TARGET CASE STUDY

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.

This business case has information of 100k orders from 2016 to 2018 made at Target in Brazil

Ques 1) Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.
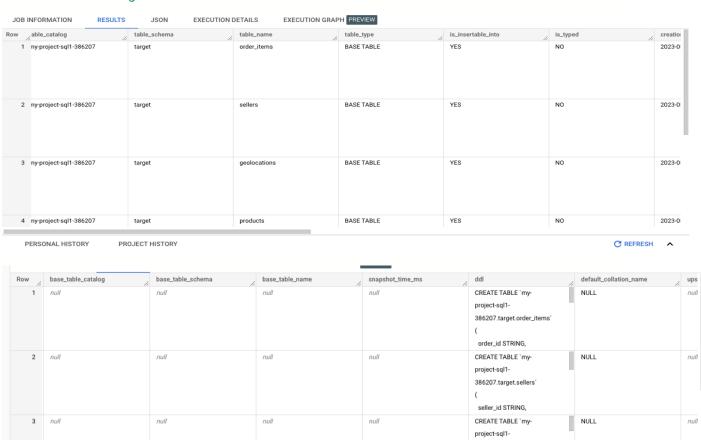
1.1)   Data type of columns in a table

Ans.
```
SELECT *
  FROM `target`.INFORMATION_SCHEMA.TABLES
```

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH PREVIEW			
Row	able_catalog	table_schema	table_name	table_type	is_insertable_into	is_typed	creatio
1	ny-project-sql1-386207	target	order_items	BASE TABLE	YES	NO	2023-0:
2	ny-project-sql1-386207	target	sellers	BASE TABLE	YES	NO	2023-0:
3	ny-project-sql1-386207	target	geolocations	BASE TABLE	YES	NO	2023-0:
4	ny-project-sql1-386207	target	products	BASE TABLE	YES	NO	2023-0:

PERSONAL HISTORY   PROJECT HISTORY   ⟳ REFRESH   ⌃

Row	base_table_catalog	base_table_schema	base_table_name	snapshot_time_ms	ddl	default_collation_name	ups
1	null	null	null	null	CREATE TABLE `my-project-sql1-386207.target.order_items` (order_id STRING,	NULL	null
2	null	null	null	null	CREATE TABLE `my-project-sql1-386207.target.sellers` (seller_id STRING,	NULL	null
3	null	null	null	null	CREATE TABLE `my-project-sql1-386207.target.geolocations ` (	NULL	null
4	null	null	null	null	CREATE TABLE `my-	NULL	null

PERSONAL HISTORY   PROJECT HISTORY   ⟳ REFRESH   ⌃

- In the complete dataset we have STRING, INT64, TIMESTAMP and FLOAT64 types of data type .

Ans.
```
SELECT MIN(order_purchase_timestamp) AS start_point,
       MAX(order_purchase_timestamp) AS end_point
       FROM `target.orders`
```

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	E

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

- The data is given from 2016-09-04 time 21:15:19 UTC to 2018-10-17 time 17:30:18 UTC.

1.3) *Cities and States of customers ordered during the given period.*

Ans.
```
SELECT c.customer_city, c.customer_state

FROM `target.customers` c INNER JOIN `target.orders` o

ON c.customer_id = o.customer_id

GROUP BY c.customer_city, c.customer_state
```

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	
11	belo horizonte	MG	

Results per page: 50 ▼   1 – 50 of 4310 |◂

- Customers ordered from 4310 different cities for the given time period.
- If we do not perform Join between customers and orders then also we get the same answer for this data but JOIN is more preferable to give proper answer.

Ques 2) 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.
```sql
SELECT DATE_TRUNC(order_purchase_timestamp,MONTH) AS month_year, count(*)
AS order_count
FROM `target.orders`
GROUP BY month_year
ORDER BY month_year
```

Row	month_year	order_count
1	2016-09-01 00:00:00 UTC	4
2	2016-10-01 00:00:00 UTC	324
3	2016-12-01 00:00:00 UTC	1
4	2017-01-01 00:00:00 UTC	800
5	2017-02-01 00:00:00 UTC	1780
6	2017-03-01 00:00:00 UTC	2682
7	2017-04-01 00:00:00 UTC	2404
8	2017-05-01 00:00:00 UTC	3700
9	2017-06-01 00:00:00 UTC	3245
10	2017-07-01 00:00:00 UTC	4026
11	2017-08-01 00:00:00 UTC	4331
12	2017-09-01 00:00:00 UTC	4285

Results per page: 50 ▼   1 – 25 of 25

- Yes we can clearly see a growing trend in orders with time in Brazil, but there is a sudden drop in order count for September & October 2018 . It would be helpful to analyze more data to get a clearer picture of the trend.

