

# Target – SQL BUSINESS CASE STUDY

*Solved By: Rahul Hipparkar*

*Email ID: hipparkarraahul18@gmail.com*

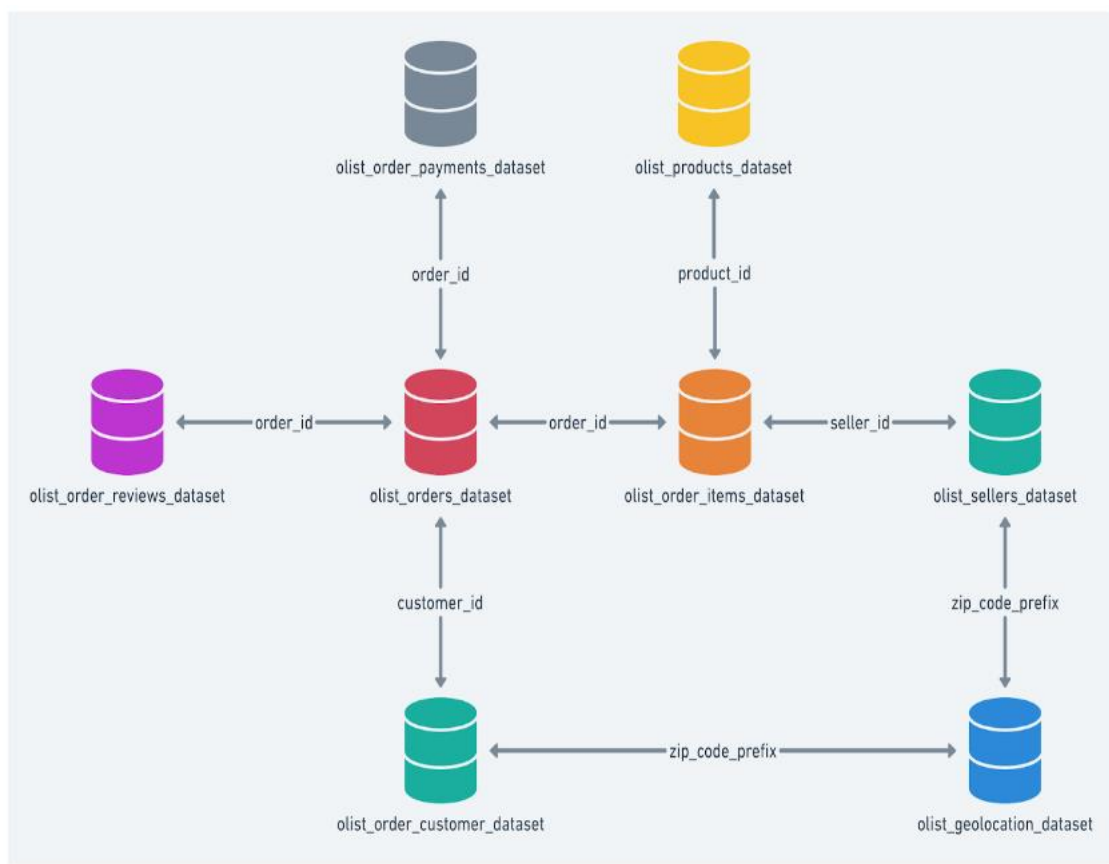
## Context:

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

## Dataset schema:



## Problems and Solutions

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

a) Data type of columns in a table:

To check data type of columns in a table we can simply use INFORMATION\_SCHEMA.COLUMNS query.

**Query:**

```
select table_name, column_name, data_type
from Target_SQL.INFORMATION_SCHEMA.COLUMNS
where table_name='customers';
```

**Solution:** Data type of columns in a Customer Table

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

b) Time period for which the data is given:

All orders are placed within the given time period:

First order date: 2016-09-04

Last order date: 2018-10-17

### Query:

```
with cte as(
select extract( date from
min(order_purchase_timestamp)) as
first_order_date,extract(date from
max(order_purchase_timestamp)) as last_order_date
from `Target_SQL.orders`
)

select *, date_diff(last_order_date,
first_order_date, year) as Time_Period_In_Years
from cte
```

### Result:

Row	first_order_date	last_order_date	Time_Period_In_Years
1	2016-09-04	2018-10-17	2

