Business Case: Target SQL Project

- 1 Import the dataset and do the usual exploratory analysis steps like checking the structure and characteristics of the data set
- 1. Data type of columns in a table
 - Tables Present in the Database

SELECT TABLE_NAME
FROM Business.INFORMATION_SCHEMA.TABLES

JOB IN	FORMATION	RESULTS	JSON
Row	TABLE_NAME ▼	//	
1	order_items		
2	sellers		
3	geolocation		
4	products		
5	orders		
6	payments		
7	customers		
8	order_reviews		

Database Schema of all the tables

Customers Table

SELECT COLUMN_NAME, DATA_TYPE
FROM Business.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'customers'

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

Sellers Table

```
SELECT COLUMN_NAME, DATA_TYPE
FROM Business.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'sellers'
```

Row	COLUMN_NAME ▼	DATA_TYPE ▼
1	seller_id	STRING
2	seller_zip_code_prefix	INT64
3	seller_city	STRING
4	seller_state	STRING

Products Table

```
SELECT COLUMN_NAME, DATA_TYPE
FROM Business.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'products'
```

Row	COLUMN_NAME ▼	DATA_TYPE ▼
1	product_id	STRING
2	product_category	STRING
3	product_name_length	INT64
4	product_description_length	INT64
5	product_photos_qty	INT64
6	product_weight_g	INT64
7	product_length_cm	INT64
8	product_height_cm	INT64
9	product_width_cm	INT64

Payments Table

SELECT COLUMN_NAME, DATA_TYPE
FROM Business.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'payments'

Row	COLUMN_NAME ▼	DATA_TYPE ▼
1	order_id	STRING
2	payment_sequential	INT64
3	payment_type	STRING
4	payment_installments	INT64
5	payment_value	FLOAT64

Orders Table

SELECT COLUMN_NAME, DATA_TYPE
FROM Business.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'orders'

Row	COLUMN_NAME ▼	DATA_TYPE ▼
1	order_id	STRING
2	customer_id	STRING
3	order_status	STRING
4	order_purchase_timestamp	TIMESTAMP
5	order_approved_at	TIMESTAMP
6	order_delivered_carrier_date	TIMESTAMP
7	order_delivered_customer_date	TIMESTAMP
8	order_estimated_delivery_date	TIMESTAMP

Order_items

```
SELECT COLUMN_NAME, DATA_TYPE
FROM Business.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'order_items'
```

Row	COLUMN_NAME ▼	DATA_TYPE ▼
1	order_id	STRING
2	order_item_id	INT64
3	product_id	STRING
4	seller_id	STRING
5	shipping_limit_date	TIMESTAMP
6	price	FLOAT64
7	freight_value	FLOAT64

Insights

- ❖ The database Schema for different tables tells about the interrelation between different tables and primary key and foreign keys
- 2. Time period for which the data is given

```
SELECT
MAX(order_purchase_timestamp) AS End_date,
MIN(order_purchase_timestamp) AS Start_date,
DATE_DIFF(MAX(order_purchase_timestamp), MIN(order_purchase_timestamp), DAY) AS
Month_Difference
FROM `Business.orders`;
```

Insights

The difference between the End date and start date is the total time period we are concerned about.

Row	End_date ▼	Start_date ▼	Month_Difference
1	2018-10-17 17:30:18 UTC	2016-09-04 21:15:19 UTC	772

3. Cities and states customers ordered during the given period

Distinct cities and states are present in the below table during the time period from

'2016-09-04 21:15:19 UTC' to '2018-10-17 17:30:18 UTC' that is 772 days.

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

2. In-depth Exploration:

1. Is there a growing trend in e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?

```
SELECT
  EXTRACT(YEAR FROM order_purchase_timestamp) AS Year,
  EXTRACT(MONTH FROM (order_purchase_timestamp)) AS Month,
  COUNT(*) AS order_frequency
FROM `Business.orders`
GROUP BY 1, 2
ORDER BY Year, Month;
```

Row	Year ▼	Month ▼	order_frequency ▼
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
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

Yearly order_frequency

SELECT

EXTRACT(YEAR FROM order_purchase_timestamp) AS Year, COUNT(*) as order_frequency

FROM 'Business.orders'

GROUP BY 1

ORDER BY 1

Row	Year ▼	//	order_frequency ▼
1		2016	329
2		2017	45101
3		2018	54011

Order Frequency monthly

```
SELECT
  EXTRACT(MONTH FROM (order_purchase_timestamp)) AS Month,
  COUNT(*) AS Order_Frequency
FROM `Business.orders`
GROUP BY Month
ORDER BY Month;
```