- We can see seasonality effect also by order count. In month of January order count suddenly increases , in March & August also we can see good growth in order counts.

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

Ans.

```sql
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) > 6 AND EXTRACT(HOUR FROM order_purchase_timestamp) <= 12
THEN "Morning"
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) > 12 AND EXTRACT(HOUR FROM order_purchase_timestamp) <= 18
THEN "Afternoon"
ELSE "Night"
END AS time_period,
COUNT(*) AS transaction_count
FROM `target.orders`
GROUP BY time_period
ORDER BY transaction_count
```

Row	time_period	transaction_count
1	Dawn	5242
2	Morning	27733
3	Night	28331
4	Afternoon	38135

- Based on these figures, we can see that the highest number of transactions occurred in the afternoon (38,135 transactions), followed by the night (28,331 transactions). The morning (27,733 transactions) also had a significant number of transactions. The dawn period had the lowest number of transactions (5,242 transactions).
- Therefore, from this data, it appears that Brazilian customers tend to make the most purchases during the afternoon and night, while the morning also sees a substantial number of transactions. The dawn period seems to have the least customer activity in terms of transactions.

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

Q 3.1) Get month on month orders by states.

Ans.

```sql
    SELECT DATE_TRUNC(EXTRACT(DATE FROM order_purchase_timestamp) ,MONTH)
AS month_year,
    c.customer_state,
    count(*) AS order_count
FROM `target.orders` o JOIN `target.customers` c
ON o.customer_id = c.customer_id
GROUP BY month_year, c.customer_state
ORDER BY c.customer_state, month_year
```

Row	month_year	customer_state	order_count
1	2017-01-01	AC	2
2	2017-02-01	AC	3
3	2017-03-01	AC	2
4	2017-04-01	AC	5
5	2017-05-01	AC	8
6	2017-06-01	AC	4
7	2017-07-01	AC	5
8	2017-08-01	AC	4
9	2017-09-01	AC	5
10	2017-10-01	AC	6
11	2017-11-01	AC	5
12	2017-12-01	AC	5

Results per page: 50 ▼   1 – 50 of 565

- We have got Month on month orders by state with result in order by customer_state and month_year. Total of 565 rows we have got in the result.

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

```sql
SELECT customer_state, COUNT(*) AS customer_count
FROM `target.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

Results per page: 50 ▼   1 – 27 of 27

- From this data, we can observe the distribution of customers across different states in Brazil. São Paulo (SP) has the highest customer count with 41,746 customers, followed by Rio de Janeiro (RJ) with 12,852 customers. Minas Gerais (MG) also has a significant number of customers with 11,635. These three states have the largest customer bases.

- In contrast, states like Pará (PA), Ceará (CE), and Pernambuco (PE) have relatively lower customer counts compared to the top states.

- Analyzing this table provides insights into the customer distribution across different Brazilian states, which can be useful for businesses to understand their customer base and tailor their strategies accordingly.

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

Q 4.1) 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(100 * (payments - LAG(t1.payments,1) OVER(ORDER BY t1._year))/
LAG(t1.payments,1) OVER(ORDER BY _year),2) AS change
FROM
(SELECT EXTRACT(YEAR FROM o.order_purchase_timestamp) _year,
SUM(payment_value) AS payments
FROM `target.orders` o JOIN `target.payments` p
ON o.order_id = p.order_id
WHERE EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
GROUP BY EXTRACT(YEAR FROM o.order_purchase_timestamp)) t1
ORDER BY t1._year
```

Row	_year	payments	change
1	2017	3669022.0	null
2	2018	8694734.0	136.98

- The cost of customer orders during the specified period in 2018 increased by approximately 136.98% compared to the same period in 2017. This indicates a significant rise in the cost of orders, which may be due to various factors such as increased prices, changes in product offerings, higher demand, or other market dynamics.

- With this information, project managers can make informed decisions regarding pricing strategies, cost management initiatives, resource allocation, and operational adjustments to ensure the project remains financially sustainable and meets its objectives.

Q 4.2) Mean & Sum of price and freight value by customer state.

