Business Case: Target SQL

Target's operations in Brazil between 2016 and 2018

Q1) Importing datasets and checking the structure and characteristics

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

   Ans:

   Query:

   ```sql
   SELECT column_name, data_type
   FROM norse-limiter-415014.target.INFORMATION_SCHEMA.COLUMNS
   WHERE table_name = 'customers';
   ```

   Snapshot:

JOB INFORMATION	RESULTS	CHART

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: NA
   Recommendations: NA

2. Time range between which the orders were placed.

   Ans:

   Query:

   ```sql
   SELECT MIN(order_purchase_timestamp) as `earliest_order_timestamp`,
          MAX(order_purchase_timestamp) as `latest_order_timestamp`
   FROM target.orders;
   ```

   Snapshot:

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

Insights:

The orders were placed between 2016 and 2018 where the earliest order took place on September 4, 2016 and the latest order placed on October 17, 2018.

Recommendations: NA

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

Ans:

Query:

```sql
SELECT count(DISTINCT customer_city) as `No of Cities`,
       count(DISTINCT customer_state) as `No of States`
FROM target.customers c inner join target.orders o on c.customer_id = o.customer_id
WHERE EXTRACT(year FROM o.order_purchase_timestamp) BETWEEN 2016 AND 2018;
```

Snapshot:

Row	No of Cities ▼	No of States ▼
1	4119	27

Insights:

The geographical distribution of customers spread across these many cities have ordered products between 2016 and 2018.

Recommendations:

We can identify the region having lower customers and expand the market based on the potential demand for the particular city/state.

Q2) In-depth Exploration

1. Growing trend in the no. of orders placed over the past years
   Ans:

Query:

```sql
SELECT EXTRACT(year from order_purchase_timestamp) as `order_year`,
       EXTRACT(month from order_purchase_timestamp) as `order_month`, count(order_id) as `order_count`
FROM target.orders
GROUP BY order_year,order_month
ORDER BY order_year,order_month;
```

Snapshot :

Row	order_year	order_month	order_count
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
11	2017	8	4331
12	2017	9	4285

Row	order_year	order_month	order_count
13	2017	10	4631
14	2017	11	7544
15	2017	12	5673
16	2018	1	7269
17	2018	2	6728
18	2018	3	7211
19	2018	4	6939
20	2018	5	6873
21	2018	6	6167
22	2018	7	6292
23	2018	8	6512
24	2018	9	16
25	2018	10	4

Insights:

The number of orders increases consistently over the years for most months and thus indicates a positive trend in order volume.

Recommendations:

Ensure there is adequate stock levels in inventory to meet fluctuating demand as there is a huge decrease in some months.

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

Ans:

Query:

```
SELECT EXTRACT(month from order_purchase_timestamp) as `order_month`,
       count(order_id) as `MonthlyOrders_OverYears`
FROM target.orders
GROUP BY order_month
ORDER BY order_month;
```

Snapshot:

Row	order_month	MonthlyOrders_OverYears
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

Insights:

The number of orders placed are peak during 5th,7th,8th months over the years where it crossed above 10,000 orders. Further, the initial 8 months recorded most number of orders and moving towards the year end the orders placed are comparatively decreased.

Recommendations:

- Launch targeted promotions and marketing campaigns towards the end of the year to stimulate demand and encourage customers to make purchases.

- Offer discounts, special deals, or limited-time offers to attract shoppers during slower periods.

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

Ans:

Query:

```sql
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"
  else "Night"
  END as `time_of_day`, count(order_id) as `order_count`
FROM target.orders
GROUP BY time_of_day
order by order_count desc;
#LIMIT 1 gives the first record as to display the first record alone where highest number of orders placed in afternoon;
```

Snapshot:

Row	time_of_day	order_count
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Insights:

The Brazilian customers mostly placed their orders in the **AFTERNOON** with **38,135** orders, followed by 28,331 orders in the night then 27,733 in the morning; finally 5242 orders which recorded least in the dawn.

