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

1. **Data type of all columns in the "customers" table.**

```
SELECT
Column_name, data_type
FROM `strong-matrix-387512.Target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers'
ORDER BY ordinal_position;
```

w	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

2. **Get the time range between which the orders were placed.**

```
SELECT format_date('%b %Y',min(order_purchase_timestamp))as first_order_date,
       format_date('%b %Y',max(order_purchase_timestamp)) as last_order_date
FROM `Target.orders`
```

first_order_date	last_order_date
Sep 2016	Oct 2018

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

```
SELECT count( distinct c.customer_state) as number_of_state,
       count(distinct c.customer_city) as number_of_city
From `Target.customers` c inner join `Target.orders` o on c.customer_id = o.customer_id
```

Query results

JOB INFORMATION | RESULTS | JSON

Row	number_of_state	number_of_city
1	27	4119

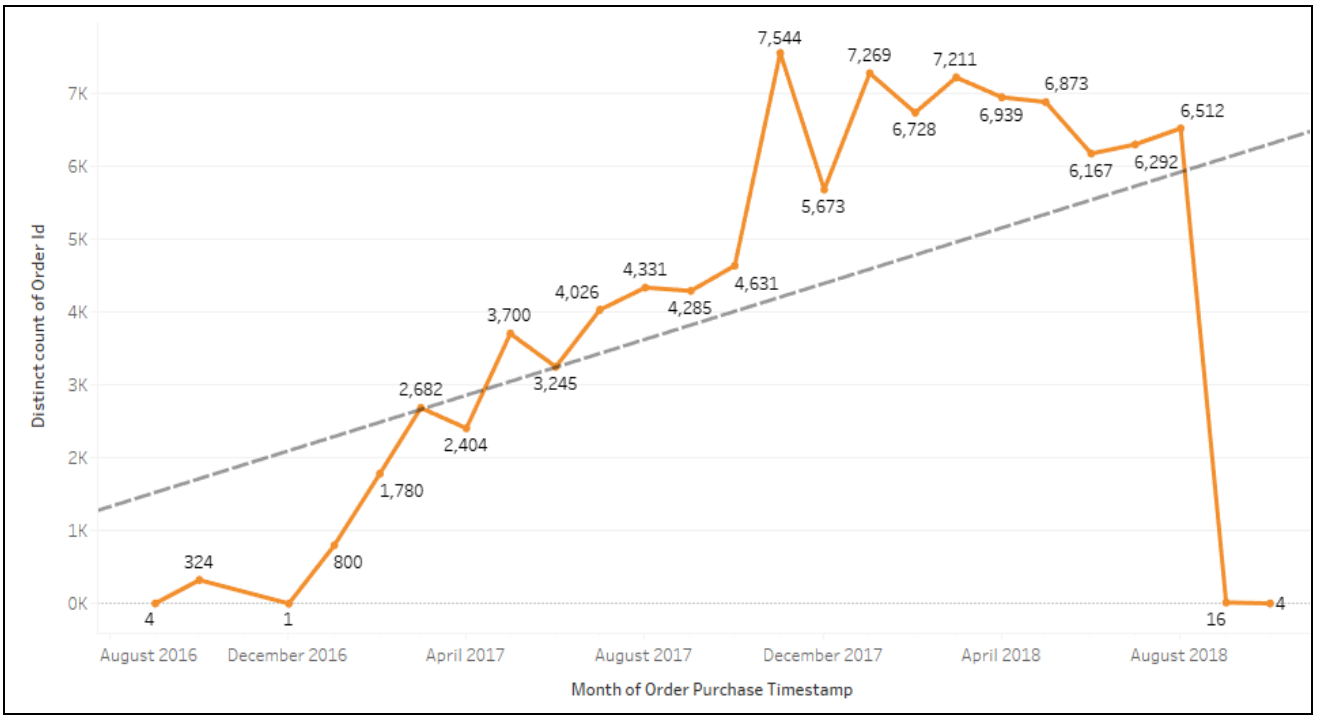
In-depth Exploration:

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

```
Select extract(year from order_purchase_timestamp) as order_year,
       extract(month from order_purchase_timestamp) as order_month,
       count(*) as num_of_orders
From `Target.orders`
group by 1,2
order by 1,2
```

Row	order_year	order_month	num_of_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

- **Insights : There is linear increase in number of orders from Sept 2016 to Oct 2016**



