Project Target-SQL

About Target:

Target is a globally renowned brand and a prominent retailer in the United States. 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 particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. Analyzing this business case helps us gain valuable insights into Target's operations in Brazil

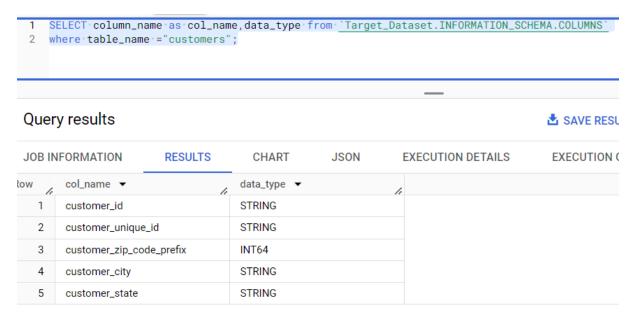
Dataset:

https://drive.google.com/drive/folders/1TGEc66YKbD443nslRi1bWgVd238gJCnb

Exploratory Analysis:

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

SELECT column_name as col_name,data_type from `Target_Dataset.INFORMATION_SCHEMA.COLUMNS` where table_name ="customers";



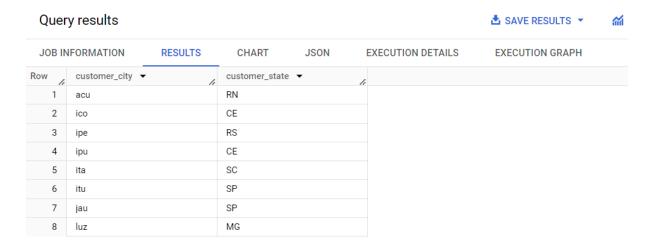
2) Get the time range between which the orders were placed.

select min(order_purchase_timestamp) as min_time ,max(order_purchase_timestamp) as max_time
from `Target Dataset.orders`;

JOB IN	IFORMATION	RESULTS	CHART	JSON	E	EXECUTION DETAILS	EXECUTION GRAPH
Row	min_time ▼	1.	max_time ▼		h		
1	2016-09-04 21:15	5:19 UTC	2018-10-17 17:3	30:18 UTC			

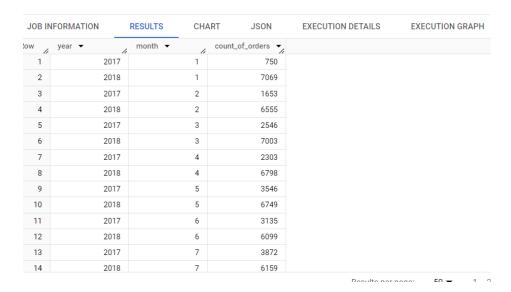
3) Count the Cities & States of customers who ordered during the given period

select c.customer_city,c.customer_state from `Target_Dataset.customers` c inner join `Target_Dataset.orders` o on c.customer_id=o.customer_id group by 1,2;



4) Is there a growing trend in the no. of orders placed over the past years?

select extract(year from order_purchase_timestamp) as year, extract(month from order_purchase_timestamp) as month, count(distinct(order_id)) as count_of_orders from `Target_Dataset.orders` where order_status="delivered" group by 1,2 order by 2,1;



Alternate way:

You can also find out the maximum orders for every year and every month where the sales were at peak.

```
create view Target_Dataset.seasonality as (select extract(year from order_purchase_timestamp ) as year, extract(month from order_purchase_timestamp) as month, count(distinct(order_id)) as count_of_orders from `Target_Dataset.orders` where order_status="delivered" group by 1,2 order by 2,1); select year,month,max(count_of_orders) as max_orders from Target_Dataset.seasonality group by year,month order by 1,2;
```

JOB IN	IFORMATION	RESULTS CHA	ART JSON	EXECUTION DETAILS
Row /	year ▼	month ▼	max_orders ▼	
1	2016	9	1	
2	2016	10	265	
3	2016	12	1	
4	2017	1	750	
5	2017	2	1653	
6	2017	3	2546	
7	2017	4	2303	
0	2017	г	2546	

D - - - - IA - - - -

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

0-6 hrs : Dawn
7-12 hrs : Mornings
13-18 hrs : Afternoon
19-23 hrs : Night

select sum(case when hour_ between 0 and 6 then orders end) as Dawn, sum(case when hour_ between 7 and 12 then orders end) as Mornings, sum(case when hour_ between 13 and 18 then orders end)as Afternoon, sum(case when hour_ between 19 and 23 then orders end)as Night from (select extract(hour from order_purchase_timestamp) as hour_, count(distinct(order_id)) as orders from `Target_Dataset.orders` where order status="delivered" group by 1)



Evolution of E-commerce orders in the Brazil region:

1)Month on month no. of orders placed in each state.

```
select c.customer_state,extract(month from o.order_purchase_timestamp) as month, count(distinct(o.order_id)) as count_of_orders from`Target_Dataset.customers` c inner join `Target_Dataset.orders` o on c.customer_id=o.customer_id group by 1,2 order by 1,2;
```

JOB IN	FORMATION	RESULTS	CHART	JSON EXECUT	ION DETAILS
Row 9	customer_state AC	▼	month ▼	count_of_orders •	
10	AC		10	6	
11	AC		11	5	
12	AC		12	5	
13	AL		1	39	
14	AL		2	39	
15	AL		3	40	

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

select count(customer_unique_id) as customer_count,customer_state from `Target_Dataset.customers` group by 2 order by 2,1