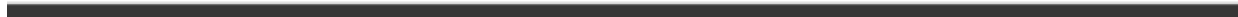
c) Cities and States of customers ordered during the given period

### Query:

```
select c.customer_city,c.customer_state
from `Target_SQL.orders` as o
inner join `Target_SQL.customers` as c
on c.customer_id=o.customer_id
limit 10;
```

**Result:** List of Cities and States of customers who placed an order during given time period:

Row	customer_city ▼	customer_state ▼	
1	rio de janeiro	RJ	
2	sao leopoldo	RS	
3	general salgado	SP	
4	brasilia	DF	
5	paranavai	PR	
6	cuiaba	MT	
7	sao luis	MA	
8	maceio	AL	
9	hortolandia	SP	
10	varzea grande	MT	



## 2) In-depth Exploration:

- a) 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?

### Query:

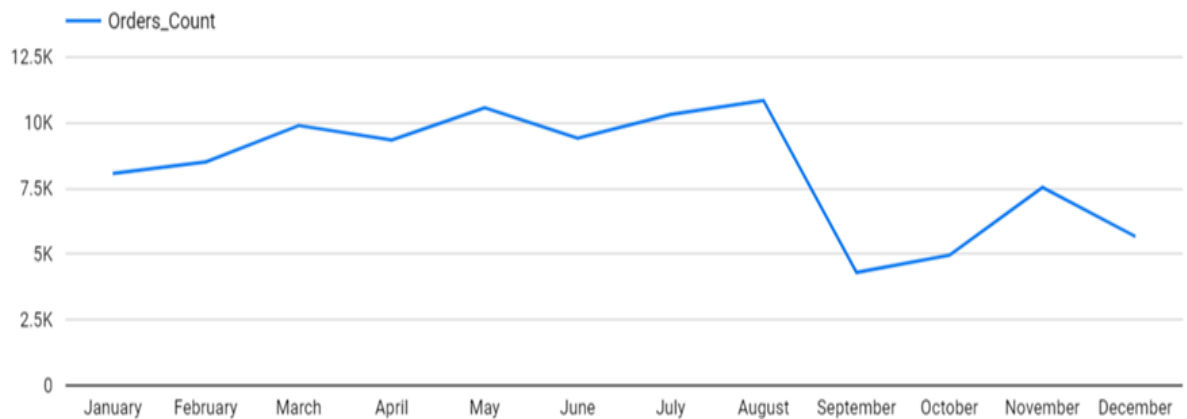
```
with cte as
(
    select extract(month from order_purchase_timestamp)
as Month,extract(year from order_purchase_timestamp)
as Year
from `Target_SQL.orders`
)

select Month,count(*) as Orders_Count
from cte
group by Month
order by Month
```

### Result:

Row	Month	Orders_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

## Looker Studio Visualization:



## Insight:

- i. There is increasing trend in number of orders placed between the months of June to August in Brazil . This is because winter season in Brazil occurs in months of June & August. The increase in sales during June to August is directly influenced by the winter season in Brazil.
- ii. There is decreasing trend in sales or number of orders during the months of August to September.

