

# PROJECT - 1

## Business Case: Target SQL

By- Mayank Jain

DSML (Jul22) Int

[mayankjain281@gmail.com](mailto:mayankjain281@gmail.com)

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

### 1.1. Data type of columns in a table

I used order to demonstrate the results of this question.

Field name	Type
order_id	STRING
customer_id	STRING
order_status	STRING
order_purchase_timestamp	TIMESTAMP
order_approved_at	TIMESTAMP
order_delivered_carrier_date	TIMESTAMP
order_delivered_customer_date	TIMESTAMP
order_estimated_delivery_date	TIMESTAMP

### 1.2. Time period for which the data is given

The time period for which the given data is:

4 September 2016 - 17 October 2018

```
SELECT
  MIN(DATE (order_purchase_timestamp)) AS startdate,
  MAX(DATE (order_purchase_timestamp)) AS enddate,
  ABS(DATE_DIFF ( DATE(MIN(order_purchase_timestamp)), DATE(MAX(order_purchase_timestamp)),MONTH)) AS
month_duration
FROM
  `norse-sector-364907.target.orders`
```

Row	startdate	enddate	month_duration
1	2016-09-04	2018-10-17	25

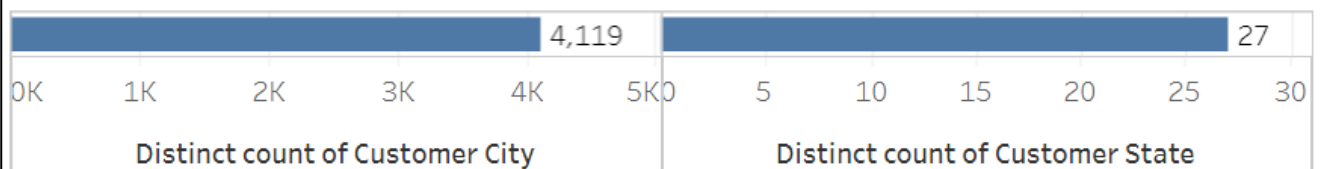
### 1.3. Cities and States covered in the dataset

There are 4119 distinct cities from 27 distinct states covered in dataset. We can get the list all distinct cities with their corresponding states by using this query -

```
SELECT
  DISTINCT(cu.customer_city),
  cu.customer_state,
FROM
  `norse-sector-364907.target.orders` AS od
JOIN
  `norse-sector-364907.target.customers` AS cu
ON
  od.customer_id=cu.customer_id
ORDER BY
  cu.customer_city
```

Row	customer_city	customer_state
1	abadia dos dourados	MG
2	abadiania	GO
3	abaete	MG
4	abaetetuba	PA
5	abaiara	CE
6	abaira	BA
7	abare	BA
8	abatia	PR
9	abdon batista	SC
10	abelardo luz	SC

