TARGET CASE STUDY

Q1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.

Q1(1) Data type of columns in a table

ANS:-

Customers Table	
Column Name	**Data type**
Customer_id	STRING
Customer_unique_id	STRING
Customer_zip_code_prefix	INTEGER
Customer_city	STRING
Customer_state	STRING

Seller Table	
Column Name	**Data type**
Seller_id	STRING
Seller_zip_code_prefix	INTEGER
Seller_city	STRING
Seller_state	STRING

Order items	
Column Name	**Data type**
Order_id	STRING
Order_item_id	INTEGER
Product_id	STRING
Seller_id	STRING
Shipping_limit_date	TIMESTAMP
price	FLOAT
Freight_value	FLOAT

Geolocations Table	
Column Name	**Data type**
Geolocation_zip_code_prefix	INTEGER
Geolocation_lat	FLOAT
Geolocation_lng	FLOAT
Geolocation_city	STRING
Geolocation_state	STRING

Payments Table	
Column Name	**Data type**
Order_id	STRING
Payment_sequential	INTEGER
Payment_type	STRING
Payment_installments	INTEGER
Payment_values	FLOAT

Orders Table	
Column Name	**Data type**
Order_id	STRING
Customer_id	STRING
Order_status	STRING
Order_purchase_timestanp	TIMESTAMP
Order_delivered_carrier_date	TIMESTAMP
Order_delivered_customer_date	TIMESTAMP
Order_estimated_delivery_date	TIMESTAMP

Reviews Table	
Column Name	**Data type**
Review_id	STRING
Order_id	STRING
Review_score	INTEGER
Review_comment_title	STRING
Review_comment_message	STRING
Review_creation_date	TIMESTAMP
Review_answer_timestamp	TIMESTAMP

Products Table	
Column Name	**Data type**
Product_id	STRING
Product_category_name	STRING
Product_name_length	INTEGER
Product_description_length	INTEGER
Product_photos_qty	INTEGER
product_weight_g	INTEGER
product_length_cm	INTEGER
product_height_cm	INTEGER
product_width_cm	INTEGER

Q1(2) Time period for which the data is given

ANS: -

```
SELECT
MIN(order_purchase_timestamp) AS Minimum_Timestamp,
MAX(order_purchase_timestamp) AS Maximum_Timestamp
FROM
  `Target_Retailers.orders`
GROUP BY
  order_id,
  customer_id
LIMIT 10
```

Row	Minimum_Timestamp	Maximum_Timestamp
1	2017-11-25 11:10:33 UTC	2017-11-25 11:10:33 UTC
2	2017-12-05 01:07:58 UTC	2017-12-05 01:07:58 UTC
3	2017-12-05 01:07:52 UTC	2017-12-05 01:07:52 UTC
4	2018-02-09 17:21:04 UTC	2018-02-09 17:21:04 UTC
5	2017-11-06 13:12:34 UTC	2017-11-06 13:12:34 UTC
6	2017-04-20 12:45:34 UTC	2017-04-20 12:45:34 UTC
7	2017-07-13 11:03:05 UTC	2017-07-13 11:03:05 UTC
8	2017-07-11 13:36:30 UTC	2017-07-11 13:36:30 UTC
9	2017-07-29 18:05:07 UTC	2017-07-29 18:05:07 UTC
10	2017-07-13 10:02:47 UTC	2017-07-13 10:02:47 UTC

INSIGHTS & RECOMMENDATIONS: -

1. This study can be used to detect seasonality, shifts in consumer preferences, or company expansion.

2. We can compare the performance indicators between the smaller time periods of the dataset, such as the months or quarters. This study can provide information regarding any trends or swings in consumer behavior, number of orders, or delivery schedules.

Q1(3) Cities and States of customers ordered during the given period

ANS: -

```
SELECT
DISTINCT
c.customer_city,
c.customer_state
FROM `Target_Retailers.orders` AS o
JOIN `Target_Retailers.customers` AS c ON o.customer_id=
c.customer_id
LIMIT 10
```

Row	customer_city	customer_state
1	rio de janeiro	RJ
2	sao leopoldo	RS
3	general salgado	SP
4	brasilia	DF
5	paranavai	PR
6	cuiaba	MT
7	sao luis	MA
8	maceio	AL
9	hortolandia	SP
10	varzea grande	MT

INSIGHTS: -

1. By doing this kind of analysis we can get to know how customers are distributed throughout various cities and states

2. To know the requirements and expectations of clients in various locations by adjusting our marketing tactics and product offers.