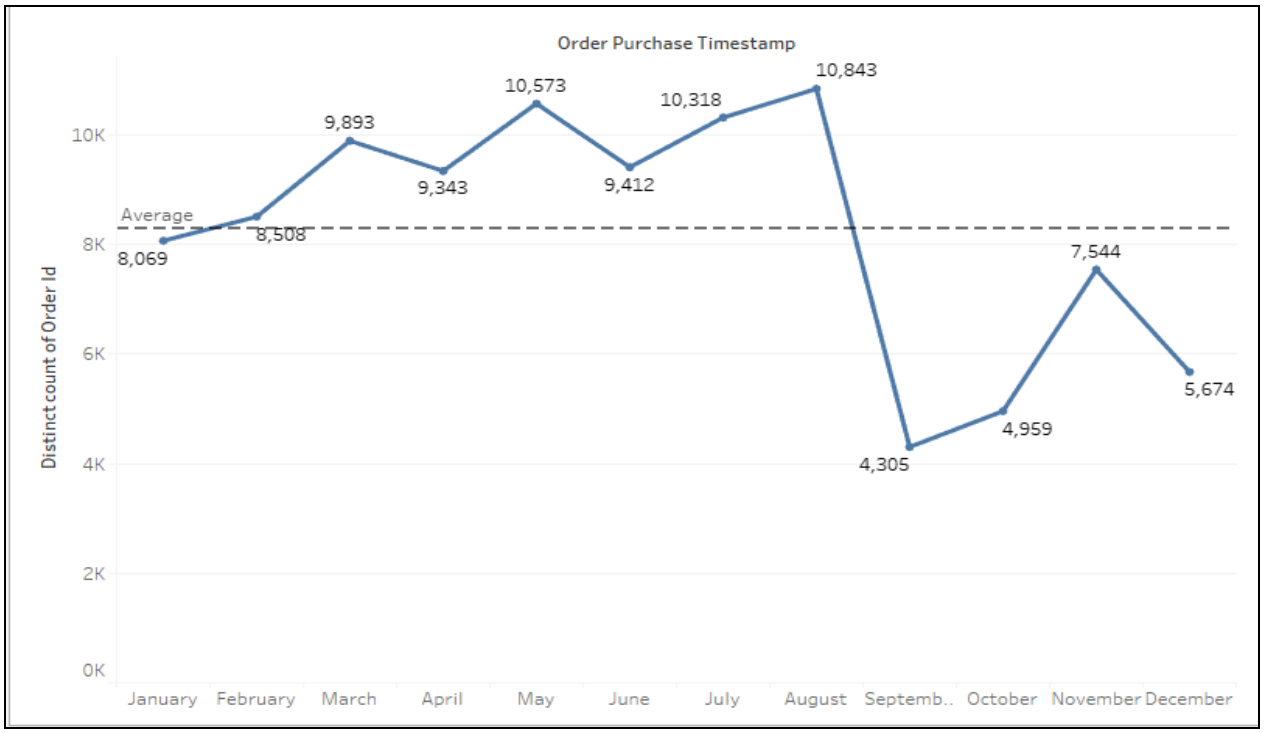
2. **Can we see some kind of monthly seasonality in terms of the no. of orders being placed?**

```
Select extract(month from order_purchase_timestamp) as order_month,
       count(*) as num_of_orders
From `Target.orders`
group by 1
order by 2 Desc
```

Row	order_month	num_of_orders
1	8	10843
2	5	10573
3	7	10318
4	3	9893
5	6	9412
6	4	9343
7	2	8508
8	1	8069
9	11	7544
10	12	5674

Insights : Highest numbers of orders were placed from March to August. In this period, the autumn season is enjoyed by people along with their major festivals like Festa Junina is celebrated across the country and vacations.

Recommendations : Target should avail enough stock of items related to festival and vacations items like clothes, shoe wear, etc.



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

```
Select distinct time_of_day, count(order_id)over(partition by time_of_day) as
count_of_time_of_day
From ( Select order_id, (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 time_of_day
From `Target.orders`
order by time_of_day desc)
```

w	time_of_day	count_of_time_of_da
1	Afternoon	38135
2	Dawn	5242
3	Mornings	27733
4	Night	28331

- **Insights :** Most orders are placed in afternoon and is followed by night and mornings.
- **Recommendations :** Target should make sure that their system and server should work without heavy workload during this periods

Evolution of E-commerce orders in the Brazil region.

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

```
Select extract(month from o.order_purchase_timestamp) as month,
          c.customer_state as state, count(*) as number_of_orders
From `Target.customers` c inner join `Target.orders` o on
c.customer_id=o.customer_id
group by 1,2     order by number_of_orders desc
```

ow	month	state	number_of_orders
1	8	SP	4982
2	5	SP	4632
3	7	SP	4381
4	6	SP	4104
5	3	SP	4047
6	4	SP	3967
7	2	SP	3357
8	1	SP	3351
9	11	SP	3012
10	12	SP	2357

- **Insights : Most numbers of orders were placed in the month of Aug by sao paulo - SP. Top 12 month wise sales were from sao-paulo.**
  The least number of orders were from Roraima (46) , Amapa(68), Acre(81) through out the years.