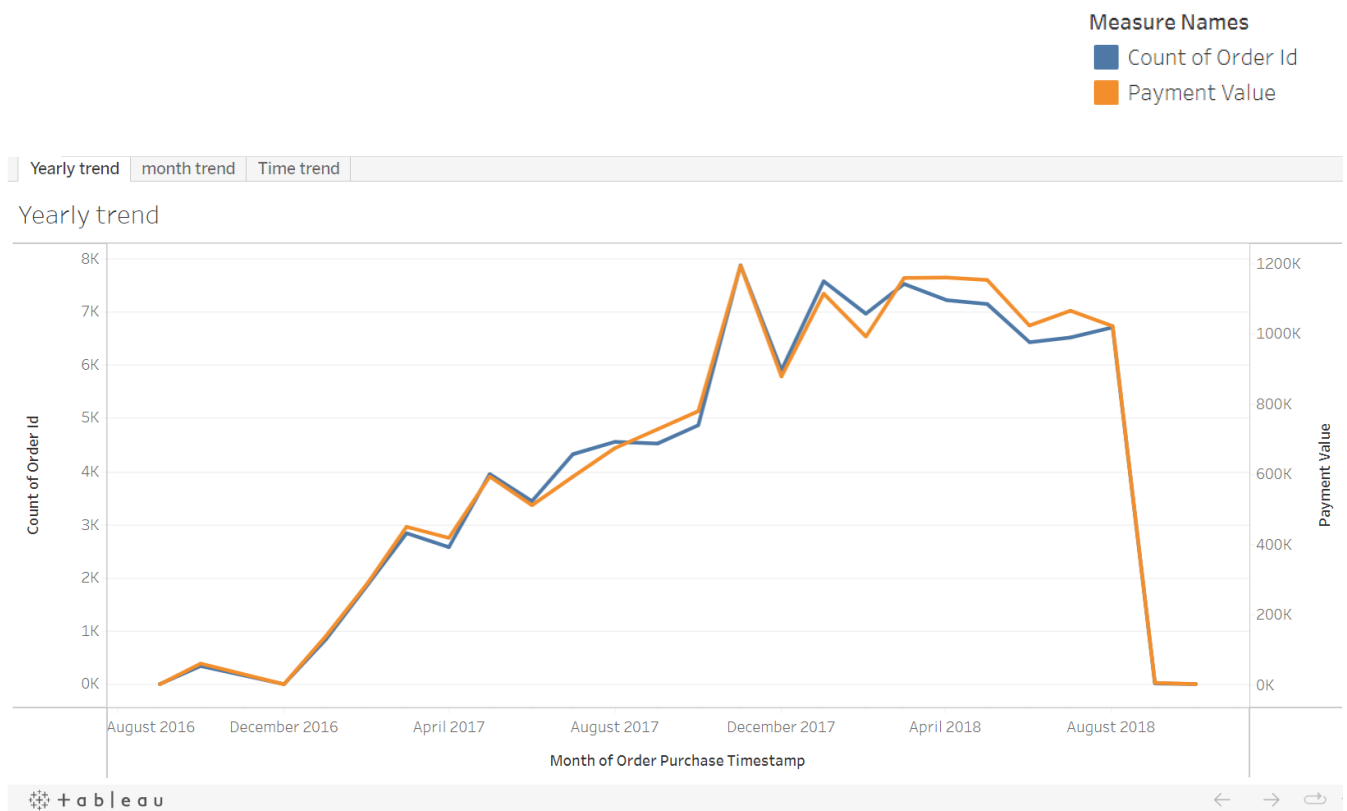
#### Cities and States count



