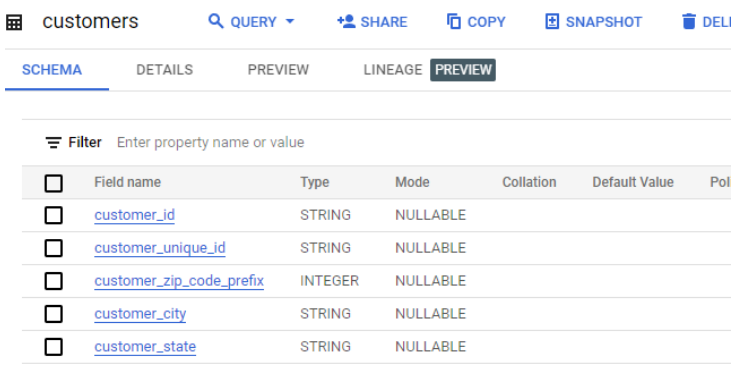
BUSINESS CASE

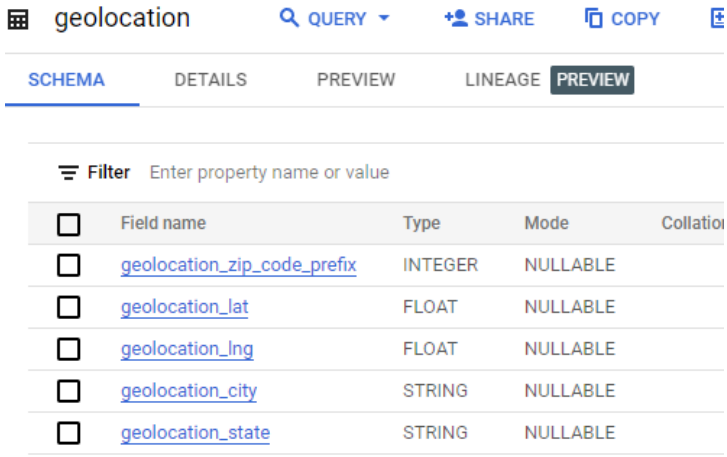
TARGET SQL

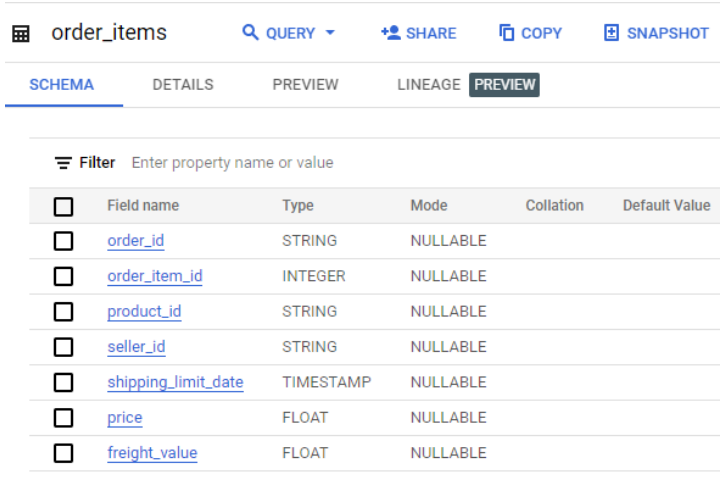
Submitted By: Rahul Kumar

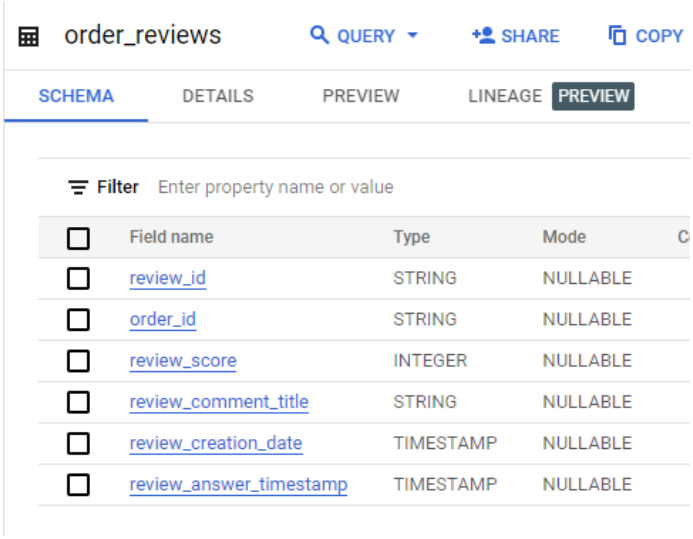
October Intermediate - 22

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









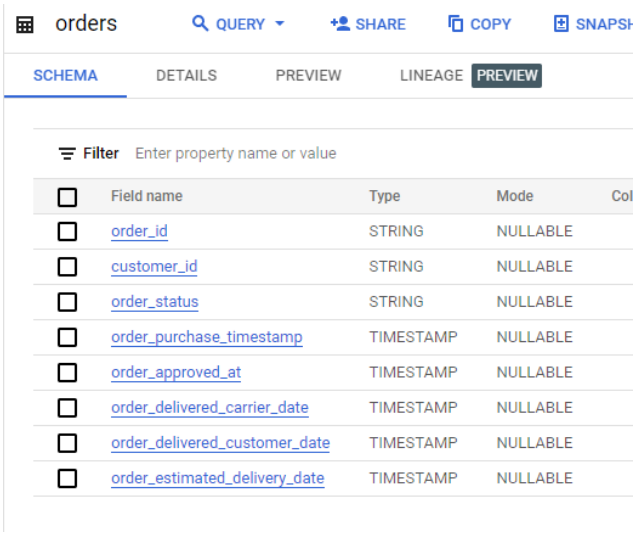
order_reviews

QUERY SHARE COPY

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Filter Enter property name or value

Field name	Type	Mode	C
review_id	STRING	NULLABLE	
order_id	STRING	NULLABLE	
review_score	INTEGER	NULLABLE	
review_comment_title	STRING	NULLABLE	
review_creation_date	TIMESTAMP	NULLABLE	
review_answer_timestamp	TIMESTAMP	NULLABLE	



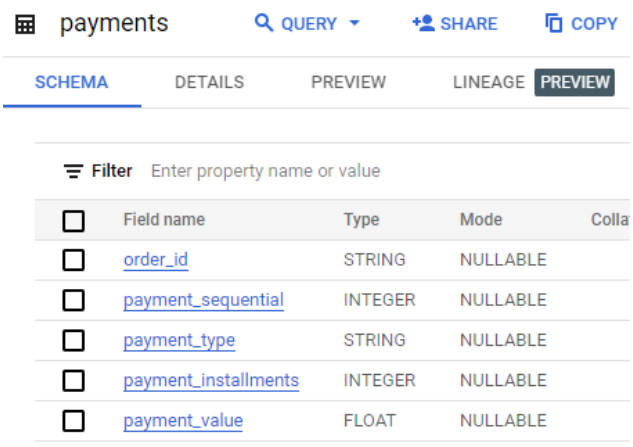
orders

QUERY SHARE COPY SNAPSH

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Filter Enter property name or value

Field name	Type	Mode	Col
order_id	STRING	NULLABLE	
customer_id	STRING	NULLABLE	
order_status	STRING	NULLABLE	
order_purchase_timestamp	TIMESTAMP	NULLABLE	
order_approved_at	TIMESTAMP	NULLABLE	
order_delivered_carrier_date	TIMESTAMP	NULLABLE	
order_delivered_customer_date	TIMESTAMP	NULLABLE	
order_estimated_delivery_date	TIMESTAMP	NULLABLE	



payments

QUERY SHARE COPY

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Filter Enter property name or value

Field name	Type	Mode	Colla
order_id	STRING	NULLABLE	
payment_sequential	INTEGER	NULLABLE	