Ans.

```sql
SELECT c.customer_state,
       ROUND(SUM(i.price),2) AS sum_price,
       ROUND(AVG(i.price),2) AS mean_price,
       ROUND(SUM(i.freight_value),2) AS sum_freight_value,
       ROUND(AVG(i.freight_value),2) AS mean_freight_value
FROM `target.order_items` i JOIN `target.orders` o
ON i.order_id = o.order_id
JOIN `target.customers` c
ON o.customer_id = c.customer_id
GROUP BY c.customer_state
```

Row	customer_state	sum_price	mean_price	sum_freight_valu	mean_freight_va
1	SP	5202955.05	109.65	718723.07	15.15
2	RJ	1824092.67	125.12	305589.31	20.96
3	PR	683083.76	119.0	117851.68	20.53
4	SC	520553.34	124.65	89660.26	21.47
5	DF	302603.94	125.77	50625.5	21.04
6	MG	1585308.03	120.75	270853.46	20.63
7	PA	178947.81	165.69	38699.3	35.83
8	BA	511349.99	134.6	100156.68	26.36
9	GO	294591.95	126.27	53114.98	22.77
10	RS	750304.02	120.34	135522.74	21.74

Results per page: 50 ▼   1 – 27 of 27

- By this table we can say SP(Sao Paulo) is having highest sum_price value(52,02,955.05) and highest sum_freight value(7,18,723.07). Higher sum_price values suggest higher, greater economic activity and revenue generation within e-commerce sector and high sum_freight_value indicates higher volumer of orders or premium shipping from those states.

- PB(Paraiba) is having highest mean_price (191.48). Mean_price reflects consumer pattern in state. Higher mean_price suggest either a preference of higher priced product or a higher purchasing power in those states.

- RR(Roraima) is having lowest sum_price and highest mean_freight_value (42.98), higher mean_freight_value means higher logistic price in this state.

Impact on Economy:

- Revenue Growth: The increase in order prices from 2017 to 2018 suggests a significant growth in revenue generated by e-commerce in Brazil. The higher order prices indicate increased transaction values, which contribute to the overall revenue growth of businesses operating in the e-commerce sector.
- Business Expansion: This growth can encourage businesses to expand their operations, invest in infrastructure, and hire more employees. The expansion of e-commerce businesses can have a positive impact on the economy by creating job opportunities and driving economic growth.
- Tax Contributions: With higher order prices, businesses in the e-commerce sector generate more revenue, which leads to increased tax contributions to the government that benefit the economy as a whole.
- Consumer Behavior: The rise in order prices reflects changes in consumer behavior, including a willingness to spend more on online purchases. This shift in consumer behavior indicates a growing acceptance of e-commerce.

Overall, the increase in order prices in Brazil's e-commerce sector suggests a positive impact on the economy, including revenue growth, business expansion, increased tax contributions, changes in consumer behavior, and enhanced market competitiveness. These factors collectively contribute to the development and dynamism of the digital economy in Brazil.

Ques 5) Analysis on sales, freight and delivery time

Q 5.1) Calculate days between purchasing, delivering and estimated delivery.
Ans.)

```sql
SELECT order_id,
    DATETIME_DIFF(order_estimated_delivery_date, order_purchase_timestamp,
DAY ) estimate_days,
    DATETIME_DIFF(order_delivered_customer_date, order_purchase_timestamp,
DAY) delivered_in,
    DATETIME_DIFF(order_delivered_customer_date,order_estimated_delivery_date,
DAY) difference
FROM `target.orders`
ORDER BY delivered_in DESC
```

Row	order_id	estimate_days	delivered_in	difference
1	ca07593549f1816d26a572e06…	28	209	181
2	1b3190b2dfa9d789e1f14c05b…	19	208	188
3	440d0d17af552815d15a9e41a…	30	195	165
4	0f4519c5f1c541ddec9f21b3bd…	32	194	161
5	285ab9426d6982034523a855f…	28	194	166
6	2fb597c2f772eca01b1f5c561b…	39	194	155
7	47b40429ed8cce3aee9199792…	15	191	175
8	2fe324febf907e3ea3f2aa9650…	22	189	167
9	2d7561026d542c8dbd8f0daea…	28	188	159
10	437222e3fd1b07396f1d9ba8c…	42	187	144
11	c27815f7e3dd0b926b5855262…	25	187	162

Results per page: 50 ▼   1 – 50 of 99441

- By this data table we can clearly see that some products were either delivered late or not delivered, this might be due to logistics issue .

- Company need to take immediate strict actions for this problem otherwise it will effect the business of the company and will loose the trust of common people. For this they need to discuss it with logistics department employees and get the things improve as soon as possible.

