Exploring E-commerce Trends and Customer Behaviour:
A Comprehensive Data Analysis

An exploratory and in-depth study of the e-commerce market in Brazil

Aditya Shinde

Introduction

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 study 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.

In this study, the given dataset has been analysed to extract valuable insights from various aspects of the business and to suggest actionable recommendations.BiqQuery was used for writing SQL queries and fetching data from dataset. Tableau and Power BI were used for creating visualisations.
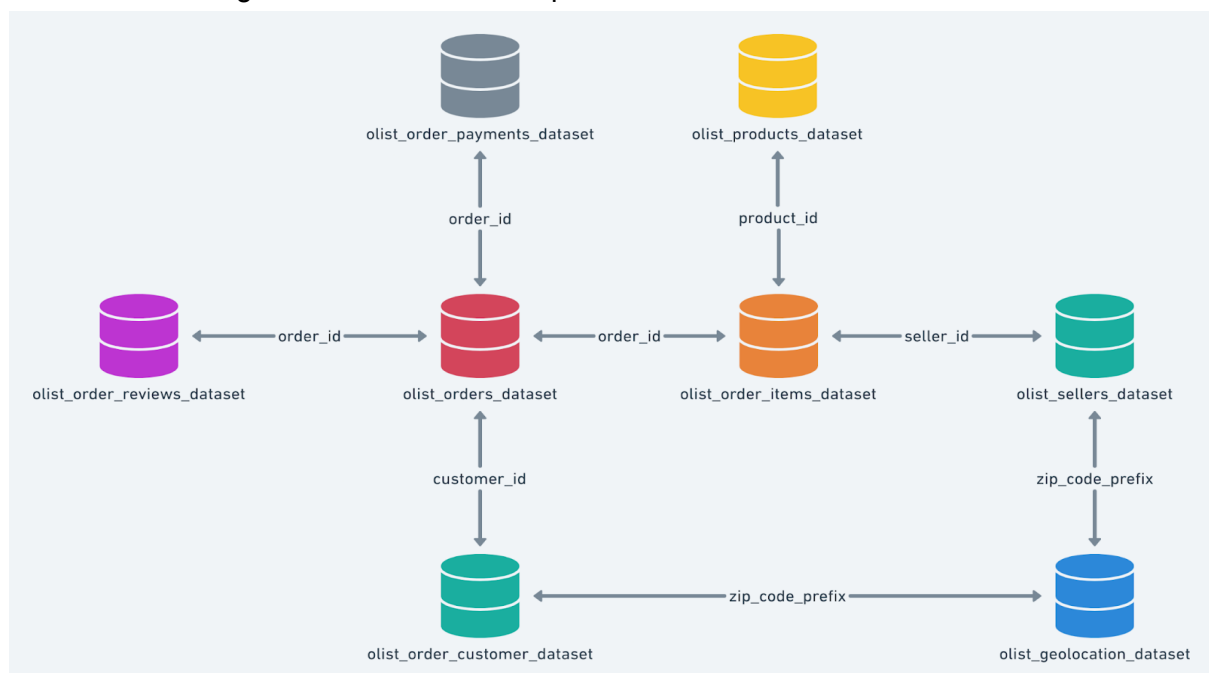
Exploratory Data Analysis

1.Structure of the dataset
The dataset contains 8 tables as mentioned below -
1. customers
2. sellers
3. order_items
4. geolocation
5. payments
6. reviews
7. orders
8. products

Below image shows the relationship between the various tables in the dataset

2. Datatype of all the columns -

Query -

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

Output -

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

Comments-

Same steps were performed for all the other tables in the dataset and the datatype for all the columns were correct and no mismatch was found

3. Time Range between which the orders were place

Query -

```sql
SELECT
  MIN(DATE(order_purchase_timestamp)) AS First_Order_Date,
  MAX(DATE(order_purchase_timestamp)) AS Latest_Order_Date,
  DATE_DIFF(MAX(DATE(order_purchase_timestamp)),MIN(DATE(order_purchase_timestamp)),month)
  AS Time_Range_Months
FROM `target_project.orders`
```

Output -

Row	First_Order_Date	Latest_Order_Date	Time_Range_Months
1	2016-09-04	2018-10-17	25

Comments-

The given dataset contains information for the orders placed between 2016 and 2018 spanning over 25 months in total.

4. Number of Cities and States in the dataset.

Query 1-

```sql
SELECT
  COUNT(DISTINCT customer_city) AS No_of_cities,
  COUNT(DISTINCT customer_state) AS No_of_states
FROM `target_project.customers`
```

Output 1 -

Row	No_of_cities ▼	No_of_states ▼
1	4119	27

Query 2-

```sql
SELECT
  COUNT(DISTINCT geolocation_city) AS No_of_cities,
  COUNT(DISTINCT geolocation_state) AS No_of_states
FROM `target_project.geolocation`
```

Output 2 -

Row	No_of_cities ▼	No_of_states ▼
1	8011	27

Insights-

We have geolocation data for about 8011 cities and 27 states in our dataset. However, after looking at the customers table we find that out of those 8011 cities our customer base is present in only 4119 cities indicating an untapped customer base of 50% in those remaining 3892 cities

In-depth Exploration

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

Query -

```sql
-- 1st cte for calculating yearly order count
WITH yearly_sales AS (
  SELECT
    EXTRACT(Year FROM o.order_purchase_timestamp) AS year,
    COUNT(o.order_id) AS order_count
  FROM `target_project.orders` AS o
  GROUP BY EXTRACT(year FROM o.order_purchase_timestamp)),

-- 2nd cte for calculating yearly order count along with previous year order count
    prev_year_sales AS(
  SELECT
    year,order_count,
    LAG(order_count) OVER(ORDER BY year) AS prev_year_count
  FROM yearly_sales ORDER BY year)

-- Final Query for calculating percent increase in order count
  SELECT p.*,ROUND(((order_count-prev_year_count)/prev_year_count) * 100) AS percent_increase
  FROM prev_year_sales p
```

Output -

Row	year	order_count	prev_year_count	percent_increase
1	2016	329	*null*	*null*
2	2017	45101	329	13609.0
3	2018	54011	45101	20.0

Insights-

- The number of orders placed is **increasing** over the past years, with a staggering increase of **13609 %** in 1st year and then a **20%** increase in 2nd year.
- This data showcases the company's ability to attract and retain customers and highlights the effectiveness of its strategies.