payment_type	STRING	NULLABLE	
payment_installments	INTEGER	NULLABLE	
payment_value	FLOAT	NULLABLE	

products QUERY SHARE COPY SNAPSHOT

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Filter Enter property name or value

Field name	Type	Mode	Collation	Default
product_id	STRING	NULLABLE		
product_category	STRING	NULLABLE		
product_name_length	INTEGER	NULLABLE		
product_description_length	INTEGER	NULLABLE		
product_photos_qty	INTEGER	NULLABLE		
product_weight_g	INTEGER	NULLABLE		
product_length_cm	INTEGER	NULLABLE		
product_height_cm	INTEGER	NULLABLE		
product_width_cm	INTEGER	NULLABLE		

sellers QUERY SHARE COPY S

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Filter Enter property name or value

Field name	Type	Mode	Collation
seller_id	STRING	NULLABLE	
seller_zip_code_prefix	INTEGER	NULLABLE	
seller_city	STRING	NULLABLE	
seller_state	STRING	NULLABLE	

2. Time period for which the data is given

```
select
min(order_purchase_timestamp) as first_purchase_date,
max(order_purchase_timestamp) as last_purchase_date
from target_data.orders;
```

Query results

JOB INFORMATION RESULTS JSON EXECUTION DETAILS EXECUTION GRAPH PREVIEW

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

3. Cities and States of customers ordered during the given period

```
select distinct customer_city, customer_state
```

from target_data.customers; select distinct customer_city, customer_state
from target_data.customers;