JOB II	NFORMATION	RESULTS CHA
Row	customer_count 🔻	customer_state ▼
1	81	AC
2	413	AL
3	148	AM
4	68	AP
5	3380	BA
6	1336	CE
7	2140	DF
8	2033	ES
9	2020	GO
10	747	MA
11	11635	MG
12	715	MS

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)

with pay as

(select extract(year from o.order_purchase_timestamp)as year,round(sum(p.payment_value),0) as payment from `Target_Dataset.payments`p inner join `Target_Dataset.orders`o

```
on p.order_id=o.order_id
where extract(month from o.order_purchase_timestamp) between 1 and 8
and extract(year from o.order_purchase_timestamp) between 2017 and 2018
group by 1
),
lagger as(
select *,lag(payment,1) over(order by year asc) as lg from pay)
select *,((payment-lg)/lg)*100 as percent_increase from lagger
```

JTION DETAILS	EXECU	JSON	ART	CHA	RESULTS	N	IFORMATIO	JOB IN
ncrease 🔻	percent_ir	6	lg ▼	▼	payment 🔻	h	year ▼	Row
null		null		69022.0	3669	2017		1
837581	136.97688	3669022.0		94734.0	8694	2018		2

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

select c.customer_state,round(sum(oi.price),1) as total_sum,round(avg(oi.price),1) as avg_price from

`Target_Dataset.customers` c inner join `Target_Dataset.orders`o on c.customer_id=o.customer_id inner join

`Target_Dataset.order_items` oi on o.order_id=oi.order_id group by c.customer state order by 1,2

Row	customer_state ▼	total_sum ▼	avg_price ▼
1	AC	15982.9	173.7
2	AL	80314.8	180.9
3	AM	22356.8	135.5
4	AP	13474.3	164.3
5	BA	511350.0	134.6
6	CE	227254.7	153.8
7	DF	302603.9	125.8
8	ES	275037.3	121.9

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

select c.customer_state,round(sum(oi.freight_value),1) as total_sum,round(avg(oi.freight_value),1) as avg_price from

`Target_Dataset.customers` c inner join `Target_Dataset.orders`o on c.customer_id=o.customer_id inner join `Target_Dataset.order_items` oi on o.order_id=oi.order_id group by c.customer_state order by 1,2

Row	customer_state ▼	total_sum ▼	avg_price ▼
1	AC	3686.7	40.1
2	AL	15914.6	35.8
3	AM	5478.9	33.2
4	AP	2788.5	34.0
5	BA	100156.7	26.4
6	CE	48351.6	32.7
7	DF	50625.5	21.0

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.

SELECT

order id,

DATE_DIFF(order_delivered_customer_date, TIMESTAMP_TRUNC(order_purchase_timestamp, DAY), DAY) AS time_to_deliver,

DATE_DIFF(order_delivered_customer_date,order_estimated_delivery_date,day) AS diff_estimated_delivery FROM Target Dataset.orders`;

Row	order_id ▼	time_to_deliver ▼	diff_estimated_delive
1	1950d777989f6a877539f5379	30	12
2	2c45c33d2f9cb8ff8b1c86cc28	31	-28
3	65d1e226dfaeb8cdc42f66542	36	-16
4	635c894d068ac37e6e03dc54e	31	-1
5	3b97562c3aee8bdedcb5c2e45	33	0
6	68f47f50f04c4cb6774570cfde	30	-1
7	276e9ec344d3bf029ff83a161c	44	4
8	54e1a3c2b97fb0809da548a59	41	4

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

```
select c.customer_state,avg(freight_value) as max_freight
from `Target_Dataset.customers` c inner join `Target_Dataset.orders` o on
c.customer_id=o.customer_id inner join `Target_Dataset.order_items` oi on
o.order_id=oi.order_id group by c.customer_state order by avg(freight_value) desc
limit 5

select c.customer_state,avg(freight_value) as min_freight
from `Target_Dataset.customers` c inner join `Target_Dataset.orders` o on
```

c.customer_id=o.customer_id inner join `Target_Dataset.order_items` oi on o.order_id=oi.order_id group by c.customer_state order by avg(freight_value) asc limit 5

JOB INFORMATION		RESULTS	CHART	JSON
Row	customer_state -		max_freight ▼	1
1	RR		42.98442307692	
2	PB		42.72380398671	
3	RO		41.06971223021	
4	AC		40.07336956521	
5	PI		39.14797047970	

Row	customer_state ▼	min_freight ▼
1	SP	15.14727539041
2	PR	20.53165156794
3	MG	20.63016680630
4	RJ	20.96092393168
5	DF	21.04135494596

Analysis based on the payments:

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

select extract(month from o.order_purchase_timestamp) as month,p.payment_type, count(distinct(o.order_id)) as count_of_orders from`Target_Dataset.payments` p inner join `Target_Dataset.orders` o on p.order_id=o.order_id group by 1,2 order by 1,2;

Row	month ▼	payment_type ▼	count_of_orders 🕶
1	1	UPI	1715
2	1	credit_card	6093
3	1	debit_card	118
4	1	voucher	337
5	2	UPI	1723
6	2	credit_card	6582
7	2	debit_card	82
8	2	voucher	288
9	3	UPI	1942
10	3	credit_card	7682

Observations made:

- 1)Target had improved its sales from the year 2016 to 2017,however there was a slight decrease in its performance in the year 2018.
- 2)Brazilian customers are more active during afternoon as they place huge number of orders. So improving marketing strategies, offers during these hours will have a positive effect on the company's performance.
- 3)It is observed that target takes a lot of time to deliver products to its customers. Improving its delivery time will increase its customer base as they would prefer target over any other retail store with more delivery time.
- 4) It is also observed that credit card sales/payments play a major role .So connecting with more credit card companies and providing credit card offers/options have a higher chance of increasing its sales.