2. Is there some kind of Monthly seasonality in terms of the no. of orders being placed?

Query -

```
SELECT * FROM (
  SELECT EXTRACT(month FROM order_purchase_timestamp) AS month,
        COUNT(order_id) AS order_count
  FROM `target_project.orders`
  GROUP BY EXTRACT(month FROM order_purchase_timestamp))a
ORDER BY a.month
```

Output -

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
11	11	7544
12	12	5674

Visualisation -
On next page

Order Count by Month
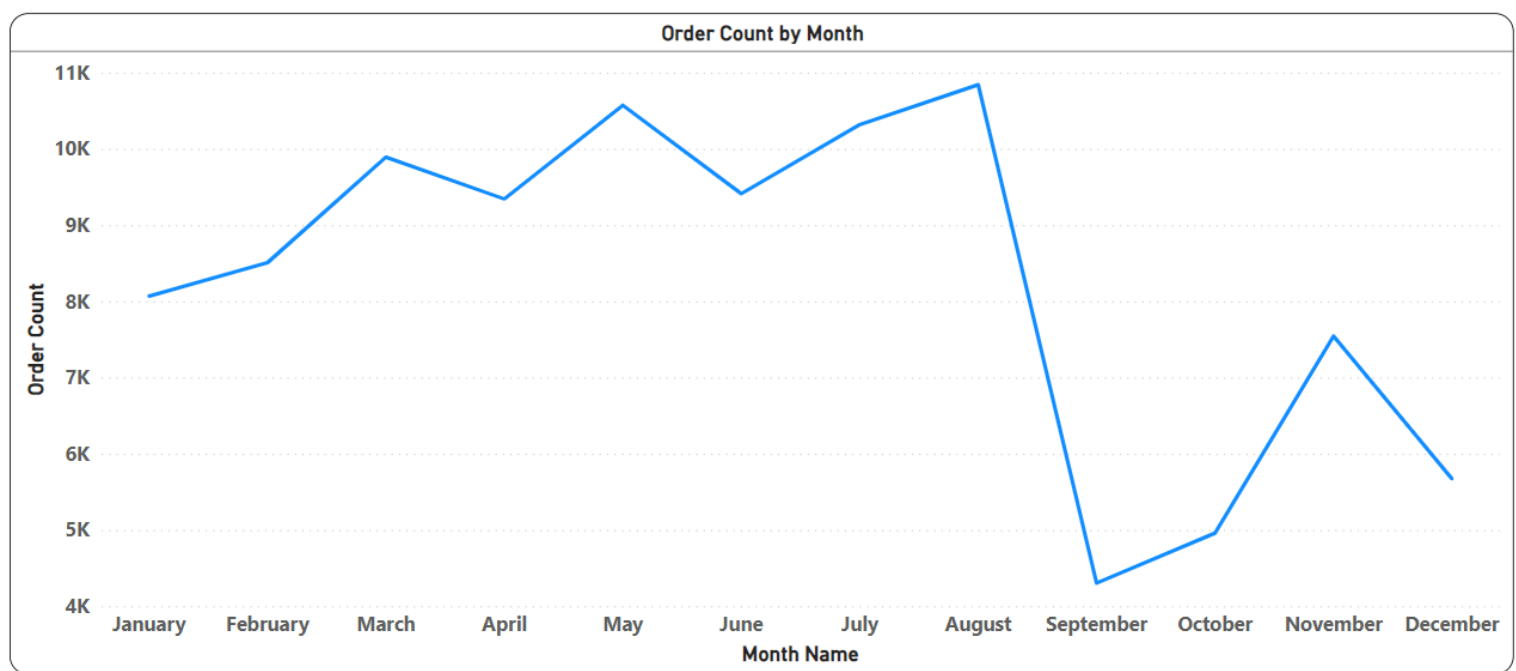
Insights-

- For the first eight months of the year, we observe a robust order count, surpassing 8,000. However, during the last four months of the year, it declines below 8,000.
- The order count begins to rise in March and continues to increase until August, which represents the peak month for order counts throughout the entire year.
- May, July, and August exhibit exceptionally high order counts, exceeding 10,000.
- September and October are the months with the lowest order counts.

Recommendation -

Weak Months:
There is a noticeable seasonal variation in the order counts, with a drop during the last four months of the year. It could be beneficial to focus on cost optimization, inventory management, and targeted marketing strategies to maintain profitability during these months.Collaborating with local influencers or running campaigns can help generate buzz and attract customers. Company should consider offering discounts or exclusive deals during these months.

March to August Peak:
The period from March to August indicates a strong demand. This suggests that the company should ensure timely replenishment to avoid stockouts, efficient supply chain management and enhance customer service to meet the increased demand. Company can also leverage this peak period to introduce new product lines, collaborate with popular brands, and offer exclusive promotions to further capitalise on the heightened demand.

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 -

```sql
WITH purchase AS (
  SELECT CASE
    WHEN EXTRACT(hour FROM order_purchase_timestamp) BETWEEN 0 AND 6 THEN 'Dawn'
    WHEN EXTRACT(hour FROM order_purchase_timestamp) BETWEEN 7 AND 12 THEN 'Mornings'
    WHEN EXTRACT(hour FROM order_purchase_timestamp) BETWEEN 13 AND 18 THEN 'Afternoon'
    ELSE 'Night' END AS purchase_time,
    order_id FROM `target_project.orders` )
SELECT purchase_time, COUNT(order_id) AS order_count
FROM purchase GROUP BY purchase_time ORDER BY order_count
```

Output -

Row	purchase_time	order_count
1	Dawn	5242
2	Mornings	27733
3	Night	28331
4	Afternoon	38135

Visualisation -
On next page

Order count by time
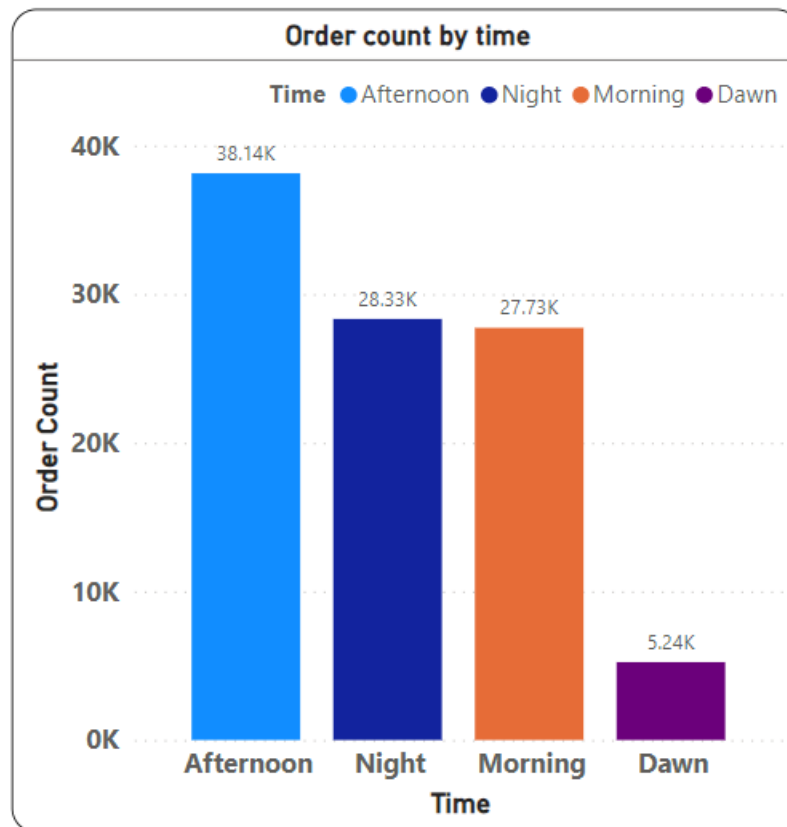
Insights and Recommendation-

- **Afternoon** (13-18 hrs) is the most popular time of the day for customers to place their orders, followed by the **night** (19-23 hrs) and the **morning** (7-12 hrs). **Dawn** (0-6 hrs) has the least number of orders.