- **Recommendation : Target should open their branch of retail store in Sau-paulo and must ensure enough product supply, transportation to meet demand of customers throughout the year. Target should focus on marketing strategy to increase order from Roraima, Amapa, Acre.**

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

```
Select customer_state, count(customer_id) as count_of_customer_statewise
From `Target.customers`
group by 1
order by count_of_customer_statewise Desc
```

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

customer_state	Percent_of total_cust
SP	41.98
RJ	12.92
MG	11.70
RS	5.50
PR	5.07
SC	3.66
BA	3.40
DF	2.15
ES	2.04
GO	2.03
PE	1.66
CE	1.34
PA	0.98
MT	0.91
MA	0.75
MS	0.72
PB	0.54
PI	0.50
RN	0.49
AL	0.42

Insights : Target have highest number of customers in Sao Paulo (41.98%) followed by Rio de Janeiro (12.92%) and Minas Gerias (11.70%).

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

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

```
Select yer,total_sum,concat(round((sum_of_2018 - total_sum)*100/total_sum),' %') as
percent_increase
from ( Select *,lead (total_sum)over(order by yer) as sum_of_2018
       From (Select distinct yer, round (sum (payment_value)over(partition by
yer),2)       as total_sum
       From ( Select *,extract(month from order_purchase_timestamp) as mnth,
       extract (year from order_purchase_timestamp) as yer,
```

```
From (SELECT p.order_id, p.payment_value,o.order_purchase_timestamp
from `Target.payments` p inner join `Target.orders` o on p.order_id = o.order_id) )
where yer =2017 or yer = 2018 and mnth between 1 and 8
order by yer ))
```

yer	total_sum	percent_increase
2017	7249746.73	20 %
2018	8694733.84	*null*

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

```
with order_details as ( Select c.customer_state as state,
                        round(sum(i.price),2) as total_amount,
                        count(distinct o.order_id) as total_orders,
   from `Target.orders` o inner join `Target.order_items` i on o.order_id = i.order_id
    inner join `Target.customers`c  on c.customer_id = o.customer_id
        group by state
                      )

        Select state,total_amount,total_orders, round(total_amount/total_orders,2) as
        avg_price
        from order_details
        order by avg_price Desc
```

Row	state	total_amount	total_orders	avg_price
1	PB	115268.08	532	216.67
2	AP	13474.3	68	198.15
3	AC	15982.95	81	197.32
4	AL	80314.81	411	195.41
5	RO	46140.64	247	186.8
6	PA	178947.81	970	184.48
7	TO	49621.74	279	177.86
8	PI	86914.08	493	176.3
9	MT	156453.53	903	173.26
10	RN	83034.98	482	172.27

- **Insights : In Paraiba, average price is highest for Target and is lowest for Sau Paulo.**

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

```
with order_details as ( Select c.customer_state as state,
                        round(sum(i.freight_value),2) as total_freight_amount,
                        count(distinct o.order_id) as total_orders,
from `Target.orders` o inner join `Target.order_items` i  on o.order_id =
i.order_id
          inner join `Target.customers`c on c.customer_id = o.customer_id
     group by state )

     Select state,total_freight_amount,total_orders,
           round(total_freight_amount/total_orders,2) as avg_freight_price
     from order_details
     order by avg_freight_price Desc
```

Row	state	total_freight_amount	total_orders	avg_freight_price
1	RR	2235.19	46	48.59
2	PB	25719.73	532	48.35
3	RO	11417.38	247	46.22
4	AC	3686.75	81	45.52
5	PI	21218.2	493	43.04
6	MA	31523.77	740	42.6
7	TO	11732.68	279	42.05
8	AP	2788.5	68	41.01
9	SE	14111.47	345	40.9
10	PA	38699.3	970	39.9

- **Insights :** **In Roraima, average freight price is highest for Target and is lowest for Sau Paulo.**

Analysis based on sales, freight and delivery time.

