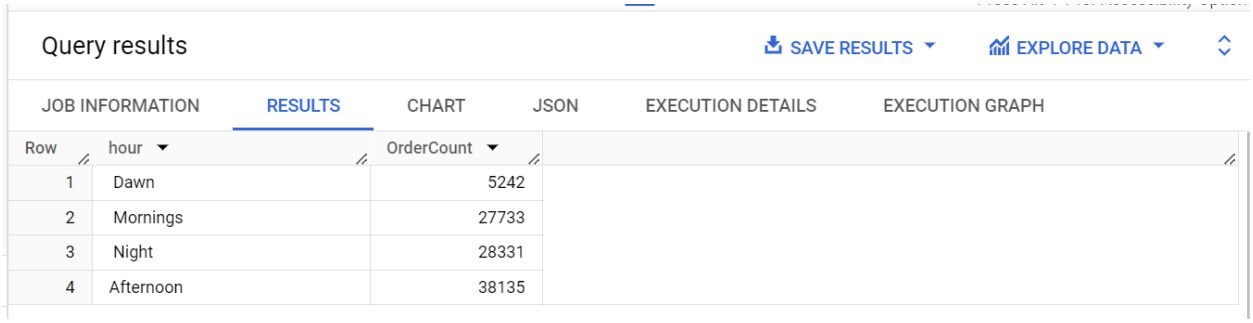
Order_count by month chart



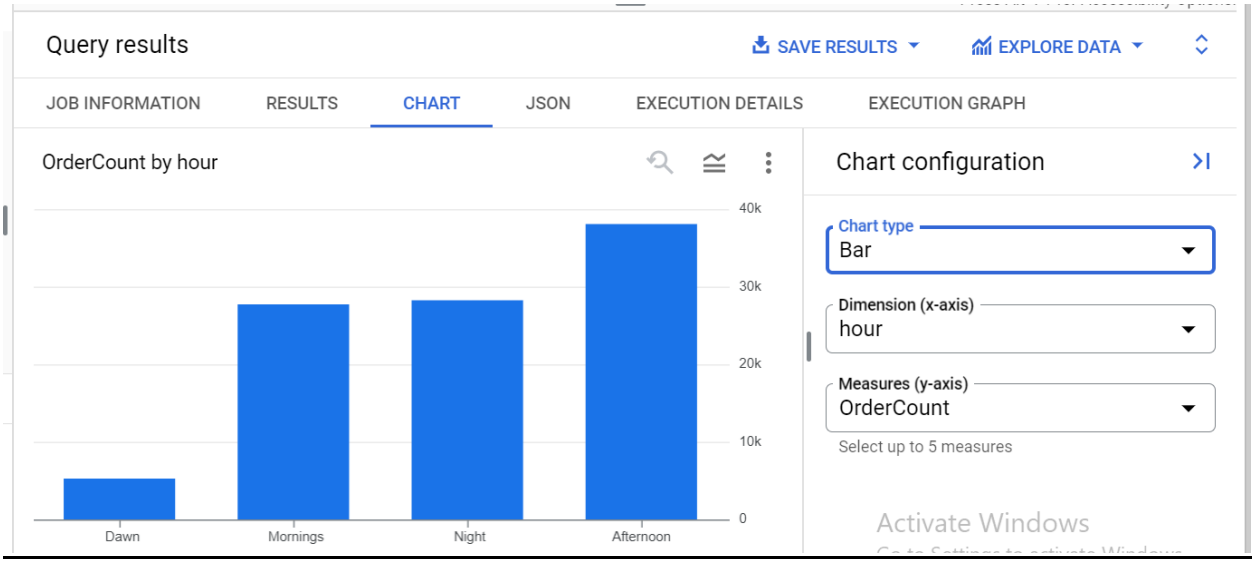
C. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

- 0-6 hrs. : Dawn
- 7-12 hrs. : Mornings
- 13-18 hrs. : Afternoon
- 19-23 hrs. : Night
- 0-6 hrs. : Dawn
- 7-12 hrs. : Mornings
- 13-18 hrs. : Afternoon
- 19-23 hrs. : Night

```
select
 case
      when extract(hour from O.order_purchase_timestamp) between 0 and
6 then ' Dawn'
      when extract(hour from O.order_purchase_timestamp) between 7 and
12 then ' Mornings'
     when extract(hour from O.order_purchase_timestamp) between 13 and
18 then 'Afternoon'
     when extract(hour from O.order_purchase_timestamp) between 19 and
23 then ' Night'
end as hour,
   count(O.order_id) as OrderCount
from `Business_Case_Target_SQL.orders` O join
`Business_Case_Target_SQL.customers` C
      on O.customer_id = C.customer_id
group by hour
order by OrderCount
```



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



Insights:

Afternoon Dominance: The majority of orders are placed during the afternoon, with 38,135 orders. This is the busiest time of day for orders.

Morning and Night Orders: Morning and night orders are nearly equal, with mornings having 27,733 orders and nights having 28,331 orders.

Dawn Orders: Dawn sees the least number of orders, with 5,242 orders. This is the quietest time for order placement.

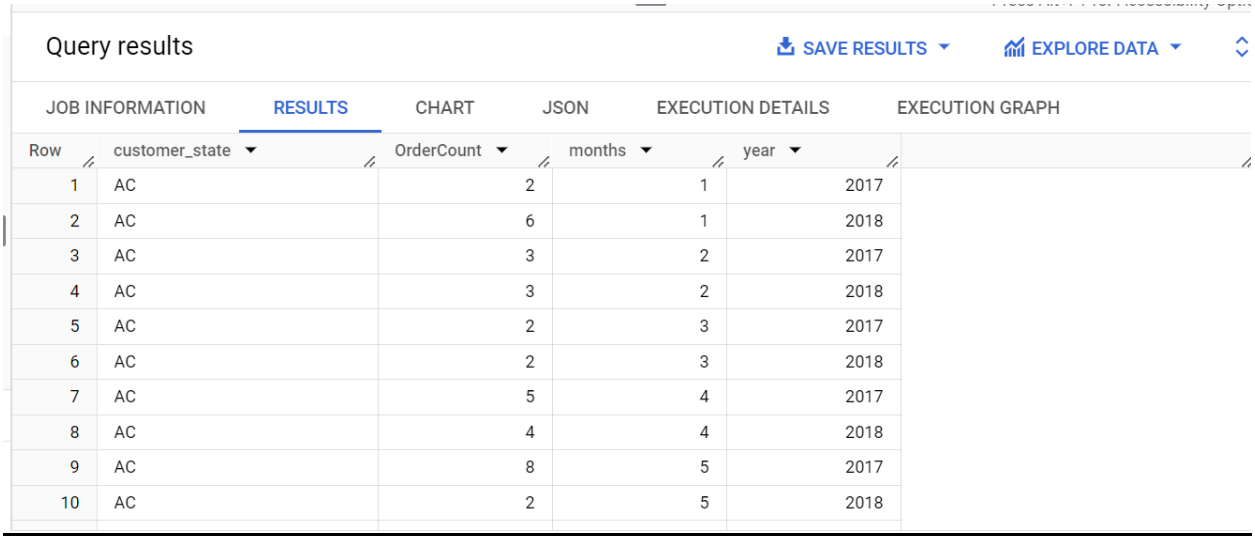
Action Plan:

