Business Case: Target SQL

About Target:

Target is one of the world's most recognized brands and one of America's leading retailers. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

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

1. Data type of columns in a table

Query**:**

```
SELECT DATA_TYPE,

column name

from Target_sql1234.INFORMATION_SCHEMA. COLUMNS

where table_name = 'customers'
```

Output:

Row	DATA_TYPE	column_name
1	STRING	customer_id
2	STRING	customer_unique_id
3	INT64	customer_zip_code_prefix
4	STRING	customer_city
5	STRING	customer_state

2. Time period for which the data is given

Query:

```
Select

date_diff(max, min, year)

from
```

```
(select

 extract(date from (min(order_purchase_timestamp))) as min1,

 extract (date from (max(order_purchase_timestamp))) as max1 from `Target_sql1234.orders`)
x;
```

Output

Row	f0_ ▼
1	2

3. Cities and States of customers ordered during the given period

Query:

```
SELECT
customer_city, customer_state, customer_unique_id
from `Target_sql1234.customers`;
```

Output

Row	customer_city ▼	customer_state ▼	customer_unique_id ▼
1	acu	RN	fcb003b1bdc0df64b4d065d9b...
2	acu	RN	46824822b15da44e983b021d...
3	acu	RN	b6108acc674ae5c99e29adc10...
4	ico	CE	402cce5c0509000eed9e77fec...
5	ico	CE	6ba00666ab7eada5ceec279b2...
6	ico	CE	796a0b1a21f597704057184a1...
7	ico	CE	05d1d2d9f0161c5f397ce7fc77...
8	ico	CE	c34585a0276ecc5e4fb03de75...
9	ico	CE	01a4fe5fc00bbdb0b0a4af5a53...
10	ico	CE	8f399f3b7ace8e6245422c9e1f...

In-depth Exploration:

1. 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:

```sql
SELECT
  Month,
  Year,
  total_orders_made
FROM
(SELECT
  Extract(MONTH FROM order_purchase_timestamp) AS Month,
  Extract(YEAR FROM order_purchase_timestamp) AS Year,
  COUNT(*) AS total_orders_made
FROM
`Target_sql1234.orders`
GROUP BY
Year,
Month
order by Year, Month) t1;
```

Output

Row	Month	Year	total_orders_made
1	9	2016	4
2	10	2016	324
3	12	2016	1
4	1	2017	800
5	2	2017	1780
6	3	2017	2682
7	4	2017	2404
8	5	2017	3700
9	6	2017	3245
10	7	2017	4026

Graph deduced:

Insights

We can conclude with the help of this graph that there is not an increasing trend in orders placed by Brazilian customers, the trend is rather fluctuating to a point. There are variations in the growing pattern. It includes both increases and decreases (a volatile trend). The trend was at peak once then it took a big dip. Also, there is no seasonality in the trend.

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

```
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 Time_of_purchase,

count(*) purchase_count

from `Target_sql1234.orders`

group by Time_of_purchase

order by purchase_count desc;
```
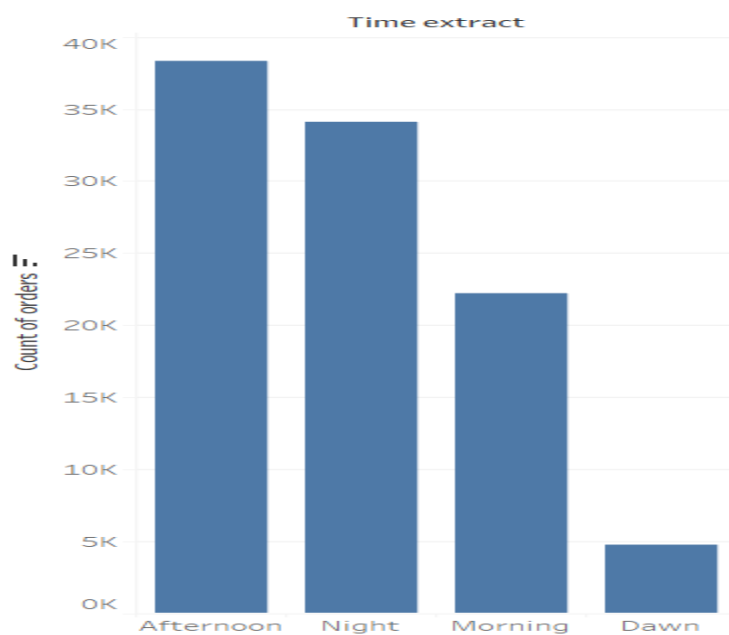
Output