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

0-6: Dawn

7-12: Morning

13-18: Afternoon

19-23: Night

**Query:**

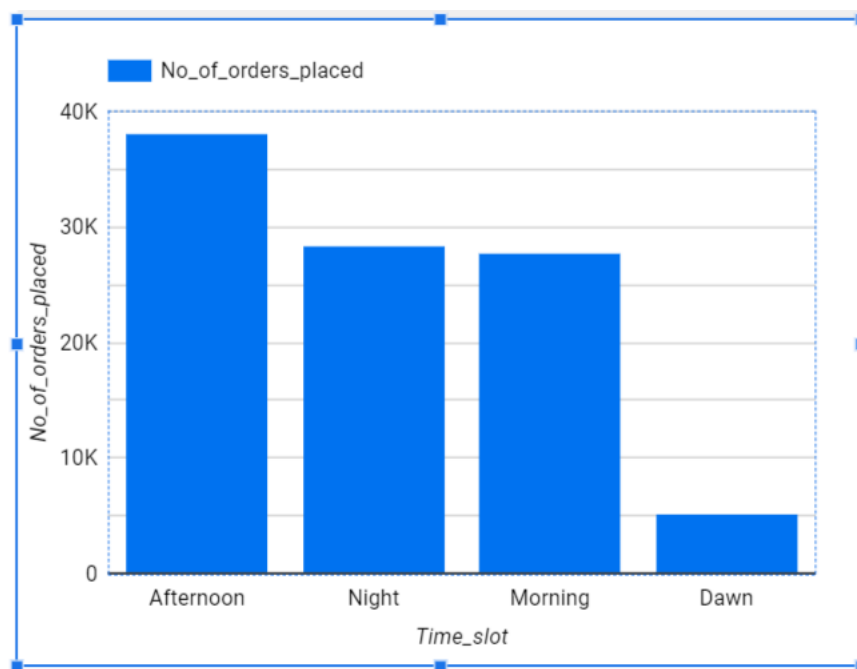
```
with cte as
(
    select *, extract(hour from
order_purchase_timestamp) as purchase_hour
    from `Target_SQL.orders`
),
cte1 as
(
    select *,
    case
        when purchase_hour between 0 and 6
        Then 'Dawn'
        when purchase_hour between 7 and 12
        Then 'Morning'
        when purchase_hour between 13 and 18
        Then 'Afternoon'
        Else 'Night'
    End as Time_Slot
    from cte
)
select distinct Time_slot, count(*) as
No_of_orders_placed
from cte1
group by Time_Slot
order by No_of_orders_placed desc
```



## Result:

Row	Time_slot ▼	No_of_orders_placed ▼
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

## Looker Studio Visualization:



**Insight:** Majority of Brazilian customers tend to buy items during the afternoon. Brazilian customers prefer afternoon for making purchases.

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

#### a) Get month on month orders by states:

##### Query:

```
with cte as(  
select *, extract(month from o.order_purchase_timestamp) as  
Month  
from `Target_SQL.orders` as o  
inner join `Target_SQL.customers` as c  
on c.customer_id=o.customer_id  
)  
  
select    Month,customer_state,count(*)  as No_of_Orders  
from cte  
group by Month,customer_state  
order by No_of_Orders desc  
Limit 10
```

##### Result :

Row	Month	customer_state	No_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

**Insight:** SP (São Paulo) state has highest number of orders placed in month of August .

b) Distribution of customers across the states in Brazil:

**Query:**

```
select customer_state, count(*) as Number_of_Customers
from `Target_SQL.customers`
group by customer_state
order by Number_of_Customers desc
```

**Result:**

Row	customer_state	Number_of_Customers
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

**Insight:** SP (São Paulo) state has highest number of customers. Because of which SP state has highest number of monthly orders as compared to other states.

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

a) Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use "payment\_value" column in payments table

**Query:**

```
with cte as
(
    select p.payment_value,extract(month from
o.order_purchase_timestamp) as Month,extract(year
from o.order_purchase_timestamp) as Year
from `Target_SQL.orders` as o
inner join `Target_SQL.payments` as p
on p.order_id=o.order_id
),
cte1 as(
select Year,round(sum(payment_value),2) as
Cost_Of_Orders
from cte
where Month not in (9,10,11,12)
group by Year
order by Year
)
select *,Round((c2.Cost_Of_Orders-
c1.Cost_Of_Orders)*100/c1.Cost_Of_Orders,2) as
Percentage_Increase
from cte1 as c1
inner join cte1 as c2
on c1.Year=2017 and c2.Year=2018
```

**Result:**

Row	Year	Cost_Of_Orders	Year_1	Cost_Of_Orders_1	Percentage_Increase
1	2017	3669022.12	2018	8694733.84	136.98

**Insight:**

There is 136.98% increase in cost of orders from 2017 to 2018 (Jan to Aug).

