Business Case: Target SQL

Prepared by: AJJAPALLY NITHISH REDDY

Date: October 5, 2024

Question 1: What does 'good' look like?

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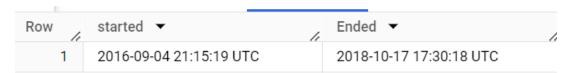
Output:

	Field name	Type	Mode	Key	Collation	Default Value
☐	customer_id	STRING	NULLABLE	-	-	-
☐	customer_unique_id	STRING	NULLABLE	-	-	-
☐	customer_zip_code_prefix	INTEGER	NULLABLE	-	-	-
☐	customer_city	STRING	NULLABLE	-	-	-
☐	customer_state	STRING	NULLABLE	-	-	-

b   Get the time range between which the orders were placed.

Query:

```
SELECT MIN(order_purchase_timestamp) AS started,
       MAX(order_purchase_timestamp) AS ended
FROM 'TargetEcommerce.orders';
```

Output:

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

Recommendations:

> Use this time range to analyze trends and predict the future sales pattern.

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

Query:

```
select count(distinct customer_city) as Cities,count(distinct customer_state)
as States
from 'TargetEcommerce.customers' c
join 'TargetEcommerce.orders' o
on c.customer_id=o.customer_id
```

Output:

Row	cities ▾	states ▾	
1	4119	27	

Insights:

This dataset consists of diverse customer locations from diverse states and cities throughout Brazil

Recommendations:

Keep the focus on regional strategies to strengthen sales in every area

Question 2: In-depth Exploration:

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

Query:

```
select extract(year from order_purchase_timestamp) as year_,count(*) as
orders_placed
    from 'TargetEcommerce.orders'
    group by 1
    order by 1
```

Output:

Row	year_ ▾	orders_placed ▾	
1	2016	329	
2	2017	45101	
3	2018	54011	

Insights:

Since Target started its retail operations in the last quarter of 2016, the initial number of orders was lower. However, from 2017 to 2018, the number of orders increased by approximately 19

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

Query:

```
select extract(month from order_purchase_timestamp)
as month_,
extract (year from order_purchase_timestamp)
as year_ ,count(*) as orders_placed
from 'TargetEcommerce.orders'
group by 1 ,2
order by 2,1
```

Output:

Row	month_	year_	orders_placed
1	11	2017	7544
2	1	2018	7269
3	3	2018	7211
4	4	2018	6939
5	5	2018	6873
6	2	2018	6728
7	8	2018	6512
8	7	2018	6292
9	6	2018	6167
10	12	2017	5673

Insights:

I have noticed that the highest number of orders is placed in the month before the holiday season. During the holiday season (December 2017), there is a slight decrease in orders, but after that, there is a rapid increase in orders starting from the first month of 2018. The order volume remains consistently stable from January to March, likely due to the back-to-school or work period after the holiday season and the peak summer months. Following this stable period, there is a decrease in orders, possibly due to climate changes and the post-holiday period

Recommendations:

Extending promotional offers during the holiday season could boost sales, while introducing off-season sales on winter clothing and home appliances may offset the decrease in revenue.

c   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:

```
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'
when (extract(hour from order_purchase_timestamp))
between 19 and 23 then 'Night' end as time_of_the_day,count(order_id) as orders
from 'TargetEcommerce.orders'
group by 1
order by 2 desc;
```

Output:

Row	time_of_the_day	orders
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Insights:

In the afternoon, Brazilian customers tend to place most of their orders

Recommendations:

So, there is a rapid decrease in orders during the midnight, i.e., dawn hours. Introducing midnight sales will help increase the orders and cover those hours.

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

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

Query:

```
select extract(month from o.order_purchase_timestamp)
as month_, c.customer_state,count(o.order_id) as
orders
from 'TargetEcommerce.orders' o
join 'TargetEcommerce.customers' c
on o.customer_id=c.customer_id
group by 1,2
order by orders desc
limit 10
```

Output:

Row	month_	customer_state	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:

São Paulo (SP) is the most populous and economically significant state, which correlates with the majority of orders being placed from that region each month.

Recommendations:

Considering that most orders come from São Paulo (SP), it might be beneficial to implement sales promotions in other states to attract more customers. Providing customized deals for different regions could potentially boost sales, and adapting strategies to local preferences may further improve sales performance.

b   How are the customers distributed across all the states?

