

Business Case: **“Biggest General Merchandise Retailer In US”**

By Kanishkar Selvakumar

***Disclaimer:**

This analysis is based on the data provided and reflects the state of the dataset as of the time of the analysis. The insights and recommendations are derived solely from my point of view and the dataset in question and do not necessarily represent the broader operations or circumstances of the company. The analysis assumes the accuracy of the data as received and has not been independently verified. Future analyses may yield different insights as new data becomes available or as business conditions change.

***Note on Results:**

Due to the large volume of results generated by this analysis, only a subset has been presented here to illustrate the key trends and patterns. Specifically, the top 10 to 15 results have been included as screenshots to provide a snapshot of the most relevant findings. For a complete view of the data and to explore additional insights, please refer to the full dataset.

Problem Statement: (All these analysis is based on this)

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:

- Data type of all columns in the "customers" table.
- Get the time range between which the orders were placed.
- Count the Cities & States of customers who ordered during the given period.

2. In-depth Exploration:

- Is there a growing trend in the no. of orders placed over the past years?
- Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
- During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
 1. 0-6 hrs : Dawn
 2. 7-12 hrs : Mornings
 3. 13-18 hrs : Afternoon
 4. 19-23 hrs : Night

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

- Get the month on month no. of orders placed in each state.
- How are the customers distributed across all the states?

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

- Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only). You can use the "payment_value" column in the payments table to get the cost of orders.
- Calculate the Total & Average value of order price for each state.
- Calculate the Total & Average value of order freight for each state.

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

- 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.

You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

1. **time_to_deliver** = order_delivered_customer_date - order_purchase_timestamp
 2. **diff_estimated_delivery** = order_delivered_customer_date - order_estimated_delivery_date
- Find out the top 5 states with the highest & lowest average freight value.
 - Find out the top 5 states with the highest & lowest average delivery time.
 - 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.

6. Analysis based on the payments:

- Find the month on month no. of orders placed using different payment types.
- Find the no. of orders placed on the basis of the payment installments that have been paid.

#Q1.1:Data type of all columns in the "customers" table.

Query

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

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	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

From this result, we can see that customers table is made of “String” and “Int64”

Recommendations

n/a

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

Query

```
SELECT MIN(order_purchase_timestamp) AS earliest_order_time,  
MAX(order_purchase_timestamp) AS latest_order_time  
FROM `ecom.orders`
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	earliest_order_time ▼	latest_order_time ▼			
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC			

Insights

From this, we can see that orders were placed between
2016-09-04 21:15:19 UTC and **2018-10-17 17:30:18 UTC**

Recommendations

n/a

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

Query

```
SELECT
  c.customer_city,
  c.customer_state,
  COUNT(DISTINCT c.customer_id) AS number_of_customers
FROM `ecom.customers` AS c
JOIN `ecom.orders` AS o ON c.customer_id = o.customer_id
GROUP BY c.customer_city, c.customer_state
ORDER BY number_of_customers DESC
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	customer_city ▼	customer_state ▼	number_of_customers ▼			
1	sao paulo	SP	15540			
2	rio de janeiro	RJ	6882			
3	belo horizonte	MG	2773			
4	brasilia	DF	2131			
5	curitiba	PR	1521			
6	campinas	SP	1444			
7	porto alegre	RS	1379			
8	salvador	BA	1245			
9	guarulhos	SP	1189			
10	sao bernardo do campo	SP	938			

Insights

From 2016-09-04 21:15:19 UTC to 2018-10-17 17:30:18 UTC, Customers belong to "Sao Paulo" has the highest order history.

This query provides a geographical breakdown of the customer base, revealing significant concentrations in specific cities and states. The data indicates that customers from "Sao Paulo" city, in "Sao Paulo" state, have the highest number of orders, significantly outpacing other regions. This suggests a strong market presence in Sao Paulo, potentially due to its large population and economic status. The disparity between "Sao Paulo" and other cities like "Rio de Janeiro" underscores varying market penetrations across different regions.

Recommendations

Market Focus: Given the high concentration of customers in "Sao Paulo", it may be beneficial to focus marketing and sales efforts in this area to further capitalize on the existing customer base. This could include localized advertising campaigns, opening new stores, or enhancing delivery logistics for faster service.

Expansion Opportunities: While "Sao Paulo" shows strong sales, there's potential for growth in other cities and states with fewer customers. Identifying reasons for lower penetration in these areas—be it lack of awareness, preference for competitors, or logistical challenges—can guide strategic expansions.

Localized Offerings: Tailor product offerings and marketing messages to the specific preferences and needs of customers in top cities and states. For regions with emerging customer bases, conduct market research to understand local preferences and customize offerings accordingly.

Infrastructure Investments: For areas with significant customer activity, consider investing in infrastructure, such as warehouses or distribution centers, to improve delivery times and reduce shipping costs.