Q2 In-depth Exploration:

Q2(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?

ANS:-

```
SELECT
  EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
  EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
  COUNT(order_id) AS order_count
FROM
  `Target_Retailers.orders`
GROUP BY
  year, month
ORDER BY
  year, month;
```

Row	year	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

INSIGHTS & RECOMMENDATIONS

1. E-commerce seems to have some seasonality, peaking during particular months. For instance, in 2018, order numbers peaked in March, then had a tiny decline in April and May before gradually increasing from June to August. This trend indicates that there may be some months or times of the year when online buying activity is stronger, maybe as a result of things like holidays, sales, or seasonal events.

2. Businesses may use forecasting models to estimate future demand and modify their strategy by using previous data on order counts and seasonality. Accurate forecasting can assist maximize workers and inventory levels.

Q2(2) What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

ANS: -

```
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 purchase_time_period,
     COUNT(order_id) AS order_count
FROM
  `Target_Retailers.orders`
GROUP BY
  purchase_time_period
ORDER BY
  purchase_time_period;
```

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

INSIGHTS & RECOMMENDATIONS

1. The biggest number of orders are placed in the afternoon, which suggests that many Brazilian consumers like to make their purchases at this time. This could be because it fits with customary working hours and breaks.

2. Dawn has the fewest orders, which shows that it is the least popular time for Brazilians to make online purchases. This may be because most clients are sleeping or otherwise occupied at these early hours.

3. As the prime period for internet shopping, the afternoon might be the focus of promotional efforts and deals. This may increase client interest and boost revenue.

4. To provide an effortless purchasing procedure for customers during peak hours, especially in the afternoon, think about optimizing the website and mobile app's performance.

Q3. Evolution of E-commerce orders in the Brazil region:

Q3(1) Get month on month orders by states

ANS: -

```
SELECT
    EXTRACT(MONTH FROM o.order_purchase_timestamp ) AS months,
    c.customer_state,
    COUNT(DISTINCT o.order_id) AS orders
FROM `Target_Retailers.orders` AS o
JOIN `Target_Retailers.customers` AS c ON o.customer_id=
c.customer_id
GROUP BY
    months,
    c.customer_state
ORDER BY
    Months
```

Row	months	customer_state	orders
1	1	RJ	990
2	1	SP	3351
3	1	DF	151
4	1	RS	427
5	1	CE	99
6	1	PE	113
7	1	PR	443
8	1	BA	264
9	1	MG	971
10	1	RN	51

Output Part 1

Row	months	customer_state	orders
28	2	DF	196
29	2	SP	3357
30	2	PA	83
31	2	MG	1063
32	2	PI	46
33	2	RJ	1176
34	2	RS	473
35	2	PE	146
36	2	RR	7
37	2	CE	101
38	2	PR	460
39	2	BA	273

Output Part 2

INSIGHTS: -

1. Sao Paulo-SP, has the highest number of orders for all the months.it suggests that it is the biggest market for e-commerce So we should check the reason for more sales so that we can implement it for other states as well.

2. Some states like Rio de Janeiro (RJ), and Minas Gerais (MG) maintained high number of orders for all months

3. From this table we can know seasonal trends as well.

Recommendations: -

1. Develop marketing campaign and know the type of customers for the states like São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) where sales are comparatively low.

Q3(2) Distribution of customers across the states in Brazil.

ANS: -

```
SELECT
     customer_state AS states,
     COUNT(customer_id)AS Total_Customers

FROM
  `Target_Retailers.customers`
GROUP BY
  customer_state
LIMIT 10;
```

Row	states	Total_Customers
1	RN	485
2	CE	1336
3	RS	5466
4	SC	3637
5	SP	41746
6	MG	11635
7	BA	3380
8	RJ	12852
9	GO	2020
10	MA	747

INSIGHTS & RECOMMENDATIONS: -

1. The state with the highest number of customers is Distrito Federal (DF), with a total of 2,140 customers. This suggests that Brazil's capital region has a sizable customer base.

2. Other states with comparatively more customers are Piauí (PI) with 495 customers, Alagoas (AL) with 413 customers, and Amazonas (AM) with 148 customers.