---

b) Mean & Sum of price and freight value by customer state:

**Query:**

```
with cte as(
select *
from `Target_SQL.order_items` as i
inner join `Target_SQL.orders` as o
on o.order_id=i.order_id
inner join `Target_SQL.customers` as c
on c.customer_id=o.customer_id
)

select customer_state,round(sum(Price),2) as
Total_Price, round(avg(Price),2) as Mean_Price,
round(sum(freight_value),2) as Total_Freight_Value,
round(avg(freight_value),2) as Mean_Freight_Value
from cte
group by customer_state
order by Total_Price desc,Mean_Price
desc,Total_Freight_Value desc, Mean_Freight_Value
desc
Limit 10
```

## Result:

Row	customer_state ▼	Total_Price ▼	Mean_Price ▼	Total_Freight_Value ▼	Mean_Freight_Value ▼
1	SP	5202955.05	109.65	718723.07	15.15
2	RJ	1824092.67	125.12	305589.31	20.96
3	MG	1585308.03	120.75	270853.46	20.63
4	RS	750304.02	120.34	135522.74	21.74
5	PR	683083.76	119.0	117851.68	20.53
6	SC	520553.34	124.65	89660.26	21.47
7	BA	511349.99	134.6	100156.68	26.36
8	DF	302603.94	125.77	50625.5	21.04
9	GO	294591.95	126.27	53114.98	22.77
10	ES	275037.31	121.91	49764.6	22.06

## Insight:

- I. SP (São Paulo) state has highest total price and total freight value. This is because SP state has largest customer base as compared to other states.
- II. SP (São Paulo) state also has favourable mean price and mean freight value among all customer states.

5) Analysis on sales, freight and delivery time:

a) Calculate days between purchasing, delivering and estimated delivery:

**Query:**

```
select
datetime_diff(order_delivered_customer_date,
order_purchase_timestamp, day) as
Actual_Delivery_Duration,
datetime_diff(order_estimated_delivery_date,
order_purchase_timestamp, day) as
Estimated_Delivery_Duration
from `Target_SQL.orders`
order by Actual_Delivery_Duration desc
Limit 10
```

**Result:**

Row	Actual_Delivery_Duration	Estimated_Delivery_Duration
1	209	28
2	208	19
3	195	30
4	194	32
5	194	39
6	194	28
7	191	15
8	189	22
9	188	28
10	187	25

**Insight:**

For an order Target took 209 days to deliver it. But for same order days estimated for delivery is 28.

2) Find time\_to\_delivery & diff\_estimated\_delivery. Formula for the same given below:

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

**Query:**

```
select
datetime_diff(order_delivered_customer_date,
order_purchase_timestamp, day) as time_to_delivery,
datetime_diff(order_estimated_delivery_date, order_delivered_customer_date, day) as diff_estimated_delivery
from `Target_SQL.orders`
where order_delivered_customer_date is not NULL
order by time_to_delivery
Limit 10;
```

**Result:**

Row	time_to_delivery	diff_estimated_delivery
1	0	7
2	0	11
3	0	16
4	0	8
5	0	25
6	0	12
7	0	27
8	0	10
9	0	19
10	0	9

**Insight:**

Target often deliver orders within 24 hours or before estimated delivery date.



3) Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery. Sort the data to get the following:

a) Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5:

**Query:**

```
with cte as(
select *,
datetime_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day) as time_to_delivery,
datetime_diff(o.order_estimated_delivery_date,o.order
_delivered_customer_date, day) as
diff_estimated_delivery
from `Target_SQL.orders` as o
inner join `Target_SQL.order_items` as i
on i.order_id=o.order_id
inner join `Target_SQL.customers` as c
on o.customer_id=c.customer_id
)
select customer_state,
round(avg(freight_value),2) as Average_Freight_Value,
round(avg(time_to_delivery),2) as
Average_Time_To_Delivery,
round(avg(diff_estimated_delivery),2) as
Average_diff_estimated_delivery
from cte
group by customer_state
order by Average_Freight_Value
Limit 5;
```