- Allocate more customer support and warehouse resources during the afternoon, as it's the peak order placement time.
- Adjust staff schedules to align with order patterns. Have a skeleton crew during dawn and optimize staffing during the afternoons.
- Schedule marketing campaigns and promotions to coincide with peak ordering hours in the afternoon.
- Consider running targeted morning and evening promotions to encourage orders during these times.
- Optimize order delivery scheduling to ensure timely deliveries during peak hours.
- For dawn orders, prioritize early morning deliveries to enhance customer satisfaction.
- Implement chat support or automated systems for customer inquiries during off-peak hours.
- Provide incentives such as discounts or loyalty points for customers who place orders during quieter times.

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

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

```
select C.customer_state,
      count(O.order_id) as OrderCount,
      extract(month from O.order_purchase_timestamp) as months,
      extract(year from O.order_purchase_timestamp) as year
from `Business_Case_Target_SQL.orders` O join
`Business_Case_Target_SQL.customers` C
      on O.customer_id = C.customer_id
group by C.customer_state,months,year
order by C.customer_state,months,year
```



Row	customer_state	OrderCount	months	year
1	AC	2	1	2017
2	AC	6	1	2018
3	AC	3	2	2017
4	AC	3	2	2018
5	AC	2	3	2017
6	AC	2	3	2018
7	AC	5	4	2017
8	AC	4	4	2018
9	AC	8	5	2017
10	AC	2	5	2018

Insights:

- The query successfully provides the monthly order counts for each state in Brazil.
- There are variations in the number of orders from month to month for each state, indicating potential seasonal trends or changes in consumer behavior.

Action Plan:

- Analyze the monthly variations to identify patterns and seasonality in different states. Use this data to plan and launch targeted promotions and marketing campaigns during high-order months to boost sales during lower-order months.
- Consider expanding your reach to states that show consistent growth in orders, potentially opening new distribution centers or marketing specifically to those regions.

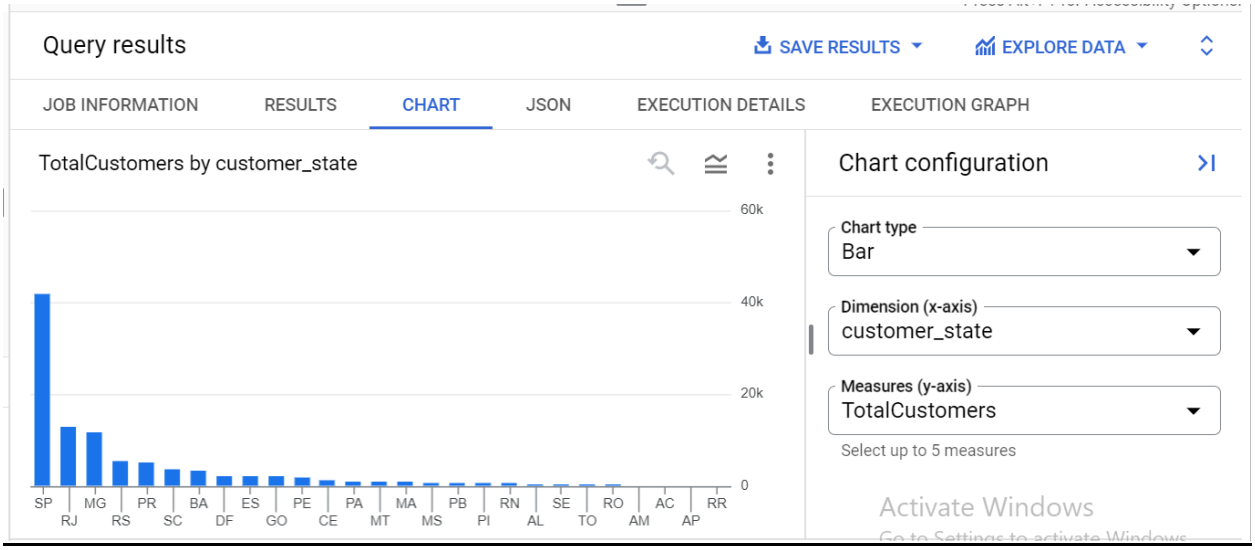
B. How are the customers distributed across all the states?

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

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

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

TotalCustomers by customer_state Chart



Insights:

1. Customer Distribution: The query successfully provides the count of customers in each state.

2. State Variations: There are significant variations in the number of customers across different states. Some states have a higher concentration of customers, while others have relatively fewer. State SP have higher concentration of customers (41746).

Action Plan:

- Consider increasing marketing efforts in states with lower customer counts to attract more customers and expand your customer base.
- Tailor marketing campaigns to the specific preferences and needs of customers in each state, taking into account regional variations.
- Optimize supply chain and logistics operations in regions with higher customer counts to meet the demand efficiently.
- Keep an eye on competitors' activities and customer acquisition strategies in different states to stay competitive.

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).

```
SELECT EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    ((SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
2018 AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND
8 THEN p.payment_value END)
    -
    SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017
AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
THEN p.payment_value END))
    /
      SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
      2017 AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1
      AND 8 THEN p.payment_value END)) * 100 AS percent_increase
FROM `Business_Case_Target_SQL.orders` o JOIN
`Business_Case_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
GROUP BY 1 ORDER BY 1
```

Query results

SAVE RESULTS | EXPLORE DATA

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

Row	month	percent_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...

percent_increase by month Chart

Insights:

- January experienced an extraordinary increase in order costs with a whopping 705.13%. This is significantly higher than the subsequent months.
- There is a clear pattern of increasing order costs from January (705.13%) to April (177.84%), followed by a gradual decrease in percentage increase from May (94.63%) to August (51.61%).