Q4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

Q4(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

ANS: -

```
SELECT

ROUND(
((payment_2018-payment_2017)/payment_2017)*100,2)
AS increase_per

FROM
(
  SELECT

 SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp)=2017
 THEN p.payment_value ELSE 0 END) AS payment_2017,

 SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp)=2018
 THEN p.payment_value ELSE 0 END) AS payment_2018


 FROM `Target_Retailers.orders` AS o
 JOIN `Target_Retailers.payments` AS 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
) AS tbl
```

Row	increase_per
1	136.98

INSIGHTS: -

1. According to the data, the cost of orders significantly increased from 2017 to 2018 which is 136.98%.

RECOMMENDATIONS: -

1. Analyze the causes of cost growth and modify pricing approach as necessary.

2. This knowledge can be Utilize to establish pricing strategies and plans and estimate future cost patterns.

3. Keep up with price trends among competitors and adjust tactics to stay competitive.

Q4(2) Mean & Sum of price and freight value by customer state

ANS: -

```
SELECT
c.customer_state,
AVG(oi.price)AS mean_of_price,
SUM(oi.price) AS sum_of_price,
AVG(oi.freight_value) AS mean_of_freight_value,
SUM(oi.freight_value) AS sum_of_freight_value

FROM `Target_Retailers.customers`AS c
JOIN `Target_Retailers.orders` AS o ON c.customer_id=o.customer_i
d
JOIN `Target_Retailers.order_items` AS oi ON o.order_id=oi.order_
id

GROUP BY
  c.customer_state
LIMIT 10;
```

INSIGHTS & RECOMMENDATIONS: -

1. Paraná (PR) has the highest total price. Which means it contributes to the most of the revenue.

2. The mean and total prices are lowest for Acre (AC). Investigating the causes of this lower price point may provide chances to boost consumer spending or diversify product lines.

3. Paraná (PR) got the highest total freight value. which means it has higher shipping cost. Strategies can be made to lower this cost.

4. Mato Grosso do Sul (MS) has the lowest mean freight value. So we can get to know reason behind it so we can apply same thing for other regions as well.

Q5. Analysis on sales, freight and delivery time

Q5(1) Calculate days between purchasing, delivering and estimated delivery

ANS: -

```
SELECT

    order_id,

DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,
Day) AS purchase_delivery_diff,
DATE_DIFF(order_estimated_delivery_date,order_purchase_timestamp,
Day) AS purchase_estimated_diff

FROM
  `Target_Retailers.orders`

WHERE
  order_delivered_customer_date IS NOT NULL
  AND order_purchase_timestamp IS NOT NULL

LIMIT 10;
```

Row	order_id	purchase_delivery_diff	purchase_estimated_diff
1	770d331c84e5b214bd9dc70a...	7	52
2	1950d777989f6a877539f5379...	30	17
3	2c45c33d2f9cb8ff8b1c86cc28...	30	59
4	dabf2b0e35b423f94618bf965f...	7	51
5	8beb59392e21af5eb9547ae1a...	10	52
6	65d1e226dfaeb8cdc42f66542...	35	52
7	c158e9806f85a33877bdfd4f60...	23	33
8	b60b53ad0bb7dacacf2989fe2...	12	7
9	c830f223aae08493ebecb52f2...	12	25
10	a8aa2cd070eeac7e4368cae3d...	7	8

INSIGHTS & RECOMMENDATIONS: -

1. Check the accuracy of the delivery dates promised to consumers. Consider changing the projected delivery dates to meet customer's reasonable expectations if there is a significant gap between the estimated and actual delivery periods.

2. Keep a close eye on delivery schedules and performance metrics to track development and spot potential improvement areas.

3. If there are delays, get in touch with consumers right away to provide them information and new delivery estimates.

Q5(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

ANS: -

```
SELECT
ABS(DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY)) AS time_to_delivery,

ABS(DATE_DIFF(order_estimated_delivery_date,order_delivered_customer_date,DAY)) AS diff_estimated_delivery

FROM `Target_Retailers.orders`

WHERE
  order_delivered_customer_date IS NOT NULL
  AND order_purchase_timestamp IS NOT NULL

LIMIT 10;
```

Row	time_to_delivery	diff_estimated_delivery
1	7	45
2	7	44
3	10	41
4	6	29
5	20	40
6	10	48
7	28	29
8	9	35
9	10	41
10	6	41

INSIGHTS & RECOMMENDATIONS: -

1. If there are delays, get in touch with consumers right away to let them know about them and to provide them updated delivery estimates.