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.

```
Select date(order_purchase_timestamp) as
purchase_date,date(order_estimated_delivery_date) as estimated_delivery_date,
       date_diff(order_estimated_delivery_date,order_purchase_timestamp,day) as
estimated_delivery_in_days,
       DATE_DIff(order_delivered_customer_date,order_purchase_timestamp,day) as
time_to_delivey_in_days,
       Date_diff(order_estimated_delivery_date,order_delivered_customer_date,day) as
diff_estimated_delivery_in_days
From `Target.orders`
order by order_id
```

Row	purchase_date	estimated_delivery_d	estimated_delivery_in_days	time_to_delivey_in_days	diff_estimated_delivery_in_days
1	2017-09-13	2017-09-29	15	7	8
2	2017-04-26	2017-05-15	18	16	2
3	2018-01-14	2018-02-05	21	7	13
4	2018-08-08	2018-08-20	11	6	5
5	2017-02-04	2017-03-17	40	25	15
6	2017-05-15	2017-06-06	21	6	14
7	2017-12-10	2018-01-04	24	8	16
8	2018-07-04	2018-07-25	20	5	15
9	2018-03-19	2018-03-29	9	9	0
10	2018-07-02	2018-07-23	20	2	18

Insights : Target delivers ordered products before estimated delivery date in most cases.

Recommendations : Target should focus to reduce estimated delivery time. It will help in not losing customers who need products before estimated delivery time which will result in an increase of customers

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

```
with order_details as ( Select c.customer_state as state,
                        round(sum(i.freight_value),2) as total_freight_amount,
                        count(distinct o.order_id) as total_orders,
from `Target.orders` o inner join `Target.order_items` i  on o.order_id = i.order_id
     inner join `Target.customers`c on c.customer_id = o.customer_id
group by state       )

Select state,round(total_freight_amount/total_orders,2) as avg_freight_price
from order_details
order by avg_freight_price Desc
Limit 5
```

ow	state	avg_freight_price
1	RR	48.59
2	PB	48.35
3	RO	46.22
4	AC	45.52
5	PI	43.04

```
with order_details as ( Select c.customer_state as state,
                        round(sum(i.freight_value),2) as total_freight_amount,
                        count(distinct o.order_id) as total_orders,
from `Target.orders` o inner join `Target.order_items` i  on o.order_id = i.order_id
     inner join `Target.customers`c on c.customer_id = o.customer_id
group by state       )
Select state,round(total_freight_amount/total_orders,2) as avg_freight_price
from order_details
order by avg_freight_price
Limit 5
```

Row	state	avg_freight_price
1	SP	17.37
2	MG	23.46
3	PR	23.58
4	DF	23.82
5	RJ	23.95

Insights : Average freight charges are highest for Roraima and lowest for Sao Paulo.This can be one of the main factors of highest number of orders from Sao Paulo.

Recommendation : Target should focus on reducing freight charges in Roraima and Paraiba which can reduce the cost and increase the number of orders from this region.

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

```
Select customer_state,round(avg(delivery_time),2) as highest_avg_delivery_time
FROM (SELECT o.order_purchase_timestamp,o.order_delivered_customer_date,c.customer_state,
date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day) as delivery_time
From `Target.orders` o inner join `Target.customers` c on o.customer_id =
c.customer_id) temp
group by 1 order by avg(delivery_time) desc
limit 5
```

w	customer_state	highest_avg_delivery_time
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

```
Select customer_state,round(avg(delivery_time),2) as lowest_avg_delivery_time
FROM (SELECT
o.order_purchase_timestamp,o.order_delivered_customer_date,c.customer_state,
date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day) as
delivery_time
From `Target.orders` o inner join `Target.customers` c on o.customer_id = c.customer_id) temp
group by 1 order by avg(delivery_time) limit 5
```

w	customer_state	lowest_avg_delivery_time
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48

Insights : In Roraima, average delivery time is highest and followed by Amapa and Amazonas. In Sao Paulo, average delivery time is lowest and is followed by Paraná and Minas Gerais.

Recommendation : Like São Paulo, Target should focus on strategy to reduce delivery time in Roraima, Amapa and Amazonas to increase the number of customers.

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

```
Select customer_state,avg_estimated_time,avg_act_delivery_time,
round((avg_estimated_time-avg_act_delivery_time),2) as Top_fast_delivery
FROM (Select customer_state,round(avg(actual_delivery_time),2) as avg_act_delivery_time ,
round(avg(esimated_delivery_time),2) as avg_estimated_time
From (SELECT customer_state,
date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day) as actual_delivery_time,
date_diff(o.order_estimated_delivery_date,o.order_purchase_timestamp,day) as esimated_delivery_time
From `Target.orders` o inner join `Target.customers` c on o.customer_id = c.customer_id) temp
group by 1)
order by (avg_estimated_time-avg_act_delivery_time) desc
limit 5
```

Row	customer_state	avg_estimated_time	avg_act_delivery_tim	Top_fast_delivery
1	AC	40.77	20.64	20.13
2	RO	38.41	18.91	19.5
3	AP	45.71	26.73	18.98
4	AM	44.76	25.99	18.77
5	RR	46.17	28.98	17.19