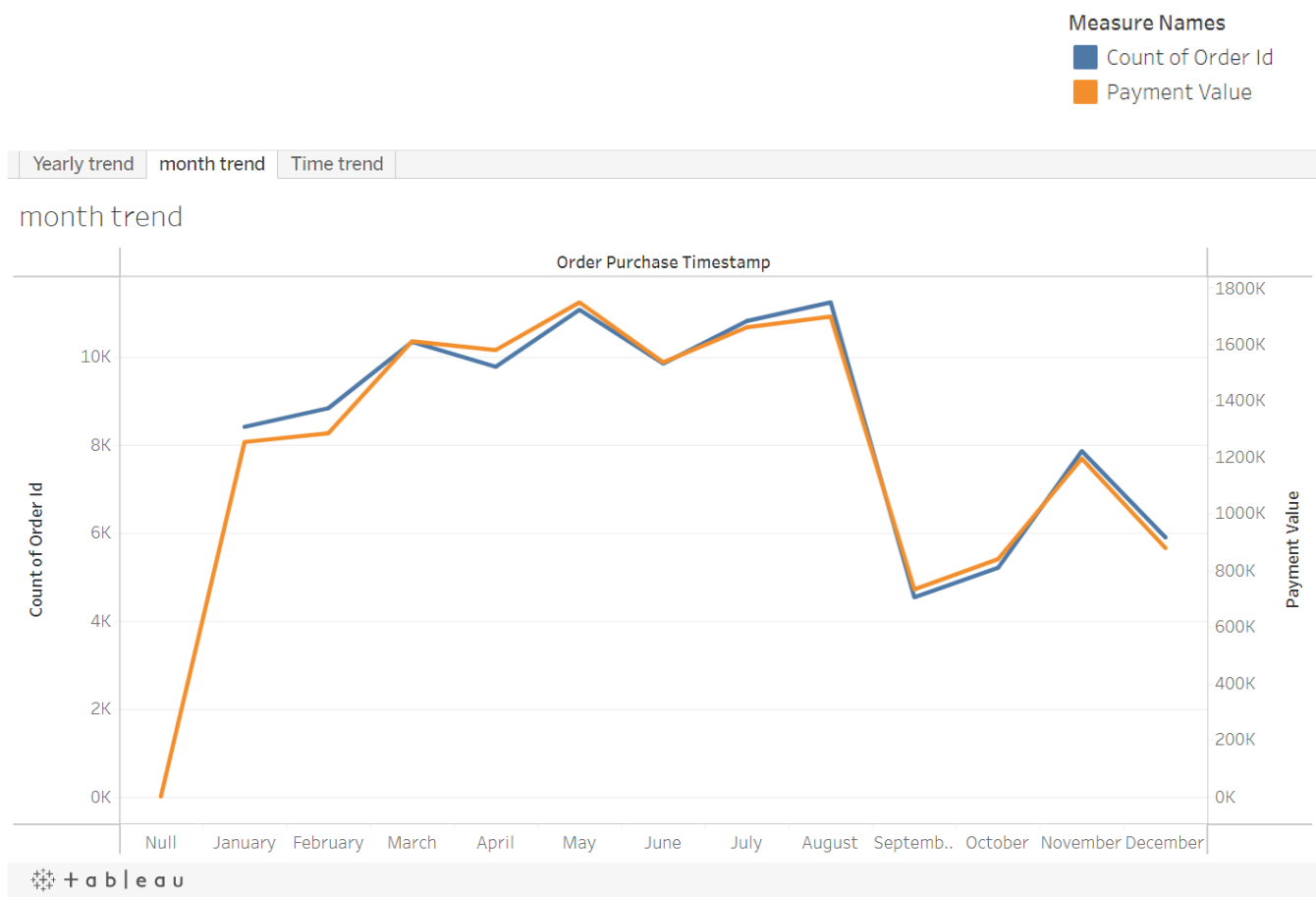
## 2. In-depth Exploration:

