Shyam Mahadev Patil

shyampatil503@gmail.com

Business Case: Target SQL

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

(1) Checking the structure & characteristics of the dataset:

QUERY 1.1

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

CODE

```
select column_name,
        data_type
from `target.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

INSIGHTS

[Data type of table is used for accurate analysis and interpretation of the dataset]

QUERY 1.2

Get the time range between which the orders were placed.

CODE

```
select
min(order_purchase_timestamp)first_order_time,
max(order_purchase_timestamp)last_order_time
from `target.orders`
```

OUTPUT

Row	first_order_time	last_order_time
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

INSIGHTS

[This provide us time range between fist order and last order.From here we determine the start and end date of the data i.e from 4th september 2016 to 17 october 2018]

QUERY 1.3

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

##when customer id is present in orders table then we say that they are the customers who made the orders

CODE

```
select
count(distinct customer_city)number_of_cities , count(distinct customer_state)number_of_states
from `target.orders` o
join `target.customers` c
on o.customer_id = c.customer_id
```

OUTPUT

Row	number_of_cities	number_of_states
1	4119	27

INSIGHTS

[Here we knew that there are 27 states and 4119 cities and this data provide the context for our analysis in Brazil]

==========

(2) In-depth Exploration:

QUERY 2.1

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

CODE

```
select
extract (year from o.order_purchase_timestamp)as year,
extract (month from o.order_purchase_timestamp)as month,
count(distinct o.order_id)number_of_orders,
round(sum(p.payment_value),2)as total_sale
from `target.orders` o
join `target.payments` p
on o.order_id = p.order_id
group by year,month
order by year,month
```

OUTPUT

Row	year	month	number_of_orders	total_sale
1	2016	9	3	252.24
2	2016	10	324	59090.48
3	2016	12	1	19.62
4	2017	1	800	138488.04
5	2017	2	1780	291908.01
6	2017	3	2682	449863.6
7	2017	4	2404	417788.03
8	2017	5	3700	592918.82
9	2017	6	3245	511276.38
10	2017	7	4026	592382.92

INSIGHTS

[Based on above analysis of number of orders ,it can be seen that there is a growing trend of target business in Brazil.It shoes overall upward trend,with some fluctuations in between.

QUERY 2.2

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

CODE

```
select
extract (month from o.order_purchase_timestamp)as month,
count(distinct o.order_id)number_of_orders,
round(sum(p.payment_value),2)total_sale
from `target.orders` o
join `target.payments`p
on o.order_id = p.order_id
group by month
order by month
```

OUTPUT

Row	month	number_of_orders	total_sale
1	1	8069	1253492.22
2	2	8508	1284371.35
3	3	9893	1609515.72
4	4	9343	1578573.51
5	5	10573	1746900.97
6	6	9412	1535156.88
7	7	10318	1658923.67
8	8	10843	1696821.64
9	9	4304	732454.23
10	10	4959	839358.03

INSIGHTS

[From the above analysis we can conclude some seasonability in the e-commerce orders. It is seen that number of orders and total sale generally increases from Marh to August.August shows peak and then it start decreasing with fluctuations.]

QUERY 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

CODE

```
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 hour,
count(order_id)number_of_orders
from `target.orders`
group by hour
order by number_of_orders desc
```

OUTPUT

Row	hour	number_of_orders
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

INSIGHTS

[From above analysis it is seen that Brazilion customers place their most of the orders in afternoon and it followed by night and least in dawn. By this the company can identify peak buying times and do their operations accordingly.]

==========

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

QUERY 3.1

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

CODE

```
select
c.customer_state,
extract (month from o.order_purchase_timestamp)as month,
count(o.order_id) number_of_orders
from `target.orders`o
join `target.customers`c
on o.customer_id = c.customer_id
group by month,customer_state
order by customer_state , month
```

OUTPUT

Row	customer_state	month	number_of_orders
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

INSIGHTS

[This provides valuable insights of number of orders on a state by state basis in any given month. SP state has the highest number of orders in any given month.]

QUERY 3.2

How are the customers distributed across all the states?

CODE

```
select

customer_state,count(customer_id)number_of_customers

from `target.customers`

group by customer_state

order by number_of_customers desc
```

OUTPUT

Row	customer_state	number_of_custome
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

INSIGHTS

[From above analysis it is seen that SP,RJ and MG are three states having highest number of customers with SP state is on top .By use of this analysis company can effectively target specific regions.]

==========

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

QUERY 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.

CODE

```
with cte as
(
select
extract(year from o.order_purchase_timestamp)year,
sum(p.payment_value)cost_of_orders
from `target.orders`o
join `target.payments`p
on o.order_id = p.order_id
where extract(month from o.order_purchase_timestamp)< 9 and
```