Recommendations: **NA**

Q3) Evolution of E-commerce orders:

1. Month on month no. of orders placed in each state

Ans:

Query:

```sql
SELECT c.customer_state,extract(year from o.order_purchase_timestamp) as `order_year`,
       extract(month from o.order_purchase_timestamp) as `order_month`,count(o.order_id) as `No of orders`
FROM target.orders o inner join target.customers c on o.customer_id = c.customer_id
GROUP BY c.customer_state,order_year,order_month
ORDER BY c.customer_state,order_year,order_month;
```

Row	customer_state	order_year	order_month	No of orders
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

Row	customer_state	order_year	order_month	No of orders
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
21	AL	2016	10	2
22	AL	2017	1	2
23	AL	2017	2	12
24	AL	2017	3	10

2. The customers distributed across all the states

Ans:

Query:

```
SELECT customer_state,count(DISTINCT customer_id) as `customer_count`
FROM target.customers
GROUP BY customer_state
ORDER BY customer_count desc;
```

Snapshot:

Row	customer_state	customer_count
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
11	PE	1652
12	CE	1336

Row	customer_state	customer_count
16	MS	715
17	PB	536
18	PI	495
19	RN	485
20	AL	413
21	SE	350
22	TO	280
23	RO	253
24	AM	148
25	AC	81
26	AP	68
27	RR	46

Insights:

The SP state has the huge number of customers where the business has large presence and strong customer base. Where in   other states business falls down because of low count of customers.

Recommendations:

1) Develop targeted marketing campaigns tailored to the preferences and characteristics of customers in each state.
2) Explore opportunities for expansion into states with lower customer counts but significant market potential.

Q4) Impact on Economy

1) The % increase in the cost of orders from year 2017 to 2018 (JAN to Aug)
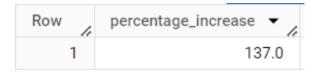
Ans:

Query:

```sql
WITH Yearly_Payment AS(
SELECT EXTRACT(YEAR FROM o.order_purchase_timestamp) AS `order_year`,
       SUM(p.payment_value) AS `total_payment`
FROM target.payments p inner join target.orders o ON p.order_id = o.order_id
WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) IN(2017,2018) AND
      EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
GROUP BY order_year)

SELECT round(((payment_2018.total_payment - payment_2017.total_payment) / payment_2017.total_payment) * 100)
       as `percentage_increase`
FROM Yearly_Payment payment_2018 inner join Yearly_Payment payment_2017
     ON payment_2018.order_year = 2018 AND payment_2017.order_year = 2017 ;
```

Snapshot:

CTE_Yearly_Payment:

Row	order_year ▼	total_payment ▼
1	2018	8694733.8399998639
2	2017	3669022.1199999228

Percentage Increase :

Row	percentage_increase ▼
1	137.0

Insights:

The percentage of the cost of orders from 2017 to 2018 with the months included from January to August has increased by 136.97 %(almost 137).

Recommendations:

Use historical data to predict future order costs and plan budgets, inventory levels, and resource allocation accordingly.

2) Total & Average value of order price for each state.
   Ans:

Query:

```sql
SELECT  c.customer_state,SUM(oi.price) AS total_order_price,AVG(oi.price) as avg_order_price
FROM target.order_items oi INNER JOIN target.orders o ON oi.order_id = o.order_id
                           INNER JOIN target.customers c on o.customer_id = c.customer_id
GROUP BY c.customer_state
order by total_order_price desc;
```

Snapshot:

Row	customer_state	total_order_price	avg_order_price
1	SP	5202955.050002…	109.6536291597…
2	RJ	1824092.669999…	125.1178180945…
3	MG	1585308.029999…	120.7485741488…
4	RS	750304.0200000…	120.3374530874…
5	PR	683083.7600000…	119.0041393728…
6	SC	520553.3400000…	124.6535775862…
7	BA	511349.9900000…	134.6012082126…
8	DF	302603.9399999…	125.7705486284…
9	GO	294591.9499999…	126.2717316759…
10	ES	275037.3099999…	121.9137012411…
11	PE	262788.0299999…	145.5083222591…
12	CE	227254.7099999…	153.7582611637…

Row	customer_state	total_order_price	avg_order_price
16	MS	116812.6399999...	142.6283760683...
17	PB	115268.0799999...	191.4752159468...
18	PI	86914.07999999...	160.3580811808...
19	RN	83034.97999999...	156.9659357277...
20	AL	80314.80999999...	180.8892117117...
21	SE	58920.85000000...	153.0411688311...
22	TO	49621.74000000...	157.5293333333...
23	RO	46140.64000000...	165.9735251798...
24	AM	22356.84000000...	135.4959999999...
25	AC	15982.94999999...	173.7277173913...
26	AP	13474.29999999...	164.3207317073...
27	RR	7829.429999999...	150.5659615384...