Peak Hours:
The majority of customers place their orders during the time span of 7 - 23 hrs. Company should focus its marketing efforts during these time periods to maximise customer engagement and conversion. Running targeted advertisements, promotions, or special offers during these peak hours to attract more customers and increase sales.The company should ensure that there is sufficient staffing and customer support available during these periods

Dawn Hours:
The lowest order count is observed during the early morning hours.Company may not need to allocate significant resources during this time.Also, company can offer discounts or incentives for orders placed during the dawn hours to attract more customers.

Evolution of E-commerce orders in the Brazil region:

1.Month on month no. of orders placed in each state.

Query -

```
-- cte for getting required data in single table
WITH cte1 AS (
  SELECT EXTRACT(year FROM order_purchase_timestamp) AS year,
    EXTRACT(month FROM order_purchase_timestamp) AS month,
    customer_State,order_id
    FROM `target_project.orders` o JOIN `target_project.customers`c
    ON o.customer_id = c.customer_id)

-- query for extracting the required information
SELECT year, month, customer_state, COUNT(order_id) AS order_count
  FROM cte1 GROUP BY year,month,customer_state
  ORDER BY customer_state,year,month
```

Output (First 10 Rows) -

Row	year	month	customer_state	order_count
1	2017	1	AC	2
2	2017	2	AC	3
3	2017	3	AC	2
4	2017	4	AC	5
5	2017	5	AC	8
6	2017	6	AC	4
7	2017	7	AC	5
8	2017	8	AC	4
9	2017	9	AC	5
10	2017	10	AC	6

Visualisation -
On next page

Monthly order count per state
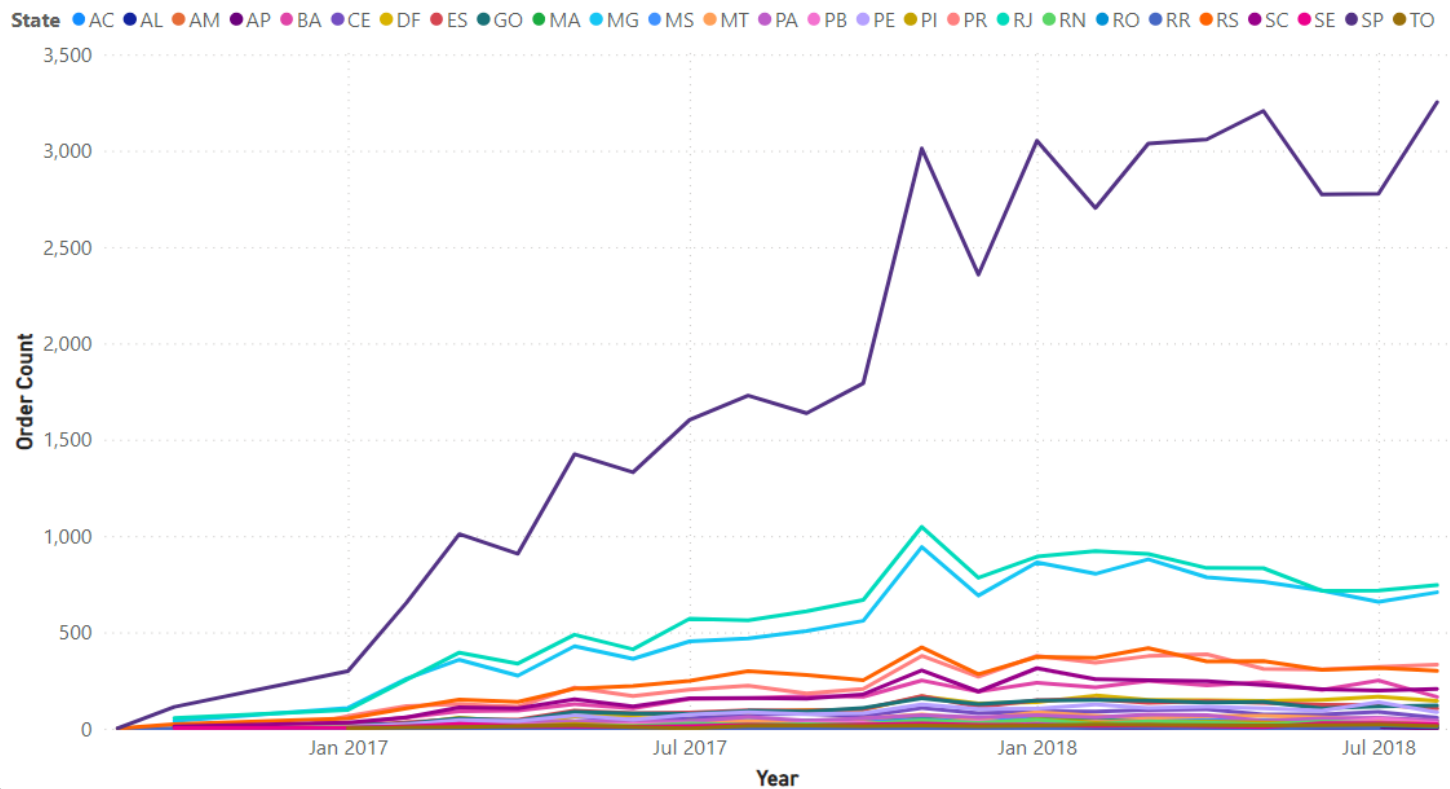
Insights and Recommendation-

- Looking at the data, there is an overall growth trend in order count. Company experienced an increase in sales during this period, suggesting positive business growth.

Focus on Strengthening Presence in Sao Paulo :
Given that a large chunk of orders came from the State of Sao Paulo and there is an overall increase in order count, it indicates a strong and growing customer base in this region.Company could think of allocating additional resources and marketing efforts to expand its presence in Sao Paulo

Investigate Slow Growth in Rio De Janeiro and Minas Gerais :
While Rio De Janeiro and Minas Gerais contribute significantly to the order count, the data suggests a slow growth rate since November 2017.It could be due to increased competition, changes in consumer preferences, or other market dynamics. Company should conduct a detailed analysis to understand the slow growth and adjust its marketing strategies based on it to regain the momentum.

Explore Opportunities in Other States:
Although other states contribute less to the total order count and have shown slow order growth, analysing the market potential in these states can uncover untapped opportunities and result in a boost to overall order growth.