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

Query

```
SELECT
  EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
  COUNT(order_id) AS total_orders
FROM `ecom.orders`
GROUP BY order_year
ORDER BY order_year
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	order_year ▼	total_orders ▼			
1	2016	329			
2	2017	45101			
3	2018	54011			

Insights

From this above image we can understand that the number of order placed over past years is increased.

The reason of low total orders in 2016 is because the data is available only for September and October month. This might be less data acquired.

The results indicate a substantial growth in the number of orders placed over the past years:

In 2016, there were 329 orders, which can be considered a starting or pilot phase.

In 2017, the number of orders increased dramatically to 45,101. This represents an increase of more than 100 times the previous year, signaling a significant uptake in the company's market presence or consumer adoption of the service.

In 2018, the growth continued, with the total orders reaching 54,011. This is an increase of approximately 20% from 2017.

Recommendations

Investigate Underlying Factors: Analyze the factors contributing to this growth. This could include marketing campaigns, expansion into new markets, improvements to the online platform, or broader economic factors.

Sustain Growth: Identify the successful strategies that have driven this growth and consider how to sustain or enhance them. This might include scaling up marketing, improving logistics and supply chain management, or expanding product ranges.

Infrastructure Scalability: Ensure that the company's infrastructure, both technical and logistical, can handle further increases in order volume. It's essential to scale operations to avoid potential bottlenecks as the company grows.

Customer Experience Focus: With the increase in order volume, maintain a focus on customer service quality. Growth should not come at the expense of customer satisfaction.

Kanishkar Selvakumar

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

Query

```
SELECT
  EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
  EXTRACT(MONTH FROM order_purchase_timestamp) AS order_month,
  COUNT(order_id) AS monthly_orders
FROM `ecom.orders`
GROUP BY order_year, order_month
ORDER BY order_year, order_month
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	order_year ▼	order_month ▼	monthly_orders ▼	
1	2016	9	4	
2	2016	10	324	
3	2016	12	1	
4	2017	1	800	
5	2017	2	1780	
6	2017	3	2682	
7	2017	4	2404	
8	2017	5	3700	
9	2017	6	3245	
10	2017	7	4026	
11	2017	8	4331	
12	2017	9	4285	
13	2017	10	4631	
14	2017	11	7544	
15	2017	12	5673	

Insights:

Initial Growth: Starting with just a few orders in September 2016, there is a rapid increase in the number of orders placed, reaching a peak in November 2017 with 7544 orders. This growth could be associated with an increased market presence, marketing campaigns, seasonal sales, or expanding product lines.

Year-End Peak: There is a notable peak towards the end of each year, especially in November and December, which could be attributed to holiday shopping for events like Black Friday, Cyber Monday, and Christmas.

Start of the Year Momentum: The momentum continues into the new year, with January 2018 also showing a high number of orders. This may reflect New Year promotions or gift card redemptions following the holiday season.

Mid-Year Plateau: From February to August, the monthly orders seem to stabilize with a slight uptrend in 2018 compared to the corresponding months in 2017, suggesting steady demand and possibly effective customer retention strategies.

Sharp Drop: There is a significant drop in September 2018, which deviates from the previous trend and continues to remain very low into October 2018. This is an anomaly that needs to be investigated.

Recommendation:

Investigate Anomalies: Look into the sharp decline in orders in September and October 2018 to understand whether it was due to external factors, operational issues, changes in consumer behavior, or data recording errors.

Capitalize on Peak Seasons: Strengthen inventory and marketing strategies around the November-December period to maximize sales during the peak shopping season.

Engagement During Off-Peak Months: Develop promotions and marketing initiatives to boost orders during traditionally lower-performing months, from February to August.

Customer Behavior Analysis: Conduct customer surveys or market research to understand the factors driving seasonality, which may include looking at economic trends, competitor activity, and changing consumer preferences.

Long-Term Data Tracking: Continue to collect and analyze data beyond October 2018 to confirm these seasonal trends and adjust strategies accordingly.

#2.3 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 time_of_day,
  COUNT(order_id) AS total_orders
FROM `ecom.orders`
GROUP BY time_of_day
ORDER BY total_orders DESC
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	time_of_day ▼	total_orders ▼		
1	Afternoon	38135		
2	Night	28331		
3	Morning	27733		
4	Dawn	5242		

Insights:

Peak Ordering Times: The majority of orders are placed in the afternoon (13-18 hrs), followed by night (19-23 hrs) and morning (7-12 hrs). This indicates that customers are more active in shopping online later in the day.

Lowest Activity in Dawn: Dawn (0-6 hrs) sees significantly fewer orders, which is expected as it's the early hours when most people are likely asleep or just starting their day.

Evening Shopping Behavior: The high volume of orders during the night suggests that many customers prefer shopping in the late hours, possibly after returning from work or during their leisure time.

Recommendations:

Targeted Promotions: Schedule flash sales or special promotions during the afternoon and night hours to capitalize on the higher traffic. This could include limited-time offers to encourage impulse purchases.

Customer Support: Ensure customer support is readily available during peak ordering times, especially in the afternoon and night, to help with any inquiries or issues that may arise during checkout.

Personalized Marketing: Use the insights from order timings to send personalized marketing emails or notifications during the morning or just before the peak times to increase engagement and remind customers of ongoing deals or items they might be interested in.

Optimize User Experience for Night Shoppers: Given the significant activity at night, ensure the website or app is optimized for nighttime browsing. This could include a dark mode feature to reduce eye strain for customers shopping in low-light conditions.

Dawn Engagement Strategies: Although dawn sees the least activity, consider experimenting with early morning offers or special discounts for early birds to potentially increase orders during this time. This could attract a niche market of early risers or those looking to place orders before starting their day.