Row	Time_of_purchase ▼	purchase_count ▼
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

Graph Deduced:

Insight

As we can see from the graph and the data extracted via the query that Brazilian customers tend to buy more during the afternoon followed by night. There could be many possible reasons for this trend to happen. People could have more free time in afternoon and then in Night after work. We can also see that there are very few numbers in dawn because of obvious reasons i.e limited availability, these are usually the sleeping hours and many more.

Evolution of E-commerce orders in the Brazil region:

1. Get month on month orders by states

Query:

```
Select
Total_orders,
year,
month,
customer_state
from
(select
Count(o.order_id) as Total_orders,
extract(year from o.order_purchase_timestamp) as year,
extract(month from o.order_purchase_timestamp) as month,
c.customer_state
from `Target_sql1234.orders` o
Join `Target_sql1234.customers` c on o.customer_id = c.customer_id
Group    by    Extract(year    from    o.order_purchase_timestamp),    extract(month    from
o.order_purchase_timestamp), c.customer_state)
order by year, month;
```

Output:

Row	Total_orders ▾	year ▾	month ▾	customer_state ▾
1	1	2016	9	RR
2	1	2016	9	RS
3	2	2016	9	SP
4	113	2016	10	SP
5	24	2016	10	RS
6	56	2016	10	RJ
7	3	2016	10	MT
8	9	2016	10	GO
9	40	2016	10	MG
10	8	2016	10	CE

Insights

we can see that the state that placed the maximum orders month over month include SP with total order of 654 in month of Feb,2017. The total orders are increasing and decreasing over the months. It keeps on fluctuating. The overall trend of orders made is fluctuating and we can safely conclude that here as well. As we can see, the number of orders increased and then took a dip and then decreased. SP has been constantly placing more orders.

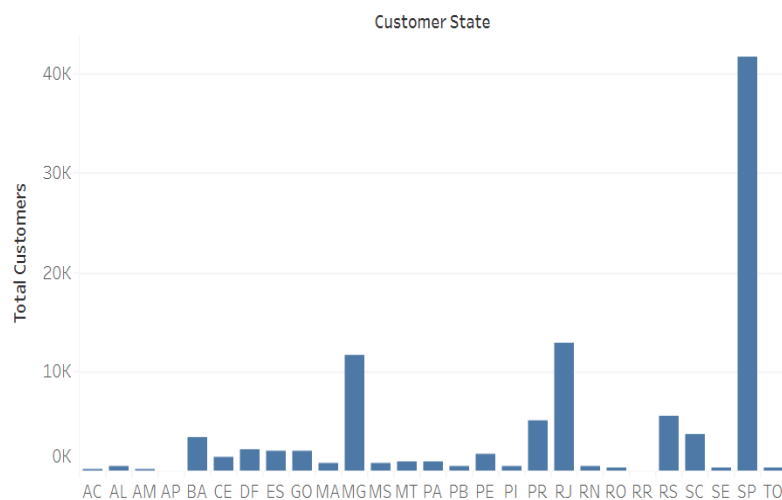Distribution of customers across the states in Brazil

Query:

```
select
customer_state,
count(*) as total_customers
from Target_sql1234.customers
group by customer_state
order by total_customers desc;
```

Output

Row	customer_state ▼	total_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

Graph deduced:

Insights

We can see that the maximum orders come from the state of SP in Brazil. It is one of the most populated state in Brazil and hence the customer base is quite large. Also, this state is the economic hub in Brazil thereby making it easier for all economic activities, easy logistics etc.