Month ▼	Order_Frequency
1	8069
2	8508
3	9893
4	9343
5	10573
6	9412
7	10318
8	10843
9	4305
10	4959
11	7544
12	5674

Insights

- ❖ Yearly Trend: As evident from the yearly order table there is a significant increase in orders from 2016 − 2018 which is a positive sign that people are ordering more from e-commerce sites and rapidly gaining momentum.
- ❖ Monthly Trend: The months from March to August have relatively higher order count compared to the rest of the months with the peak month being August where is order count is maximum this is due to the seasonal patterns as these months corresponds to autumn and Winter months in Brazil, people prefer to shop online due to reasons such as weather

conditions, holiday or cultural factors and exclusive offers and deals during winter month attracts the customers to promote online shopping.

In the months from September to January e-commerce activity is low due to seasonal variations and the transition from winter to spring. Consumers might be less inclined to make online purchases during this time as they plan for year-end expenses and engage in other seasonal activities. Additionally, significant shopping events like Black Friday and Christmas lead consumers to postpone purchases and wait for discounts and promotions. Economic factors and the presence of a robust traditional retail sector with physical stores further contribute to a relatively lower proportion of online sales as consumers opt for inperson shopping experiences during the holiday season.

Recommendation:

- The months from March to August give an opportunity to expand Target's presence in Brazil as these are the months where there is increased e-commerce activity, providing offers and discounts, right marketing strategies to attract customers, and having a user-friendly interface for the customers to get notified about the deals and sale prices by following these practices Target can get maximum profit and become customers favorite.
- The months from September onwards can be a crucial turning point as the e-commerce activity is low target can analyse the products that customers purchased the most and try to offer periodic discounts on them and can focus more on marketing and bring back-to-school offers so that customers continue to buy so that we can retain the customers.
 - 2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

```
WITH order_hours AS (SELECT Extract(HOUR FROM order_purchase_timestamp) as hour,

COUNT(*) AS order_count

FROM `Business.orders`

GROUP BY 1)

SELECT

CASE

WHEN hour >= 0 and hour< 5 THEN 'Dawn'

WHEN hour>= 5 and hour< 12 THEN 'Morning'

WHEN hour>= 12 and hour< 17 THEN 'Afternoon'

WHEN hour>= 17 and hour< = 23 THEN 'Night'

END as time,

SUM(order_count) as total_count

FROM order hours
```

Row	time ▼	total_count ▼
1	Dawn	4552
2	Morning	22428
3	Afternoon	32211
4	Night	40250

Insights:

❖ With the above table it is evident that the customers are mainly active in the afternoon and night period that corresponds to work hours, free time after the work hours. So, the company should focus on these hours because it is the peak buying time.

Recommendations:

- ❖ Target can focus on giving sales and discounts in order to get max sales and assist can assist this with email marketing sending notifications at these times which can benefit the company for more sales.
- ❖ In order to increase the orders in the morning and Dawn period Target can go with flash sales for 1 hour for specific items so that the order count can increase.
- ❖ Analysing customer buying patterns can positively affect the company sales.

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

3.1 Get month-on-month orders by states

```
SELECT C.customer_state, FORMAT_DATETIME("%B",
DATETIME(0.order_purchase_timestamp)) as Month,
COUNT(0.order_id) as order_count
FROM `Business.customers` C
JOIN `Business.orders` O ON C.customer_id = 0.customer_id
GROUP BY 1,2
order by 1,2
```

Row	customer_state ▼	Month ▼	order_count ▼
1	AC	April	9
2	AC	August	7
3	AC	December	5
4	AC	February	6
5	AC	January	8
6	AC	July	9
7	AC	June	7
8	AC	March	4
9	AC	May	10
10	AC	November	5

Row	customer_state ▼	Month ▼	order_count ▼
9	AC	May	10
10	AC	November	5
11	AC	October	6
12	AC	September	5
13	AL	April	51
14	AL	August	34
15	AL	December	14
16	AL	February	39
17	AL	January	39

Insights:

❖ The order counts differ majorly from state to state, SP, RJ, MG, and BA have consistent and good numbers of order counts, and throughout the year,

there is not much variation which indicates the strong presence of ecommerce in these areas.

The months from April to August and January have higher sales due to the festival and winter season and the new year accounts for high sales in January and offers that come by with it.