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

Query-

```sql
-- For calculating customer count by each state
WITH cte AS
(SELECT
  customer_state AS state,
  COUNT(DISTINCT customer_unique_id) AS customer_count,
FROM `target_project.customers` GROUP BY customer_state
)

-- For calculating customer count along with percent of total by each state
SELECT
  state,customer_count,
  ROUND(customer_count/(SELECT COUNT(DISTINCT customer_unique_id) total_count FROM
`target_project.customers`)*100) AS percent_of_total
FROM cte ORDER BY customer_count DESC
```

Output (First 10 Rows)-

Row	state	customer_count	percent_of_total
1	SP	40302	42.0
2	RJ	12384	13.0
3	MG	11259	12.0
4	RS	5277	5.0
5	PR	4882	5.0
6	SC	3534	4.0
7	BA	3277	3.0
8	DF	2075	2.0
9	ES	1964	2.0
10	GO	1952	2.0

Visualisation
On next page-
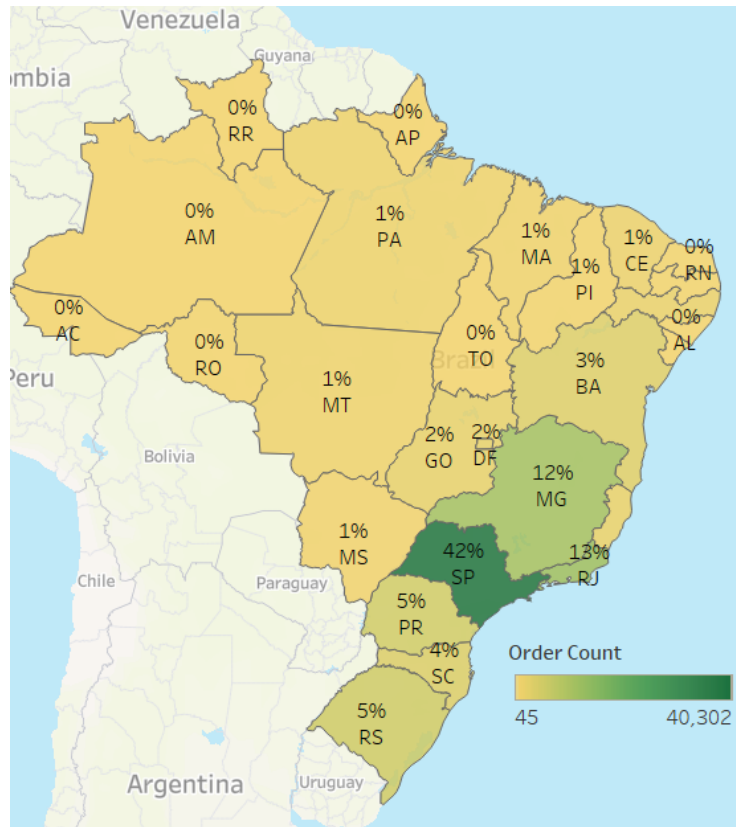
Insights and Recommendation-

As expected the customer spread follows the state wise order count trend derived in the previous analysis, The southern and southeastern regions have higher customer presence, while the northern and northeastern regions have relatively lower customer presence.

Southern and Southeastern States:
The southeastern states of Brazil namely Sao Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) collectively contribute 67% of the total customer. Additionally, The states of Rio Grande do Sul (RS), Parana (PR), and Santa Catarina (SC) have a notable customer presence with a combined total of 14%. Overall, this region represents 81% of the company's customer base, indicating a strong foothold..

The company can leverage this advantage to expand its customer presence beyond Sao Paulo in this southeastern and south region.

Other States:
While the majority of orders are concentrated in the aforementioned states, there is still a huge untapped customer base in other parts of brazil. The company has an opportunity to tap into this potential by investing in logistics infrastructure and improving the delivery network. Additionally, implementing targeted marketing strategies tailored to these regions can help attract and engage customers in these areas of Brazil.

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

1.The % increase in the cost of orders from 2017 to 2018 (Months between Jan to Aug only).

Query-

```sql
-- 1st cte for getting the data in one table
WITH cte1 AS (
  SELECT EXTRACT(year FROM order_purchase_timestamp) AS year,
    EXTRACT(month FROM order_purchase_timestamp) AS month,
    payment_value
  FROM `target_project.orders` o JOIN `target_project.payments`p ON o.order_id = p.order_id),

--2nd cte for getting the total payment value for both years
cte2 AS (
SELECT year,
  SUM(payment_value) AS total_payment,
  LAG(SUM(payment_value),1) OVER(ORDER BY year) AS prev_year
FROM cte1 WHERE month BETWEEN 1 AND 8
GROUP BY year ORDER BY year)

-- Query FOR getting % increase from 2017 to 2018
SELECT ROUND(((total_payment - prev_year)/total_payment)*100) AS percentage_increase
FROM cte2 WHERE prev_year IS NOT NULL
```

Output (First 10 Rows)-

Row	percentage_increase
1	137.0

Insights -

137% increase in payment value from 2017 to 2018 -
The substantial increase in payment value suggests that the business experienced strong revenue growth between 2017 and 2018. This growth can be attributed to various factors, such as increased sales volume, higher average order value, successful marketing and sales strategies

2.The Total & Average value of order price for each state.

Query-

```sql
SELECT
  customer_state,
  ROUND(SUM(payment_value)) AS total_order_value,
  ROUND(AVG(payment_value)) AS average_value
FROM `target_project.payments` p
LEFT JOIN `target_project.orders` o ON p.order_id = o.order_id
LEFT JOIN `target_project.customers` c ON o.customer_id = c.customer_id
GROUP BY customer_state ORDER BY total_order_value desc
```

Output (First 10 Rows)-

Row	customer_state	total_order_value	average_value
1	SP	5998227.0	138.0
2	RJ	2144380.0	159.0
3	MG	1872257.0	155.0
4	RS	890899.0	157.0
5	PR	811156.0	154.0
6	SC	623086.0	166.0
7	BA	616646.0	171.0
8	DF	355141.0	161.0
9	GO	350092.0	166.0
10	ES	325968.0	155.0

Visualisation
On next page-

Order Value and Average Order Value by State
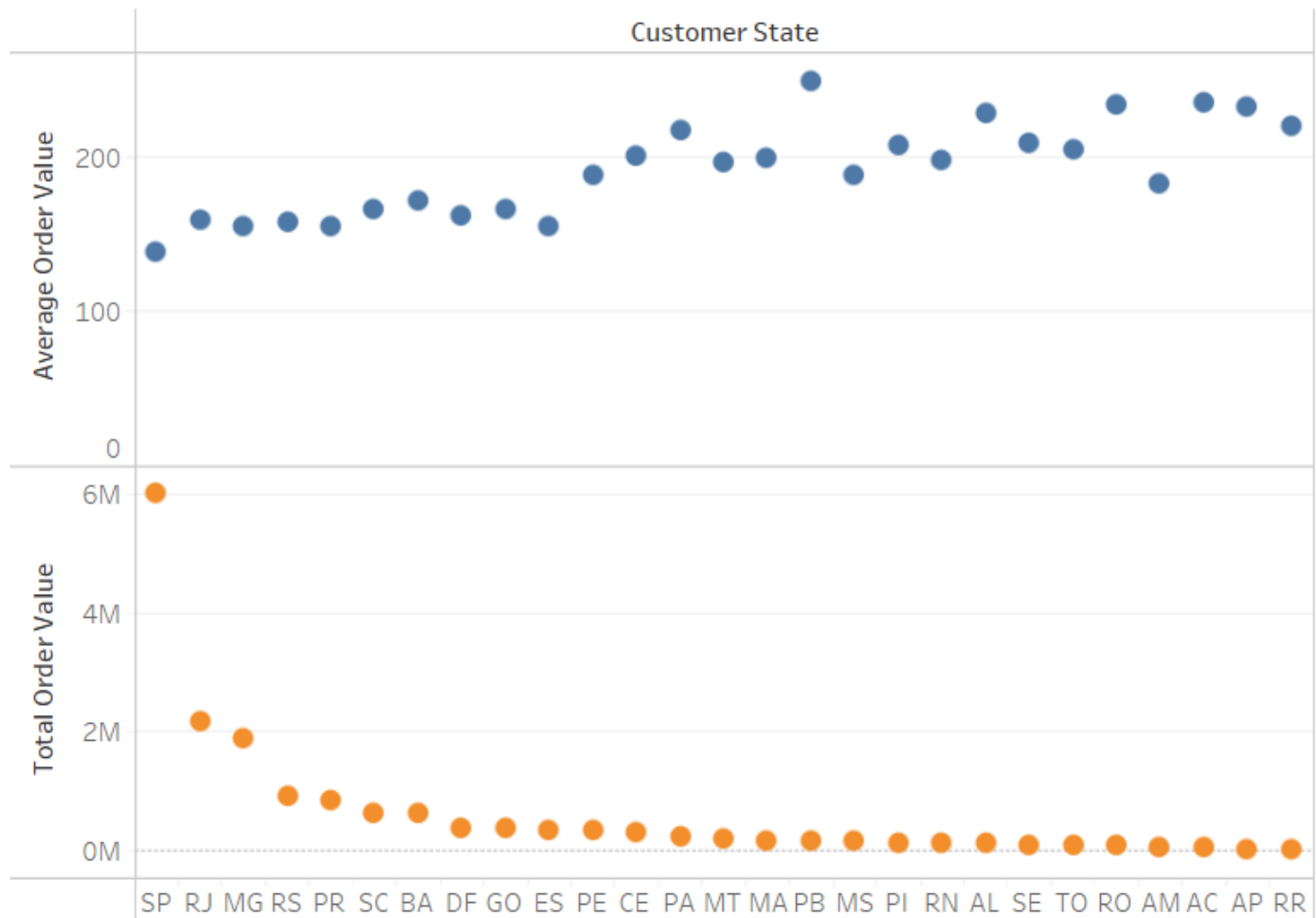
Insights -