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

1. 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:

```
SELECT

(ROUND((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

Else 0

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

Else 0

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

ELSE 0

End) * 100),2) as percentage_increase_in_cost

from

`Target_sql1234.orders` o

Join
```

```sql
`Target_sql1234.payments` p on

o.order_id = p.order_id

where

(Extract  (Year  from  o.order_purchase_timestamp)  =  2017  AND  Extract(month  from
o.order_purchase_timestamp) between 1 and 8)

OR

(extract(year  from  o.order_purchase_timestamp)  =  2018  and  extract(month  from
o.order_purchase_timestamp) between 1 and 8);
```

Output

Row	perc..._field_1 ▼	perc..._field_2 ▼
1	137.0	2

Insights

We can see that there is about 137% increase in the cost of orders from 2017 to 2018   which is a
significant increase in cost during this period. There can be multiple factors contributing and some
of those maybe market conditions changing or due to inflation

Mean & Sum of price and freight value by customer state

Query:

```sql
select

customer_state,

ROUND(AVG(freight_value),2) as mean_freight_value,

ROUND(Avg(Price),2) as mean_price,

ROUND(sum(freight_value),2) as Sum_freight_value,

ROUND(Sum(Price),2) as Sum_price,

From

Target_sql1234.orders o

Join

`Target_sql1234.order_items` oi
```

```
on o.order_id = oi.order_id

join

`Target_sql1234.customers` c on

o.customer_id = c.customer_id

group by

customer_state

order by customer_state;
```

Output

Row	customer_state	mean_freight_value	mean_price	Sum_freight_value	Sum_price
1	AC	40.07	173.73	3686.75	15982.95
2	AL	35.84	180.89	15914.59	80314.81
3	AM	33.21	135.5	5478.89	22356.84
4	AP	34.01	164.32	2788.5	13474.3
5	BA	26.36	134.6	100156.68	511349.99
6	CE	32.71	153.76	48351.59	227254.71
7	DF	21.04	125.77	50625.5	302603.94
8	ES	22.06	121.91	49764.6	275037.31
9	GO	22.77	126.27	53114.98	294591.95
10	MA	38.26	145.2	31523.77	119648.22

Insights

We can see that the maximum contributor is the state SP and the lowest contributor is RR for both Average and sum of price and freight value.

Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery.

Query:

```
Select

order_id,

date_diff(order_estimated_delivery_date,order_purchase_timestamp,day)as Estimated_delivery,

date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as Actual_delivery

from `target-sql-388013.Target_sql1234.orders`

where order_status= 'delivered';
```

Output

Row	order_id	Estimated_delivery	Actual_delivery
1	635c894d068ac37e6e03dc54e...	32	30
2	3b97562c3aee8bdedcb5c2e45...	33	32
3	68f47f50f04c4cb6774570cfde...	31	29
4	276e9ec344d3bf029ff83a161c...	39	43
5	54e1a3c2b97fb0809da548a59...	36	40
6	fd04fa4105ee8045f6a0139ca5...	35	37
7	302bb8109d097a9fc6e9cefc5...	28	33
8	66057d37308e787052a32828...	32	38
9	19135c945c554eebfd7576c73...	33	36
10	4493e45e7ca1084efcd38ddeb...	33	34

Insights

We can see that for most of the occurrences the actual delivery is done before the estimated delivery. Target has an efficient network of logistics services and that could be the possible reason for the same. We can also see that planning and forecasting has done effectively to for this to happen and for someday when the case is opposite we can contribute that to bad weather or unviability of certain product etc.