Recommendation:

- ❖ Target should focus more on the states which have a high level of sales such as the ones mentioned in insights these are states with a strong ecommerce presence and should focus on customer satisfaction.
- Target should launch marketing campaigns in these states and analyze customer requirements by doing surveys and focusing on finding out the reason and promoting products that interest people.
- Target should focus on seasonal trends like the peak month of August which is the time when all the people buy the most therefore introducing offers, sales, and different marketing methods to gain a stronghold
 - 3.2 Distribution of customers across the states in Brazil

```
SELECT customer_state , COUNT (DISTINCT customer_id)
FROM `Business.customers`
GROUP BY 1
ORDER BY 2 DESC
```

Row	customer_state ▼	f0_ ▼
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
13	PA	975
14	MT	907
15	MA	747
16	MS	715
17	PB	536
18	PI	495

Insights

- States like SP, RJ, and MG have the highest number of customers as the ecommerce presence is quite strong in these areas.
- The states of AC, AP, and RR have the lowest count and the people are not so confident in e-commerce shopping.

Recommendations:

❖ To increase customer count for Target in the states of RR, AP, and AL in Brazil, implement targeted marketing campaigns, establish local partnerships, offer region-specific products, enhance the customer experience, and optimize the online presence with localized content and promotions. Understanding the local market is essential.

- ❖ To increase customer count in RJ, MG, and SP, Target should prioritize improving the store experience, tailoring product selection, enhancing online presence, implementing competitive pricing and promotions, and engaging with local communities. These strategies aim to create a positive shopping environment, cater to regional preferences, optimize the online shopping experience, attract customers with attractive pricing, and build customer loyalty through community engagement.
 - 4. Impact on the Economy: Analyse the money movement by e-commerce by looking at order prices, freight and others.
 - 4.1 Get a % increase in the cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use "payment_value" column in payments table

```
WITH values 2017 AS
(SELECT round(sum(payment value),2) AS values 2017
FROM 'Business.payments' P
JOIN Business.orders as O
ON P.order id = O.order id
WHERE EXTRACT(YEAR FROM order purchase timestamp) = 2017
and EXTRACT(MONTH FROM order purchase timestamp) BETWEEN 1 and 8
ORDER BY 1),
values 2018 AS
(SELECT round(sum(payment value),2) AS values 2018
FROM 'Business.payments' P
JOIN Business.orders as O
ON P.order id = O.order id
WHERE EXTRACT(YEAR FROM order purchase timestamp)= 2018
and EXTRACT(MONTH FROM order purchase timestamp) BETWEEN 1 and 8
ORDER BY 1)
SELECT values 2017, values 2018,
round(((values 2018.values 2018 -
values 2017.values 2017.values
FROM values 2017, values 2018;
```

Row	vvalues_2017 ▼ //	vvalues_2018 ▼	increase ▼
1	3669022.12	8694733.84	136.98

Insights:

- ❖ There is a 136% increase in the cost of orders from previous years we can conclude that e-commerce is rapidly growing and becoming a necessity in Brazil.
- The higher cost of orders indicates increased demand and potentially higher average order values.
- ❖ E-commerce growth can be attributed to several factors, including improved internet access, increased smartphone penetration, convenience, a wider range of products, and competitive pricing. As more people embrace online shopping and businesses expand their digital presence, the e-commerce industry in Brazil is likely to continue growing.

Recommendation:

- Improve the online shopping experience with a user-friendly interface and efficient checkout process.
- ***** Enhance customer service through prompt responses and personalized assistance.
- **Expand product assortment to cater to a wider range of customer preferences.**
- Strengthen marketing efforts to increase brand visibility and attract more customers.
- ❖ Invest in logistics, data analytics, and customer retention strategies to support growth and maintain competitiveness.

4.2 Mean & Sum of price and freight value by a customer state

```
SELECT C.customer_state,
avg(price) as average_price,
avg(freight_value) as average_freight_value,
sum(price) as sum_price,
sum(freight_value) as sum_freight
FROM `Business.customers` C
JOIN `Business.orders` O ON C.customer_id = O.customer_id
JOIN `Business.order_items` OT ON O.order_id =
OT.order_id
GROUP BY 1
ORDER BY 1;
```

Row	customer_state ▼	average_price ▼	average_freight_valu	sum_price ▼	sum_freight ▼
1	AC	173.7277173913	40.07336956521	15982.94999999	3686.749999999
2	AL	180.8892117117	35.84367117117	80314.81000000	15914.589999999
3	AM	135.4959999999	33.20539393939	22356.84000000	5478.889999999
4	AP	164.3207317073	34.00609756097	13474.299999999	2788.500000000
5	BA	134.6012082126	26.36395893656	511349.9900000	100156.6799999
6	CE	153.7582611637	32.71420162381	227254.70999999	48351.589999999
7	DF	125.7705486284	21.04135494596	302603.9399999	50625.499999999
8	ES	121.9137012411	22.05877659574	275037.30999999	49764.59999999
9	GO	126.2717316759	22.76681525932	294591.9499999	53114.979999999