- Customer experience is most important for e-commerce.

Q 5.2) Find time_to_delivery & diff_estimated_delivery .

Ans.)

```
SELECT order_id,
DATETIME_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY)
time_to_delivery,
DATETIME_DIFF(order_estimated_delivery_date,order_delivered_customer_date,
DAY) diff_estimate_delivery
FROM `target.orders`
```

Row	order_id	time_to_delivery	diff_estimate_de
1	1950d777989f6a877539f5379…	30	-12
2	2c45c33d2f9cb8ff8b1c86cc28…	30	28
3	65d1e226dfaeb8cdc42f66542…	35	16
4	635c894d068ac37e6e03dc54e…	30	1
5	3b97562c3aee8bdedcb5c2e45…	32	0
6	68f47f50f04c4cb6774570cfde…	29	1
7	276e9ec344d3bf029ff83a161c…	43	-4
8	54e1a3c2b97fb0809da548a59…	40	-4
9	fd04fa4105ee8045f6a0139ca5…	37	-1
10	302bb8109d097a9fc6e9cefc5…	33	-5
11	66057d37308e787052a32828…	38	-6

Results per page: 50 ▼   1 – 50 of 99441

- By this query we get time to delivery of an order & difference in estimated time to delivery and actual delivery date of a product. Negative diff_estimated_date means delayed in product delivery, whereas positive means early product delivery. 0 diff_estimated_delivery means delivery on the same day as expected.

- A negative diff_estimate_delivery is a negative sign for the company as well and vice versa.

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

Ans.)

```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),4) avg_time_to_delivery,
ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,order_delivered_customer
_date, DAY)),4)  avg_diff_estimate_delivery ,
    ROUND(AVG(i.freight_value),4) avg_freight_value
FROM `target.orders` o
JOIN `target.order_items` i
ON o.order_id = i.order_id
JOIN `target.customers` c
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
```

Row	customer_state	avg_time_to_del	avg_diff_estimat	avg_freight_valu
1	MT	17.5082	13.6393	28.1663
2	MA	21.2038	9.11	38.257
3	AL	23.993	7.9766	35.8437
4	SP	8.2596	10.2656	15.1473
5	MG	11.5155	12.3972	20.6302
6	PE	17.7921	12.5521	32.9179
7	RJ	14.6894	11.1445	20.9609
8	DF	12.5015	11.2747	21.0414
9	RS	14.7083	13.203	21.7358
10	SE	20.9787	9.1653	36.6532
11	PR	11.4808	12.5339	20.5317
12	PA	23.3017	13.3748	35.8327

Results per page: 50 ▼   1 – 27 of 27

Q 5.4) Sort the data to get the following:
Q 5.5) Top 5 states with highest average freight value - sort in desc limit .
Ans.)

```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),4) avg_time_to_delivery,
ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,order_delivered_customer
_date, DAY)),4)  avg_diff_estimate_delivery ,
    ROUND(AVG(i.freight_value),4) avg_freight_value
FROM `target.orders` o
JOIN `target.order_items` i
ON o.order_id = i.order_id
JOIN `target.customers` c
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_freight_value DESC
LIMIT 5
```

Row	customer_state	avg_time_to_deli	avg_diff_estimat	avg_freight_valu
1	RR	27.8261	17.4348	42.9844
2	PB	20.1195	12.1502	42.7238
3	RO	19.2821	19.0806	41.0697
4	AC	20.3297	20.011	40.0734
5	PI	18.9312	10.6826	39.148

- RR(Roraima) is having the highest average freight value (42.9844) followed by PB(paraiba), RO(Rondônia), AC(Acre), PI(Piauí)

- By this table we can say when avg_time_to_delivery decreases then avg_freight_value also decreases, which means in some states average logistics cost is high due to time taken for delivery.

Q) *Top 5 states with lowest average freight value - sort in asc limit .*

Ans.)

```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),4) avg_time_to_delivery,
ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,order_delivered_customer
_date, DAY)),4)  avg_diff_estimate_delivery ,
    ROUND(AVG(i.freight_value),4) avg_freight_value
FROM `target.orders` o
JOIN `target.order_items` i
ON o.order_id = i.order_id
JOIN `target.customers` c
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_freight_value ASC
LIMIT 5
```

Row	customer_state	avg_time_to_delivery	avg_diff_estimate_delivery	avg_freight_value
1	SP	8.2596	10.2656	15.1473
2	PR	11.4808	12.5339	20.5317
3	MG	11.5155	12.3972	20.6302
4	RJ	14.6894	11.1445	20.9609
5	DF	12.5015	11.2747	21.0414

- By this table we can say that Sp (São Paulo) is having lowest avg_freight_value (15.1473) followed by PR(Paraná).
- By analyzing this table we can say when the avg_time_to_delivery is less, then avg_freight_value is also less, means delivery charges are less for the places where average time to delivery is less.

