

Business Case: Target SQL

Q.1 Initial exploration like checking the structure & characteristics of the data

1). Data type of column 'customers'.

<input type="checkbox"/>	Field name	Type
<input type="checkbox"/>	customer_id	STRING
<input type="checkbox"/>	customer_unique_id	STRING
<input type="checkbox"/>	customer_zip_code_prefix	INTEGER
<input type="checkbox"/>	customer_city	STRING
<input type="checkbox"/>	customer_state	STRING

2). Time range between which the orders were placed.

```
select
min(order_purchase_timestamp) as first_order_date,
max(order_purchase_timestamp) as last_order_date
from `target_sql.orders`
```

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

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

```
select
count(distinct c.customer_state) as state,
count(distinct c.customer_city) as city
from `target_sql.customers` as c
inner join `target_sql.orders` as o
on c.customer_id = o.customer_id
```

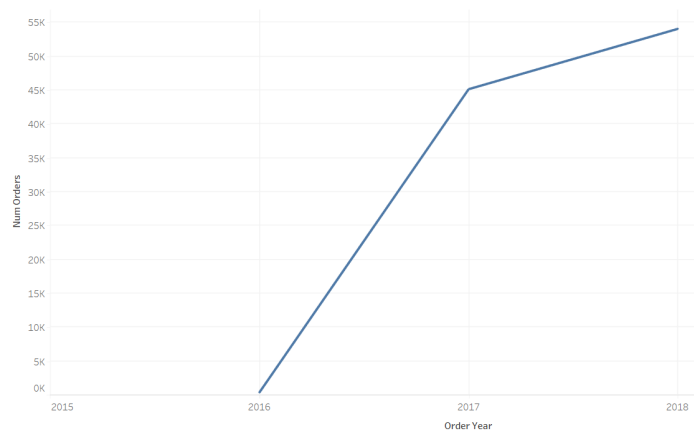
Row	state	city
1	27	4119

Q.2 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,
count(*) as num_orders
from `target_sql.orders`
group by 1
order by 1
```

Row	order_year	num_orders
1	2016	329
2	2017	45101
3	2018	54011

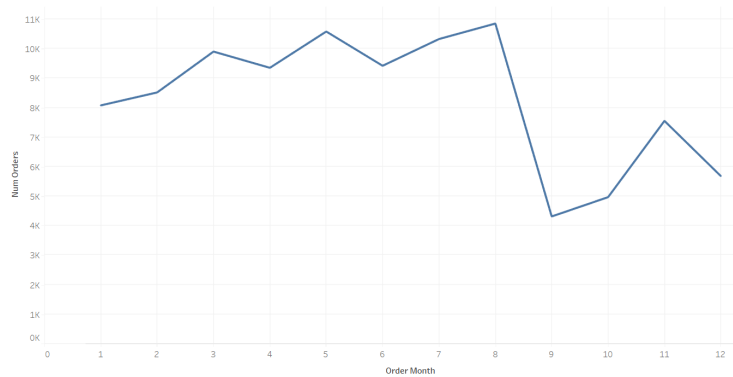


As we can see number of orders increased from previous year

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_orders
from `target_sql.orders`
group by 1
order by num_orders desc
```

Row	order_month	num_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



As we can see from chart that number of orders in month 8,5,7 are greater than other months

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

```

SELECT
  CASE
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 0 AND 6 THEN 'Dawn'
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 7 AND 12 THEN 'Morning'
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 13 AND 18 THEN
      'Afternoon'
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 19 AND 23 THEN 'Night'
  END AS Time_of_day,
  COUNT(o.order_id) AS Number_of_orders
FROM
  target_sql.orders o
JOIN
  target_sql.customers c
ON o.customer_id = c.customer_id
GROUP BY
  1
ORDER BY
  Number_of_orders DESC;

```

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

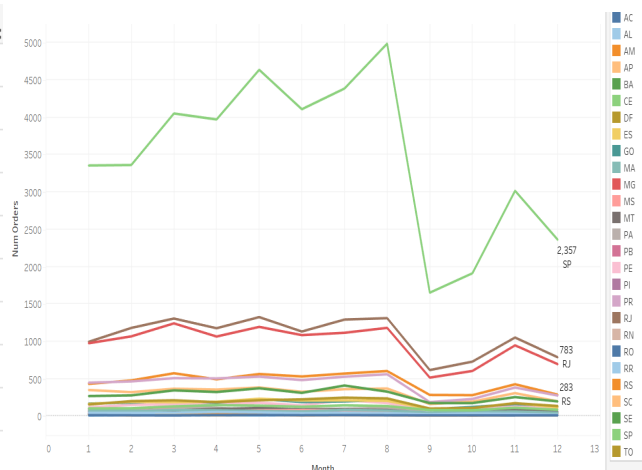
As we can see Brazilian customers placed their orders mostly in afternoon So its better to keep store open at Afternoon