Query:

```
select customer_state,count(distinct customer_id) as customers
from 'TargetEcommerce.customers'
group by 1
order by 2 desc
limit 10
```

Output:

Row	customer_state	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

Insights:

Data indicates that 42% of customers are located in the state of São Paulo (SP)

Recommendations:

To attract new customers, it is important to expand our focus beyond São Paulo (SP), where the majority of our customers are located. We can achieve this by customizing our marketing campaigns and targeting regional events and carnivals. Implementing sales promotions that align with these events may help improve our overall sales performance.

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

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

Query:

```
with cte_1 as(
 select extract(year from order_purchase_timestamp)
 as year_ ,round(sum(p.payment_value)) as payment
 from `TargetEcommerce.payments` p
 join `TargetEcommerce.orders` o
 on p.order_id=o.order_id
 where extract(month from order_purchase_timestamp)
 between 1 and 8
 group by 1),

 cte_2  as (
 select *,lead(payment,1)over(order by year_ desc)
 as previous_year   from cte_1)

 select round(((payment*100)/previous_year)-100) as percent_increase
 from cte_2
 where year_=2018
 and previous_year is not null
```

Output:

Row	percent_increase ▼
1	137.0

Insights:

The cost of orders from January to August 2017 was 3669022, and in the same months of 2018, it increased to 8694734. This shows a rapid increase of 137%. Previously, in 2017, sales were lower during the non-holiday season, but in 2018, there was a 137%

Recommendations:

Continuous analysis of the latest trends and adaptive strategies will be crucial for effectively responding to changes in customer behavior and market conditions, given the increase in sales

compared to the previous year.

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

Query:

```
 with cte_1 as(
select   c.customer_state,sum(oi.price) as total_value,count(distinct oi.order_id)
as total_orders
from  'TargetEcommerce.order_items' oi
join 'TargetEcommerce.orders' o
on oi.order_id = o.order_id
join 'TargetEcommerce.customers' c
on c.customer_id=o.customer_id
group by 1
order by total_value desc)

select customer_state,total_value,
total_value/total_orders as avg_value
from cte_1
order by total_value desc
limit 10
```

Output:

Row	customer_state	total_value	avg_value
1	SP	5202955.050001…	125.7511794562…
2	RJ	1824092.669999…	142.9315679360…
3	MG	1585308.029999…	137.3274454261…
4	RS	750304.0200000…	138.1266605301…
5	PR	683083.7600000…	136.6714205682…
6	SC	520553.3400000…	144.1177574750…
7	BA	511349.9900000…	152.2781387730…
8	DF	302603.9399999…	142.4018541176…
9	GO	294591.9499999…	146.7822371699…
10	ES	275037.3099999…	135.8208938271…

Insights:

While São Paulo (SP) has the highest total order value, Bahia (BA) has the highest average order value. This suggests that even though São Paulo has a large number of customers making

smaller purchases due to its high population, states like Bahia (BA) and Santa Catarina (SC) have fewer customers but with higher average purchase values. This indicates that customers in these states tend to make higher-value purchases.

Recommendations:

Implement strategies to encourage bulk purchases, such as offering discounts for bulk purchases in São Paulo (SP) to increase the average order value. Similarly, focus on enhancing sales efforts in states like Espírito Santo (ES) and Goiás (GO). Additionally, promote premium products to increase total sales value in states such as Bahia (BA) and Santa Catarina (SC)

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

Query:

```
with cte_1 as(
select c.customer_state,sum(oi.freight_value)
as total_value,count(distinct oi.order_id)
as total_orders
from  'TargetEcommerce.order_items' oi
join 'TargetEcommerce.orders' o
on oi.order_id = o.order_id
join 'TargetEcommerce.customers' c
on c.customer_id=o.customer_id
group by 1)

select customer_state,total_value,
total_value/total_orders as avg_value
from cte_1
order by avg_value desc
limit 10
```

Output:

Row	customer_state	total_value	avg_value
1	RR	2235.19	48.59108695652...
2	PB	25719.73000000...	48.34535714285...
3	RO	11417.37999999...	46.22421052631...
4	AC	3686.749999999...	45.51543209876...
5	PI	21218.20000000...	43.03894523326...
6	MA	31523.77000000...	42.59968918918...
7	TO	11732.68000000...	42.05261648745...
8	AP	2788.500000000...	41.00735294117...
9	SE	14111.46999999...	40.90281159420...
10	PA	38699.30000000...	39.89618556701...

