

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

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

<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

The customer table consists of 5 columns namely: customer_id, customer_unique_id, customer_zip_code_prefix, customer_city and customer_state with data types string, string, integer, string and string respectively. This column is essentially contains the basic details of the customers who have ordered in the time span between 2016-09-04 and 2018-10-17.

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

```
SELECT MIN(order_purchase_timestamp) AS min_order_date, MAX(order_purchase_timestamp) AS  
max_order_date  
FROM `project.orders`
```

Row	min_order_date ▼	max_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_city) AS city_count, COUNT(DISTINCT c.customer_state) AS  
state_count  
FROM `project.customers` AS c JOIN `project.orders` AS o  
ON c.customer_id = o.customer_id  
WHERE DATE(order_purchase_timestamp) BETWEEN '2016-09-04' AND '2018-10-17'
```

Row	city_count ▼	state_count ▼
1	4119	27

INSIGHTS

It can be seen that customers from 4119 cities of all 27 states of Brazil have placed orders within the time span between 2016-09-04 and 2018-10-17. Therefore, it is quite evident from the data that the Target Retail has a widespread coverage in Brazil and attracts the customers of different parts of the country.

RECOMMENDATIONS

Since the Target Retail has a wide acceptance from people of different regions of the country, they should try implementing marketing strategies understanding the customer movement, thereby creating an engaging environment in the market.

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 order_count
FROM `project.orders`
GROUP BY order_year
ORDER BY order_year
```

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

INSIGHTS

A substantial increase in the number of orders placed is clearly evident from the data.

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 months, COUNT(*) AS order_count
FROM project.orders
GROUP BY months
ORDER BY months
```

Row	months	order_count
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

It can be seen that there is spike in the orders in the months of March, May, July, August whereas, there is a significant drop in the number of orders in months of September, October, December.

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

- 1. 0-6 hrs : Dawn**
- 2. 7-12 hrs : Mornings**
- 3. 13-18 hrs : Afternoon**
- 4. 19-23 hrs : Night**

```
SELECT
CASE
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 0 AND EXTRACT(HOUR FROM
order_purchase_timestamp) <=6 THEN 'Dawn'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 7 AND EXTRACT(HOUR FROM
order_purchase_timestamp)<=12 THEN 'Morning'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >=13 AND EXTRACT(HOUR FROM
order_purchase_timestamp)<=18 THEN 'Afternoon'
ELSE 'Night'
END AS times_of_day,
COUNT(*) AS order_count
FROM `project.orders`
GROUP BY times_of_day
ORDER BY order_count
```

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

INSIGHTS

Most people of Brazil tend to place their orders at Afternoon. Thereby implying that customers are more likely to place orders on Afternoon followed by Night and Morning and are less likely to place orders on Dawn

RECOMMENDATIONS

Over the time, due to the advancement in the technology people tend to prefer online shopping more than window shopping. This is clearly evident from the spike in the number of orders over the year. Also, a fluctuating trend in the purchases can be seen, in the months of March, May, July, August there is an increase in the order volume and declines significantly in the months of September, October, December.

Therefore, during the months and time of day with least purchases it is recommended of carry out promotional campaigns like limited time discounts for most demanded products and more offers in the months of festivals, so that there would be reach among all strata of the society.

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

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

```
SELECT x.month,x.customer_state,COUNT(x.order_purchase_timestamp) AS order_count
FROM(SELECT *, EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month
FROM `project.orders` AS o JOIN `project.customers` AS c
ON o.customer_id = c.customer_id) AS x
GROUP BY x.customer_state,x.month
ORDER BY x.month
```

Row	month	customer_state	order_count
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
11	1	PA	82
12	1	MA	66

INSIGHTS

We get an insight about the purchasing behaviour of people of different states of Brazil. The month-on-month variation in the order volume of each states paves a way to the conclusion that there is a fluctuating trend in the purchases.

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

```
SELECT customer_state, COUNT(DISTINCT customer_id) AS customer_count
FROM `project.customers`
GROUP BY customer_state
ORDER BY customer_count DESC
```

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

INSIGHTS

The analysis throughs light on the purchasing behaviour of the Brazilian customers. It is seen that the need for goods vary from state-to-state. SP being the state with the greatest number of order counts.

RECOMMENDATIONS

It is recommended to implement region-wise marketing strategies especially in those states and cities where there are least purchases. Promoting the new products in the market through social media campaigns and regional commercial advertisements would increase the customer coverage.

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

1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
You can use the "payment_value" column in the payments table to get the cost of orders.