Q.3 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 num_orders
from `target_sql.customers` as c
inner join `target_sql.orders` as o
on c.customer_id = o.customer_id
group by 1,2
order by num_orders desc
```

Row	month	state	num_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



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

```

select
customer_state as state,
count(distinct customer_id) as no_of_customers
from `target_sql.customers`
group by 1
order by no_of_customers desc

```



The data reveals that the state of São Paulo (SP) has the highest number of customers

Q.4 Impact on Economy

1). Get the % increase in the cost of orders from year 2017 to 2018

```

SELECT
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
(
(
SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018 AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 THEN
p.payment_value END)
-
SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 THEN
p.payment_value END)
)
/

```

```

SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 THEN
p.payment_value END)
)*100 AS percent_increase
FROM
`target_sql.orders` o
JOIN
`target_sql.payments` p ON o.order_id = p.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 1
ORDER BY 1;

```

Row	month	percent_increase
1	1	705.1266954171...
2	2	239.9918145445...
3	3	157.7786066709...
4	4	177.8407701149...
5	5	94.62734375677...
6	6	100.2596912456...
7	7	80.04245463390...
8	8	51.60600520477...

January shows the highest percentage increase, followed by February and April.

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

```

SELECT
c.customer_state,
ROUND(SUM(i.price), 2) AS total_price,
ROUND(AVG(i.price), 2) AS Avg_price,
FROM
`target_sql.orders` o
JOIN
`target_sql.order_items` i ON o.order_id = i.order_id
JOIN
`target_sql.customers` c ON o.customer_id = c.customer_id

```

GROUP BY

c.customer_state

order by Avg_price desc

Row	customer_state	total_price	Avg_price
1	PB	115268.08	191.48
2	AL	80314.81	180.89
3	AC	15982.95	173.73
4	RO	46140.64	165.97
5	PA	178947.81	165.69
6	AP	13474.3	164.32
7	PI	86914.08	160.36
8	TO	49621.74	157.53
9	RN	83034.98	156.97
10	CE	227254.71	153.76

The state of Paraíba (PB) has the highest average price value and average freight value.

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

SELECT

c.customer_state,

ROUND(SUM(i.freight_value), 2) AS total_freight_value,

ROUND(AVG(i.freight_value), 2) AS Avg_freight_value

FROM

`target_sql.orders` o

JOIN

`target_sql.order_items` i ON o.order_id = i.order_id

JOIN

`target_sql.customers` c ON o.customer_id = c.customer_id

GROUP BY

c.customer_state

order by Avg_freight_value desc

Row	customer_state	total_freight_value	Avg_freight_value
1	RR	2235.19	42.98
2	PB	25719.73	42.72
3	RO	11417.38	41.07
4	AC	3686.75	40.07
5	PI	21218.2	39.15
6	MA	31523.77	38.26
7	TO	11732.68	37.25
8	SE	14111.47	36.65
9	AL	15914.59	35.84
10	PA	38699.3	35.83

RR State has the highest avg_freight_value

Q.5 Analysis 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
  order_id,
  DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY)
  AS time_to_delivery,
  DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date, DAY)
  AS diff_estimated_delivery
FROM
  `target_sql.orders`
WHERE
  DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) IS NOT NULL
ORDER BY
  time_to_delivery;

```


Row	order_id	time_to_delivery	diff_estimated_delive
1	e65f1eeee1f52024ad1dcd034...	0	9
2	bb5a519e352b45b714192a02f...	0	25
3	434cecee7d1a65fc65358a632...	0	19
4	d3ca7b82c922817b06e5ca211...	0	11
5	1d893dd7ca5f77ebf5f59f0d20...	0	10
6	d5fbedc85190ba88580d6f82...	0	7
7	79e324907160caea526fd8b94...	0	8
8	38c1e3d4ed6a13cd0cf612d4c...	0	16
9	8339b608be0d84fca9d8da68b...	0	27
10	f349cdb62f69c3fae5c4d7d3f3...	0	12

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

```

SELECT
  c.customer_state,
  ROUND(AVG(i.freight_value), 2) AS avg_freight_value,
FROM
  `target_sql.orders` o
JOIN
  `target_sql.order_items` i ON o.order_id = i.order_id
JOIN
  `target_sql.customers` c ON o.customer_id = c.customer_id
GROUP BY
  c.customer_state
order by avg_freight_value desc

```

Row	customer_state	avg_freight_value			
1	RR	42.98	23	DF	21.04
2	PB	42.72	24	RJ	20.96
3	RO	41.07	25	MG	20.63
4	AC	40.07	26	PR	20.53
5	PI	39.15	27	SP	15.15

(SP) has the lowest mean freight value, while Roraima (RR) has the highest mean freight value.