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

Query

```
SELECT
  c.customer_state AS state,
  EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
  COUNT(o.order_id) AS total_orders
FROM `ecom.customers` AS c
JOIN `ecom.orders` AS o ON c.customer_id = o.customer_id
GROUP BY state, order_year, order_month
ORDER BY state, order_year, order_month
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	state ▼	order_year ▼	order_month ▼	total_orders ▼		
1	AC	2017	1	2		
2	AC	2017	2	3		
3	AC	2017	3	2		
4	AC	2017	4	5		
5	AC	2017	5	8		
6	AC	2017	6	4		
7	AC	2017	7	5		
8	AC	2017	8	4		
9	AC	2017	9	5		
10	AC	2017	10	6		

Insights

Growth Trends: Most states show an upward trend in the number of orders placed, indicating expanding market penetration and increasing customer base. This is particularly noticeable in major states like SP, RJ, and MG, where there's a clear growth trajectory from 2016 to 2018.

Seasonal Peaks: There's a pronounced spike in orders in November across many states, aligning with Black Friday and the pre-Christmas shopping season. This suggests customers are highly responsive to sales and holiday promotions.

Regional Variances: Different states exhibit varying growth rates and order volumes, indicating diverse market dynamics. For example, SP (São Paulo) consistently has the highest volume of orders, reflecting its status as a major economic hub, while smaller states like AC (Acre) and AP (Amapá) show much lower order volumes.

Emerging Markets: States with initially low order volumes, such as AP and RR, show significant percentage growth year over year, suggesting that these could be emerging markets with potential for further expansion.

Recommendations

Targeted Marketing Campaigns: Tailor marketing strategies to the unique characteristics of each state. High-growth and high-volume states may benefit from campaigns designed to reinforce brand loyalty and increase order frequency, while emerging markets may require strategies focused on brand awareness and acquisition.

Seasonal Promotions: Capitalize on the observed November peak with targeted promotions and sales, ensuring adequate inventory and logistical readiness to handle the surge in orders.

Regional Expansion: Focus on expanding market presence in emerging states with growing order volumes. Consider localized strategies that cater to the specific preferences and needs of these markets.

Market Segmentation: Analyze the demographic and shopping behavior data to further segment the market, allowing for more personalized and effective marketing strategies. This could include age, gender, income level, and previous purchase history.

Infrastructure Optimization: For states showing significant growth or high order volumes, ensure that logistics and delivery infrastructure are optimized to handle demand efficiently, reducing delivery times and improving customer satisfaction.

#3.2. How are the customers distributed across all the states?

Query

```
SELECT
  customer_state AS state,
  COUNT(DISTINCT customer_unique_id) AS unique_customers
FROM `ecom.customers`
GROUP BY state
ORDER BY unique_customers DESC
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	state	unique_customers		
1	SP	40302		
2	RJ	12384		
3	MG	11259		
4	RS	5277		
5	PR	4882		
6	SC	3534		
7	BA	3277		
8	DF	2075		
9	ES	1964		
10	GO	1952		
11	PE	1609		
12	CE	1313		
13	PA	949		
14	MT	876		

Insights

State Concentration: A significant concentration of customers is located in SP (São Paulo), RJ (Rio de Janeiro), and MG (Minas Gerais), which together account for a substantial portion of the customer base. This highlights the economic and population density in these regions.

Regional Differences: Southern and Southeastern states like RS (Rio Grande do Sul), PR (Paraná), and SC (Santa Catarina) also have a significant customer presence, indicating a broader appeal and market penetration in these affluent regions.

Emerging Markets: Northern and Northeastern states, including PA (Pará), MT (Mato Grosso), MA (Maranhão), and states with traditionally smaller e-commerce footprints like AC (Acre), AP (Amapá), and RR (Roraima), show lower customer numbers. These areas represent emerging markets with potential growth opportunities.

Market Saturation vs. Potential: The varying customer distribution suggests areas of both market saturation and untapped market potential. States with lower customer counts may indicate opportunities for expansion, assuming the barriers to entry (like logistics) can be addressed.

Recommendations

Focus on High-Density Areas: Continue to strengthen market presence in SP, RJ, and MG through targeted marketing campaigns, localized promotions, and ensuring superior logistics and customer service to maintain dominance and customer loyalty in these competitive regions.

Expand in Emerging Markets: Develop strategies to increase brand awareness and customer acquisition in states with fewer customers. This could include partnerships with local businesses, targeted social media campaigns, and offering incentives for first-time customers.

Customize Offerings: Consider regional preferences and cultural nuances in product offerings. Tailored marketing and product selection can enhance relevance and appeal in diverse markets.

Improve Logistics in Less Penetrated Areas: Addressing logistical challenges in remote or less-served areas can open up new markets. Investment in local distribution centers or partnerships with local delivery services can reduce delivery times and costs.

Utilize Data for Personalization: Leverage customer data to personalize shopping experiences. Custom recommendations, localized marketing messages, and region-specific offers can increase engagement and conversion rates.

Monitor and Respond to Market Changes: Keep a close eye on changing demographics, consumer preferences, and competitive dynamics across different states. Agility in responding to market changes can provide a competitive edge.

#4.1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
 #You can use the "payment_value" column in the payments table to get the cost of orders.