Q5(3) Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery

ANS:-

```
SELECT
c.customer_state,
AVG(oi.freight_value) AS mean_of_freight_value,

AVG(ABS(DATE_DIFF(o.order_delivered_customer_date,o.order_purcha
se_timestamp,DAY))) AS time_to_delivery,

AVG(ABS(DATE_DIFF(o.order_estimated_delivery_date,o.order_deliver
ed_customer_date,DAY))) AS diff_estimated_delivery

 FROM `Target_Retailers.customers` AS c
 JOIN `Target_Retailers.orders` AS o ON c.customer_id=o.customer_
id
 JOIN `Target_Retailers.order_items` AS oi ON o.order_id=oi.order
_id

WHERE
  o.order_delivered_customer_date IS NOT NULL
  AND o.order_purchase_timestamp IS NOT NULL

GROUP BY
  c.customer_state
LIMIT 10;
```

Row	customer_state	mean_of_freight_value	time_to_delivery	diff_estimated_delivery
1	RJ	20.909784391347...	14.6893821...	14.2255054432348...
2	MG	20.6258372687155	11.5155221...	13.1347836184872...
3	SC	21.506627623230...	14.5209858...	12.1178623718887...
4	SP	15.114994078763...	8.25960855...	10.9855737140150...
5	GO	22.562867808519...	14.9481774...	12.8950373298199...
6	RS	21.61427034077937	14.7082993...	14.4588292842002...
7	BA	26.487556339940...	18.7746402...	12.9242465381482...
8	MT	27.996914175506...	17.5081967...	14.88717454194793
9	SE	36.573173333333...	20.9786666...	13.9066666666666...
10	PE	32.693333333333...	17.7920962...	14.6884306987399...

INSIGHTS & RECOMMENDATIONS: -

1. If we just consider first 10 rows, the average freight value varies from 15.11 to 38.49 depending on the status of the consumer.

2. Examine the variables, such as distance, shipping companies, or transportation infrastructure, that affect freight costs in each state. We can find ways to reduce shipping costs and increase cost-effectiveness.

3. Investigate the trends and patterns unique to each client state. Determine whether states have more or lesser mean freight values, longer or shorter transit times, and greater or lesser variations from projected delivery. This study might reveal regional differences in customer behavior, market dynamics, or logistics, which can help states develop tailored initiatives.

Q5(4). Sort the data to get the following:

Q5(5). Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

ANS: -

```
SELECT
c.customer_state,
ROUND(AVG(oi.freight_value),2) AS Max_freight_value

FROM `Target_Retailers.customers` AS c
JOIN `Target_Retailers.orders` AS o ON c.customer_id=o.customer_i
d
JOIN `Target_Retailers.order_items` AS oi ON o.order_id=oi.order_
id
GROUP BY c.customer_state
ORDER BY
Max_freight_value DESC

LIMIT 5;
```

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

INSIGHTS & RECOMMENDATIONS: -

1. Look into whether there are any particular product categories or order kinds that considerably increase the high freight values in some states. Understanding client preferences and order characteristics may be used to develop marketing strategies, promotions, and product lineups that are responsive to local needs.

2. Examine the causes of increased freight values in states with RR, PB, RO, AC, and PI. To cut expenses and boost customer satisfaction, optimize logistics and delivery techniques.

Q5(6) Top 5 states with highest/lowest average time to delivery

ANS: - for highest

```
SELECT
  c.customer_state,
  AVG(DATE_DIFF(order_delivered_customer_date,
order_purchase_timestamp,DAY)) AS avg_time_del

FROM `Target_Retailers.customers` AS c
JOIN `Target_Retailers.orders` AS o ON
c.customer_id=o.customer_id

WHERE
  o.order_delivered_customer_date IS NOT NULL
  AND o.order_purchase_timestamp IS NOT NULL

GROUP BY
  c.customer_state
ORDER BY
  avg_time_del DESC
LIMIT 5;
```

Row	customer_state	avg_time_del
1	RR	28.975609756...
2	AP	26.731343283...
3	AM	25.986206896...
4	AL	24.040302267...
5	PA	23.316067653...