Insights:

A revenue of 52,02,955 price value has been generated by the customers of the state SP which is huge and followed by the state RJ nearly 18,24000.The customers from the state RR recorded a very low revenue through orders which is of approx 8000.

Recommendations:

- Adjust pricing strategies based on regional demand, competition, and customer purchasing power to maximize revenue and profitability.

- Tailor marketing campaigns and promotions to specific states based on their order price metrics.

- Focus marketing efforts on states with lower average order prices to stimulate sales and encourage larger purchases, while offering incentives to high-value states to maintain customer loyalty.

3) Total & Average value of order freight for each state.
   Ans:

   Query:

```sql
SELECT  c.customer_state,
        SUM(oi.freight_value) AS total_order_freight_value,
        AVG(oi.freight_value) as avg_order_freight_value
FROM target.order_items oi INNER JOIN target.orders o ON oi.order_id = o.order_id
        inner join target.customers c on o.customer_id = c.customer_id
GROUP BY c.customer_state
order by total_order_freight_value desc;
```

Snapshot:

Row	customer_state	total_order_freight_value	avg_order_freight_value
1	SP	718723.0699999833	15.147275390419248
2	RJ	305589.31000000035	20.96092393168248
3	MG	270853.46000000357	20.630166806306541
4	RS	135522.74000000212	21.735804330392945
5	PR	117851.68000000139	20.531651567944248
6	BA	100156.67999999883	26.363958936562248
7	SC	89660.260000000431	21.470368773946436
8	PE	59449.6599999999	32.917862679955796
9	GO	53114.979999999865	22.766815259322794
10	DF	50625.499999999811	21.041354945968383
11	ES	49764.599999999889	22.058776595744682
12	CE	48351.589999999924	32.714201623815995

Row	customer_state	total_order_freight_value	avg_order_freight_value
16	PB	25719.730000000029	42.723803986710941
17	PI	21218.200000000033	39.147970479704767
18	MS	19144.030000000006	23.374884004884006
19	RN	18860.100000000013	35.652362948960295
20	AL	15914.589999999991	35.843671171171152
21	SE	14111.469999999983	36.653168831168855
22	TO	11732.680000000013	37.246603174603187
23	RO	11417.379999999996	41.069712230215842
24	AM	5478.8899999999967	33.205393939393936
25	AC	3686.7499999999991	40.073369565217405
26	AP	2788.5000000000009	34.006097560975618
27	RR	2235.19	42.984423076923093

Insights:

- The price rate at which a product is delivered from one point to another that is freight value is high in the state SP ie., 718723.

- This might be because of regional variances in shipping distances, logistics infrastructure, or carrier availability.

Recommendations:

- Tailor shipping strategies to minimize freight costs while meeting customer expectations.

- Strategically locate warehouses and distribution centers to minimize shipping distances and reduce freight costs for orders shipped to different states.

Q5) Analysis based on sales, freight and delivery time

1. The no. of days taken to deliver each order from the order's purchase date . Also, calculate the difference (in days) between the estimated & actual delivery date of an order.

Ans:

Query:

```sql
SELECT
    order_id,
    DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) AS delivery_time,
    DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date, DAY) AS diff_estimated_delivery
FROM
    target.orders
WHERE order_delivered_customer_date IS NOT NULL
    AND order_estimated_delivery_date IS NOT NULL
ORDER BY delivery_time desc;
#the analysis focuses on completed orders with delivery information available.
```

Snapshot:

Row	order_id	delivery_time	diff_estimated_delivery
1	ca07593549f1816d26a572e06...	209	181
2	1b3190b2dfa9d789e1f14c05b...	208	188
3	440d0d17af552815d15a9e41a...	195	165
4	0f4519c5f1c541ddec9f21b3bd...	194	161
5	285ab9426d6982034523a855f...	194	166
6	2fb597c2f772eca01b1f5c561b...	194	155
7	47b40429ed8cce3aee9199792...	191	175
8	2fe324febf907e3ea3f2aa9650...	189	167

Insights:

- Rows with null values in the calculated date difference indicate incomplete delivery information for those orders.
- This could be due to various reasons such as ongoing orders, missing delivery data, or orders that haven't been fulfilled yet.
- It could signify delays in order fulfillment or discrepancies between estimated and actual delivery dates.