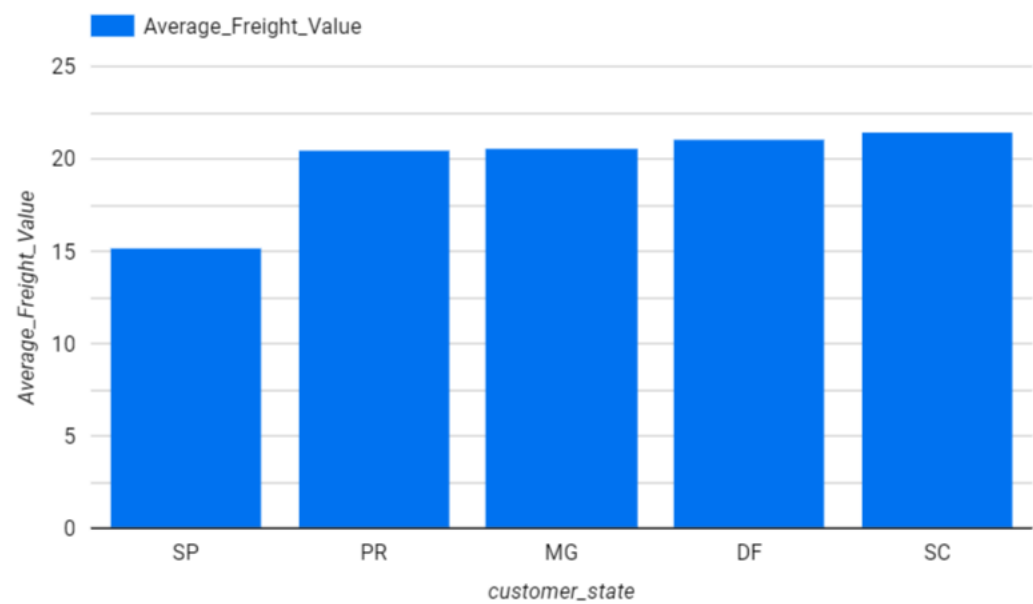
Result:

Row	customer_state ▾	Average_Freight_Value ▾	Average_Time_To_Delivery ▾	Average_diff_estimated_delivery ▾
1	SP	15.15	8.26	10.27
2	PR	20.53	11.48	12.53
3	MG	20.63	11.52	12.4
4	RJ	20.96	14.69	11.14
5	DF	21.04	12.5	11.27

Insight:

SP(São Paulo) state has lowest average freight value.

Looker Studio Visualization:



b)Top 5 states with highest/lowest average time to delivery:

**Query:**

```
with cte as(
select *,
datetime_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day) as time_to_delivery,
datetime_diff(o.order_estimated_delivery_date,o.order
_delivered_customer_date, day) as
diff_estimated_delivery
from `Target_SQL.orders` as o
inner join `Target_SQL.order_items` as i
on i.order_id=o.order_id
inner join `Target_SQL.customers` as c
on o.customer_id=c.customer_id
)
select customer_state,
round(avg(freight_value),2) as Average_Freight_Value,
round(avg(time_to_delivery),2) as
Average_Time_To_Delivery,
round(avg(diff_estimated_delivery),2) as
Average_diff_estimated_delivery
from cte
group by customer_state
order by Average_Time_To_Delivery
Limit 5;
```

## Result:

Row	customer_state ▾	Average_Freight_Value ▾	Average_Time_To_Delivery ▾	Average_diff_estimated_delivery ▾
1	SP	15.15	8.26	10.27
2	PR	20.53	11.48	12.53
3	MG	20.63	11.52	12.4
4	DF	21.04	12.5	11.27
5	SC	21.47	14.52	10.67

## Insight:

In SP (São Paulo) state Target requires average 8 days to deliver an order to the customer.

## Looker Studio Visualization:



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

**Query:**

```
with cte as(
select *,
datetime_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day) as time_to_delivery,
datetime_diff(o.order_estimated_delivery_date,o.order
_delivered_customer_date, day) as
diff_estimated_delivery
from `Target_SQL.orders` as o
inner join `Target_SQL.order_items` as i
on i.order_id=o.order_id
inner join `Target_SQL.customers` as c
on o.customer_id=c.customer_id
)

select customer_state,
round(avg(freight_value),2) as Average_Freight_Value,
round(avg(time_to_delivery),2) as
Average_Time_To_Delivery,
round(avg(diff_estimated_delivery),2) as
Average_diff_estimated_delivery
from cte
group by customer_state
order by
Average_Time_To_Delivery,Average_diff_estimated_deliv
ery desc
Limit 5;
```