Query

```
WITH total_payments_2017 AS (
  SELECT
    SUM(payment_value) AS total_2017
  FROM `ecom.payments` p
  JOIN `ecom.orders` o ON p.order_id = o.order_id
  WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017
    AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
),
total_payments_2018 AS (
  SELECT
    SUM(payment_value) AS total_2018
  FROM `ecom.payments` p
  JOIN `ecom.orders` o ON p.order_id = o.order_id
  WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018
    AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
)
SELECT
  total_2017,
  total_2018,
  ((total_2018 - total_2017) / total_2017) * 100 AS percentage_increase
FROM total_payments_2017, total_payments_2018
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	total_2017 ▼	total_2018 ▼	percentage_increase ▼		
1	3669022.1200000118	8694733.83999979	136.97687164665447		

Insights

The data reveals a substantial 136.98% increase in the total cost of orders from January to August, comparing the years 2017 and 2018. This is a significant jump in e-commerce spending and suggests a rapidly growing market.

Market Expansion: The large increase in payment values indicates significant market expansion. This could be due to an increase in the number of customers, a rise in the average order value, or a combination of both.

Consumer Confidence: The rise in spending could also reflect increased consumer confidence in e-commerce platforms, suggesting that customers are more comfortable making larger and more frequent purchases online.

Product Offerings and Pricing: The e-commerce platform may have expanded its product offerings, or there could have been an adjustment in pricing strategies, including premium product listings and inflationary effects.

Economic Factors: External economic factors could have contributed to this increase, such as greater disposable income, economic growth in the region, or increased internet penetration.

Recommendations

Customer Analysis: Perform detailed customer segmentation to understand who the new customers are and what is driving the increased spending.

Optimize Pricing Strategy: Given the rise in order values, analyze the pricing strategy to ensure it remains competitive while also maximizing profitability.

Enhance User Experience: Invest in improving the platform's user experience to sustain the growth in order volume and cost, ensuring that the increased traffic can be effectively managed without compromising service quality.

Expand Logistics Capacity: Consider scaling up logistics and distribution capacities to handle the increased order flow without delay, ensuring customer satisfaction with delivery times.

Market Diversification: Explore opportunities to diversify the product range even further, catering to the evident increase in demand and varied customer preferences.

Retention Programs: Develop customer loyalty and retention programs to keep the growing customer base engaged and reduce churn.

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

Query

```
SELECT
  c.customer_state AS state,
  SUM(p.payment_value) AS total_order_value,
  AVG(p.payment_value) AS average_order_value
FROM `ecom.customers` AS c
JOIN `ecom.orders` AS o ON c.customer_id = o.customer_id
JOIN `ecom.payments` AS p ON o.order_id = p.order_id
GROUP BY state
ORDER BY total_order_value DESC
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	state	total_order_value	average_order_value		
1	SP	5998226.9599999283	137.50462977396811		
2	RJ	2144379.6899999999	158.52588822355338		
3	MG	1872257.2600000079	154.70643364733127		
4	RS	890898.5399999998	157.18040578687359		
5	PR	811156.37999999791	154.15362599771953		
6	SC	623086.4299999959	165.97933670751203		
7	BA	616645.82000000367	170.8160166204986		
8	DF	355141.08000000048	161.13479128856639		
9	GO	350092.30999999988	165.76340435606087		
10	ES	325967.55000000092	154.70695301376369		

Insights

High Spending States: São Paulo (SP) leads with the highest total order value, which is consistent with its status as the most populous and economically influential state in Brazil.

Average Order Value Variance: The average order value varies significantly between states, with states like PB, CE, and AC having relatively high average order values despite a smaller total order value, suggesting that fewer customers are placing larger orders.

Regional Economic Indicators: Higher average order values in certain states may indicate a higher purchasing power or preference for premium products, while higher total order values reflect the overall market size and e-commerce activity.

Market Penetration and Saturation: States with low average and total order values might be less penetrated by e-commerce services, or there could be other factors influencing lower e-commerce activity, such as access to alternative shopping methods or lower internet penetration.

Recommendations

Targeted Marketing: Customize marketing strategies for states with high average order values to enhance customer value, focusing on premium products and upselling opportunities.

Market Expansion: In states with low total order values but higher averages, there may be opportunities to increase market penetration by improving logistics, payment options, and localized marketing efforts to attract more customers.

Customer Research: Conduct detailed customer behavior studies in states with lower average order values to understand barriers to higher spending and to adapt the product mix to local tastes and purchasing power.

Promotions and Discounts: For states where the average order value is low, consider introducing promotions, discounts, or low-cost product lines to increase the average spend.

Logistics Optimization: Ensure that logistics and delivery systems are efficient in states with high total order values to maintain customer satisfaction and encourage repeat business.

Local Partnerships: In regions with lower e-commerce adoption, consider partnering with local businesses to increase brand presence and build trust with local customers.