### 2.1. Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?

If we do yearly sales analysis, e-commerce in Brazil was in growing trend from September 2016 till august 2017, But since august 2017 till October 2018 it have been falling trend.



If we do monthly sales analysis, We can analyse that there is high sales in summer season. The peak months are March, May and August (sales above 1600k).

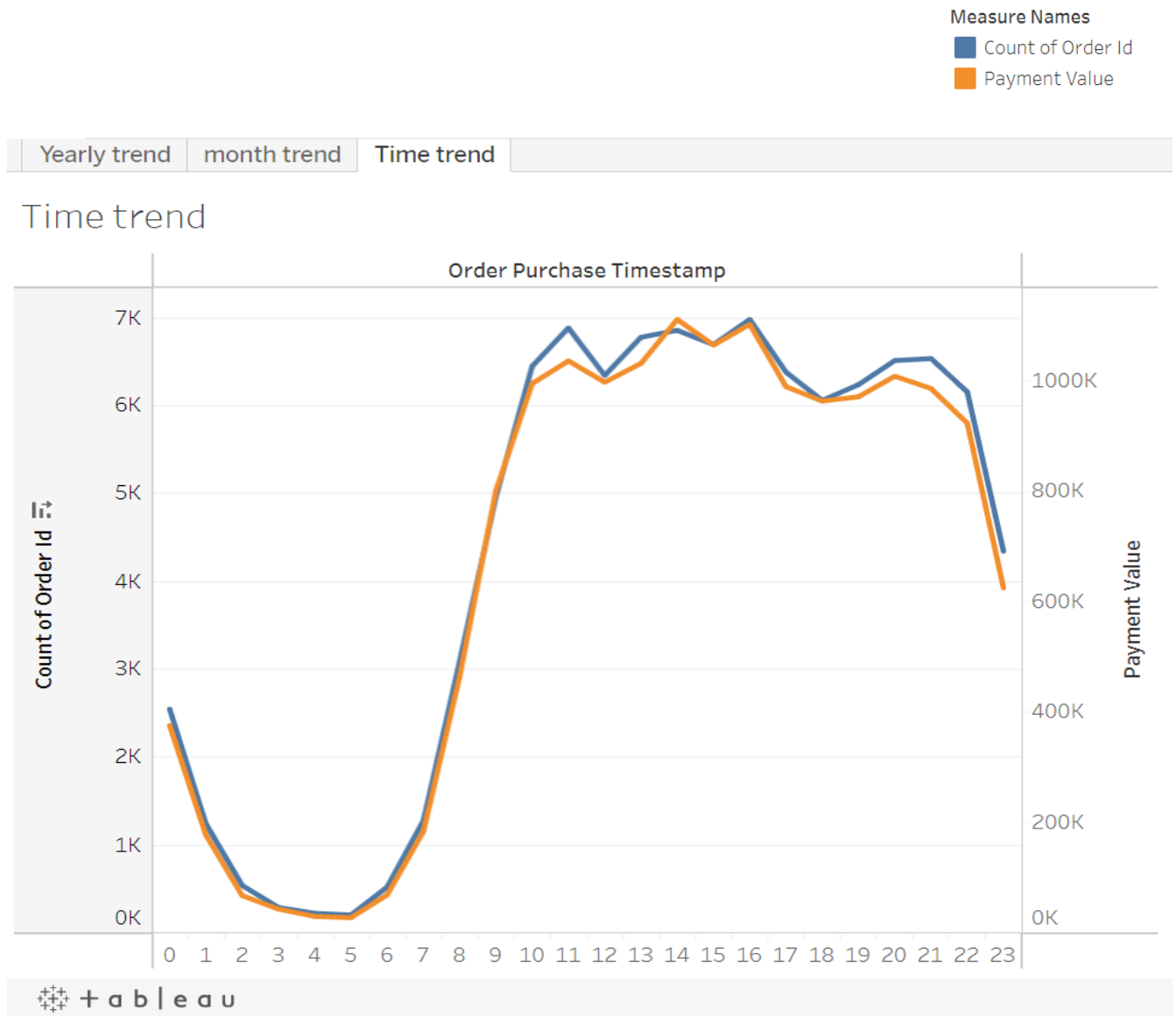


## 2.2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

Brazilian customers more tend to buy in Afternoon.

2pm and 4pm are peak hours for sales, crossed 1.1 Million payment value.

From 10am to 10pm, every hour more than 6000 Order purchase recorded.



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

#### 3.1. Get month on month orders by region, states

(Region information is not available in dataset)

##### a. Aggregated month orders by states. (January to December)

In order to count every product purchased within same order\_id, we considered duplicate order\_id values as well.

```
SELECT
  c.customer_state,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS MONTH,
  COUNT(order_id) AS ORDER_COUNT
FROM target.customers AS c
JOIN target.orders AS o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state, MONTH
ORDER BY c.customer_state, MONTH
```

Row	customer_state	MONTH	ORDER_COUNT
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6
11	AC	11	5
12	AC	12	5
		(total orders in AC state) =	81

b. Orders by states across the range of available years(2016-2018), month-by-month level of granularity.

```
SELECT
  c.customer_state,
  EXTRACT(YEAR FROM o.order_purchase_timestamp) AS YEAR,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS MONTH,
  COUNT(order_id) AS ORDER_COUNT
FROM target.customers AS c
JOIN target.orders AS o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state, YEAR, MONTH
ORDER BY c.customer_state, YEAR, MONTH
```

Row	customer_state	YEAR	MONTH	ORDER_COUNT
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
11	AC	2017	11	5
12	AC	2017	12	5
13	AC	2018	1	6
14	AC	2018	2	3
15	AC	2018	3	2
16	AC	2018	4	4
17	AC	2018	5	2
18	AC	2018	6	3
19	AC	2018	7	4
20	AC	2018	8	3
			(total orders in AC state) =	81



## 3.2. How are customers distributed in Brazil

Here you can see the list of customer distribution in Brazil over 27 states.

```
SELECT
  customer_state,
  COUNT(customer_id) AS customer_count
FROM target.customers
GROUP BY customer_state
ORDER BY customer_state
```

Row	customer_state	customer_count
1	AC	81
2	AL	413
3	AM	148
4	AP	68
5	BA	3380
6	CE	1336
7	DF	2140
8	ES	2033
9	GO	2020
10	MA	747
11	MG	11635
12	MS	715
13	MT	907
14	PA	975
15	PB	536
16	PE	1652
17	PI	495
18	PR	5045
19	RJ	12852
20	RN	485
21	RO	253
22	RR	46
23	RS	5466
24	SC	3637
25	SE	350
26	SP	41746
27	TO	280