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

```

SELECT
    c.customer_state,
    ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp,
DAY)), 2)
    AS avg_time_to_delivery,
FROM
    `target_sql.orders` o
JOIN
    `target_sql.order_items` i ON o.order_id = i.order_id
JOIN
    `target_sql.customers` c ON o.customer_id = c.customer_id
GROUP BY
    c.customer_state
ORDER BY
    avg_time_to_delivery desc

```

Row	customer_state	avg_time_to_delivery			
1	RR	27.83	23	SC	14.52
2	AP	27.75	24	DF	12.5
3	AM	25.96	25	MG	11.52
4	AL	23.99	26	PR	11.48
5	PA	23.3	27	SP	8.26

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

```

SELECT
    c.customer_state,
    ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,o.order_estimated_delivery_date,
DAY)), 2)
    AS diff_estimated_delivery
FROM
    `target_sql.orders` o
JOIN
    `target_sql.order_items` i ON o.order_id = i.order_id
JOIN
    `target_sql.customers` c ON o.customer_id = c.customer_id
GROUP BY
    c.customer_state
ORDER By diff_estimated_delivery

```

Row	customer_state	diff_estimated_delivery
1	AC	-20.01
2	RO	-19.08
3	AM	-18.98
4	AP	-17.44
5	RR	-17.43

AC has faster order delivery compared to estimated delivery

Q.6 Analysis based on the payments

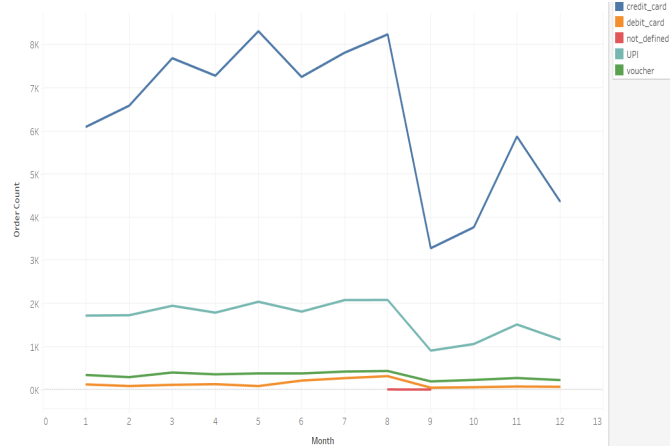
1). Find the month on month no. of orders placed using different payment types.

```

SELECT
    p.payment_type,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    COUNT(DISTINCT o.order_id) AS order_count
FROM
    `target_sql.orders` o
JOIN
    `target_sql.payments` p
ON
    o.order_id = p.order_id
GROUP BY
    1, 2
ORDER BY
    1, 2;

```

Row	payment_type	month	order_count
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



The analysis shows an overall uptrend from January to August and another uptrend from September to November. Credit card transactions are the most popular payment method, followed by UPI. Debit card transactions are the least preferred option.

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

```

SELECT
    p.payment_installments,
    COUNT(o.order_id) AS order_count
FROM
    `target_sql.orders` o
JOIN
    `target_sql.payments` p
ON
    o.order_id = p.order_id
WHERE
    o.order_status != 'canceled'
GROUP BY
    1
ORDER BY
    2 DESC;

```

Row	payment_installment	order_count
1	1	52184
2	2	12353
3	3	10392
4	4	7056
5	10	5292
6	5	5209
7	8	4239
8	6	3898
9	7	1620
10	9	638

The analysis reveals that the majority of orders (maximum count) have only one payment installment. The highest number of installments is 24, which is associated with 18 orders.

Q.7 Actionable Insights & Recommendations

Actionable Insights

1. The data reveals that the state of SP has significantly more orders than the next five states combined. This indicates an opportunity for improvement in the other states. Focusing on these states can help increase the number of orders and expand the customer base.
2. Seasonal variations in sales are observed, with increased sales during festive periods. Businesses should plan their marketing and sales strategies accordingly to capitalize on these peak periods and enhance customer satisfaction, resulting in overall sales growth.
3. Improving delivery times in areas with longer delivery durations can have a positive impact on customer satisfaction and encourage repeat purchases. Streamlining logistics and implementing efficient shipping processes are key to achieving this.

4. States like SP and RJ already have high order counts. To further boost sales and foster brand loyalty, it is recommended to focus on customer retention strategies, such as personalized marketing campaigns, loyalty programs, and exceptional customer service experiences.
5. Analyzing customer demographics can provide valuable insights for tailoring products and marketing strategies to specific target audiences. This customization can lead to increased sales and customer satisfaction.

Recommendations

1. Improve logistics and shipping processes to reduce delivery times and enhance customer satisfaction. This includes optimizing warehouse operations, refining shipping routes, and partnering with reliable courier services.
2. Implement customer retention strategies to encourage repeat purchases and foster loyalty. This can be achieved through loyalty programs, referral rewards, and personalized offers.
3. Evaluate pricing and freight fees to ensure competitiveness in the market while maximizing revenue and profitability. Consider increasing prices or adjusting freight fees as appropriate.
4. Invest in technology and infrastructure to enhance the e-commerce experience. This includes implementing chatbots for customer support, improving website performance, and offering personalized product recommendations based on customer behavior.
5. Collaborate with sellers to expand product offerings and improve product quality, catering to diverse customer needs and preferences.