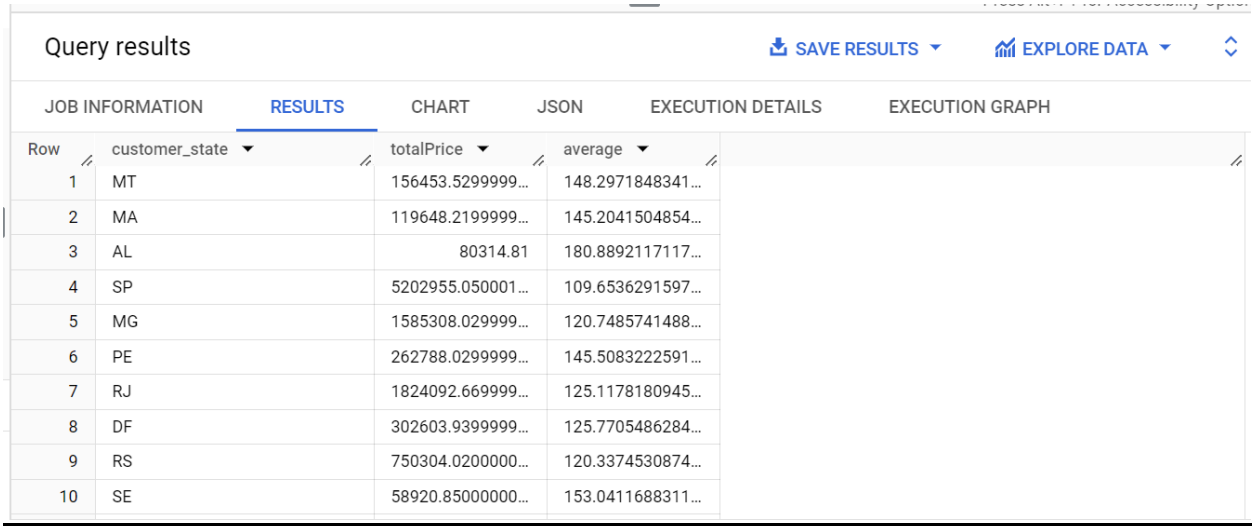
Action Plan:

- The massive increase in January needs immediate attention. Investigate the specific reasons for such a dramatic spike. It could be attributed to special promotions, clearance sales, or other factors. Analyze if this spike can be repeated in subsequent years.
- Examine the factors contributing to the consistent growth in order costs from January to April. Identify whether these factors can be extended or adapted to other months to maintain a steady increase in revenue.

- While the percentage increase decreases from May onwards, it is essential to sustain growth by implementing effective strategies. Consider introducing new products, enhancing customer engagement, or expanding marketing efforts during these months.
- Focus on building customer loyalty throughout the year. Implement loyalty programs or incentives to encourage repeat purchases and maintain steady revenue growth.

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

```
select C.customer_state, sum(O.price) as totalPrice,
avg(O.price) as average
from `Business_Case_Target_SQL.orders` ORD join
`Business_Case_Target_SQL.order_items` O on ORD.order_id = O.order_id
join `Business_Case_Target_SQL.customers` C on ORD.customer_id =
C.customer_id
group by C.customer_state
```