The Total & Average value of order price for each state -

States like Sao Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) have higher total order values, indicating a larger customer base. The average order value varies between 248 to 138 currency units. It follows a inverse relationship with total order value for each state i.e. the state with the highest order value has the least average order value and vice versa

3.The Total & Average value of order freight for each state.

Query-

```sql
SELECT
  customer_state,
  ROUND(SUM(freight_value)) AS total_frieght_value,
  ROUND(AVG(freight_value)) AS average_value
FROM `target_project.order_items` oi
LEFT JOIN `target_project.orders` o ON oi.order_id = o.order_id
LEFT JOIN `target_project.customers` c ON o.customer_id = c.customer_id
GROUP BY customer_state ORDER BY total_frieght_value desc
```
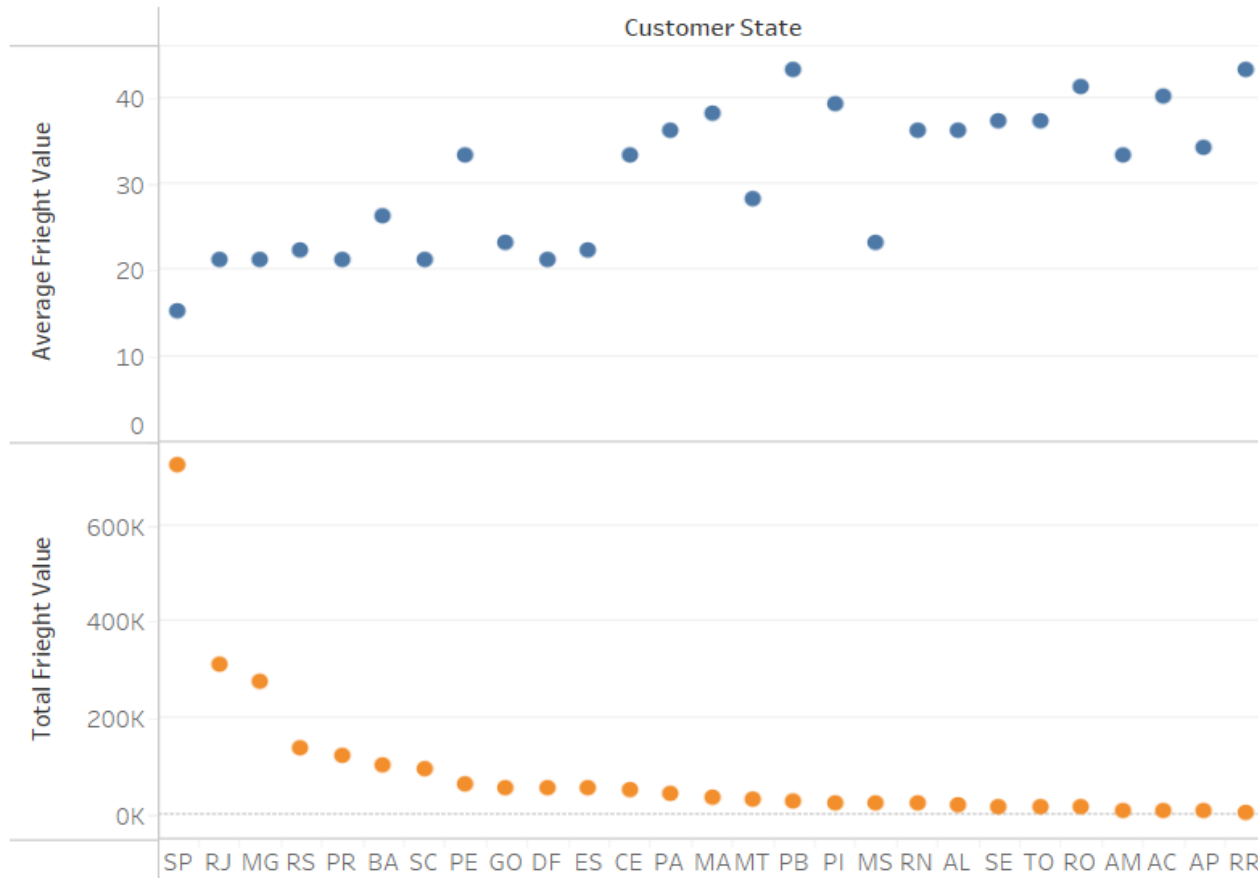
Output (First 10 Rows)-

Row	customer_state	total_frieght_value	average_value
1	SP	718723.0	15.0
2	RJ	305589.0	21.0
3	MG	270853.0	21.0
4	RS	135523.0	22.0
5	PR	117852.0	21.0
6	BA	100157.0	26.0
7	SC	89660.0	21.0
8	PE	59450.0	33.0
9	GO	53115.0	23.0
10	DF	50625.0	21.0

Visualisation -
On next page-

Frieght Value and Average Frieght Value by State

Insights and Recommendation-

The Total & Average value of order freight for each state. -

The data gives insights into the pricing and shipping strategy of the company.States with higher average freight charges tend to have lower order values. This indicates an opportunity for the company to consider expanding its presence in these states by investing in logistics and targeting the untapped customer base. By optimising shipping costs or introducing incentives, the company can attract more customers and increase order values in these regions.

Analysis based on sales, freight and delivery time.

1. No. of days taken to deliver each order from the order's purchase date
    Also, the difference (in days) between the estimated & actual delivery
    date of an order.