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

Query

```
SELECT
  c.customer_state AS state,
  SUM(oi.freight_value) AS total_freight_value,
  AVG(oi.freight_value) AS average_freight_value
FROM `ecom.customers` AS c
JOIN `ecom.orders` AS o ON c.customer_id = o.customer_id
JOIN `ecom.order_items` AS oi ON o.order_id = oi.order_id
GROUP BY state
ORDER BY state
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	state ▼	total_freight_value ▼		average_freight_value ▼	
1	AC	3686.7499999999986		40.0733695652174	
2	AL	15914.589999999997		35.843671171171167	
3	AM	5478.8899999999994		33.20539393939395	
4	AP	2788.5000000000018		34.006097560975626	
5	BA	100156.679999999858		26.363958936562188	
6	CE	48351.589999999887		32.714201623816017	
7	DF	50625.49999999984		21.041354945968457	
8	ES	49764.599999999831		22.058776595744643	
9	GO	53114.979999999829		22.766815259322811	
10	MA	31523.770000000019		38.257002427184418	

Insights

Variation in Freight Costs: There is a considerable variation in average freight costs across states, with AC and RR having the highest average freight values, and SP having the lowest. This could be due to the geographical distance from the shipping origin, the accessibility of the region, or the density of orders.

Economies of Scale: São Paulo (SP) has the lowest average freight cost, which could be attributed to economies of scale due to high order volume, as well as better logistics infrastructure and proximity to distribution centers.

Logistical Challenges: States with higher average freight costs may present logistical challenges or have lower order densities, which increases the cost per shipment.

Shipping Efficiency: States like DF, despite being centrally located, show a higher average freight value than some others, which might indicate inefficiencies or higher delivery costs that need investigation.

Recommendations

Review Shipping Partners: For states with higher average freight values, consider reviewing contracts with shipping partners or exploring alternative delivery methods to reduce costs.

Bulk Delivery Discounts: Negotiate bulk delivery discounts with carriers, especially for states with high order volumes, to reduce the average freight cost further.

Distribution Center Placement: Consider the placement of new distribution centers in regions with high freight values to reduce the distance packages need to travel, which can lower costs and improve delivery times.

Order Aggregation: Implement order aggregation strategies in states with higher freight costs to consolidate shipments and reduce the number of trips required.

Freight Cost Analysis: Conduct a more detailed analysis of freight costs to identify specific factors contributing to the high average values in certain states. This should include an evaluation of weight, distance, package dimensions, and carrier efficiency.

Customer Shipping Options: Offer varied shipping options to customers, including economy shipping with longer delivery times but lower costs, which can be especially appealing in regions with high average freight costs.

Kanishkar Selva

#5.1. 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.

You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
 $\text{time_to_deliver} = \text{order_delivered_customer_date} - \text{order_purchase_timestamp}$
 $\text{diff_estimated_delivery} = \text{order_delivered_customer_date} - \text{order_estimated_delivery_date}$

Query

```
SELECT
    order_id,
    DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY)
AS time_to_deliver,
    DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date,
DAY) AS diff_estimated_delivery
FROM `ecom.orders`
WHERE order_status = 'delivered'
```

Screenshots

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	order_id	time_to_deliver		diff_estimated_delivery	
1	635c894d068ac37e6e03dc54e...	30		-1	
2	3b97562c3aee8bdedcb5c2e45...	32		0	
3	68f47f50f04c4cb6774570cfde...	29		-1	
4	276e9ec344d3bf029ff83a161c...	43		4	
5	54e1a3c2b97fb0809da548a59...	40		4	
6	fd04fa4105ee8045f6a0139ca5...	37		1	
7	302bb8109d097a9fc6e9cefc5...	33		5	
8	66057d37308e787052a32828...	38		6	
9	19135c945c554eebfd7576c73...	36		2	
10	4493e45e7ca1084efcd38ddeb...	34		0	

Insights:

Delivery Performance: The time_to_deliver values vary from 29 to 43 days. This range suggests that delivery times can be quite long, possibly due to logistics complexity or operational delays.

Estimated vs. Actual Delivery: The diff_estimated_delivery column contains both negative and positive values. Negative values indicate orders were delivered before the estimated delivery date, whereas positive values indicate that delivery was later than estimated.

Early Deliveries: Orders with a negative diff_estimated_delivery reflect well on the logistics and supply chain efficiency. Customers receiving orders earlier than expected can lead to higher customer satisfaction and trust in the service.

Late Deliveries: Positive values in the diff_estimated_delivery column suggest that these orders were delivered later than Target estimated to customers. Late deliveries can lead to customer dissatisfaction and potentially impact the company's reputation.

Recommendations:

Improve Logistics: For orders with long delivery times and late deliveries, a review of the logistics process is advisable. Improving partnerships with courier services or optimizing routes can reduce delivery times.

Communication with Customers: Maintain clear and open communication regarding estimated delivery times. If early or late deliveries are a possibility, customers should be kept informed throughout the process.

Review Estimation Process: If orders are consistently delivered early or late relative to estimates, the system used for generating delivery estimates should be reviewed and adjusted for accuracy.

Customer Feedback: For orders with large discrepancies in estimated and actual delivery dates, follow up with customers to assess the impact on satisfaction. Use this feedback to improve processes and customer service policies.

Performance Incentives: Implement incentive programs for logistics partners to encourage on-time deliveries.

Kanishkar Selvakumar

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

```
WITH StateFreight AS (  
  SELECT  
    c.customer_state AS state,  
    AVG(oi.freight_value) AS average_freight_value  
  FROM `ecom.customers` AS c  
  JOIN `ecom.orders` AS o ON c.customer_id = o.customer_id  
  JOIN `ecom.order_items` AS oi ON o.order_id = oi.order_id  
  GROUP BY state  
)
```