Insights and Recommendation:

- Average Order Value: The average price across different customer states varies significantly, ranging from 109.65 (SP) to 191.48 (PB). Target could focus on increasing the average order value by implementing strategies such as upselling, cross-selling, and offering bundled or complementary products.
- Freight Costs: The average freight value also varies across states, with AL having the highest average at 35.84 and SP having the lowest at 15.15. Target could optimize logistics and negotiate better shipping rates to reduce freight costs, which would make their products more attractive to customers and potentially increase sales.
- ❖ Market Potential: Consider the sum of prices and sum of freight values to gauge market potential. SP has the highest sum of prices (5,202,955.05) and sum of freight (718,723.07), indicating a large market with significant revenue potential. Target could focus on expanding its presence and marketing efforts in SP to tap into this lucrative market.
- Regional Opportunities: Identify states with higher average prices and lower competition, such as PB (191.48) and AL (180.89). Target could strategically target these regions with tailored marketing campaigns and promotions to attract customers who are willing to spend more, potentially increasing sales and profitability.
- Customer Retention: Analyze customer behavior in each state to identify opportunities for improving customer retention. For instance, states like MG and RJ have significantly higher sums of prices and freight values, indicating potential loyal customer bases. Target could implement loyalty programs, personalized offers, and exceptional customer service to retain customers in these regions and encourage repeat purchases.

5. Analysis of sales, freight, and delivery time

- 5.1 Calculate days between purchasing, delivering and estimated delivery
- 5.2 Find time to delivery, diff estimated delivery.

Row	order_id ▼	delivery_time ▼	estimated_delivery
1	bfbd0f9bdef84302105ad712d	54	18
2	3b697a20d9e427646d925679	23	23
3	be5bc2f0da14d8071e2d45451	24	34
4	65d1e226dfaeb8cdc42f66542	35	52
5	a41c8759fbe7aab36ea07e038	30	56
6	d207cc272675637bfed0062ed	27	50
7	cd3b8574c82b42fc8129f6d50	10	50
8	ae8a60e4b03c5a4ba9ca0672c	30	58
9	ef1b29b591d31d57c0d733746	28	52

5.3 Group data by state, take mean of freight_value, time_to_delivery, diff estimated delivery

Row	customer_state ▼	mean_freight ▼	delivery_time ▼	estimated_delivery
1	RJ	18.57	7	52
2	MG	20.01	30	17
3	SC	18.51	30	59
4	SP	13.79	7	51
5	RJ	19.38	10	52
6	RJ	14.11	35	52
7	GO	21.01	23	33
8	SP	10.5	12	7
9	RS	21.76	12	25

Row	customer_state ▼	mean_freight ▼ //	delivery_time ▼	estimated_delivery
1	ES	15.78	209	28
2	RJ	17.26	208	19
3	PA	25.12	195	30
4	PI	27.88	194	32
5	SE	27.75	194	28
6	PI	105.19	194	39
7	SP	54.33	191	15
8	SP	16.05	189	22
9	SE	20.8	188	28

Insights:

❖ The maximum time taken for delivery is 209 days whereas the estimated date is 28 which will create a bad reputation for the company this trend is followed in multiple cases the customer satisfaction will be low and could impact the image of the company.

Recommendation:

- ❖ Optimize operations: Streamline warehouse and inventory management systems, collaborate with reliable shipping partners, and establish multiple fulfillment centers strategically to improve delivery efficiency.
- ❖ Enhance last-mile delivery: Focus on improving the final leg of the delivery process by utilizing local delivery services, optimizing routes, and implementing real-time tracking and communication with customers.
- Prioritize customer experience: Offer express delivery options for faster service, maintain clear and proactive communication with customers regarding order updates and potential delays, and utilize predictive analytics to forecast demand and optimize inventory levels.