Query-

```sql
SELECT
  order_id,
  DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,day) AS time_to_deliver,
  DATE_DIFF(order_estimated_delivery_date,order_delivered_customer_date,day) AS diff_estimated_delivery
FROM `target_project.orders`
WHERE order_delivered_customer_date IS NOT NULL
ORDER BY time_to_deliver DESC
```

Output (First 10 Rows)-

Row	order_id	time_to_deliver	diff_estimated_delive
1	ca07593549f1816d26a572e06...	209	-181
2	1b3190b2dfa9d789e1f14c05b...	208	-188
3	440d0d17af552815d15a9e41a...	195	-165
4	0f4519c5f1c541ddec9f21b3bd...	194	-161
5	285ab9426d6982034523a855f...	194	-166
6	2fb597c2f772eca01b1f5c561b...	194	-155
7	47b40429ed8cce3aee9199792...	191	-175
8	2fe324febf907e3ea3f2aa9650...	189	-167
9	2d7561026d542c8dbd8f0daea...	188	-159
10	437222e3fd1b07396f1d9ba8c...	187	-144

Visualisation -
On next page-

Distribution of Delivery Time

Delivery Type
- Early
- Late
- On Time

7% | 4%

89%
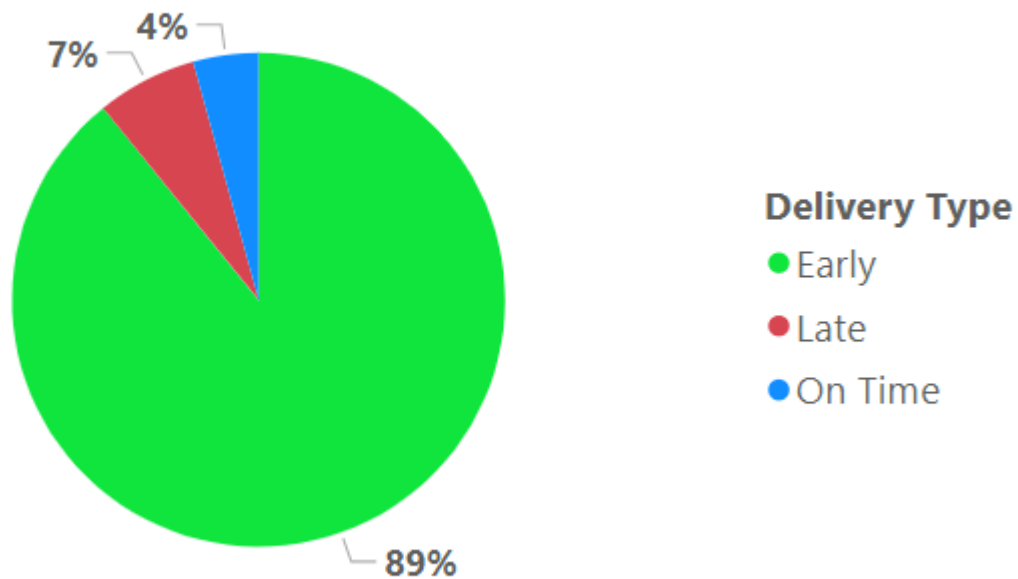
Insights -

- The **mean delivery time** for all orders was **13 days**, while the **median delivery time** was **10 days.** These figures indicate the typical duration it took for orders to be delivered.
- The company performed commendably in terms of order delivery. Specifically, **89%** of the total orders were **delivered ahead** of the estimated delivery date, demonstrating the company's efficiency in fulfilling customer expectations. Additionally, **4%** of the orders were delivered precisely **on the estimated delivery date.**
- A small portion of the orders, comprising **7%**, experienced **delays in delivery**. While this percentage represents a relatively minor occurrence, it highlights an area for potential improvement in order fulfilment operations.

2. The top 5 states with the highest & lowest average freight value.

Query-

```sql
-- cte for getting the required data in one table
WITH cte AS
(SELECT
  customer_state,
  ROUND(AVG(freight_value),2) AS average_freight_value
FROM `target_project.order_items` oi
LEFT JOIN `target_project.orders` o ON oi.order_id = o.order_id
LEFT JOIN `target_project.customers` c ON o.customer_id = c.customer_id
GROUP BY customer_state),

-- cte for getting top 5 low freight charges state
bottom_5 AS (SELECT *,"low freight charges" as category from cte ORDER BY average_freight_value LIMIT 5),

-- cte for getting top 5 high freight charges state
top_5 AS (SELECT *,"high freight charges" from cte ORDER BY average_freight_value DESC LIMIT 5)

-- final result
SELECT * FROM bottom_5
UNION ALL
SELECT * FROM top_5 ORDER BY average_freight_value
```

Output-

Row	customer_state	average_freight_value	category
1	SP	15.15	low freight charges
2	PR	20.53	low freight charges
3	MG	20.63	low freight charges
4	RJ	20.96	low freight charges
5	DF	21.04	low freight charges
6	PI	39.15	high freight charges
7	AC	40.07	high freight charges
8	RO	41.07	high freight charges
9	PB	42.72	high freight charges
10	RR	42.98	high freight charges

3. The top 5 states with the highest & lowest average delivery time.

Query-

```sql
-- cte for getting the required data in one table
WITH cte AS
(SELECT
  customer_state,
  ROUND(AVG(DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,day))) AS time_to_deliver,
FROM `target_project.orders` o
LEFT JOIN `target_project.customers` c ON o.customer_id = c.customer_id
GROUP BY customer_state),

-- cte for getting top 5 states
bottom_5 AS (SELECT *,"Top 5" as category from cte ORDER BY time_to_deliver LIMIT 5),

-- cte for getting top 5 states
top_5 AS (SELECT *,"Bottom 5" from cte ORDER BY time_to_deliver DESC LIMIT 5)

-- final result
SELECT * FROM bottom_5
UNION ALL
SELECT * FROM top_5 ORDER BY time_to_deliver
```
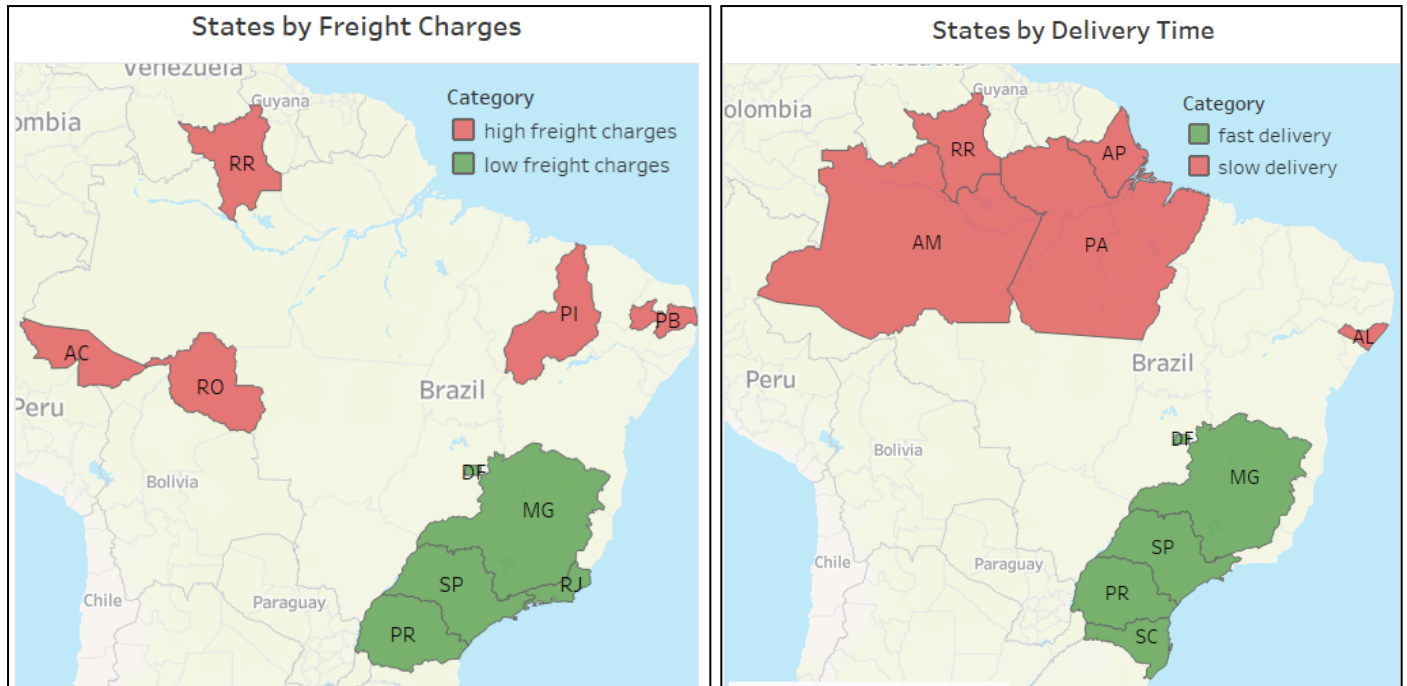
Output-