```
SELECT * FROM (  
  SELECT  
    state,  
    average_freight_value  
  FROM StateFreight  
  ORDER BY average_freight_value DESC  
  LIMIT 5  
)
```

UNION ALL

```
SELECT * FROM (  
  SELECT  
    state,  
    average_freight_value  
  FROM StateFreight  
  ORDER BY average_freight_value ASC  
  LIMIT 5  
)
```

ScreenShot

Query results

JOB INFORMATION		RESULTS	CHART
Row	state ▼	average_freight_value ▼	
1	RR	42.984423076923093	
2	PB	42.723803986710941	
3	RO	41.069712230215842	
4	AC	40.073369565217405	
5	PI	39.147970479704767	
6	SP	15.147275390419248	
7	PR	20.531651567944248	
8	MG	20.630166806306541	
9	RJ	20.96092393168248	
10	DF	21.041354945968383	

Insights:

Geographical Challenges: The states with the highest average freight values—RR (Roraima), PB (Paraíba), RO (Rondônia), AC (Acre), and PI (Piauí)—are likely to be those that face geographical challenges. They might be further from distribution centers or harder to access due to terrain, leading to higher transportation costs.

Economies of Scale: SP (São Paulo) has the lowest average freight value, which is consistent with it being a major logistic hub with high order volumes. The economies of scale and better infrastructure can lead to lower average shipping costs.

Inefficiencies in Delivery Network: High average freight values can also indicate inefficiencies in the delivery network, potentially arising from a lack of competition among freight services or inadequate logistics planning.

Variability in Freight Pricing: The significant variability between the highest and lowest freight values suggests that shipping costs are a major factor in the overall cost structure of e-commerce operations.

Recommendations:

Review Freight Partnerships: For states with high average freight values, review shipping partnerships and negotiate better terms or seek alternative providers that may offer more competitive rates.

Logistic Improvements: Invest in logistics improvements in states with the highest freight costs. This could involve opening new distribution centers or using alternative modes of transport, like water or air, where feasible.

Regional Warehousing: Consider setting up regional warehouses or fulfillment centers closer to the states with the highest freight values to reduce shipping distances and costs.

Dynamic Pricing Strategy: Implement a dynamic pricing strategy for shipping that considers the average freight value by state, which could be used to either incentivize customers through promotional shipping rates or to ensure profitability.

Cost Absorption and Transparency: Depending on the competitive landscape and customer expectations, consider absorbing part of the freight cost in states with high values or being transparent with customers about the reasons for higher charges.

Incentives for Bulk Orders: In states with high freight values, offer incentives for bulk purchases that can be shipped more efficiently, reducing the average freight cost per item.

#5.3. Find out the top 5 states with the #highest average delivery time.

Query

```
WITH DeliveryTime AS (  
  SELECT c.customer_state,  
         AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,  
o.order_purchase_timestamp, DAY)) AS average_delivery_days  
  FROM `ecom.orders` o  
  JOIN `ecom.customers` c ON o.customer_id = c.customer_id  
  WHERE o.order_delivered_customer_date IS NOT NULL  
  GROUP BY c.customer_state  
)  
SELECT customer_state AS state,  
       average_delivery_days  
FROM DeliveryTime  
ORDER BY average_delivery_days DESC  
LIMIT 5
```

lowest average delivery time.

Query

```
WITH DeliveryTime AS (  
  SELECT c.customer_state,  
         AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,  
o.order_purchase_timestamp, DAY)) AS average_delivery_days  
  FROM `ecom.orders` o  
  JOIN `ecom.customers` c ON o.customer_id = c.customer_id  
  WHERE o.order_delivered_customer_date IS NOT NULL  
  GROUP BY c.customer_state  
)  
SELECT customer_state AS state,  
       average_delivery_days  
FROM DeliveryTime  
ORDER BY average_delivery_days ASC  
LIMIT 5
```

ScreenShots (Highest, Lowest respectively)

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	state	average_delivery_days		
1	RR	28.975609756097562		
2	AP	26.731343283582085		
3	AM	25.986206896551728		
4	AL	24.040302267002513		
5	PA	23.316067653276981		

Query results

JOB INFORMATION		RESULTS	CH.
Row	state	average_delivery_days	
1	SP	8.2980614890725874	
2	PR	11.526711354864908	
3	MG	11.543813298106569	
4	DF	12.509134615384616	
5	SC	14.479560191711331	

Insights:

The states with the lowest average delivery times are among the most economically developed and urbanized in the country, with better logistics infrastructure and transportation links.

In contrast, the states with the highest average delivery times are generally less developed, with more challenging geography and infrastructure constraints that can impact delivery efficiency.

Recommendations:

For states with high average delivery times, consider reviewing and optimizing logistics routes or partnering with local carriers that have better regional infrastructure.

Investigate the potential of regional distribution centers in states with longer delivery times to reduce transit distance and time.

Analyze the demand patterns in states with high delivery times to optimize inventory stocking, ensuring that products are closer to the end consumer.

For the states with the shortest delivery times, continue to maintain the high standard of logistics performance, as this is likely contributing to customer satisfaction.

Use the positive delivery performance in the most efficient states as a benchmark and seek to replicate the successful practices in states where performance is lagging.

Evaluate customer feedback from areas with longer delivery times to assess the impact on customer satisfaction and repeat business, implementing strategies to manage customer expectations where necessary.

#5.4. 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 DeliverySpeed AS (  
  SELECT  
    c.customer_state,  
    AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,  
o.order_estimated_delivery_date, DAY)) AS avg_days_early  
  FROM  
    `ecom.orders` o  
  JOIN  
    `ecom.customers` c ON o.customer_id = c.customer_id  
  WHERE  
    o.order_status = 'delivered' AND  
    o.order_delivered_customer_date IS NOT NULL AND  
    o.order_estimated_delivery_date IS NOT NULL  
  GROUP BY  
    c.customer_state  
)  
SELECT  
  customer_state AS state,  
  avg_days_early  
FROM  
  DeliverySpeed  
WHERE  
  avg_days_early < 0  
ORDER BY  
  avg_days_early  
LIMIT 5;
```