2. Find time_to_delivery & diff_estimated_delivery. Formula for the same given below:
- time_to_delivery = order_delivered_customer_date-order_purchase_timestamp
- diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date

Query:

```
Select

order_id,

date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)as
diff_estimated_delivery,

date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as time_to_delivery

from `target-sql-388013.Target_sql1234.orders`

where order_status= 'delivered';
```

Output-

Row	order_id	diff_estimated_delive	time_to_delivery
1	635c894d068ac37e6e03dc54e...	1	30
2	3b97562c3aee8bdedcb5c2e45...	0	32
3	68f47f50f04c4cb6774570cfde...	1	29
4	276e9ec344d3bf029ff83a161c...	-4	43
5	54e1a3c2b97fb0809da548a59...	-4	40
6	fd04fa4105ee8045f6a0139ca5...	-1	37
7	302bb8109d097a9fc6e9cefc5...	-5	33
8	66057d37308e787052a32828...	-6	38
9	19135c945c554eebfd7576c73...	-2	36
10	4493e45e7ca1084efcd38ddeb...	0	34

3. Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery

Query:

```
select

c.customer_state,

ROUND(Avg(oi.freight_value),2) as Avg_freight_value,

ROUND(Avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp, day)),2) as
avg_time_to_delivery,

ROUND(Avg(date_DIFF(o.order_estimated_delivery_date,o.order_delivered_customer_date,
day)),2) as avg_diff_estimated_delivery

From

Target_sql1234.orders o

Join

Target_sql1234.order_items oi

on o.order_id = oi.order_id

Join

Target_sql1234.customers c

on o.customer_id = c.customer_id

group by c.customer_state;
```

Output

Row	customer_state	Avg_freight_value	avg_time_to_delivery	avg_diff_estimated_
1	MT	28.17	17.51	13.64
2	MA	38.26	21.2	9.11
3	AL	35.84	23.99	7.98
4	SP	15.15	8.26	10.27
5	MG	20.63	11.52	12.4
6	PE	32.92	17.79	12.55
7	RJ	20.96	14.69	11.14
8	DF	21.04	12.5	11.27
9	RS	21.74	14.71	13.2
10	SE	36.65	20.98	9.17

Insights:

We can see that average estimated delivery and delivery time have a positive correlation with average freight. There is a positive correlation between freight value and delivery time

4. Sort the data to get the following:

Top 5 states with highest average freight value - sort in desc/asc limit 5

Query:

```
Select

c.customer_state,

ROUND(Avg(oi.freight_value),2) as avg_freight_value

from

Target_sql1234.orders o

Join

Target_sql1234.order_items oi

on o.order_id = oi.order_id

Join

Target_sql1234.customers c
```

```
on o.customer_id = c.customer_id

group by

c.customer_state

order by

avg_freight_value desc

limit 5;
```

Output

Row	customer_state	avg_freight_value
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15

Top 5 states with lowest average freight value - sort in desc/asc limit 5

Query:

```
Select

c.customer_state,

ROUND(Avg(oi.freight_value),2) as avg_freight_value

from

Target_sql1234.orders o

Join

Target_sql1234.order_items oi

on o.order_id = oi.order_id

Join

Target_sql1234.customers c

on o.customer_id = c.customer_id

group by

c.customer_state
```

order by

avg_freight_value

limit 5;

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

Top 5 states with highest average time to delivery

Query:

```
SELECT

c.customer_state,

ROUND(Avg(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY))) as avg_time_to_delivery

from

Target_sql1234.orders o

join

Target_sql1234.customers c

on o.customer_id = c.customer_id

group by

c.customer_state

order by

avg_time_to_delivery desc

limit 5;
```

Output

Row	customer_state ▼	avg_time_to_delivery
1	RR	29.0
2	AP	27.0
3	AM	26.0
4	AL	24.0
5	PA	23.0

Insights

We can see that these states have the highest average days that it took to deliver the order there. With RR(Roraima) being on top. This could be possibly because of its remote location. The logistics route may not be as developed as other populated states in Brazil. This could be true for all other 4 states too as these are the least populated states in BR and have poor infrastructure and logistics

Top 5 states with lowest average time to delivery

Query:

```
SELECT

c.customer_state,

ROUND(Avg(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY))) as
avg_time_to_delivery

from

Target_sql1234.orders o

join

Target_sql1234.customers c

on o.customer_id = c.customer_id

group by

c.customer_state

order by

avg_time_to_delivery

limit 5;
```