Q 5.6) Top 5 states with highest average time to delivery

Ans.)

```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),4) avg_time_to_delivery,
FROM `target.orders` o
JOIN `target.order_items` i
ON o.order_id = i.order_id
JOIN `target.customers` c
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_time_to_delivery DESC
LIMIT 5
```

Row	customer_state	avg_time_to_delivery
1	RR	27.8261
2	AP	27.7531
3	AM	25.9632
4	AL	23.993
5	PA	23.3017

- By this table we can say that RR (Roraima) is having highest avg_time_to_delivery with 27+ Days, followed by AP (Amapá) also for 27+ Days.

- Company should work on these states delivery process to get it done fast, as 27 days are very much for delivery .

- Slower / delayed delivery in some states may effect the business in those sates.

Q) Top 5 states with lowest average time to delivery

Ans.)

```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),4) avg_time_to_delivery,
FROM `target.orders` o
JOIN `target.order_items` i
ON o.order_id = i.order_id
JOIN `target.customers` c
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_time_to_delivery ASC
LIMIT 5
```

Row	customer_state	avg_time_to_delivery
1	SP	8.2596
2	PR	11.4808
3	MG	11.5155
4	DF	12.5015
5	SC	14.521

- By analyzing this table we can say that SP (São Paulo) is having the least avg_time_to_delivery 8+ Days, which means fastest delivery, followed by state PR(Paraná) with 11+ Days.

- Fast delivery gains trust of customers and in result more orders for the company from those states.

Ques 5.7) Top 5 states where delivery is really fast compared to estimated date
Ans.)
```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,
order_purchase_timestamp, DAY )),2) estimate_days,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),2) delivered_in,
    ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,
order_delivered_customer_date, DAY)),2) difference
FROM `target.orders` o JOIN `target.customers` c
ON o.customer_id = c.customer_id
GROUP BY c.customer_state
ORDER BY difference ASC
LIMIT 5
```

Row	customer_state	estimate_days	delivered_in	difference
1	AL	32.23	24.04	7.95
2	MA	30.11	21.12	8.77
3	SE	30.35	21.03	9.17
4	ES	25.27	15.33	9.62
5	BA	29.04	18.87	9.93

- Difference column gives us information about products delivered before or after the estimated delivery date. A positive value defines delivery before estimated delivery day and vice versa.

- AL (Alagoas) is having the minimum positive difference with 7.97 Days ~ 8 Days. Which means fastest delivery in this state , followed by MA , SE, ES, BA states.

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

Ans.)

```sql
SELECT c.customer_state,
    ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,
order_purchase_timestamp, DAY )),2) estimate_days,
    ROUND(AVG(DATETIME_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)),2) delivered_in,
    ROUND(AVG(DATETIME_DIFF(order_estimated_delivery_date,
order_delivered_customer_date, DAY)),2) difference
FROM `target.orders` o JOIN `target.customers` c
ON o.customer_id = c.customer_id
GROUP BY c.customer_state
ORDER BY difference DESC
LIMIT 5
```

Row	customer_state	estimate_days	delivered_in	difference
1	AC	40.77	20.64	19.76
2	RO	38.41	18.91	19.13
3	AP	45.71	26.73	18.73
4	AM	44.76	25.99	18.61
5	RR	46.17	28.98	16.41

- AC (Acre) , RO (Rondônia) are having maximum average delivery day with approx 19 Days, followed by AP, AM, RR. Although the average difference is still positive for these states which means on an average products are delivered before estimated date of delivery , but it less as compared to other states.