Row	customer_state	time_to_deliver	category
1	SP	8.0	fast delivery
2	MG	12.0	fast delivery
3	PR	12.0	fast delivery
4	DF	13.0	fast delivery
5	SC	14.0	fast delivery
6	PA	23.0	slow delivery
7	AL	24.0	slow delivery
8	AM	26.0	slow delivery
9	AP	27.0	slow delivery
10	RR	29.0	slow delivery

Visualisation -

Insights based on above analysis of freight value and delivery time by each state -

Freight Charges -
- The states with low average freight charges include SP (Sao Paulo), PR (Parana), MG (Minas Gerais), RJ (Rio de Janeiro), and DF (Distrito Federal). It is a good indicator as states like SP (Sao Paulo),MG (Minas Gerais) and RJ (Rio de Janeiro) are top 3 states by order value.By maintaining low freight charges in these states, we can enhance customer satisfaction and retention.
- Conversely, there are states with high average freight charges, such as PI (Piauí), AC (Acre), RO (Rondonia), PB (Paraíba), and RR (Roraima). Notably, these states have a lower order count. As a result, their higher freight charges can be attributed to the logistical challenges associated with serving these areas.

Delivery Time -
- The states with relatively faster delivery times include SP (Sao Paulo), MG (Minas Gerais), PR (Parana), DF (Distrito Federal), and SC (Santa Catarina). Remarkably, these states also have low freight charges, indicating a significant logistical advantage of the company in these regions.
- On the other hand, there are states with longer delivery times, namely PA (Para), AL (Alagoas), AM (Amazonas), AP (Amapa), and RR (Roraima). These states are located in the northern region, which presents geographical challenges due to their distance and lower order count. Consequently, the delivery time in these areas tends to be higher.

4. The top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

Query-

```sql
-- cte for getting the required data in one table
WITH cte AS
(SELECT
  customer_state,
  ROUND(AVG(DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,day))) AS
avg_time_to_deliver,
  ROUND(AVG(DATE_DIFF(order_estimated_delivery_date,order_purchase_timestamp,day))) AS
avg_estimated_time_to_deliver
FROM `target_project.orders` o
LEFT JOIN `target_project.customers` c ON o.customer_id = c.customer_id
WHERE order_delivered_customer_date IS NOT NULL
GROUP BY customer_state)

-- query for fetching result
SELECT
  customer_state,
  (avg_estimated_time_to_deliver - avg_time_to_deliver) AS delivery_diff
  FROM cte ORDER BY delivery_diff DESC LIMIT 5
```

Output -

Row	customer_state	delivery_diff
1	AC	20.0
2	AP	19.0
3	RO	19.0
4	AM	19.0
5	RR	17.0

Visualisation -

24
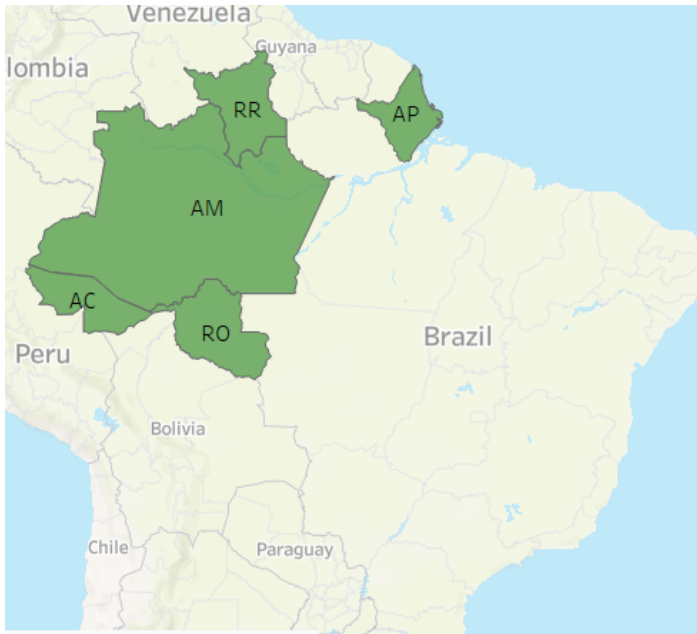
Insights -

Based on the provided data, it can be inferred that the company maintains a substantial buffer for estimated delivery times in northern states due to their geographical distance. However, it is noteworthy that the company consistently fulfills its delivery commitments well in advance of the estimated delivery dates.

Analysis based on the payments

1. Month on month no. of orders placed using different payment types.

Query-

```sql
-- 1st cte for getting the required details in one table
WITH cte AS (
  SELECT EXTRACT(year FROM order_purchase_timestamp) AS year,
    EXTRACT(month FROM order_purchase_timestamp) AS month,
    payment_type
  FROM `target_project.payments` p LEFT JOIN `target_project.orders` o ON p.order_id = o.order_id ),

-- 2nd cte for getting the details of credit card count
  credit_card AS (
  SELECT year,month,COUNT(payment_type) AS credit_card_count
  FROM cte WHERE payment_type = 'credit_card'
  GROUP BY year,month,payment_type),

-- 3rd cte for getting the details of voucher count
  voucher AS (
  SELECT year,month,COUNT(payment_type) AS voucher_count
  FROM cte WHERE payment_type = 'voucher'
  GROUP BY year,month,payment_type),

-- 4th cte for getting the details of debit card count
  debit_card AS (
  SELECT year,month, COUNT(payment_type) AS debit_card_count
  FROM cte WHERE payment_type = 'debit_card'
  GROUP BY year,month,payment_type),

-- 5th cte for getting the details of upi count
  upi AS (
  SELECT year,month,COUNT(payment_type) AS upi_count
  FROM cte WHERE payment_type = 'UPI'
  GROUP BY year,month,payment_type)

-- query for getting final result
SELECT
  IFNULL(c.year,v.year) AS year,
  IFNULL(c.month,v.month) AS month,
  credit_card_count,debit_card_count,upi_count,voucher_count
FROM credit_card c
FULL OUTER JOIN debit_card d ON c.year = d.year AND c.month = d.month
FULL OUTER JOIN upi u ON c.year = u.year AND c.month = u.month
FULL OUTER JOIN voucher v ON c.year = v.year AND c.month = v.month
ORDER BY year,month
```