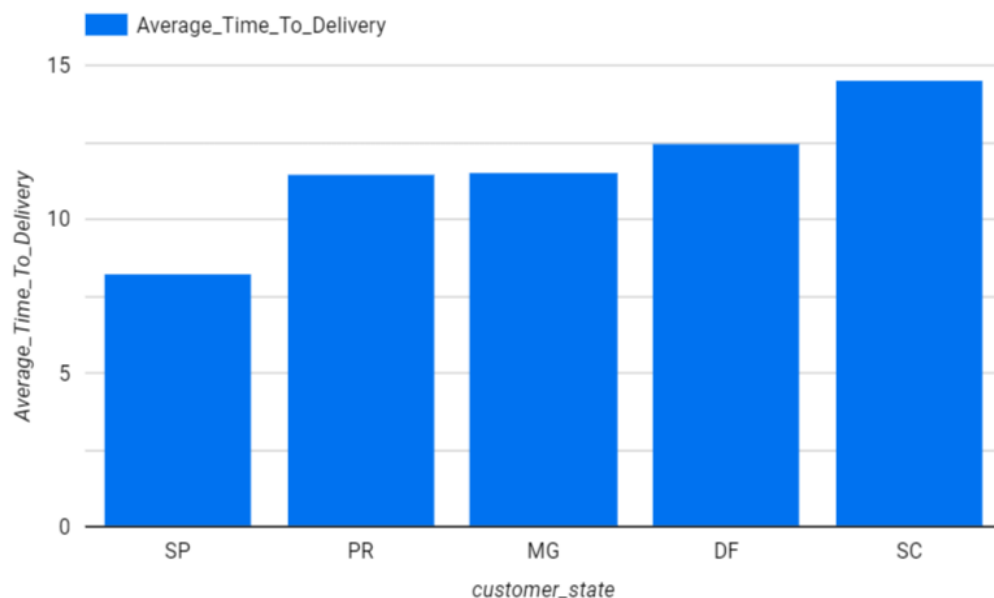
## Result:

Row	customer_state ▼	Average_Freight_Value ▼	Average_Time_To_Delivery ▼	Average_diff_estimated_delivery ▼
1	SP	15.15	8.26	10.27
2	PR	20.53	11.48	12.53
3	MG	20.63	11.52	12.4
4	DF	21.04	12.5	11.27
5	SC	21.47	14.52	10.67

## Insight:

- In the state of São Paulo (SP), delivery is remarkably fast, as Target takes average 8 days to deliver an order their. Moreover, orders in São Paulo are typically delivered, on average, 10 days before the estimated delivery date.
- In above given list of states delivery is really fast. Also orders are delivered in these states before estimated delivery date.

## Looker Studio Visualization:



6) Payment type analysis:

a) Month over Month count of orders for different payment types:

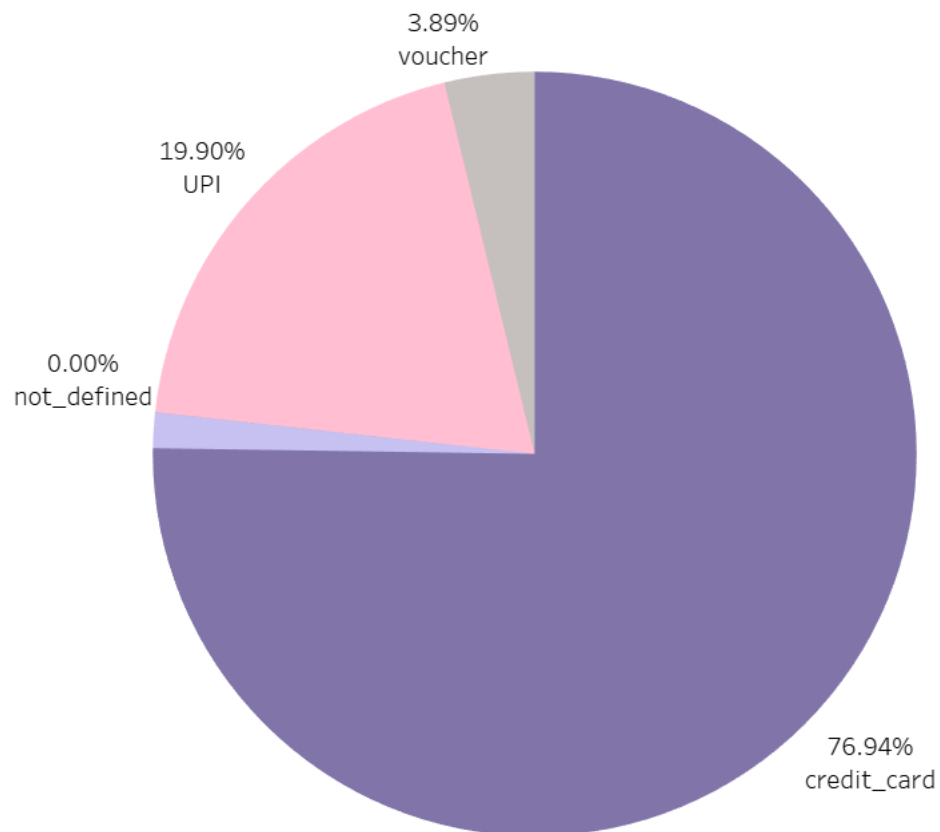
**Query:**

```
with cte as
(
    select *,
        extract(month from o.order_purchase_timestamp) as
Month
from `Target_SQL.orders` as o
inner join `Target_SQL.payments` as p
on p.order_id=o.order_id
)
select Month,payment_type, count(*) as
Count_of_orders
from cte
group by Month,payment_type
order by Month,payment_type;
```