```
SELECT DISTINCT year,
ROUND((((SUM(p.payment_value)) -
(LAG(SUM(p.payment_value)) OVER(ORDER BY year)))/
(LAG(SUM(p.payment_value)) OVER(ORDER BY year))*100,2) AS percentage_cost_increase
FROM
(SELECT
order_id,
EXTRACT (YEAR FROM order_purchase_timestamp) AS year,
EXTRACT (MONTH FROM order_purchase_timestamp) AS month
FROM `project.orders`) AS o
JOIN `project.payments` AS p
ON o.order_id=p.order_id
WHERE o.year BETWEEN 2017 AND 2018
AND o.month BETWEEN 1 AND 8
GROUP BY o.year
ORDER BY o.year
```

Row	year	percentage_cost_increase
1	2017	null
2	2018	136.98

INSIGHTS

There is a remarkable increase of 136.98% in the cost of orders.

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

```
SELECT c.customer_state,
ROUND(SUM(p.payment_value),2) AS total_order_price,
ROUND(AVG(p.payment_value),2) AS average_order_price
FROM `project.customers` AS c JOIN `project.orders` AS o
ON c.customer_id = o.customer_id
JOIN `project.payments` AS p
ON o.order_id = p.order_id
```


GROUP BY c.customer_state
ORDER BY total_order_price

Row	customer_state	total_order_price	average_order_price
1	RR	10064.62	218.8
2	AP	16262.8	232.33
3	AC	19680.62	234.29
4	AM	27966.93	181.6
5	RO	60866.2	233.2
6	TO	61485.33	204.27
7	SE	75246.25	208.44
8	AL	96962.06	227.08
9	RN	102718.13	196.78
10	PI	108523.97	207.11

INSIGHTS

An insight about the total order price and average order price for each state is received.

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

```
SELECT c.customer_state,
ROUND(SUM(oi.freight_value),2) AS total_freight_value,
ROUND(AVG(oi.freight_value),2) AS avg_freight_value
FROM `project.customers` AS c JOIN `project.orders` AS o
ON c.customer_id = o.customer_id
JOIN `project.order_items` AS oi
ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY total_freight_value
```

Row	customer_state	total_freight_value	avg_freight_value
1	RR	2235.19	42.98
2	AP	2788.5	34.01
3	AC	3686.75	40.07
4	AM	5478.89	33.21
5	RO	11417.38	41.07
6	TO	11732.68	37.25
7	SE	14111.47	36.65
8	AL	15914.59	35.84
9	RN	18860.1	35.65
10	MS	19144.03	23.37

INSIGHTS

An insight about the total freight value and average freight value for each state is received.

RECOMMENDATIONS

As there is a visible growth in the spending patterns of the customers, introducing new offers, providing freebies and more reward points for the customers with the highest spending pattern would attract more customers. Providing exclusive offers and additional services to these customers would encourage them shopping with the brand.

5. Analysis based 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.
Do this in a single query.

You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

- $\text{time_to_deliver} = \text{order_delivered_customer_date} - \text{order_purchase_timestamp}$
- $\text{diff_estimated_delivery} = \text{order_estimated_delivery_date} - \text{order_delivered_customer_date}$

```
SELECT order_id,  
DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) AS time_to_deliver,  
DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date, DAY) AS diff_estimated_delivery  
FROM `project.orders`  
ORDER BY order_id
```

Row	order_id	time_to_deliver	diff_estimated_delivery
1	00010242fe8c5a6d1ba2dd792...	7	8
2	00018f77f2f0320c557190d7a1...	16	2
3	000229ec398224ef6ca0657da...	7	13
4	00024acbcdcf0a6daa1e931b03...	6	5
5	00042b26cf59d7ce69dfabb4e...	25	15
6	00048cc3ae777c65dbb7d2a06...	6	14
7	00054e8431b9d7675808bcb8...	8	16
8	000576fe39319847cbb9d288c...	5	15
9	0005a1a1728c9d785b8e2b08...	9	0
10	0005f50442cb953dcd1d21e1f...	2	18

INSIGHTS

An insight about the days taken to deliver each order from the order's purchase date and the difference between the estimated and actual delivery date of the orders is received.

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

TOP 5 STATES WITH HIGHEST FREIGHT VALUE

```
SELECT c.customer_state, ROUND(AVG(oi.freight_value),2) AS highest_avg_freight
FROM `project.customers` AS c JOIN `project.orders` AS o
ON c.customer_id = o.customer_id
JOIN `project.order_items` AS oi
ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY highest_avg_freight DESC
LIMIT 5
```

Row	customer_state	highest_avg_freight
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 FREIGHT VALUE