Ques 6) Payment type analysis:

Q 6.1) Month over Month count of orders for different payment types.
Ans.)
```
SELECT FORMAT_TIMESTAMP('%Y-%m', TIMESTAMP_TRUNC(order_purchase_timestamp,
MONTH)) AS month_on_month_date,
p.payment_type, COUNT(*) AS order_count
FROM `target.payments` p JOIN `target.orders` o
ON p.order_id = o.order_id
GROUP BY month_on_month_date, p.payment_type
ORDER BY month_on_month_date
```

Row	month_on_month_date	payment_type	order_count
1	2016-09	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	2017-01	voucher	61
8	2017-01	UPI	197
9	2017-01	credit_card	583
10	2017-01	debit_card	9
11	2017-02	credit_card	1356
12	2017-02	voucher	119

Results per page: 50 ▾   1 – 50 of 90

- After analyzing this table we can say maximum order counts(76,795) occurred when customers choose to use credit card as a payment type.
- UPI was the second(19,784) most preferable payment mode by customers, followed by vouchers with order counts(5,775).
- Debit card over-all order counts were 1,529 .
- As a company we are on a right track, as when customer pay from credit card then they tend to buy mode items, we need to keep on providing good offers for credit card.

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

Ans.)

```sql
SELECT payment_installments,
       COUNT(order_id) AS count_of_orders
FROM `target.payments`
GROUP BY payment_installments
ORDER BY count_of_orders DESC
```

Row	payment_installments	count_of_orders
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
11	12	133

Results per page: 50 ▼   1 – 24 of 24

- By this table we can say maximum customers opted for 1 payment_installments with 52,546 order counts.

- Least customers opted for 22, 23 months of payment_installments with only 1-1 order counts.

Insights :-

1. Growing Trend: The analysis indicates a consistent growth trend in orders over time for TARGET in Brazil, with the exception of a sudden drop in order count in September and October 2018. This growth trend showcases the potential of the e-commerce market in Brazil and indicates a positive response from customers towards online shopping.

2. Seasonality Effect: The analysis reveals seasonality in order counts, with higher order counts in January, March, and August. Understanding and leveraging these seasonal peaks can help TARGET optimize marketing efforts, inventory management, and resource allocation during these periods to maximize sales and customer engagement.

3. Customer Distribution: The data highlights the distribution of customers across different states in Brazil, with São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) having the largest customer bases. This information can be utilized to prioritize marketing campaigns, establish regional strategies, and tailor product offerings to cater to the preferences and demands of customers in these states.

4. Delivery Performance: The analysis identifies areas for improvement in delivery performance. Instances of late or undelivered products indicate potential logistics issues that need immediate attention. By addressing these issues, TARGET can enhance customer satisfaction, build trust, and maintain a competitive advantage in the market.

5. Regional Economic Activity: The analysis of sum_price and sum_freight_value by state provides insights into economic activity within the e-commerce sector. São Paulo stands out as a state with the highest sum_price and sum_freight_value, suggesting significant economic activity and revenue generation. This information can guide strategic decisions regarding investment, expansion, and resource allocation.

Recommendations :-

1. Investigate the Drop in Order Count: Further analysis and investigation should be conducted to identify the factors behind the sudden drop in order count in September and October 2018. This investigation can involve exploring external factors, evaluating marketing strategies, and reviewing customer feedback to uncover the root causes. Based on the findings, appropriate actions can be taken to mitigate such drops in the future.

2. Seasonal Marketing Campaigns: Develop targeted marketing campaigns aligned with the observed seasonal peaks to capitalize on increased customer engagement and purchase intent. Create promotions, discounts, and exclusive offers that resonate with customers during high-demand months such as January, March, and August. This can help drive sales and enhance customer loyalty.

3. Optimize Logistics and Delivery: Address the issues related to late or undelivered products by working closely with the logistics department. Improve coordination, streamline processes, and invest in efficient logistics systems to ensure timely and reliable deliveries. Implement tracking mechanisms to provide customers with real-time updates on their orders, enhancing their overall experience.

4. Focus on High-Value Regions: Prioritize efforts and allocate resources to states with larger customer bases, such as São Paulo, Rio de Janeiro, and Minas Gerais. Customize marketing strategies, product offerings, and customer support to cater to the specific preferences and needs of customers in these regions. This targeted approach can yield higher customer satisfaction and increased sales.

5. Evaluate Pricing and Shipping Strategies: Analyze mean_price and mean_freight_value variations across states to understand customer preferences and price sensitivity. Consider adjusting pricing strategies and exploring partnerships with logistics providers to optimize shipping costs. Additionally, offer competitive shipping options and transparent pricing to attract customers and improve their shopping experience.

6. Enhance Payment Options: Encourage the use of alternative payment methods such as UPI and credit cards by offering incentives and exclusive discounts. Highlight the convenience and security of these payment options to build customer trust and drive higher conversion rates. Collaborate with payment service providers to ensure a seamless and user-friendly checkout experience.

THANKYOU