Output (First 10 Rows) -

Row	year	month	credit_card_count	debit_card_count	upi_count	voucher_count
1	2016	9	3	null	null	null
2	2016	10	254	2	63	23
3	2016	12	1	null	null	null
4	2017	1	583	9	197	61
5	2017	2	1356	13	398	119
6	2017	3	2016	31	590	200
7	2017	4	1846	27	496	202
8	2017	5	2853	30	772	289
9	2017	6	2463	27	707	239
10	2017	7	3086	22	845	364

Visualisation -

Month on month no. of orders placed using different payment types
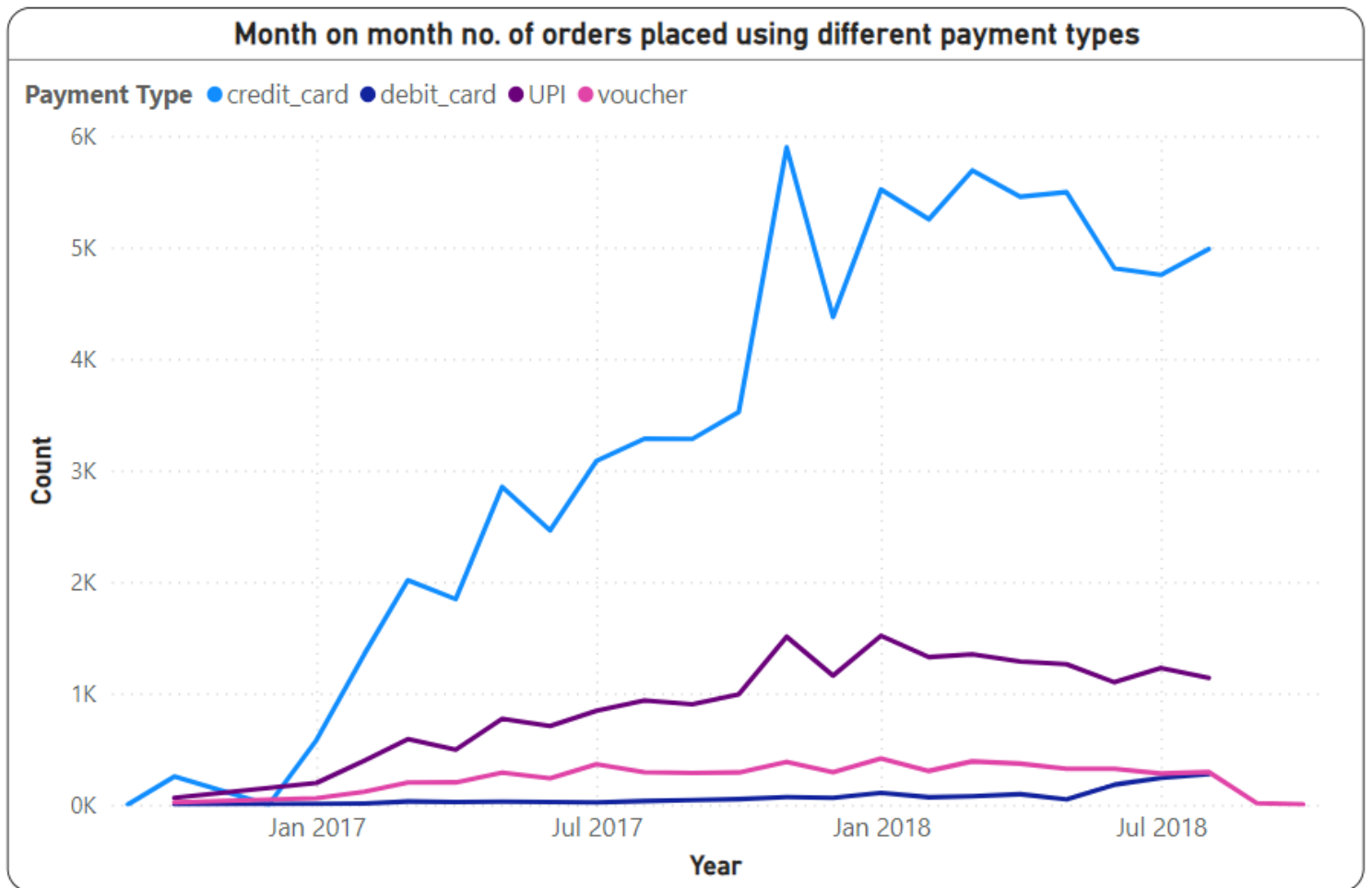
Insights and recommendation-

Credit Card -
- Credit card payment method is the most preferred payment method by the customers with **74%** of the total payments during this time done by credit card
- By looking at month on month data reveals a substantial growth in the number of payments made through credit cards.

UPI -
- UPI is the second most preferred method accounting for **19%** of the total payments
- It has also experienced significant growth, indicating a shift towards digital payment options among customers.

Voucher and Debit Card
- Voucher is the 3rd most preferred type of payment accounting for **6%** of the total payments
- Debit cards, on the other hand, are the least favoured option, comprising a mere **1%** of the total payments.

Recommendation -
Considering the insights gathered, the company could leverage incentives or attractive offers specifically targeting credit card and UPI-based payments. By doing so, the aim is to enhance sales and encourage greater adoption of these payment methods, which are preferred by a substantial portion of customers.

2. No. of orders placed on the basis of the payment instalments that have been paid.

Query-

```sql
SELECT
  payment_installments,
  COUNT(DISTINCT order_id) AS count
FROM `target_project.payments`
WHERE payment_installments > 1 AND payment_sequential >=1
GROUP BY payment_installments
ORDER BY count DESC
```

Output(First 10 Rows) -

Row	payment_installment	count
1	2	12389
2	3	10443
3	4	7088
4	10	5315
5	5	5234
6	8	4253
7	6	3916
8	7	1623
9	9	644
10	12	133

Insights and Recommendations

Most popular EMI instalments:
- The table indicates that EMI instalments of **2** and **3** are the most popular options, with counts of 12,389 and 10,443, respectively. This suggests that customers prefer shorter repayment durations for their purchases
- The number of orders tends to decrease as the number of instalments increases beyond 4.
- Company can focus on promoting and offering more attractive incentives for shorter-term EMI options to align with customer preferences..

Potential demand for longer-term instalments:
- Although the number of orders decreases with increasing instalments, there is still a notable number of orders for instalments beyond 4, such as 10, 5, and 8.
- This suggests that there is a segment of customers who prefer longer-term repayment plans. Company could explore strategies to target and cater to this specific segment