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	customer_city	customer_state
17	buri	SP
18	cacu	GO
19	caem	BA
20	catu	BA
21	chui	RS
22	cipo	BA
23	codo	MA
24	cruz	CE
25	embu	SP
26	iacu	BA
27	iapu	MG
28	iati	PE
29	ibia	MG
30	icem	SP
31	ichu	BA
32	iepe	SP
33	ijui	RS
34	imbe	RS

Load more

1. In-depth Exploration:
    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?

```
select distinct extract(month from order_purchase_timestamp) as month,
extract(year from order_purchase_timestamp) as year,
count (order_id) as order_per_month
from target_data.orders
group by month, year
order by year, month;
```

Row	month	year	order_per_month
1	9	2016	4
2	10	2016	324
3	12	2016	1
4	1	2017	800
5	2	2017	1780
6	3	2017	2682
7	4	2017	2404
8	5	2017	3700
9	6	2017	3245
10	7	2017	4026
11	8	2017	4331
12	9	2017	4285
13	10	2017	4631
14	11	2017	7544
15	12	2017	5673
16	1	2018	7269
17	2	2018	6728
18	3	2018	7211
19	4	2018	6939

From the deried result we can conclude that there is a growing trend of e-commerce in brazil with peak seaonality during end months and beginning months on an year.

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

```
select distinct extract(hour from order_purchase_timestamp) as hour,
count (order_id) as order_per_month
from target_data.orders
group by hour
order by hour;
```

Query results

JOB INFORMATION | RESULTS

Row	hour	order_per_month
1	0	2394
2	1	1170
3	2	510
4	3	272
5	4	206
6	5	188
7	6	502
8	7	1231
9	8	2967
10	9	4785

Query results

JOB INFORMATION | RESULTS

Row	hour	order_per_month
11	10	6177
12	11	6578
13	12	5995
14	13	6518
15	14	6569
16	15	6454
17	16	6675
18	17	6150
19	18	5769
20	19	5982

Query results

JOB INFORMATION | RESULTS | JSON

Row	hour	order_per_month
21	20	6193
22	21	6217
23	22	5816
24	23	4123

From the derived results we can conclude that Brazilians customers are more active in the evening and night hours and the footfall decreases in the late night and early morning timings.

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

1. Get month on month orders by states

```
select  customer_state, extract(month from order_purchase_timestamp) as month,
extract(year from order_purchase_timestamp) as year,
count(order_id) as total_no_of_orders
from target_data.customers as a
join target_data.orders as b
on  a.customer_id = b.customer_id
group by customer_state, month, year
order by customer_state, year, month;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTI

Row	customer_state	month	year	total_no_of_orde
1	AC	1	2017	2
2	AC	2	2017	3
3	AC	3	2017	2
4	AC	4	2017	5
5	AC	5	2017	8
6	AC	6	2017	4
7	AC	7	2017	5
8	AC	8	2017	4
9	AC	9	2017	5
10	AC	10	2017	6
11	AC	11	2017	5
12	AC	12	2017	5
13	AC	1	2018	6
14	AC	2	2018	3
15	AC	3	2018	2
16	AC	4	2018	4
17	AC	5	2018	2
18	AC	6	2018	3
19	AC	7	2018	4
20	AC	8	2018	3

2. Distribution of customers across the states in Brazil

```
select customer_state, count(distinct customer_id) as statewise_customers
from target_data.customers
group by customer_state;
```

Query results

JOB INFORMATION | RESULTS | JSON | EX

Row	customer_state	statewise_custo
1	RN	485
2	CE	1336
3	RS	5466
4	SC	3637
5	SP	41746
6	MG	11635
7	BA	3380
8	RJ	12852

Load more

4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.
    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
    2. Mean & Sum of price and freight value by customer state

```
select customer_state, sum(freight_value) as sum_fv,
sum(price) as sum_price, avg(freight_value) as avg_fv, avg(price) as avg_price
from target_data.customers as a
join target_data.orders as b
on a.customer_id = b.customer_id
join target_data.order_items as c
on b.order_id = c.order_id
group by customer_state;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	customer_state	sum_fv	sum_price	avg_fv	avg_price
1	MT	29715.4300...	156453.529...	28.1662843...	148.297184...
2	MA	31523.7700...	119648.219...	38.2570024...	145.204150...
3	AL	15914.5899...	80314.81	35.8436711...	180.889211...
4	SP	718723.069...	5202955.05...	15.1472753...	109.653629...
5	MG	270853.460...	1585308.02...	20.6301668...	120.748574...
6	PE	59449.6599...	262788.029...	32.9178626...	145.508322...
7	RJ	305589.310...	1824092.66...	20.9609239...	125.117818...
8	DF	50625.4999...	302603.939...	21.0413549...	125.770548...
9	RS	135522.740...	750304.020...	21.7358043...	120.337453...
10	SE	14111.4699...	58920.8500...	36.6531688...	153.041168...
11	PR	117851.680...	683083.760...	20.5316515...	119.004139...
12	PA	38699.3000...	178947.809...	35.8326851...	165.692416...
13	BA	100156.679...	511349.990...	26.3639589...	134.601208...
14	CE	48351.5899...	227254.709...	32.7142016...	153.758261...
15	GO	53114.9799...	294591.949...	22.7668152...	126.271731...
16	ES	49764.5999...	275037.309...	22.0587765...	121.913701...
17	SC	89660.2600...	520553.340...	21.4703687...	124.653577...
18	PI	21218.2000...	86914.0800...	39.1479704...	160.358081...
19	PB	25719.7300...	115268.079...	42.7238039...	191.475215...
20	RN	18860.1000...	83034.9800...	35.6523629...	156.965935...