Insights:

Roraima (RR) and Paraíba (PB) have higher average shipping costs, likely because they are located in remote areas.

Recommendations:

To optimize logistics in areas with high freight costs, consider forming partnerships with local delivery companies or establishing regional warehouses. These strategies can help reduce shipping expenses and improve delivery efficiency.

Question 5: Analysis based on sales, freight and delivery time.

a   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.

Query:

```
select  order_purchase_timestamp,order_estimated_delivery_date,
order_delivered_customer_date,
date_diff(order_delivered_customer_date, order_purchase_timestamp,day ) as
time_to_deliver,
timestamp_diff(order_delivered_customer_date,
order_estimated_delivery_date,day)
as diff_estimated_deivery
from 'TargetEcommerce.orders'
order by order_delivered_customer_date desc
limit 10
```

Output:

Row	order_purchase_timestamp	order_estimated_delivery_date	order_delivered_customer_date	time_to_deliver	diff_estimated_deive
1	2018-06-02 18:37:14 UTC	2018-07-13 00:00:00 UTC	2018-10-17 13:22:46 UTC	136	96
2	2018-05-21 06:48:46 UTC	2018-06-27 00:00:00 UTC	2018-10-11 16:41:14 UTC	143	106
3	2018-07-30 09:08:06 UTC	2018-08-14 00:00:00 UTC	2018-10-02 00:18:50 UTC	63	49
4	2018-07-22 09:54:03 UTC	2018-08-17 00:00:00 UTC	2018-09-27 02:24:33 UTC	66	41
5	2018-08-05 17:11:44 UTC	2018-08-20 00:00:00 UTC	2018-09-25 00:47:25 UTC	50	36
6	2018-07-15 02:11:15 UTC	2018-08-06 00:00:00 UTC	2018-09-21 23:46:29 UTC	68	46
7	2018-08-01 19:43:06 UTC	2018-08-23 00:00:00 UTC	2018-09-21 15:55:02 UTC	50	29
8	2018-07-23 10:22:26 UTC	2018-08-13 00:00:00 UTC	2018-09-20 16:08:33 UTC	59	38
9	2018-02-23 14:57:35 UTC	2018-03-15 00:00:00 UTC	2018-09-19 23:24:07 UTC	208	188
10	2018-08-06 14:32:27 UTC	2018-08-17 00:00:00 UTC	2018-09-19 16:44:44 UTC	44	33

Insights:

Most orders take longer than the estimated time for delivery.

Recommendations:

Prioritize reducing delivery times in states with longer wait times by partnering with more reliable local carriers

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

Query:

```
(select distinct c.customer_state,
avg(a.freight_value)over(partition by customer_state)as avg_,
'Top-5' as rank_type
from 'TargetEcommerce.order_items' a
join 'TargetEcommerce.orders' b
on a.order_id=b.order_id
join 'TargetEcommerce.customers' c
on b.customer_id=c.customer_id
order by 2 desc
limit 5)
union all
(select distinct c.customer_state,
avg(a.freight_value)over(partition by customer_state)as avg_,
'Bottom-5' as rank_type
from 'TargetEcommerce.order_items' a
join 'TargetEcommerce.orders' b
on a.order_id=b.order_id
join 'TargetEcommerce.customers' c
on b.customer_id=c.customer_id
order by 2 asc
```

Row	order_purchase_timestamp	order_estimated_delivery_date	order_delivered_customer_date	time_to_deliver	diff_estimated_deive

```
limit 5)
order by 2 desc
```

Output:

Row	customer_state	avg_	rank_type
1	SP	15.14727539041...	Top-5
2	PR	20.53165156794...	Top-5
3	MG	20.63016680630...	Top-5
4	RJ	20.96092393168...	Top-5
5	DF	21.04135494596...	Top-5
6	PI	39.14797047970...	Bottom-5
7	AC	40.07336956521...	Bottom-5
8	RO	41.06971223021...	Bottom-5
9	PB	42.72380398671...	Bottom-5
10	RR	42.98442307692...	Bottom-5

Insights:

Roraima (RR) and Paraíba (PB) have the highest average freight costs, whereas São Paulo (SP), Paraná (PR), and Minas Gerais (MG) have the lowest.