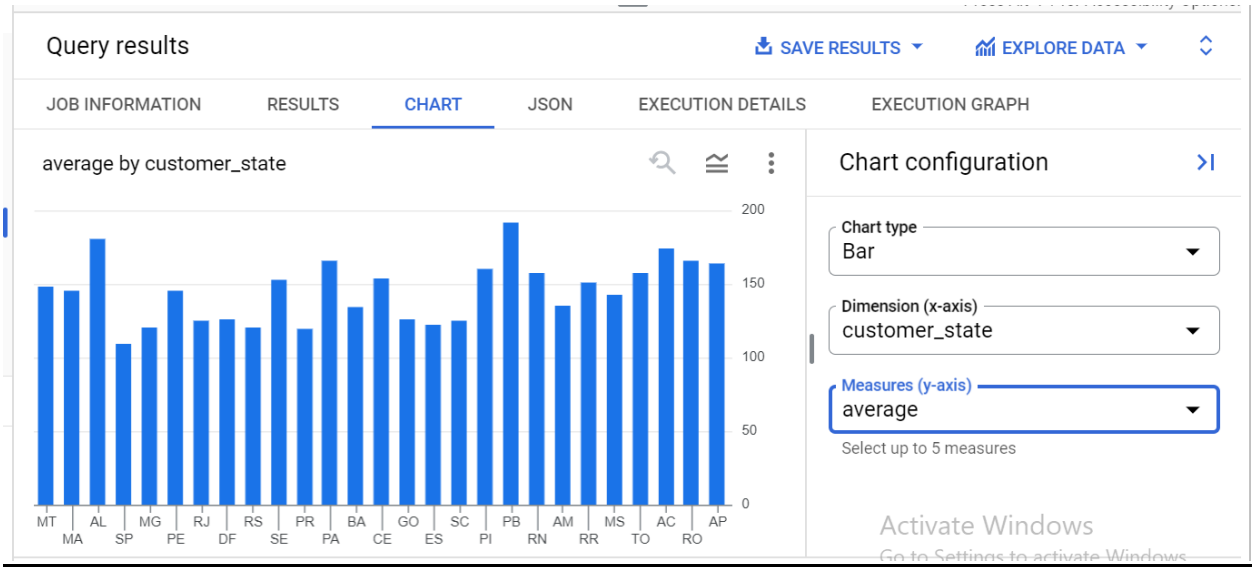


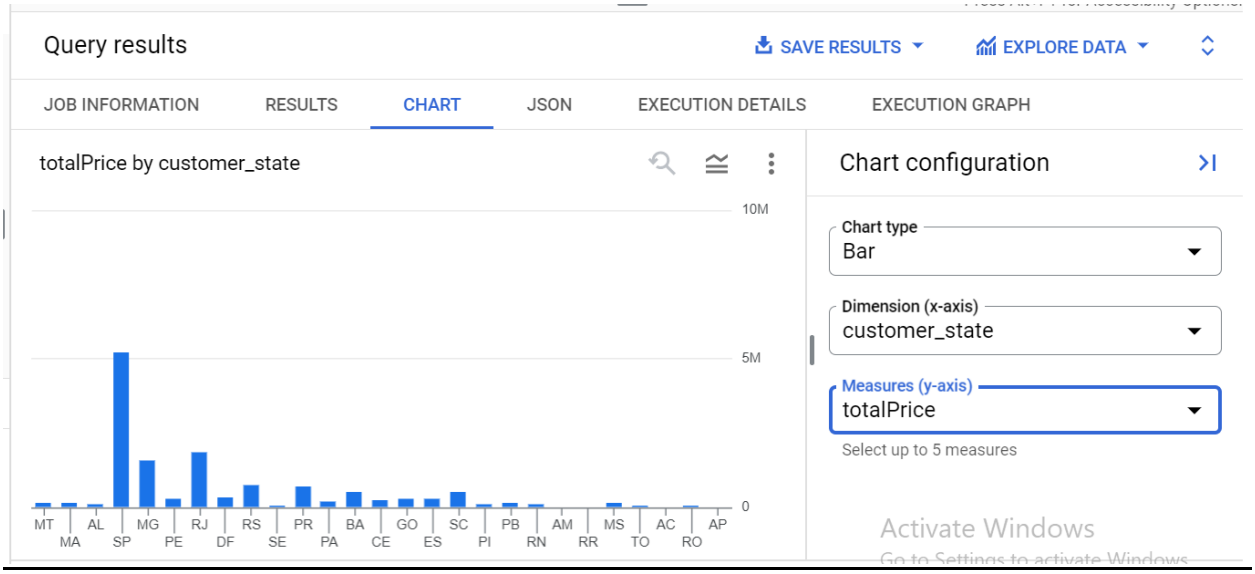
Query results

SAVE RESULTS | EXPLORE DATA

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

Row	customer_state	totalPrice	average
1	MT	156453.5299999...	148.2971848341...
2	MA	119648.2199999...	145.2041504854...
3	AL	80314.81	180.8892117117...
4	SP	5202955.050001...	109.6536291597...
5	MG	1585308.029999...	120.7485741488...
6	PE	262788.0299999...	145.5083222591...
7	RJ	1824092.669999...	125.1178180945...
8	DF	302603.9399999...	125.7705486284...
9	RS	750304.0200000...	120.3374530874...
10	SE	58920.85000000...	153.0411688311...



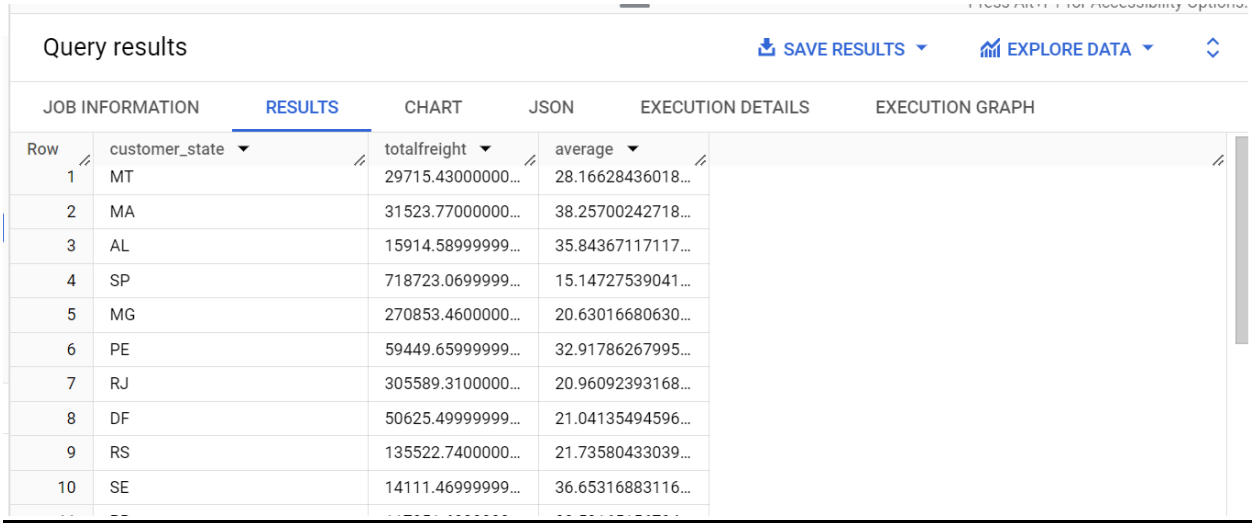


Insights:

- The state of (SP) has the highest total order value.
- Average order values vary across states

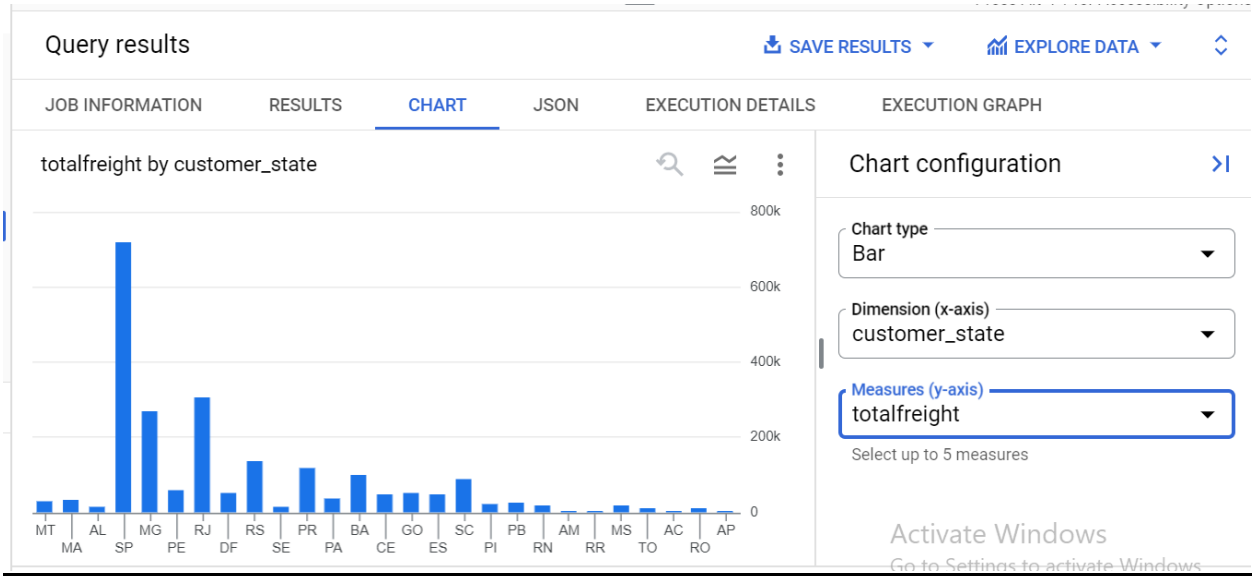
C. Calculate the Total & Average value of order freight for each state.