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

4.1. Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only)

237% increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only)

```
SELECT
  sale_2017,
  sale_2018,
  Round((sale_2018/sale_2017)*100) AS percentage_increase
FROM
  (SELECT
    SUM(payment_value) AS sale_2017
  FROM `target.payments` AS p
  Join `target.orders` AS o
  ON p.order_id= o.order_id
  WHERE EXTRACT(MONTH FROM order_purchase_timestamp) BETWEEN 1 AND 8 AND EXTRACT(YEAR FROM order_purchase_timestamp) = 2017),
  (SELECT
    SUM(payment_value) AS sale_2018
  FROM `target.payments` AS p
  Join `target.orders` AS o
  ON p.order_id= o.order_id
  WHERE EXTRACT(MONTH FROM order_purchase_timestamp) BETWEEN 1 AND 8 AND EXTRACT(YEAR FROM order_purchase_timestamp) = 2018)
```

Row	sale_2017	sale_2018	percentage_increase
1	3669022.12	8694733.84	237.0

## 4.2. Mean & Sum of price and freight value by customer state

```
SELECT c.customer_state,
       AVG(price+freight_value) AS mean,
       SUM(price+freight_value) AS sum
FROM `target.order_items` AS oi
JOIN `target.orders` AS o
ON oi.order_id=o.order_id
Join `target.customers` AS c
ON c.customer_id= o.customer_id
GROUP BY c.customer_state
ORDER BY c.customer_state
```

Row	customer_state	mean	sum
1	AC	213.801087	19669.7
2	AL	216.7328829	96229.4
3	AM	168.7013939	27835.73
4	AP	198.3268293	16262.8
5	BA	160.9651671	611506.67
6	CE	186.4724628	275606.3
7	DF	146.8119036	353229.44
8	ES	143.9724778	324801.91
9	GO	149.0385469	347706.93
10	MA	183.4611529	151171.99
11	MG	141.378741	1856161.49
12	MS	166.0032601	135956.67
13	MT	176.4634692	186168.96
14	PA	201.5251019	217647.11
15	PB	234.1990199	140987.81
16	PE	178.4261849	322237.69
17	PI	199.5060517	108132.28
18	PR	139.5357909	800935.44
19	RJ	146.078742	2129681.98
20	RN	192.6182987	101895.08
21	RO	207.0432374	57558.02
22	RR	193.5503846	10064.62
23	RS	142.0732574	885826.76
24	SC	146.1239464	610213.6
25	SE	189.6943377	73032.32
26	SP	124.8009046	5921678.12
27	TO	194.7759365	61354.42

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

### 5.1. Create columns:

- $\text{time\_to\_delivery} = \text{order\_purchase\_timestamp} - \text{order\_delivered\_customer\_date}$
- $\text{diff\_estimated\_delivery} = \text{order\_estimated\_delivery\_date} - \text{order\_delivered\_customer\_date}$

```
SELECT
  order_id,
  DATE_DIFF(order_estimated_delivery_date, order_purchase_timestamp, DAY) time_to_delivery,
  DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date, DAY) diff_estimated_delivery ,
FROM `target.orders`
ORDER BY order_purchase_timestamp
```

Row	order_id	time_to_delivery	diff_estimated_delivery
1	2e7a8482f6fb09756ca50c10d7bfc047	45	null
2	e5fa5a7210941f7d56d0208e4e071d35	52	null
3	809a282bbd5dbcabb6f2f724fca862ec	16	null
4	bfb0f9bdef84302105ad712db648a6c	18	36
5	71303d7e93b399f5bcd537d124c0bcfa	22	null
6	3b697a20d9e427646d92567910af6d57	23	0
7	be5bc2f0da14d8071e2d45451ad119d9	34	-10
8	65d1e226dfaeb8cdc42f665422522d14	52	-16
9	a41c8759fbe7aab36ea07e038b2d4465	56	-25
10	d207cc272675637bfed0062edffd0818	50	-22