INSIGHTS & RECOMMENDATIONS: -

1. RR has the highest Average time to delivery

2. Examine the effects of geographic elements on delivery timeframes in various states, such as remoteness or difficult terrain. To get over these difficulties, think about establishing specialised distribution methods.

-For lowest

```
SELECT
  c.customer_state,
  AVG(DATE_DIFF(order_delivered_customer_date,
order_purchase_timestamp,DAY)) AS avg_time_del
FROM `Target_Retailers.customers` AS c
JOIN `Target_Retailers.orders` AS o ON
c.customer_id=o.customer_id

WHERE
  o.order_delivered_customer_date IS NOT NULL
  AND o.order_purchase_timestamp IS NOT NULL

GROUP BY
  c.customer_state
ORDER BY
  avg_time_del ASC
LIMIT 5;
```

Row	customer_state	avg_time_del
1	SP	8.298061489...
2	PR	11.52671135...
3	MG	11.54381329...
4	DF	12.50913461...
5	SC	14.47956019...

Q5(7) Top 5 states where delivery is really fast/ not so fast compared to estimated date

ANS: -

```
SELECT
  c.customer_state,
  AVG(DATE_DIFF(o.order_estimated_delivery_date,o.order_delivered
_customer_date , DAY)) AS avg_del_speed
FROM
  `Target_Retailers.orders` AS o
JOIN
  `Target_Retailers.customers` AS c ON o.customer_id =
c.customer_id
WHERE
  o.order_delivered_customer_date IS NOT NULL
  AND o.order_estimated_delivery_date IS NOT NULL
GROUP BY
  c.customer_state
ORDER BY
  avg_del_speed DESC
LIMIT 5;
```

Row	customer_state	avg_delivery_speed
1	AC	19.76250000000...
2	RO	19.13168724279...
3	AP	18.73134328358...
4	AM	18.60689655172...
5	RR	16.41463414634...

INSIGHTS & RECOMMENDATIONS: -

1. The states AC, RO, AP, AM, and RR have the quickest delivery times compared to the expected arrival date. This suggests that these states have effective delivery systems.

2. Analyze the success of the delivery alliances and partnerships with regional carriers in these states. By doing this we can think about extending or replicating successful relationships in other areas.

3. To identify any difficulties in the delivery process, look at states where delivery times are slower than the estimated delivery date. Boost carrier coordination, deploy technological solutions to track and manage deliveries more efficiently to address these difficulties.

Q6. Payment type analysis:

Q6(1) Month over Month count of orders for different payment types

ANS: -

```
  SELECT
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
  p.payment_type,
  COUNT(o.order_id) AS no_of_order
FROM
  `Target_Retailers.orders` AS o
JOIN
  `Target_Retailers.payments` AS p ON o.order_id = p.order_id
GROUP BY
  month,
  p.payment_type
ORDER BY
  month ASC,
  p.payment_type;
```

Row	month	payment_type	no_of_order
1	1	UPI	1715
2	1	credit_card	6103
3	1	debit_card	118
4	1	voucher	477
5	2	UPI	1723
6	2	credit_card	6609
7	2	debit_card	82
8	2	voucher	424
9	3	UPI	1942
10	3	credit_card	7707

INSIGHTS & RECOMMENDATIONS: -

1. The payment method with the highest number of orders is credit card.

2. Order counts for more recent payment options like UPI can reveal the level of consumer uptake and acceptance of these technologies. Decisions on increasing payment choices can be informed by tracking their development over time.

3. Following order counts for various payment methods can reveal market trends at a more general level and changes in consumer payment preferences in the retail sector.

Q6(2) Count of orders based on the no. of payment installments

ANS: -

```
SELECT
  payment_installments,
  COUNT(order_id) AS order_count
FROM
  `Target_Retailers.payments`
GROUP BY
  payment_installments
ORDER BY
  order_count DESC;
```

Row	payment_installments	order_count
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 & RECOMMENDATIONS: -

1. Higher count installments suggest that the clients have chosen this option. Order counts in this instance are substantially larger for payment installments with values of 15, 17, 20, 21, 23, and 24.

2. There are payment installments with extremely few or no orders, including 2, 3, and 22. This shows that either these payment choices are not well offered or that clients may not find them to be acceptable.

3. Consider launching educational activities if clients are unaware of the available payment installment choices or don't fully understand their advantages.