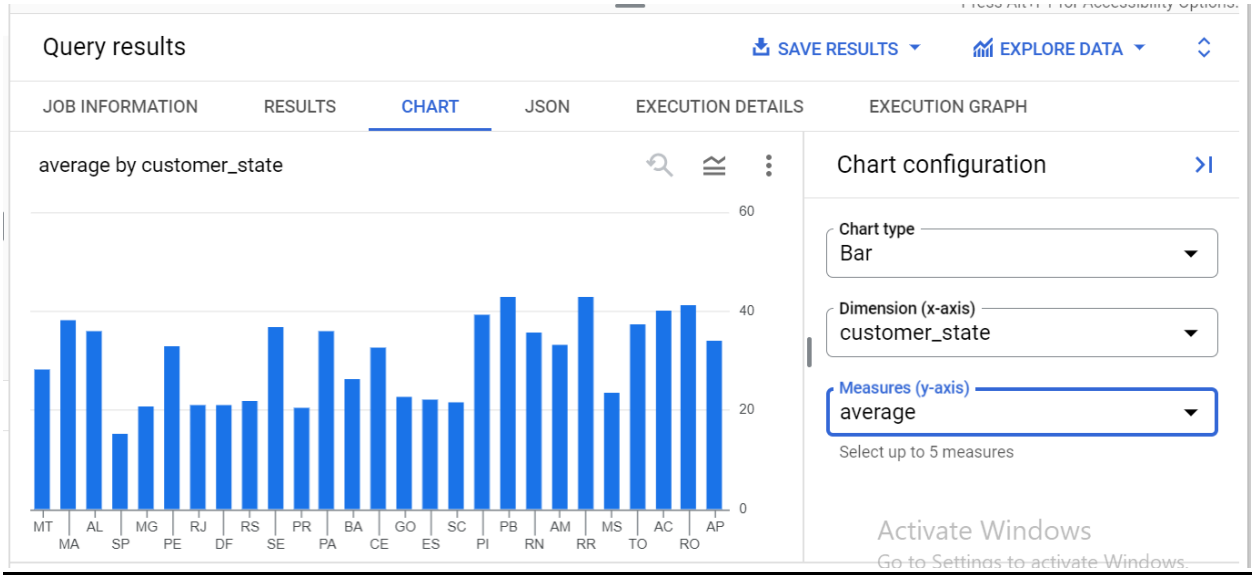
```
select C.customer_state,
sum(i.freight_value) as totalfreight,
avg(i.freight_value) as average
from `Business_Case_Target_SQL.orders` O
join `Business_Case_Target_SQL.order_items` i on O.order_id =
i.order_id
join `Business_Case_Target_SQL.customers` C on O.customer_id =
C.customer_id
group by C.customer_state
```



Row	customer_state	totalfreight	average
1	MT	29715.43000000...	28.16628436018...
2	MA	31523.77000000...	38.25700242718...
3	AL	15914.58999999...	35.84367117117...
4	SP	718723.0699999...	15.14727539041...
5	MG	270853.4600000...	20.63016680630...
6	PE	59449.65999999...	32.91786267995...
7	RJ	305589.3100000...	20.96092393168...
8	DF	50625.49999999...	21.04135494596...
9	RS	135522.7400000...	21.73580433039...
10	SE	14111.46999999...	36.65316883116...

Totalfreight by customer_state chart





Insights:

- Variation in Total Order Prices: The total order prices vary significantly across different states, (SP) having the highest total order price
- While (SP) has the highest total price value and total freight value, it surprisingly has the lowest average price value and average freight value among all states. On the other hand, the state of (PB) has the highest average price value and average freight 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.

```
select order_id,
date_diff(order_delivered_customer_date, order_purchase_timestamp,
day) as time_to_deliver,
date_diff(order_estimated_delivery_date,order_purchase_timestamp,day)
as EstimatedDeliveryDays,
date_diff(order_estimated_delivery_date,order_delivered_customer_date,
day) as diff_estimated_delivery
from `Business_Case_Target_SQL.orders`
where
date_diff(order_delivered_customer_date,order_purchase_timestamp,day)
is not null
```

Query results

SAVE RESULTS | EXPLORE DATA

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

Row	order_id	time_to_deliver	EstimatedDeliveryDa	diff_estimated_delive
1	1950d777989f6a877539f5379...	30	17	-12
2	2c45c33d2f9cb8ff8b1c86cc28...	30	59	28
3	65d1e226dfaeb8cdc42f66542...	35	52	16
4	635c894d068ac37e6e03dc54e...	30	32	1
5	3b97562c3aee8bdedcb5c2e45...	32	33	0
6	68f47f50f04c4cb6774570cfde...	29	31	1
7	276e9ec344d3bf029ff83a161c...	43	39	-4
8	54e1a3c2b97fb0809da548a59...	40	36	-4
9	fd04fa4105ee8045f6a0139ca5...	37	35	-1
10	302bb8109d097a9fc6e9cefc5...	33	28	-5

Insight:

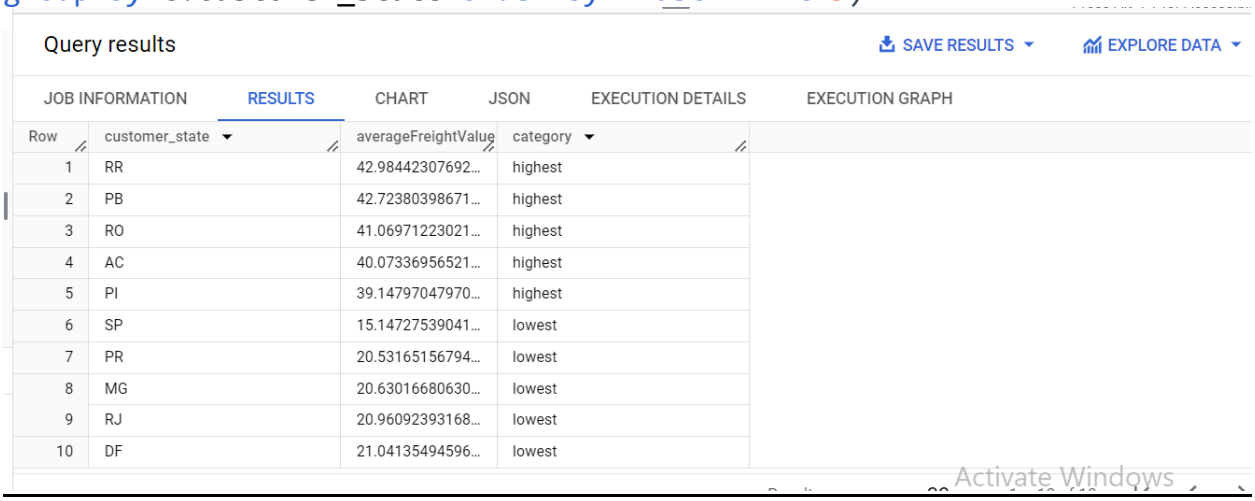
- **Delivery Time:** The time_to_deliver column shows the actual time it took to deliver each order from the purchase date. It varies for different orders, ranging from a few days to several weeks.
- **Estimated vs. Actual Delivery:** The EstimatedDeliveryDays column indicates the estimated number of days for delivery from the purchase date. The diff_estimated_delivery column represents the difference between the estimated and actual delivery days. Negative values indicate orders delivered earlier than estimated, while positive values suggest delays.

Action Plan:

- Analyze orders with longer delivery times to identify bottlenecks or inefficiencies in the delivery process. Streamline logistics and improve delivery speed where possible.
- For orders with significant delays (diff_estimated_delivery > 0), prioritize communication with affected customers. Notify them of the delay, provide reasons, and offer solutions or compensation as necessary to maintain customer satisfaction.
- Use historical data on estimated vs. actual delivery times to enhance delivery forecasting accuracy. Implement predictive models to better estimate delivery dates, reducing the occurrence of delays.
- Ensure that orders are dispatched promptly and that delivery routes are optimized to minimize delays. Regularly evaluate the performance of logistics partners and make improvements where needed.
- Consider implementing tracking and notification systems that allow customers to monitor the status of their orders in real-time. Proactive communication can help manage expectations and reduce inquiries related to delivery.

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

```
(select C.customer_state, avg(i.freight_value) as averageFreightValue,
'highest' as category
from `Business_Case_Target_SQL.orders` O
join `Business_Case_Target_SQL.customers` C on O.customer_id =
C.customer_id
join `Business_Case_Target_SQL.order_items` i on O.order_id =
i.order_id
group by C.customer_state order by 2 desc limit 5)
union all
(select C.customer_state, avg(i.freight_value) as averageFreightValue,
'lowest' as category
from `Business_Case_Target_SQL.orders` O
join `Business_Case_Target_SQL.customers` C on O.customer_id =
C.customer_id
join `Business_Case_Target_SQL.order_items` i on O.order_id =
i.order_id
group by C.customer_state order by 2 asc limit 5)
```



Query results

SAVE RESULTS | EXPLORE DATA

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

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

Insights:

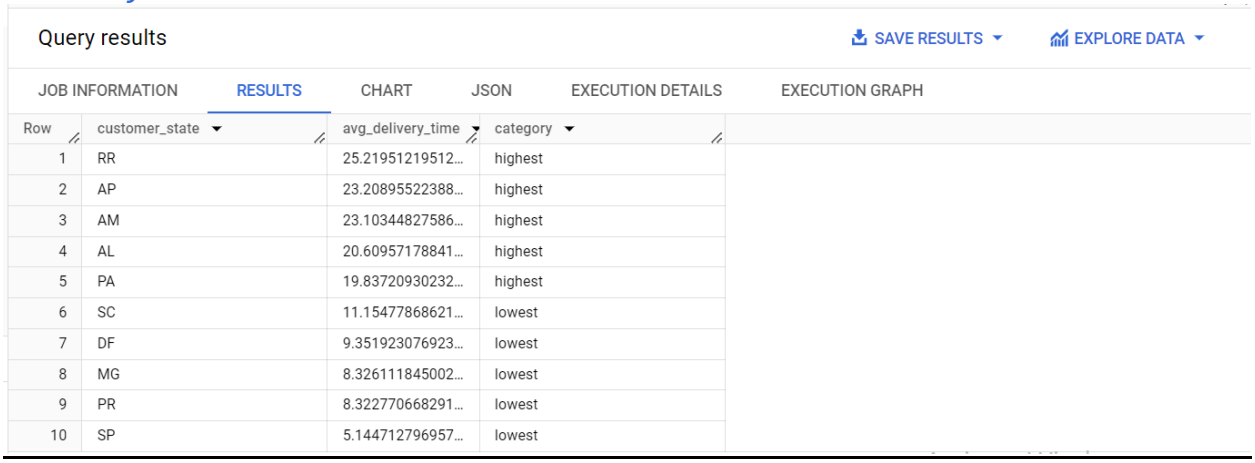
- Highest Average Freight Value States: The states with the highest average freight values are RR , PB , RO , AC , and PI . These states tend to have higher average freight costs for orders, which may be due to factors such as distance, logistics, or other regional considerations.
- Lowest Average Freight Value States: The states with the lowest average freight values are SP, PR, MG, RJ, and DF. These states have relatively lower average freight costs, possibly due to better transportation infrastructure and proximity to distribution centers.

Action Plan:

- High Freight Cost States: For the states with the highest average freight values (RR, PB, RO, AC, and PI), the company should consider optimizing its logistics and supply chain operations. This could involve negotiating better freight rates with carriers, exploring more efficient delivery routes
- Low Freight Cost States: In states with the lowest average freight values (SP, PR, MG, RJ, and DF), the company can focus on maintaining and potentially expanding its customer base. Lower freight costs can be an attractive feature for customers, so marketing efforts could highlight this advantage to attract more orders from these regions.