## 5.2. Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery

```
SELECT c.customer_state,
       ROUND(AVG(freight_value)) AS freight_value_mean,
       ROUND(AVG(DATE_DIFF(order_estimated_delivery_date, order_purchase_timestamp, DAY))) AS time_to_delivery_mean,
       ROUND(AVG(DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date, DAY))) AS diff_estimated_delivery_mean
FROM `target.order_items` AS oi
JOIN `target.orders` AS o
ON oi.order_id=o.order_id
Join `target.customers` AS c
ON c.customer_id= o.customer_id
GROUP BY c.customer_state
ORDER BY c.customer_state
```

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	AC	40	41	-20
2	AL	36	32	-8
3	AM	33	45	-19
4	AP	34	45	-17
5	BA	26	29	-10
6	CE	33	31	-10
7	DF	21	24	-11
8	ES	22	25	-10
9	GO	23	27	-11
10	MA	38	30	-9

## 5.1. Calculate days between purchasing, delivering and estimated delivery

```

SELECT
  order_id,
  expected_delivery_days,
  actual_delivery_days,
  CASE WHEN actual_delivery_days > expected_delivery_days THEN 'late delivery' WHEN actual_delivery_days IS NULL THEN 'not delivered' ELSE 'on time delivery' END AS summary
FROM(
  SELECT
    order_id,
    order_purchase_timestamp,
    order_estimated_delivery_date,
    order_delivered_customer_date,
    DATE_DIFF(order_estimated_delivery_date, order_purchase_timestamp, DAY) expected_delivery_days,
    DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) actual_delivery_days,
  FROM `target.orders`)
ORDER BY order_purchase_timestamp

```

Row	order_id	expected_delivery_days	actual_delivery_days	summary
1	2e7a8482f6fb09756ca50c10d7bfc047	45	null	not delivered
2	e5fa5a7210941f7d56d0208e4e071d35	52	null	not delivered
3	809a282bbd5dbcabb6f2f724fca862ec	16	null	not delivered
4	bfbdf0f9bdef84302105ad712db648a6c	18	54	late delivery
5	71303d7e93b399f5bcd537d124c0bcfa	22	null	not delivered
6	3b697a20d9e427646d92567910af6d57	23	23	on time delivery
7	be5bc2f0da14d8071e2d45451ad119d9	34	24	on time delivery
8	65d1e226dfaeb8cdc42f665422522d14	52	35	on time delivery
9	a41c8759fbe7aab36ea07e038b2d4465	56	30	on time delivery
10	d207cc272675637bfed0062edffd0818	50	27	on time delivery

## 5.2. Sort the data to get the following:

- Top 5 states with highest average freight value - sort in desc limit 5

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	PB	43	33	-12
2	RR	43	46	-17
3	RO	41	39	-19
4	AC	40	41	-20
5	PI	39	30	-11

- Top 5 states with lowest average freight value - sort in asc limit 5

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	SP	15	19	-10
2	PR	21	24	-13
3	RJ	21	26	-11
4	DF	21	24	-11
5	MG	21	24	-12

- Top 5 states with highest average time to delivery - - sort in desc limit 5

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	RR	43	46	-17
2	AP	34	45	-17
3	AM	33	45	-19
4	AC	40	41	-20
5	RO	41	39	-19

- Top 5 states with lowest average time to delivery - - sort in asc limit 5

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	SP	15	19	-10
2	PR	21	24	-13
3	DF	21	24	-11
4	MG	21	24	-12
5	ES	22	25	-10

- Top 5 states where delivery is really fast/ not so fast compared to estimated date

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	AC	40	41	-20
2	AM	33	45	-19
3	RO	41	39	-19
4	RR	43	46	-17
5	AP	34	45	-17

- Top 5 states where delivery is really fast/ not so fast compared to estimated date

Row	customer_state	freight_value_mean	time_to_delivery_mean	diff_estimated_delivery_mean
1	AL	36	32	-8
2	SE	37	30	-9
3	MA	38	30	-9
4	SP	15	19	-10
5	BA	26	29	-10



## 6. Payment type analysis:

### 6.1. Month over Month count of orders for different payment types

```
SELECT
EXTRACT(YEAR FROM o.order_purchase_timestamp) AS YEAR,
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS MONTH,
payment_type,
COUNT(p.order_id) AS orders_count
FROM `target.payments` AS p
JOIN `target.orders` AS o
ON p.order_id=o.order_id
GROUP BY YEAR, MONTH, payment_type
ORDER BY YEAR, MONTH
```