ScreenShot

Query results

JOB INFORMATION		RESULTS	CHART
Row	state ▼	avg_days_early ▼	
1	AC	-19.762500000000006	
2	RO	-19.13168724279836	
3	AP	-18.731343283582088	
4	AM	-18.60689655172413	
5	RR	-16.414634146341463	

Insights:

The results indicate that on average, orders in the states of AC, RO, AP, AM, and RR are delivered significantly earlier than the estimated delivery dates, with the average delivery being earlier by approximately 16 to 20 days.

The negative value of avg_days_early means that the actual delivery date occurred before the estimated delivery date. A negative difference here is a positive outcome, as it indicates a faster delivery than what was promised to the customer.

AC (Acre) shows the fastest delivery compared to the estimated dates, followed closely by RO (Rondônia), AP (Amapá), AM (Amazonas), and RR (Roraima).

Interestingly, some of these states also appeared in the list of states with the longest delivery times. This suggests that while deliveries in these states take longer overall, they are still arriving earlier than the estimates provided to customers.

Recommendations:

Review Estimation Algorithms: The estimation process for delivery times to these states might be overly conservative. Review the algorithms or methods used to calculate estimated delivery dates, ensuring they are as accurate as possible to set realistic expectations for customers.

Understand Logistics Efficiency: Investigate why deliveries in these states are consistently faster than estimated. There may be efficiencies or best practices in these delivery routes or logistics operations that can be applied to other areas.

Customer Communication: While it is excellent that deliveries are faster than estimated, overestimating delivery times may lower customer satisfaction or affect their purchase decisions. Improving the accuracy of estimates can enhance customer trust and satisfaction.

Balancing Delivery Performance: Given that some of these states also have the longest delivery times, work on strategies to reduce the actual delivery time while maintaining or improving the estimation accuracy.

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

```
SELECT
  EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
  p.payment_type,
  COUNT(DISTINCT o.order_id) AS number_of_orders
FROM
  `ecom.orders` o
JOIN
  `ecom.payments` p ON o.order_id = p.order_id
WHERE
  o.order_status = 'delivered'
GROUP BY
  order_year,
  order_month,
  p.payment_type
ORDER BY
  order_year,
  order_month,
  p.payment_type;
```

Screenshot

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	E
Row	order_year	order_month	payment_type	number_of_orders		
1	2016	10	UPI	51		
2	2016	10	credit_card	208		
3	2016	10	debit_card	2		
4	2016	10	voucher	9		
5	2016	12	credit_card	1		
6	2017	1	UPI	188		
7	2017	1	credit_card	541		
8	2017	1	debit_card	9		
9	2017	1	voucher	32		
10	2017	2	UPI	371		
11	2017	2	credit_card	1249		
12	2017	2	debit_card	13		
13	2017	2	voucher	63		
14	2017	3	UPI	565		
15	2017	3	credit_card	1901		

Insights:

Credit Card Dominance: Credit card payments seem to be the most popular payment type across the board, with consistently high numbers month over month. For example, in March 2017, credit card payments accounted for 1901 orders, indicating strong customer preference or trust in credit card transactions.

Emergence of UPI: There is a notable presence of UPI (Unified Payments Interface) transactions, suggesting a trend toward mobile or digital payments. UPI payments show significant numbers, such as 188 orders in January 2017 and an increase to 565 by March 2017, indicating growing adoption.

Decrease in Other Forms: Debit cards and vouchers are used less frequently compared to credit cards and UPI. Vouchers do have a slight increase in use, from 9 in October 2016 to 63 in March 2017, which might indicate promotional activities or loyalty programs influencing this payment method.

Recommendations:

Encourage Diverse Payment Options: Continue to offer multiple payment options to cater to different customer preferences. Consider promotions or incentives for less used payment types if there is a business interest in diversifying payment methods.

Promote UPI for Faster Transactions: Given the increasing trend of UPI transactions, consider marketing campaigns or incentives specifically around UPI payments, which could benefit from faster checkout times and potentially lower processing fees.

Optimize for Credit Card Users: Since credit cards are the most used payment type, ensure that the credit card payment process is optimized for ease and security, as this will directly affect the largest segment of your customer base.