Recommendations:

- Implement robust tracking mechanisms to monitor the progress of orders from purchase to delivery.
- Utilize tracking technologies, status updates, and notifications to keep customers informed about the status of their orders and reduce uncertainty.
- Provide regular updates on order progress, potential delays, and revised delivery estimates to manage customer expectations effectively.

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

Ans:

Query:

```sql
WITH StateFreight AS(
SELECT c.customer_state,AVG(oi.freight_value) AS avg_freight_value
FROM target.order_items oi INNER JOIN target.orders o ON oi.order_id = o.order_id
        inner join target.customers c on o.customer_id = c.customer_id
GROUP BY c.customer_state)

SELECT customer_state,avg_freight_value
FROM
    (
        SELECT customer_state,avg_freight_value,
            ROW_NUMBER() OVER (ORDER BY avg_freight_value DESC) AS rn_high,
            ROW_NUMBER() OVER (ORDER BY avg_freight_value ASC) AS rn_low
        FROM StateFreight
    )
WHERE rn_high <= 5 or rn_low <= 5
ORDER BY avg_freight_value;
```

Snapshot:

Row	customer_state	avg_freight_value
1	SP	15.147275390419132
2	PR	20.531651567944269
3	MG	20.630166806306651
4	RJ	20.960923931682483
5	DF	21.041354945968422
6	PI	39.147970479704838
7	AC	40.073369565217362
8	RO	41.069712230215814
9	PB	42.723803986710969
10	RR	42.984423076923072

Insights:

States with lower average freight values may indicate more efficient logistics operations, optimized shipping routes, or closer proximity to distribution centers. Conversely, higher average freight values may suggest logistical challenges or inefficiencies that contribute to increased shipping costs.

Recommendations:

- Tailor shipping strategies to address regional variations in freight costs.
- Explore options such as zone-based pricing, carrier negotiations, and fulfillment center placement to minimize shipping expenses and improve cost-effectiveness.
- Evaluate the placement of warehouses and distribution centers to reduce shipping distances and lower freight costs.

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

Ans:

Query:

```
WITH StateDeliveryTime AS (
    SELECT c.customer_state,
    AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY))
        AS avg_delivery_time
    FROM target.order_items oi INNER JOIN target.orders o ON oi.order_id = o.order_id inner join
        target.customers c on o.customer_id = c.customer_id
    GROUP BY c.customer_state
)
SELECT customer_state, round(avg_delivery_time) as avg_delivery_time
FROM
    (
        SELECT customer_state, avg_delivery_time,
            ROW_NUMBER() OVER (ORDER BY avg_delivery_time ASC) AS rn_lowest,
            ROW_NUMBER() OVER (ORDER BY avg_delivery_time DESC) AS rn_highest
        FROM StateDeliveryTime
    )
WHERE rn_lowest <= 5 OR rn_highest <= 5
ORDER BY avg_delivery_time ASC;
```

Snapshot:

Row	customer_state	avg_delivery_time
1	SP	8.0
2	PR	11.0
3	MG	12.0
4	DF	13.0
5	SC	15.0
6	PA	23.0
7	AL	24.0
8	AM	26.0
9	RR	28.0
10	AP	28.0

Insights:

- States with the lowest average delivery times, which indicate regions where orders are delivered more quickly.
- This suggests efficient logistics operations, optimized shipping routes, and effective coordination with carriers.
- States with the highest average delivery times, which may indicate logistical challenges or inefficiencies impacting delivery performance.
- States with shorter delivery times are likely to provide better customer experiences, leading to higher levels of customer loyalty and positive brand perception.

Recommendations:

- Invest in route optimization technologies, real-time tracking systems, and local delivery hubs to expedite order fulfillment and improve delivery speed.
- Collaborate closely with reliable carriers and logistics providers to improve delivery performance and reliability.