```
SELECT c.customer_state, ROUND(AVG(oi.freight_value),2) AS lowest_avg_freight
FROM `project.customers` AS c JOIN `project.orders` AS o
ON c.customer_id = o.customer_id
JOIN `project.order_items` AS oi
ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY lowest_avg_freight
LIMIT 5
```

Row	customer_state	lowest_avg_freight
1	SP	15.15
2	PR	20.53
3	MG	20.63
4	RJ	20.96
5	DF	21.04

INSIGHTS

RR, PB, RO, AC and PI are the states with high freight value. On the other hand SP,PR,MG,RJ,DF are the states with lowest freight value

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

TOP-5 STATES WITH HIGHEST AVERAGE DELIVERY TIME

```
SELECT c.customer_state,  
ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp, HOUR)),2)  
AS avg_time_to_delivery,  
FROM `project.customers` AS c JOIN `project.orders` AS o  
ON c.customer_id = o.customer_id  
JOIN `project.order_items` AS oi  
ON o.order_id=oi.order_id  
GROUP BY c.customer_state  
ORDER BY avg_time_to_delivery DESC  
LIMIT 5
```

Row	customer_state	avg_time_to_delivery
1	RR	676.98
2	AP	676.46
3	AM	632.85
4	AL	587.23
5	PA	569.6

TOP-5 STATES WITH LOWEST AVERAGE DELIVERY TIME

```
SELECT c.customer_state,  
ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp, HOUR)),2)  
AS avg_time_to_delivery,  
FROM `project.customers` AS c JOIN `project.orders` AS o  
ON c.customer_id = o.customer_id  
JOIN `project.order_items` AS oi  
ON o.order_id=oi.order_id  
GROUP BY c.customer_state  
ORDER BY avg_time_to_delivery  
LIMIT 5
```

Row	customer_state	avg_time_to_delivery
1	SP	208.87
2	PR	286.24
3	MG	287.11
4	DF	310.52
5	SC	359.53

INSIGHTS

RR,AP,AM,AL,PA are the states with highest delivery time whereas SP,PR,MG,DF,SC are the states with lowest average delivery time.

- Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

SELECT

```
x.customer_state,x.avg_dif,DENSE_RANK() OVER(ORDER BY x.avg_dif) AS rank
FROM (
SELECT c.customer_state,
ROUND(AVG(date_diff(o.order_estimated_delivery_date,o.order_delivered_customer_date,DAY)),2) AS avg_dif
FROM `project.customers` AS c JOIN `project.orders` AS o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state) AS x
ORDER BY rank
LIMIT 5
```

Row	customer_state	avg_dif	rank
1	AL	7.95	1
2	MA	8.77	2
3	SE	9.17	3
4	ES	9.62	4
5	BA	9.93	5

INSIGHTS

AL,MA,SE,ES,BA are the states whose order delivery is fast as compared to the estimated date of delivery

RECOMMENDATIONS

Opt for the means in which the orders get delivered faster. Customers tends to opt for the retailers who deliver their goods faster and within the estimated delivery date. Provide the precise tracking details of their orders and provide lucid return and exchange platform. Adopt best strategies analysing the constraints in each areas for the delay in delivery.

6. Analysis based on the payments:

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

```
SELECT x.month,x.payment_type, COUNT(x.order_purchase_timestamp) AS order_count
FROM (SELECT *, EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month
FROM `project.payments` AS p JOIN `project.orders` AS o
ON p.order_id = o.order_id) AS x
GROUP BY x.payment_type,x.month
ORDER BY x.month
```

Row	month	payment_type	order_count
1	1	voucher	477
2	1	credit_card	6103
3	1	debit_card	118
4	1	UPI	1715
5	2	credit_card	6609
6	2	voucher	424
7	2	UPI	1723
8	2	debit_card	82
9	3	voucher	591
10	3	credit_card	7707
11	3	UPI	1942
12	3	debit_card	109

INSIGHTS

The data gives us an idea about number of purchases made in each months using different payment types namely: vouchers, credit card, debit card and UPI. This throws a light on the contribution of internet banking in the shopping.

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

```
SELECT x.payment_installments, COUNT(x.order_purchase_timestamp) AS order_count
FROM (SELECT *
FROM `project.payments` AS p JOIN `project.orders` AS o
ON p.order_id = o.order_id) AS x
```


GROUP BY x.payment_installments

Row	payment_installment	order_count
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

INSIGHTS

The results show that the majority of items were purchased with full payment. However, a fair number of items were purchased in his 2,3,4 instalments. Thereby implying instalments are preferred. Additionally, some products were purchased in up to 24 instalments, implies there are customers opting for longer instalments.

RECOMMENDATIONS

Internet banking is becoming popular day by day, collaborating with banks and payment interfaces with cashbacks and extra discount would attract the customers. Hassle free transactions and wide options for EMI's would increase the sales.