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

```
(select C.customer_state,
  avg(date_diff(O.order_delivered_customer_date,O.order_delivered_carr
ier_date,day)) as avg_delivery_time,
  'highest' as category
  from `Business_Case_Target_SQL.orders` O
  join `Business_Case_Target_SQL.customers` C on O.customer_id =
C.customer_id
  where O.order_delivered_carrier_date is not null AND
O.order_delivered_customer_date is not null
  group by 1 order by 2 desc limit 5)
union all
(select C.customer_state,
avg(date_diff(O.order_delivered_customer_date,O.order_delivered_carrie
r_date,day)) as avg_delivery_time,
'lowest' as category
from `Business_Case_Target_SQL.orders` O
join `Business_Case_Target_SQL.customers` C on O.customer_id =
C.customer_id
where O.order_delivered_carrier_date is not null AND
O.order_delivered_customer_date is not null
group by 1 order by 2 asc limit 5)
order by 2 desc
```



Query results

SAVE RESULTS | EXPLORE DATA

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

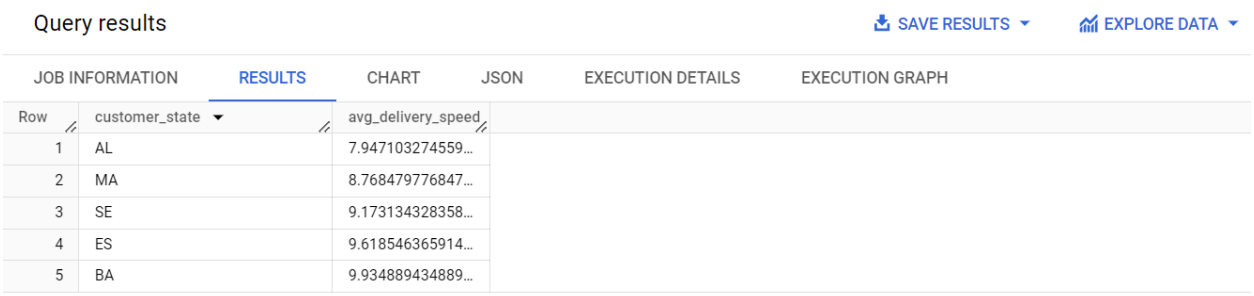
Row	customer_state	avg_delivery_time	category
1	RR	25.21951219512...	highest
2	AP	23.20895522388...	highest
3	AM	23.10344827586...	highest
4	AL	20.60957178841...	highest
5	PA	19.83720930232...	highest
6	SC	11.15477868621...	lowest
7	DF	9.351923076923...	lowest
8	MG	8.326111845002...	lowest
9	PR	8.322770668291...	lowest
10	SP	5.144712796957...	lowest

Insights:

- Highest Average Delivery Times: States such as RR, AP, AM, AL, and PA have the highest average delivery times. This suggests that customers in these states generally experience longer delays in receiving their orders.
- Lowest Average Delivery Times: In contrast, states like SC, DF, MG, PR, and SP have significantly lower average delivery times. Customers in these states experience relatively faster deliveries.

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.

```
select C.customer_state,
avg(date_diff(O.order_estimated_delivery_date,O.order_delivered_custom
er_date,day)) as avg_delivery_speed
from `Business_Case_Target_SQL.orders` O
join `Business_Case_Target_SQL.customers` C on O.customer_id =
C.customer_id
where O.order_estimated_delivery_date is not null AND
O.order_delivered_customer_date is not null
group by 1
having avg(date_diff(O.order_estimated_delivery_date,
O.order_delivered_customer_date,day)) > 0  order by 2 limit 5
```

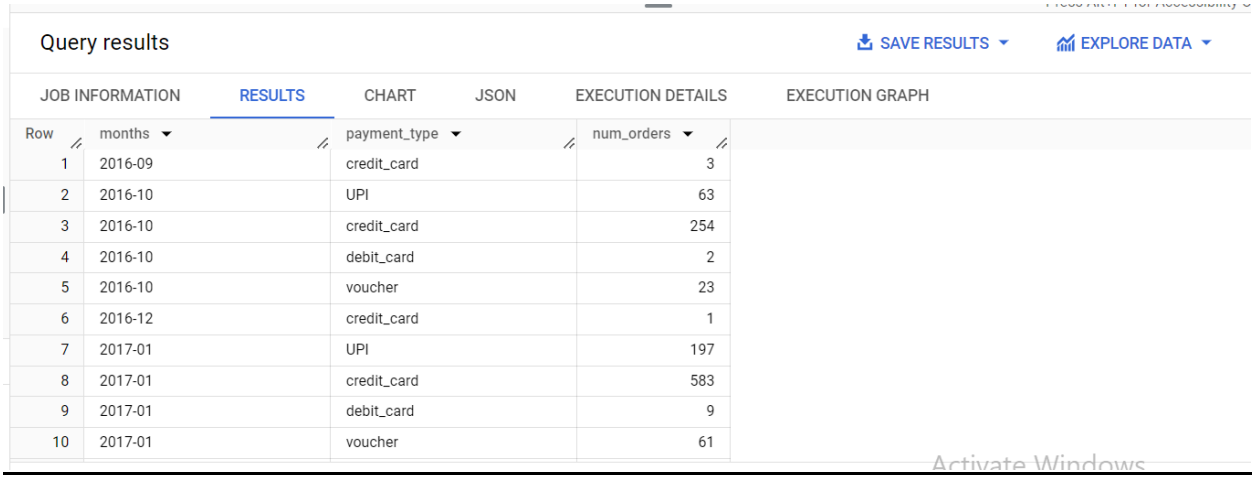


Row	customer_state	avg_delivery_speed
1	AL	7.947103274559...
2	MA	8.768479776847...
3	SE	9.173134328358...
4	ES	9.618546365914...
5	BA	9.934889434889...

- (AL) - State with the fastest delivery speed, indicating an efficient logistics and supply chain management system.
- (MA) - follows closely with quick deliveries compared to the estimated dates, which is beneficial for customer satisfaction.
- (SE) - Also shows a fast delivery speed, making it a favorable state for online shoppers.
- (ES) - demonstrates efficient delivery operations, contributing to customer satisfaction in the region.
- (BA) - is another state where deliveries are faster than expected, improving the overall customer experience.