Leverage Data for Loyalty Programs: The usage of vouchers is increasing, which could be tied to loyalty programs. Analyze the data to see if vouchers are leading to repeat purchases and consider enhancing loyalty programs accordingly.

Monitor Payment Trends: Keep monitoring the trends in payment types, as this can inform you about consumer behavior changes over time, and adapt your payment processing infrastructure to prioritize the most popular methods.

User Experience and Trust: Work on user experience aspects related to payment types to enhance trust and satisfaction. The dominance of credit cards suggests that customers may prefer the perceived security and benefits they offer.

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

Query

```
SELECT
  p.payment_installments,
  COUNT(DISTINCT o.order_id) AS number_of_orders
FROM
  `ecom.payments` p
JOIN
  `ecom.orders` o ON p.order_id = o.order_id
WHERE
  o.order_status = 'delivered'
GROUP BY
  p.payment_installments
ORDER BY
  p.payment_installments;
```

ScreenShot

Query results

JOB INFORMATION		RESULTS	CHART
Row	payment_installment	number_of_orders	
1	0	2	
2	1	47586	
3	2	12052	
4	3	10147	
5	4	6882	
6	5	5090	
7	6	3800	
8	7	1560	
9	8	4122	
10	9	618	
11	10	5137	
12	11	22	
13	12	128	
14	13	15	
15	14	14	

Insights:

Preference for Fewer Installments: The majority of orders are paid for with just one installment, indicating that customers prefer to pay in full at the time of purchase. This is shown by the highest number of orders, 47,586, being in this category.

Decline as Installments Increase: There is a clear decline in the number of orders as the number of

installments increases. After the one-time payments, the next most popular installment option is 2 installments (12,052 orders), followed by 3 installments (10,147 orders).

Some Large Installment Plans: Notably, there are spikes at higher installment numbers such as 8 and 10, which suggests that for larger purchases, customers prefer these specific installment plans.

Low Utilization of Middle to High Installments: Installments between 11 and 23 are rarely used, with the number of orders for each falling mostly under 30. This may indicate a lack of interest or awareness of these options, or a potential mismatch with customer purchase behaviors.

Very Long-Term Installments Are Least Popular: Very long-term installment options (over 18 months) are the least utilized, with very few orders choosing these plans.

Recommendations:

Promote Full Payments When Possible: Given the clear preference for one-time payments, consider offering small discounts or incentives for paying in full to encourage this behavior, which may help with cash flow and reduce processing fees.

Evaluate and Adjust Installment Plans: Analyze whether all installment options are beneficial for the company. It may be more efficient to streamline the number of installment options based on their popularity.

Marketing for Popular Installment Options: Marketing campaigns could highlight the most popular installment options (2, 3, 8, and 10 installments) to align with customer preferences.

Educate Customers: For the installment plans that are rarely used, determine if there is a lack of awareness and, if these plans are valuable, consider educating customers about them.

Assess Credit Risk: For higher installment numbers, assess the credit risk and potential financial impact on the company. It may be beneficial to have stricter credit checks for long-term installment plans.

Simplify Offering: Consider simplifying the installment offering if certain numbers of installments are not being utilized. Too many options can confuse customers and complicate the purchasing process.

Customer Research: Conduct customer research to understand the reasons behind these trends. This can provide insights into customer financial planning and could influence the credit terms the company offers.

Flexibility for Larger Purchases: Ensure that installment plans are flexible for larger purchases where customers tend to prefer more installments (like the spikes at 8 and 10).

***Disclaimer and Limitations**

This analysis is conducted based on data provided to us up to the point of this analysis. The insights, conclusions, and recommendations herein reflect our interpretation of this dataset and do not necessarily represent the views, strategies, or operational status of the company or entity involved. We have relied on the data's accuracy as provided, without independent verification. Future analyses may lead to different insights as new data becomes available or if business conditions evolve.

- **Data Privacy and Confidentiality:** This analysis respects privacy and confidentiality norms and complies with applicable data protection laws, including, but not limited to, the handling of personally identifiable information (PII). No PII has been disclosed in this report.
- **Limitations and Assumptions:** This report makes several assumptions for analytical purposes, and its findings are subject to these assumptions' validity. The analysis may exclude certain data segments due to availability, relevance, or other factors deemed appropriate by the analyst. We acknowledge potential biases in data collection and analysis methods and have strived to minimize their impact on our conclusions.
- **Intended Audience:** This report is intended for internal stakeholders and strategic decision-making purposes. It should not be considered as financial or legal advice, nor should it be used as the sole basis for any investment decisions.
- **Use of External Data or Sources:** Where external data sources or analytical tools have been employed, they are cited accordingly. The inclusion of such sources does not imply endorsement or their accuracy by us. Users are encouraged to consult the original sources for further information.
- **Revision Policy:** Given the dynamic nature of data and business environments, this analysis is subject to revisions. Updates or corrections may be issued as new information becomes available or in response to feedback from stakeholders.
- **Non-Endorsement:** The mention of any software, platforms, or tools used in the analysis does not constitute an endorsement. These are selected based on their utility for the purposes of this analysis only.
- **Use at Own Risk:** The insights and recommendations provided are based on the data and assumptions noted. Stakeholders are encouraged to use this information responsibly and in conjunction with other sources of information, acknowledging the inherent risks and uncertainties in data-driven decision-making.

Thank you!