```
extract(year from o.order_purchase_timestamp) = 2017

group by year

union all

select

extract(year from o.order_purchase_timestamp)year,

sum(p.payment_value)cost_of_orders

from `target.orders`o

join `target.payments`p

on o.order_id = p.order_id

where extract(month from o.order_purchase_timestamp)< 9 and

extract(year from o.order_purchase_timestamp) = 2018

group by year

)

select

round((((max(cost_of_orders)-
min(cost_of_orders))/min(cost_of_orders))*100),2)percentage_increase_in_cost_of_orders

from cte
```

OUTPUT

Row	percentage_increase
1	136.98

INSIGHTS

[The overall percentage increase in the cost of orders from 2017 to 2018 which include months from january to august is 136.98]

QUERY 4.2

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

CODE

```
select
c.customer_state, round(sum(ot.price),2)total_order_price, round(avg(ot.price),2)avg_order_price
from `target.order_items`ot
join `target.orders`o
on ot.order_id=o.order_id
join `target.customers`c
on c.customer_id=o.customer_id
group by customer_state
order by total_order_price desc
```

OUTPUT

Row	customer_state	total_order_price	avg_order_price
2	RJ	1824092.67	125.12
3	MG	1585308.03	120.75
4	RS	750304.02	120.34
5	PR	683083.76	119.0
6	SC	520553.34	124.65
7	BA	511349.99	134.6
8	DF	302603.94	125.77
9	GO	294591.95	126.27
10	ES	275037.31	121.91

INSIGHTS

[It shows SP state has the highest total price value.By this we analyse cost trends and price with state wise patterns]

QUERY 4.3

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

CODE

```
select

c.customer_state, round(sum(ot.freight_value),2)total_freight_value,
round(avg(ot.freight_value),2)avg_freight_value

from `target.order_items`ot

join `target.orders`o

on ot.order_id=o.order_id

join `target.customers`c

on c.customer_id=o.customer_id

group by customer_state

order by customer_state
```

OUTPUT

Row	customer_state	total_freight_value	avg_freight_value
1	AC	3686.75	40.07
2	AL	15914.59	35.84
3	AM	5478.89	33.21
4	AP	2788.5	34.01
5	BA	100156.68	26.36
6	CE	48351.59	32.71
7	DF	50625.5	21.04
8	ES	49764.6	22.06
9	GO	53114.98	22.77
10	MA	31523.77	38.26

INSIGHTS

[It shows comprensive analysis of freight values which are state wise and gives valuable insights for strategies.]

---

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

QUERY 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:

--time_to_deliver = order_delivered_customer_date - order_purchase_timestamp

--diff_estimated_delivery = order_delivered_customer_date - order_estimated_delivery_date

CODE

```
select
distinct order_id,
date_diff(order_delivered_customer_date,order_purchase_timestamp,day)time_to_deliver,
date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)diff_estimated_delivery
from `target.orders`
where date_diff(order_delivered_customer_date,order_purchase_timestamp,day) is not null
```

OUTPUT

Row	order_id	time_to_deliver	diff_estimated_delive
1	1950d777989f6a877539f5379...	30	-12
2	2c45c33d2f9cb8ff8b1c86cc28...	30	28
3	65d1e226dfaeb8cdc42f66542...	35	16
4	635c894d068ac37e6e03dc54e...	30	1
5	3b97562c3aee8bdedcb5c2e45...	32	0
6	68f47f50f04c4cb6774570cfde...	29	1
7	276e9ec344d3bf029ff83a161c...	43	-4
8	54e1a3c2b97fb0809da548a59...	40	-4
9	fd04fa4105ee8045f6a0139ca5...	37	-1
10	302bb8109d097a9fc6e9cefc5...	33	-5

INSIGHTS

[Above analysis gives us idea about time duration between purchasing an order,its delivery,and estimated delivery and we try to optimize business operations]

QUERY 5.2

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

CODE

```
with cte as
(select c.customer_state,
round(avg(ot.freight_value),2)avg_freight_value
from `target.customers`c
join `target.orders`o
on c.customer_id = o.customer_id
join `target.order_items`ot
on o.order_id = ot.order_id
group by c.customer_state
)

select customer_state,avg_freight_value
from
(
(select customer_state,avg_freight_value
from
(select*,
dense_rank()over(order by avg_freight_value desc)rnk
from cte
```

```
)
where rnk<=5
order by avg_freight_value  desc)
union all
(select customer_state,avg_freight_value
from
(select*,
dense_rank()over(order by avg_freight_value asc)rnk
from cte
)
where rnk<=5
order by avg_freight_value asc)
)
order by avg_freight_value asc
```

OUTPUT

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

(Here we combine top 5 states with highest avg freight value and 5 states with lowest avg freight value in single output table for better comparision)

INSIGHTS

[From above analysis it is seen that SP state has lowest avg freight value among all states ,followed by PR state, while RR state has highest avg freight value followed by PB]

QUERY 5.3

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

CODE