Output

Row	customer_state	avg_time_to_delivery
1	SP	8.0
2	MG	12.0
3	PR	12.0
4	DF	13.0
5	SC	14.0

Top 5 states where delivery is really fast compared to estimated date

Query:

```
Select

c.customer_state,

ROUND(AVG(date_DIFF(o.order_estimated_delivery_date, order_delivered_customer_date, day)))
as avg_time_to_delivery

FROM

Target_sql1234.orders o

join

Target_sql1234.customers c on o.customer_id = c.customer_id

group by

c.customer_state

order by

avg_time_to_delivery desc

limit 5;
```

Output:

Row	customer_state	avg_time_to_delivery
1	AC	20.0
2	RO	19.0
3	AM	19.0
4	AP	19.0
5	RR	16.0

Top 5 states where delivery is not so fast compared to estimated date

Query:

```sql
Select

c.customer_state,

ROUND(AVG(date_DIFF(o.order_estimated_delivery_date, order_delivered_customer_date, day)))
as avg_time_to_delivery

FROM

Target_sql1234.orders o

join

Target_sql1234.customers c on o.customer_id = c.customer_id

group by

c.customer_state

order by

avg_time_to_delivery asc

limit 5;
```

Output

Row	customer_state ▼	avg_time_to_delivery
1	AL	8.0
2	MA	9.0
3	SE	9.0
4	SP	10.0
5	BA	10.0

5. Payment type analysis:

Month over Month count of orders for different payment types

Query:

```sql
SELECT
extract( year from o.order_purchase_timestamp) as order_year,
extract( month from o.order_purchase_timestamp) as order_month,
payment_type,
COUNT(*) as order_count
from `Target_sql1234.orders`o
join
`Target_sql1234.payments` p on
o.order_id=p.order_id
group by
order_year,
 order_month,
 payment_type
order by
 order_year,
 order_month;
```

Output

Row	order_year	order_month	payment_type	order_count
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

Count of orders based on the no. of payment instalments

Query:

```sql
select

payment_installments,

count(*) total_no_of_orders

from `Target_sql1234.payments`

group by payment_installments;
```

Output

Row	payment_installment	total_no_of_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

Insights:

As we can see that, most of the people is paying one instalment for their orders which is followed by 2 and 3 instalments. People prefer paying for their orders mostly in 3 maximum payments. While, these higher payment instalments i.e. 22, 23 and 24 are least preferred by people and there are only few orders with that number of instalments.

Actionable Insights and Recommendations

- Target should monitor the market trends in Brazil as the trend is quite fluctuating. They should put more marketing efforts and promote in Brazil, for the customers to buy from there.
- They should also focus on behavioural targeting which means tailoring the needs as per the people of Brazil.
- As Afternoon and night hours are the peak hours for shopping. They should plan there marketing/promotional offers around that time. Discounts can also be offered during that specific time period of afternoon or night to increase the sales.
- There were volatile trends throughout in as we compare the state wise data, Target should show the tailor-made products as per the different state and their behavioural preferences.
- Target should identify the top selling products and high revenue products and consider going full throttle on both of those to increase the revenue.

- There were regional sales disparity too, so Target should be making more advertising and marketing efforts towards those state which has lesser order volumes.
- Conducting market research to know the reasons behind the lower order volumes and then taking actions accordingly. They can offer special discounts or promotional offers to get the sales up and running in the states of DE, ES GO etc.
- For the increasing overall cost, we need to identify the factors contributing to the increase. Target should think about partnering with local vendors or negotiate with existing partners and crack up some beneficial deals to reduce the cost.
- For the delivery disparity, they can plan to work with logistics services. Considering to partner with local courier service and setting up state-wise/region-wise distribution system to lower the difference between actual and estimated delivery. This disparity in delivery times should be reduced.
- Delivering fast to the states where it takes a lot of time will eventually increase the sales from that state as well.
- Target should promote the flexible payment system and encourage customer for this. They might consider partnering with regional financial institutes.