5.3 Top 5 states with highest/lowest average time to delivery

Row	customer_state ▼	avg_delivery ▼
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

5.4Top 5 states with highest/lowest average freight value Highest Freight value

Row	customer_state ▼	avg_freight ▼
1	PB	43.09
2	RR	43.09
3	RO	41.33
4	AC	40.05
5	PI	39.12

```
SELECT customer_state,round(avg(freight_value),2) as avg_freight FROM `Business.customers` c

JOIN Business.orders o ON c.customer_id = o.customer_id

JOIN `Business.order_items` ot ON o.order_id = ot.order_id

WHERE order_delivered_customer_date IS NOT NULL

GROUP BY 1

ORDER BY 2 DESC

LIMIT 5
```

Lowest freight Value:

Row	customer_state	· //	avg_freight	• //
1	SP			15.11
2	PR			20.47
3	MG			20.63
4	RJ			20.91
5	DF			21.07

```
SELECT customer_state,round(avg(freight_value),2) as avg_freight
FROM `Business.customers` c

JOIN Business.orders o ON c.customer_id = o.customer_id

JOIN `Business.order_items` ot ON o.order_id = ot.order_id

WHERE order_delivered_customer_date IS NOT NULL

GROUP BY 1

ORDER BY 2

LIMIT 5
```

5.5Top 5 states where delivery is really fast/ not so fast compared to estimated date

```
SELECT customer_state,
    round(avg(DATE_DIFF(order_estimated_delivery_date,order_purchase_timestamp,DAY)),2) as
avg_estimate_delviery
FROM `Business.customers` c

JOIN Business.orders o ON c.customer_id = o.customer_id

WHERE order_estimated_delivery_date IS NOT NULL

GROUP BY 1

ORDER BY 2

LIMIT 5
```

Row	customer_state ▼	avg_estimate_delvier
1	SP	18.81
2	DF	24.06
3	MG	24.22
4	PR	24.25
5	ES	25.27

Not fast:

Row	customer_state	•	avg_estimate_delvie
1	RR		46.17
2	AP		45.71
3	AM		44.76
4	AC		40.77
5	RO		38.41

```
SELECT customer_state,
    round(avg(DATE_DIFF(order_estimated_delivery_date,order_purchase_timesta
mp,DAY)),2) as avg_estimate_delviery
FROM `Business.customers` c

JOIN Business.orders o ON c.customer_id = o.customer_id

WHERE order_estimated_delivery_date IS NOT NULL

GROUP BY 1

ORDER BY 2 DESC

LIMIT 5
```

6. Payment type analysis

```
SELECT EXTRACT(YEAR FROM O.order_purchase_timestamp) as YEAR,

EXTRACT(Month FROM O.order_purchase_timestamp) as Month,

payment_type,count(O.order_id) as orders

FROM `Business.orders` O

JOIN Business.payments P ON O.order_id = P.order_id

GROUP BY 1,2,3

ORDER BY 1,2
```

Row	YEAR ▼	Month ▼	payment_type ▼	orders ▼
1	2016	9	credit_card	3
2	2016	10	credit_card	254
3	2016	10	voucher	23
4	2016	10	debit_card	2
5	2016	10	UPI	63
6	2016	12	credit_card	1
7	2017	1	voucher	61
8	2017	1	UPI	197
9	2017	1	credit_card	583

Insights:

- ❖ We can observe that year by year credit card payments are increasing followed by UPI payments as it is easy and can be used for faster transactions.
- Debit card payments are not much popular as there is a very low probability that customer uses a debit card
- Vouchers are increasingly used by people more and more

Recommendation:

- ❖ To retain credit card customers Target can provide offers on credit cards frequently so that the credit card payments can increase more and tie up with banks to issue Target Credit cards which can offer cash backs
- Simplify the payment process and remove any unnecessary steps or friction points to make it easy and convenient for customers to use debit card and UPI payments.
- Promote the benefits of debit card and UPI payments, highlighting their convenience, security, and instant transfer capabilities, and offer incentives such as cashbacks, discounts, or rewards to encourage customers to choose these payment options.
- Optimize the mobile experience, collaborate with banks and UPI providers, leverage social media and email marketing, and provide excellent customer

support to further enhance the adoption of debit card and UPI payments for Target e-commerce.

6.2 Count of orders based on the no. of payment installments

SELECT payment_installments,COUNT(order_id) as orders FROM `Business.payments`

GROUP BY 1

ORDER BY 1,2

Row	payment_installment	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

Insights:

- ❖ The majority share is taken by one-time payments followed by 2 installments and 3 installments payments.
- ❖ Instalments 4 to 8 are considerably less as people don't want long-term Emi.

Recommendation:

- Offer flexible payment options such as one-time payments and installment plans at the checkout to cater to different customer preferences and financial situations.
- Clearly communicate the pricing, financing terms, and any associated fees or interest rates for installment plans to ensure transparency and build trust with customers.
- ❖ Provide exclusive promotions, discounts, or incentives for customers who choose installment plans, encouraging them to opt for this payment method.
- Empower customers by allowing them to customize their payment schedules and choose payment options that suit their needs.
- By providing flexible payment options and incentives, you can increase ecommerce sales for Target and attract customers who may have been hesitant to make larger purchases.