VI. Analysis based on the payments: A. Find the month on month no. of orders placed using different payment types.

```
select format_datetime("%Y-%m",O.order_purchase_timestamp) as months,
P.payment_type,
count(*) as num_orders
from `Business_Case_Target_SQL.orders` O
join `Business_Case_Target_SQL.payments` P on O.order_id = P.order_id
group by months,P.payment_type
order by months,P.payment_type
```



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

Insights:

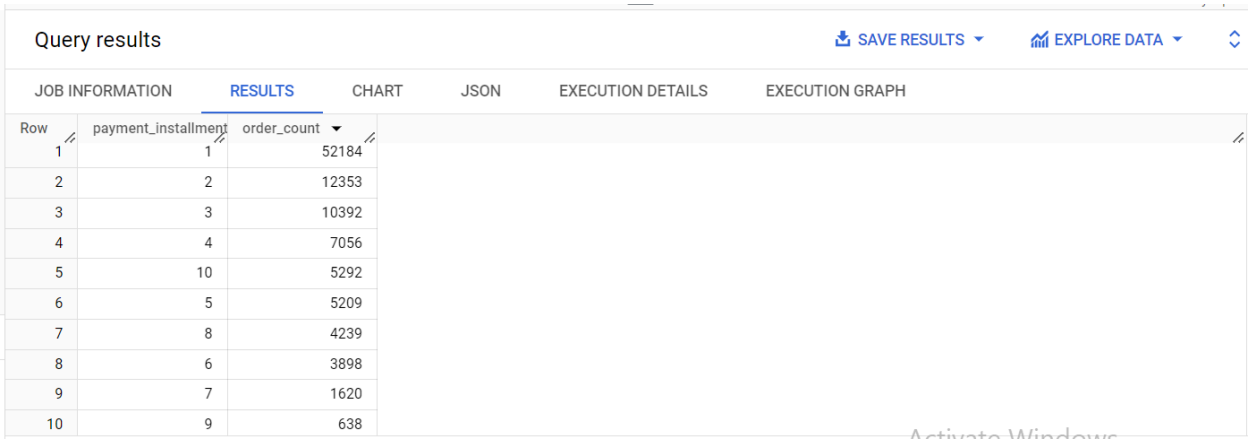
- In September 2017, "credit_card" was the most popular payment method, with 3,283 orders, while "debit_card" and "voucher" had significantly fewer orders.
- In October 2017, "credit_card" remained popular, with 3,524 orders, and "UPI" had 993 orders. "debit_card" and "voucher" had relatively low numbers.
- In November 2017, "credit_card" continued to dominate with 5,897 orders, followed by "UPI" with 1,509 orders.

Action Plan:

- The company can consider offering promotions or incentives to encourage the use of "UPI," "debit_card," and "voucher" payment methods, which have fewer orders. This may help diversify payment options and reduce dependence on a single payment method.
- Educate customers about the benefits and ease of using different payment methods, especially those with lower usage. Offering discounts or cashback for certain payment methods can be an effective strategy.

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 `Business_Case_Target_SQL.orders` o
JOIN `Business_Case_Target_SQL.payments` p ON o.order_id = p.order_id
WHERE o.order_status != 'canceled'
GROUP BY 1
ORDER BY 2 DESC
```



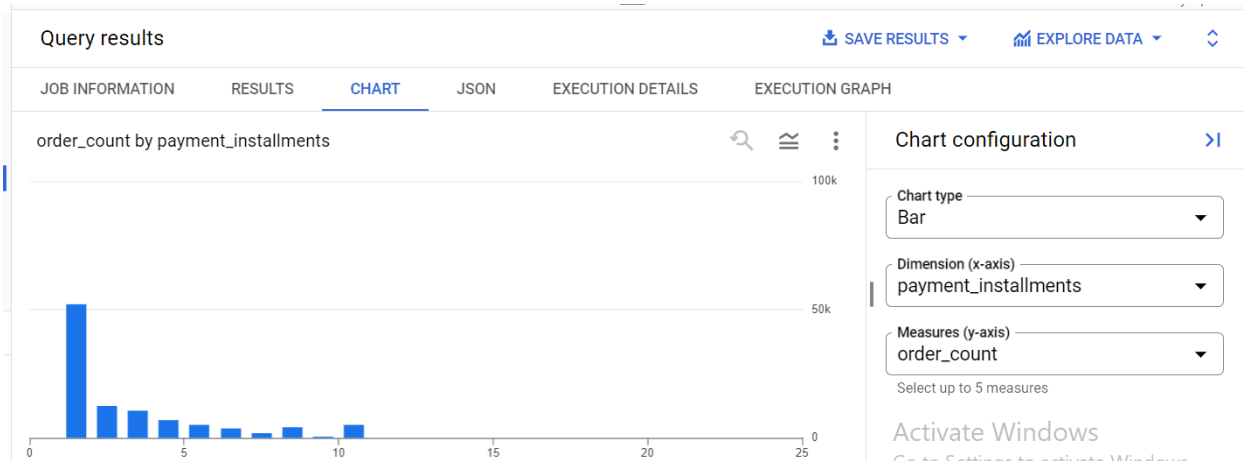
Row	payment_installments	order_count
1	1	52184
2	2	12353
3	3	10392
4	4	7056
5	10	5292
6	5	5209
7	8	4239
8	6	3898
9	7	1620
10	9	638

Insights:

The majority of orders are placed with 12 payment instalments, with 133 orders falling into this category.

Instalments of 15 and 18 follow, with 74 and 27 orders, respectively. These are the second and third most common choices for payment instalments.

Orders with 0, 2, 5, and 8 payment instalments are the least common, with only a few orders in each category.



Conclusion:-

The analysis of e-commerce data in the Brazilian market provides valuable insights into customer buying patterns, sales trends, payment preferences, and delivery experiences.
By understanding these patterns and trends, businesses can make informed decisions and implement strategies to optimize their operations and drive growth.
The state of SP dominates the e-commerce market in Brazil, indicating the need to focus on other states for potential growth opportunities.
Analyzing customer demographics can help tailor products and marketing strategies to specific target audiences, leading to increased sales.
Offering discounts during off-peak seasons can encourage customers and boost sales during the low sale seasons.

Recommendations:

1. Need some check up on the estimated time for delivery algorithm, by analyzing the data we see that there's significant gap in the actual delivery time and estimated delivery time. This might increase the online sales because consider if you see 50 days estimated delivery time instead of (10+5 buffer) 15 days, would you buy it or go to the store and bring it yourself?
2. Float more offers on the time period of 7-11PM, seems good time for increasing traffic for sales.
3. Provide more offers on are 'Bed Table Bath', 'Furniture Decoration', 'Health Beauty' and increase sales for these categories.
4. Had to go through top highest freight value states as there is no state that is in top state for orders, if we can find a reason for the higher cost of freight value we should optimize the cost (if possible).
5. Provide more offers on credit card purchase as most of the people are buying with these.
6. Provide the above mentioned offers mostly in Sao Paulo, Rio de Janeiro, Belo Horizonte cities as most buying population is in these cities.
7. Need to look further in the states SP, RJ, MG because the highest number of both fastest deliveries and the slowest deliveries occur in these states.