Recommendations:

In order to improve operational efficiency and cut costs, it's important to strategically allocate resources to enhance logistics in regions where shipping costs are high. This can be achieved through partnerships with local logistics providers. Furthermore, consider offering expedited shipping options in areas with lower shipping costs to improve customer satisfaction even more.

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

Query:

```
(select c.customer_state,round(avg(timestamp_diff(order_delivered_customer_date,
order_purchase_timestamp,day))) as avg_delivery_time,'top-5' as rank_type
from 'TargetEcommerce.orders' o
join 'TargetEcommerce.customers' c
on c.customer_id=o.customer_id
group by 1
order by 2
limit 5)

union all
```

```
(select c.customer_state,round(avg(timestamp_diff(order_delivered_customer_date,
order_purchase_timestamp,day))) as avg_delivery_time,'bottom-5' as rank_type
from 'TargetEcommerce.orders' o
join 'TargetEcommerce.customers' c
on c.customer_id=o.customer_id
group by 1
order by 2 desc
limit 5)
order by avg_delivery_time
```

Output:

Row	customer_state	avg_delivery_time	rank_type
1	SP	8.0	top-5
2	MG	12.0	top-5
3	PR	12.0	top-5
4	DF	13.0	top-5
5	SC	14.0	top-5
6	PA	23.0	bottom-5
7	AL	24.0	bottom-5
8	AM	26.0	bottom-5
9	AP	27.0	bottom-5
10	RR	29.0	bottom-5

Insights:

Delivery times vary significantly by state, with some regions experiencing delays.

Recommendations:

Prioritize reducing delivery times in states with longer wait times by improving supply chain efficiency or partnering with more reliable local carriers

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

Query:

```
select c.customer_state,
round(avg(timestamp_diff(order_delivered_customer_date,
order_estimated_delivery_date,day)),2)
```

```
as avg_delivery_days
from 'TargetEcommerce.orders' o
join 'TargetEcommerce.customers'  c
on o.customer_id=c.customer_id
where order_status='delivered'
group by 1
order by 2
limit 5
```

Output:

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

Insights:

Here "-" indicates the days before the estimated delivery time. Certain states regularly outperform the estimated delivery times.

Recommendations:

Highlight fast delivery times in marketing campaigns for these regions to turn it into a competitive advantage and attract more customers.

Question 6:Analysis based on the payments:

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

Query:

```
with cte as (
select o.order_id,payment_type,extract(year from order_purchase_timestamp)
as year_,
extract(month from order_purchase_timestamp) as month_
from 'TargetEcommerce.payments' p
```

```
join 'TargetEcommerce.orders' o
on o.order_id=p.order_id)
select year_, month_,payment_type,count(*) as orders
from cte
group by 1,2,3
order by 4 desc
limit 10
```

Output:

Row	year_	month_	payment_type	orders
1	2017	11	credit_card	5897
2	2018	3	credit_card	5691
3	2018	1	credit_card	5520
4	2018	5	credit_card	5497
5	2018	4	credit_card	5455
6	2018	2	credit_card	5253
7	2018	8	credit_card	4985
8	2018	6	credit_card	4813
9	2018	7	credit_card	4755
10	2017	12	credit_card	4377

Insights:

Credit cards remain the most popular payment method among customers at all times.

Recommendations:

During busy months, offer special deals for specific payment methods to encourage customers to try other ways to pay. This will help increase the variety of payment options used.

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

Query:

```
select p.payment_installments as no_of_installments,count(distinct o.order_id)
as orders
from 'TargetEcommerce.payments' p
join 'TargetEcommerce.orders' o
on p.order_id=o.order_id
where payment_installments != 0
group by 1
order by 1 desc
```

```
limit 10
```

Output:

Row	no_of_installments	orders
1	1	49060
2	2	12389
3	3	10443
4	4	7088
5	10	5315
6	5	5234
7	8	4253
8	6	3916
9	7	1623
10	9	644

Insights:

Most customers prefer to pay in a single installment, reflecting a preference for upfront payment of the full amount.

Recommendations:

Maintain the option for single-installment payments while also highlighting flexible installment plans for higher-value purchases to attract customers who prefer spreading payments over time.