```
with cte as
(select
c.customer_state,
round(avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day)),2)avg_delivery_time
from `target.orders`o
join `target.customers`c
on o.customer_id = c.customer_id
group by customer_state)

select customer_state,avg_delivery_time
from
(
```

```
(select
customer_state,avg_delivery_time
from
(select customer_state,avg_delivery_time,
dense_rank()over(order by avg_delivery_time asc)rnk
from cte
order by avg_delivery_time asc
)
where rnk <= 5 )
union all
(select
customer_state,avg_delivery_time
from
(select customer_state,avg_delivery_time,
dense_rank()over(order by avg_delivery_time desc)rnk
from cte
order by avg_delivery_time desc
)
where rnk <= 5)
)
order by avg_delivery_time asc
```

OUTPUT

Row	customer_state	avg_delivery_time
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48
6	PA	23.32
7	AL	24.04
8	AM	25.99
9	AP	26.73
10	RR	28.98

(Here we combine top 5 states with highest avg delivery time and 5 states with lowest avg delivery time in single output table for better comparision)

INSIGHTS

[From above analysis it is seen that SP state has the lowest average delivery time,while RR state has the highest average delivery time]

QUERY 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.

CODE

```
select customer_state , diff_estimated_delivery

from

(select *,

dense_rank()over(order by diff_estimated_delivery desc)rnk

from

(

select  c.customer_state,

round(avg(date_diff(o.order_estimated_delivery_date,o.order_delivered_customer_date,day)),2)diff_estimated_delivery

from `target.orders`o

join `target.customers`c

on o.customer_id = c.customer_id

group by c.customer_state

))

where rnk<= 5

order by diff_estimated_delivery desc
```

OUTPUT

Row	customer_state	diff_estimated_deliv
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

INSIGHTS

[From above analysis we seen state wise data with comparison of actual and estimated delivery date.It deals with required for customer satisfaction and crucial for business optimization]

---

(6) Analysis based on the payments:

QUERY 6.1

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

CODE

```
select  p.payment_type,
extract(month from o.order_purchase_timestamp)as month,
count(o.order_id) as number_of_orders
from `target.orders` o
join `target.payments` p
on o.order_id = p.order_id
group by p.payment_type , month
order by p.payment_type , month
```

OUTPUT

Row	payment_type	month	number_of_orders
1	UPI	1	1715
2	UPI	2	1723
3	UPI	3	1942
4	UPI	4	1783
5	UPI	5	2035
6	UPI	6	1807
7	UPI	7	2074
8	UPI	8	2077
9	UPI	9	903
10	UPI	10	1056

INSIGHTS

[From above analysis it is seen that CREDIT CARD payment method is associated with large number of orders,followed by UPI . Also we seen that the DEBIT CARD is the least preferred method.]

QUERY 6.2

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

CODE

```
select p.payment_installments,

count(o.order_id) as number_of_orders

from `target.orders` o
```

```
join `target.payments` p
on o.order_id = p.order_id
group by p.payment_installments
order by number_of_orders desc
```

OUTPUT

Row	payment_installment	number_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

INSIGHTS

[From above analysis it is seen that the maximum number of orders have only one payment installment,while the highest number of installment is 24 and it related to 18 orders]

Actionable Insights

1 By focusing on the states with lesser number of orders there is a chance of expansion of customers base. States like SP have high order counts for further improvement need to focus on customer service experiences.

2. By improving delivery times in areas with longer delivery durations with efficient shipping process can create some positive impact on customer satisfaction and it encourage to them for repeat purchases.

3. Here it shows seasonality and this will help to increase the overall sales growth by marketing strategies during peak periods in mind.

4. There is some specific day time in which large number of order's were placed and with this data there is a chance to increase customer base by marketing strategies and offering discounts accordingly.

5. Money movement by e-commerce with order prices, freight and others going to create impact on economic conditions of business. Here this data help to identify areas of improvement during economic fluctuations.

6. There is a state wise analysis, which shows number of orders in any given months for each state and customers distribution across all states. This can provide valuable insights for marketing strategies to deal with specific target customers, which leads in growth of sales.

Recommendations

1. Improve the customer service experience by offering different support services .Also gives proper responses to customers regarding their inquiries.

2. Use of different marketing strategies like offering discounts , proper pricing in comparison with competitor .

3. Increase the brand awareness using social media will create impact on customers.

4. Make the customers e-commerce experience more delight by improving website performance and product recommendation based on their activity and their searching.

5. Take care of valid retention strategies because it makes influence on purchasing decisions.

6. Improve customer satisfaction by proper shipping process and faster deliveries.

Conclusion

Overall this data provides valuable insights regarding Target's operations in Brazilian market. The information can shed light on various aspects of the business and this is useful for optimize business operations, business strategies and to create positive impact on companies sales growth.

==========