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

Ans:
Query:

```sql
WITH StateDeliverySpeed AS (
    SELECT
        c.customer_state,
        AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY)) AS avg_delivery_speed
    FROM target.order_items oi INNER JOIN target.orders o ON oi.order_id = o.order_id inner join
        target.customers c on o.customer_id = c.customer_id
    WHERE o.order_status ='delivered'
    GROUP BY c.customer_state
)
SELECT
    customer_state,
    round(avg_delivery_speed) as avg_delivery_speed
FROM
    (
        SELECT customer_state,avg_delivery_speed,
            ROW_NUMBER() OVER (ORDER BY avg_delivery_speed ASC) AS rn
        FROM StateDeliverySpeed
    )
WHERE rn <= 5
ORDER BY avg_delivery_speed;
```

<u>**Snapshot**</u>:

Row	customer_state ▼	avg_delivery_speed
1	AC	-20.0
2	RO	-19.0
3	AM	-19.0
4	AP	-17.0
5	RR	-17.0

<u>**Insights**</u>:

- Orders are consistently delivered ahead of schedule.
- A negative sign in the avg_delivery_speed indicates that, on average, orders in that state were delivered before the estimated delivery date.
- Customers in states with negative delivery speed values are likely to perceive the brand positively and may exhibit higher levels of loyalty and repeat purchases.

<u>**Recommendations**</u>:

- Continue to prioritize order fulfillment efficiency and maintain high service standards to sustain fast delivery times.
- Regularly assess and optimize logistics processes, warehouse operations, and transportation networks to minimize delivery lead times.
- Continue to prioritize order fulfillment efficiency and maintain high service standards to sustain fast delivery times.
- Regularly assess and optimize logistics processes, warehouse operations, and transportation networks to minimize delivery lead times.

Q6) Analysis based on the payments:

1) The month on month no. of orders placed using different payment types.

<u>**Ans:**</u>

Query:

```sql
SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
    p.payment_type,
    COUNT(*) AS num_orders
FROM target.orders o INNER JOIN
    target.payments p ON o.order_id = p.order_id
GROUP BY order_year,order_month,payment_type
ORDER BY order_year,order_month,num_orders DESC;
```

Snapshot:

Row	order_year	order_month	payment_type	num_orders
1	2016	9	credit_card	3
2	2016	10	credit_card	254
3	2016	10	UPI	63
4	2016	10	voucher	23
5	2016	10	debit_card	2
6	2016	12	credit_card	1
7	2017	1	credit_card	583
8	2017	1	UPI	197
9	2017	1	voucher	61
10	2017	1	debit_card	9
11	2017	2	credit_card	1356
12	2017	2	UPI	398

Row	order_year	order_month	payment_type	num_orders
79	2018	7	credit_card	4755
80	2018	7	UPI	1229
81	2018	7	voucher	281
82	2018	7	debit_card	242
83	2018	8	credit_card	4985
84	2018	8	UPI	1139
85	2018	8	voucher	295
86	2018	8	debit_card	277
87	2018	8	not_defined	2
88	2018	9	voucher	15
89	2018	9	not_defined	1
90	2018	10	voucher	4

Insights:

- Over the years the customers tends to use different mode of payments like using credit card,debit card,voucher,UPI transactions etc.
- The credit card usage recorded more number of orders over the years from 2016 to 2018.

Recommendations:

- Prioritize payment security and fraud prevention measures to safeguard customer transactions and build trust.
- Tailor promotional campaigns and marketing initiatives to capitalize on trends in payment method usage.
- Offer exclusive discounts, incentives, or rewards for customers using specific payment methods to encourage adoption and drive sales.

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

Ans:

Query:

```
SELECT p.payment_installments,
       COUNT(o.order_id) AS num_orders
FROM target.payments p INNER JOIN
     target.orders o ON p.order_id = o.order_id
GROUP BY p.payment_installments
ORDER BY p.payment_installments;
```

Snapshot:

Row	payment_installments	num_orders
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
11	10	5328
12	11	23

Row	payment_installments	num_orders
13	12	133
14	13	16
15	14	15
16	15	74
17	16	5
18	17	8
19	18	27
20	20	17
21	21	3
22	22	1
23	23	1
24	24	18

Insights:

The more number of customers chose payment installment "1" may be because of their convenience which recorded 52546 orders followed by installment "2" and then '3'

Recommendations:

- Highlight installment payment options during the checkout process to increase awareness and encourage utilization.
- Provide a variety of installment plans with different terms and conditions to cater to diverse customer preferences and financial situations.