Row	YEAR	MONTH	payment_type	orders_count
1	2016	9	credit_card	3
2	2016	10	credit_card	254
3	2016	10	voucher	23
4	2016	10	debit_card	2
5	2016	10	UPI	63
6	2016	12	credit_card	1
7	2017	1	voucher	61
8	2017	1	UPI	197
9	2017	1	credit_card	583
10	2017	1	debit_card	9

## 6.2. Distribution of payment installments and count of orders

```
SELECT
payment_installments,
COUNT(order_id) AS orders_count
FROM `target.payments`
GROUP BY payment_installments
ORDER BY payment_installments
```

Row	payment_installments	orders_count
1	0	2
2	1	52546
3	2	12413
4	3	10461
5	4	7098
6	5	5239
7	6	3920
8	7	1626
9	8	4268
10	9	644

## 7. Actionable Insights and Recommendations

SP, RJ, MG state has most customer count (above 10000)

RR, AP, AC state has the least customer count (below 100)

We can use new customers making strategies:

1. Ask for referrals, Networking.
2. Marketing and Advertisement.
3. Offer discounts and incentives for new customers only.
4. Re-contact old customers.
5. Improve website.
6. Partner with complementary businesses.
7. Promote expertise.
8. Use online reviews to your advantage.

We can use customer retention strategies

1. Create a strong onboarding experience.
2. Provide personalized customer experiences.
3. Build trust with your customers.
4. Implement a customer feedback loop.
5. Maintain a customer communication calendar.
6. Send a company newsletter.
7. Start a customer education program.
8. Offer unique services.
9. Start a customer retention program.

In AL, MA, SE state the actual delivery duration is very fast, so we can lower the expected delivery duration to promise customer fast delivery.

In AC, AM, RO state have the very least mean difference between actual delivery duration and expected delivery duration. Either we can increase the duration of expected delivery or we can improvise the delivery system for those states by using some optimize supply chain strategies.

1. Assess realistic demands. Same-day delivery requires smart inventory management practices.
2. Utilize distributed order management (DOM) systems.
3. Ensure regular audits of the supply chain.
4. Enable efficient distribution model.
5. Prepare an effective warehouse.
6. Automate your supply chain processes
7. Use real-time data analysis for continuous improvement
8. Implement an effective inventory management process

Around 76% orders paid by using the credit card payment type and around 19% paid by using UPI payment type. if company have more benefits from taking UPI payment then we can educate customers UPI payments, can give discounts, vouchers for one time payment to increase UPI payment type percentage.

By analysing monthly trend we can notice that company sales more products in summer season and when the winter season comes the sales falls down. To maintain sales in whole year company should increase more items in inventory that belongs to winter season and can make the store/website more attractive in winter festivals like Christmas and new year.

Can choose product categories that are in high demand during winter-

1. Decorations: This is a must-buy for winter! From Christmas to New Year, roads are full of lights. People buy so many Christmas themed decorations, Christmas trees, different kinds of lights and lanterns, party decorations etc.
2. Winter wear: Stores selling fashionable and trendy winter wear see a considerable increase in sales on their online stores
3. Electronics: This is the season for gifting your friends and family, and usually this category is a favourite all year round.
4. Personalised gifts: This is a category that receives extra attention during gifting season. Personalised gifts like embossed pens, utensils, pillows etc. are unique to every person. Online stores that sell handmade products also show higher revenues because of Christmas and New year. People love receiving gifts that have been made by hand as it shows more love and appreciation.

Can utilise social media effectively and follow the winter season theme on website -

1. use interacting features like reels, stories, polls, live, etc to increase brand reach
2. Post attractive winter-themed visuals so that you get more traffic
3. Change the theme of your social media platform to match the winter mood
4. Put up more pictures and videos that are winter-related
5. Change the cover photos on your social media platforms
6. Add decorative accessories to the main product to make it an attractive
7. Change the main website layout to make it more cheerful and festive
8. Change product descriptions to match winter needs or gifts
9. Use winter themed wrapping paper or packaging material