5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery

```
select date_diff(order_delivered_customer_date, order_purchase_timestamp, day) as paurchase_to_delivery,
date_diff(order_delivered_customer_date, order_estimated_delivery_date, day) as actual_to_expected
from target_data.orders;
```

Query results

JOB INFORMATION | RESULTS | JSON | E

Row	paurchase_to_delivery	actual_to_expected
1	30	12
2	30	-28
3	35	-16
4	30	-1
5	32	0
6	29	-1
7	43	4
8	40	4
9	37	1
10	33	5
11	38	6
12	36	2
13	34	0
14	42	11
15	35	3
16	32	7
17	31	9
18	29	0
19	30	0
20	30	8

2. Find time_to_delivery & diff_estimated_delivery.

```
select customer_state,
date_diff(order_delivered_customer_date, order_purchase_timestamp, day) over(partition by) as
time_to_delivery,
date_diff(order_delivered_customer_date, order_estimated_delivery_date, day) as diff_estimated_delivery
from target_data.customers as a
inner join target_data.orders as b
on a.customer_id = b.customer_id
inner join target_data.order_items as c
on b.order_id = c.order_id
group by customer_state;
```

4. Sort the data to get the following:
5. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

```
select customer_state, avg(freight_value) as avg_freight
from target_data.customers as a
inner join target_data.orders as b
```

```
on a.customer_id = b.customer_id
inner join target_data.order_items as c
on b.order_id = c.order_id
group by customer_state
order by avg_freight asc
limit 5;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DE

Row	customer_state	avg_freight
1	SP	15.1472753...
2	PR	20.5316515...
3	MG	20.6301668...
4	RJ	20.9609239...
5	DF	21.0413549...

6. Payment type analysis:

1. Month over Month count of orders for different payment types

```
select extract(month from order_purchase_timestamp) as order_month,
extract(year from order_purchase_timestamp) as order_year,
payment_type,
count(a.order_id) as total_order
from target_data.orders as a
inner join target_data.payments as b
on a.order_id = b.order_id
group by payment_type, order_month, order_year
order by order_year, order_month;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	order_month	order_year	payment_type	total_order
1	9	2016	credit_card	3
2	10	2016	credit_card	254
3	10	2016	voucher	23
4	10	2016	debit_card	2
5	10	2016	UPI	63
6	12	2016	credit_card	1
7	1	2017	voucher	61
8	1	2017	UPI	197
9	1	2017	credit_card	583
10	1	2017	debit_card	9
11	2	2017	credit_card	1356
12	2	2017	voucher	119
13	2	2017	UPI	398
14	2	2017	debit_card	13

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

```
select payment_installments,
count(order_id) as total_orders
from target_data.payments
group by payment_installments;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAI

Row	payment_installments	total_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
10	9	644
11	10	5328
12	11	23
13	12	133
14	13	16
15	14	15
16	15	74
17	16	5
18	17	8
19	18	27

Actionable Insights

The performed analysis on the dataset “Target SQL” gives various observation worthy details on the case of 100k orders from 2016 to 2018 made at Target in Brazil.

- The growth of e-commerce is up hill in Brazil and is showing a good increase every year.
- Sales tend to increase in the festive season of year-end and is continued till New Year.
- All states in Brazil show good amount of orders, so we can say Target is performing well sales-wise pan Brazil.
- Customers tend to buy more on credit than upfront pay using debit-card, UPI or cash.
- Most of the credits have few no. of installments ranging from mostly from 1 to 6 with a few outliers up to 10.

Recommendations

As a data analyst some of the prominent recommendations to Target may be:

- While to growth in customers is going up, it is always good to increase customer acquisition.
- Taking measures to accelerate sales in the mid-year/ off-season.
- Payments made on vouchers should be promoted to increase customer acquisition.
- Few states are showing lesser number of customers, measures to boost sales in those regions should be taken.
- Cost of delivery can be reduced.
- The gap between estimated delivery and actual delivery can be bridged to a better number.