Insights : In Acre, order delivery is the fastest as compared to other states and is followed by Rondonia and Amapa

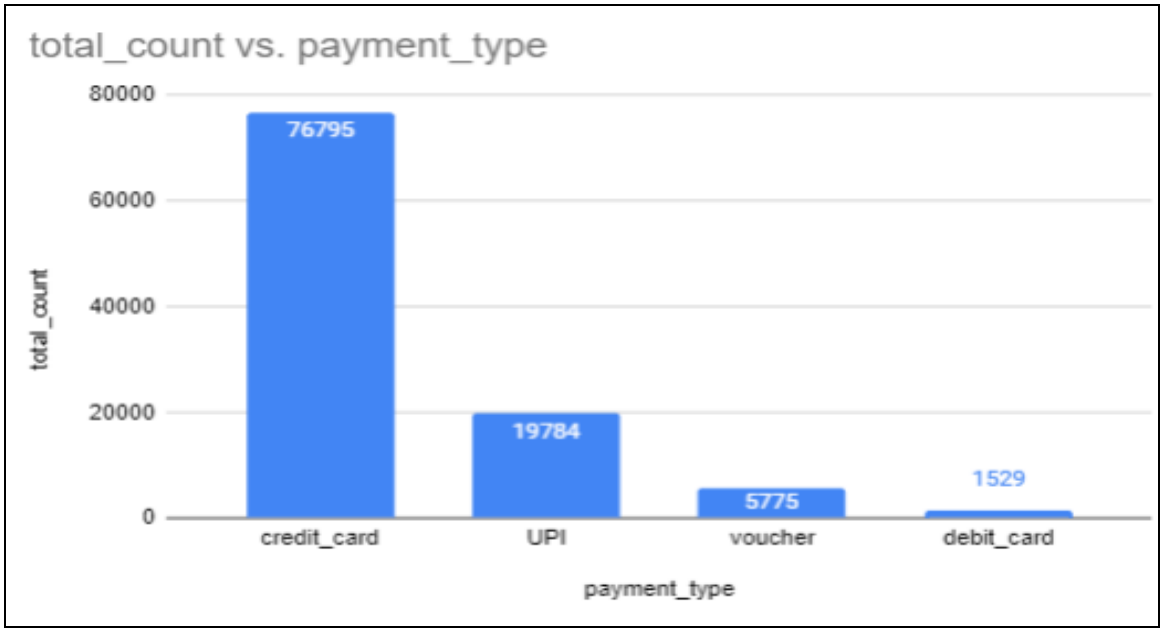
Recommendation : In Acre, Rondonia, Amapa estimated delivery time should be reduced by Target to increase the number of customers.

Analysis based on the payments:

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

```
SELECT extract(month from o.order_purchase_timestamp) as purchase_month, p.payment_type,
       count(o.order_id)as total_count
from `Target.payments` p inner join `Target.orders` o on p.order_id = o.order_id
group by 1,2
order by purchase_month,3 Desc
```

purchase_month	payment_type	total_count
1	credit_card	6103
1	UPI	1715
1	voucher	477
1	debit_card	118
2	credit_card	6609
2	UPI	1723
2	voucher	424
2	debit_card	82
3	credit_card	7707
3	UPI	1942



- **Insights :** In every month, most orders were placed with credit card and least orders were placed by debit card.

- **Recommendations :** Target can try different offers and rewards on using debit cards.

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

```
Select payment_installments,count(order_id) as total_orders
from `Target.payments`
group by 1
order by 2 Desc
```

ow	payment_installments	total_orders
1	1	52546
2	2	12413
3	3	10461
4	4	7098
5	10	5328
6	5	5239
7	8	4268
8	6	3920
9	7	1626
10	9	644

Insights : Most orders were placed with EMI of 1 to 4 months by customers.

General Insights :

- Maximum sales happened between march and august during festivals like Festa Junina and autumn vacations.
- Highest numbers of orders were placed by Sao Paulo and lowest were from Roraima, Amapa and Acre.
- Sao Paulo has lowest average freight charges and delivery_time is minimum whereas
- Roraima has highest average freight charges and also delivery_time is maximum.

Recommendation :

- Target should ensure that enough stock should be available for Sao Paulo to meet customer's demand.
- Marketing strategy need to be updated for Roraima, Amapa and Acre to get more orders from this state.
- Target should reduce estimated freight charges to ensure no loss of customers who need product before time.
- For Roraima, Target should look for other alternative for reducing freight charges and reducing delivery time.