**Result:**

Row	Month	payment_type	Count_of_orders
1	1	UPI	1715
2	1	credit_card	6103
3	1	debit_card	118
4	1	voucher	477
5	2	UPI	1723
6	2	credit_card	6609
7	2	debit_card	82
8	2	voucher	424
9	3	UPI	1942
10	3	credit_card	7707

## Visualization:



## Insights:

- i. Credit card is most preferred payment type in Brazil for making purchases.
- ii. There is a noticeable increasing trend in the usage of credit cards for making purchases during the months of June to August. This trend coincides with the winter season in Brazil, suggesting a possible correlation between seasonal factors and the preference for credit card transactions.



b) Count of orders based on the no. of payment installments

**Query:**

```
with cte as
(
    select *
    from `Target_SQL.orders` as o
    inner join `Target_SQL.payments` as p
    on p.order_id=o.order_id
)
select payment_installments, count(*) as
Count_of_orders
from cte
group by payment_installments
order by Count_of_orders desc
Limit 10;
```

**Result:**

Row	payment_installments	Count_of_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

**Insight:**

In Brazil 1 month payment installments is mostly preferred for making purchases.

# Actionable Insights

1. Decreasing trend in sales during the months of August to September. To counter this decreasing trend Target can offer targeted promotions as well as gather customer feedback to improve sales.
2. RR (Roraima) state has lowest customer count and monthly sales. To increase sales in this region Target should focus on customer acquisition through targeted marketing efforts and tailored promotions. Target can also implement ad campaigns specifically for peoples of Roraima state.
3. Roraima(RR) state has highest freight value. To reduce freight value in this state Target can optimize its logistics and shipping services in this region.
4. Target takes 28 days on average to deliver an order in Roraima(RR) state. To optimize delivery in this region Target need to focus on its logistics and shipping services. Target can also setup warehouses and delivery centers in this region to reduce overall delivery time .
5. Debit card is least preffered payment method for making purchases in Brazil. To increase the usage of debit card , Target can provide discounts and offers on the debit card. Target can also provide EMI services on debit cards so that customers can have an option to use it for making purchases.

# Recommendations

1. Capitalize on the increasing trend of orders during the winter months (June to August) in Brazil by strategically promoting and offering products or services that align with the preferences and needs of customers during this season.
2. To counter the declining sales trend in August to September, Target can offer targeted promotions and gather customer feedback for potential improvements.
3. To improve customer acquisition and monthly sales in some regions Target should focus on targeted marketing efforts and tailored promotions. Target can also implement intuitive ad campaigns to improve its monthly sales in the least profitable regions.
4. To reduce freight value Target can explore opportunities to optimize shipping and logistics processes. Like Amazon or Myntra Target can provides its own logistics and shipping services to reduce freight value .
5. By optimizing logistics and shipping servicesTarget can reduce average delivery time also it can setup its own warehouses and delivery centers for better inventory management.
6. Target can collaborate with football leagues to promote their brand as football is most famous sport in Brazil.