Business Case: Target SQL

**Note that in the output in which there were more rows , I have used the limit function to take photos of only 10 rows.

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

A. Data type of all columns in the “customers” table.

Ans-
```
SELECT *,data_type
FROM `scaler-dsml-sql93.Buisness_case1.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers';
```

Query results

SAVE RE

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GR/

Row	column_name	ordinal_position	is_nullable	data_type
1	customer_id	1	YES	STRING
2	customer_unique_id	2	YES	STRING
3	customer_zip_code_prefix	3	YES	INT64
4	customer_city	4	YES	STRING
5	customer_state	5	YES	STRING

Insights- this table mainly consist of location of customers, only zip code is integer other are string values

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

```
select min(order_purchase_timestamp)as min_d,max(order_purchase_timestamp)as max_d
from `Buisness_case1.orders`
```

Query results

SAVE RESULTS

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

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

Insights- the company first order was placed on 4th of september 2016 and last order was placed on 17th of october 2018

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

```
select count(distinct customer_city) as city,count(distinct customer_state) as
state from `Buisness_case1.orders` o join `Buisness_case1.customers` c
```

```
on o.customer_id=c.customer_id;
```

Query results

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

Row	city	state
1	4119	27

Insights- this company has customer in 4119 cities and 27 state.

Q 2)In-depth Exploration:

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

Ans-

```
SELECT * FROM
    (select EXTRACT(YEAR FROM order_purchase_timestamp) AS year_start ,
    EXTRACT(MONTH FROM order_purchase_timestamp) AS ord_month,
    COUNT(order_purchase_timestamp) AS no_of_orders from `Buisness_case1.orders`
    GROUP BY 1,2
    )TT
    ORDER BY 1,2
    limit 10;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION G

Row	year_start	ord_month	no_of_orders
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

Insights- the company order started from 2016 in month of november and significant increase in no of orders have increased in every month ,but in 2016 december there have been a sudden fall may be due to app issue,any bug or any new competitor in the market.

B)Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

```
Ans- select extract(YEAR FROM order_purchase_timestamp ) as YEAR ,extract(month from
order_purchase_timestamp)as month ,COUNT(order_purchase_timestamp) AS no_of_orders
from `Buisness_case1.orders`
 GROUP BY YEAR,month
 order by year,month asc
 limit 10;
```

Query results

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

Row	YEAR	month	no_of_orders
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

Insights- the company order started from 2016 in month of november and it has started at good pace then has less order in month of december may be due to season or app issue but from january,february & may 2017 the company came back strongly with increase in orders may be it has also applied some offers for new customer.

Q C)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
```

```
Ans- SELECT
CASE
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 0 AND EXTRACT(HOUR FROM
    order_purchase_timestamp) < 7 THEN 'Dawn'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 7 AND EXTRACT(HOUR FROM
    order_purchase_timestamp) < 13 THEN 'Morning'
```

```
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 13 AND EXTRACT(HOUR FROM
    order_purchase_timestamp) < 19 THEN 'Afternoon'
ELSE 'Night'
END AS time_of_day,
COUNT(*) AS order_count
FROM
`Buisness_case1.orders`
GROUP BY
time_of_day
ORDER BY
Time_of_day;
```

Query results

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

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

Insights- As from output we can clearly see that the maximum order were placed in afternoon ,least were placed at dawn and morning and night time the order count is almost same.

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

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

Ans-

```
select extract(month from order_purchase_timestamp)as month,c.customer_state
,COUNT(order_purchase_timestamp) AS no_of_orders
from `Buisness_case1.orders` o join `Buisness_case1.customers` c
on o.customer_id=c.customer_id
GROUP BY 1,2
order by 1,3 desc
limit 10;
```

Query results

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

Row	month	customer_state	no_of_orders
1	1	SP	3351
2	1	RJ	990
3	1	MG	971
4	1	PR	443
5	1	RS	427
6	1	SC	345
7	1	BA	264
8	1	GO	164
9	1	ES	159
10	1	DF	151

Insights- As we can see from the table that maxm order have came from the SP-State
May be they have given good offers or their head quarters is located in that state or delivery time was time was also good thats why more no of orders were placed.

Q B). How are the customers distributed across all the states?

```
ANS-select count(distinct customer_id)as no_of_cudtomers, customer_state from
`Buisness_case1.customers`
  group by 2
  order by 1 desc
  limit 10
```

Query results

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

Row	no_of_cudtomers	customer_state
1	41746	SP
2	12852	RJ
3	11635	MG
4	5466	RS
5	5045	PR
6	3637	SC
7	3380	BA
8	2140	DF
9	2033	ES
10	2020	GO

Insights- As we can see from the table that maxm customers are from the SP-State May be they have given good offers or their head quarters is located in that state.

Q IV). Impact on Economy: Analyse the money movement by e-commerce by looking at order prices, freight and others.

A. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).

Ans-

```
WITH YearlyCost AS (
SELECT
EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
SUM(p.payment_value) AS total_cost
FROM
`Buisness_case1.payments` p join `Buisness_case1.orders`o
on p.order_id=o.order_id
WHERE
EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017, 2018)
AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
GROUP BY
EXTRACT(YEAR FROM o.order_purchase_timestamp),
EXTRACT(MONTH FROM o.order_purchase_timestamp)
)

SELECT
2018 AS year,
2017 AS previous_year,
SUM(CASE WHEN order_year = 2018 THEN total_cost ELSE 0 END) AS current_year_cost,
SUM(CASE WHEN order_year = 2017 THEN total_cost ELSE 0 END) AS previous_year_cost,
(SUM(CASE WHEN order_year = 2018 THEN total_cost ELSE 0 END) - SUM(CASE WHEN
order_year = 2017 THEN total_cost ELSE 0 END))
/ SUM(CASE WHEN order_year = 2017 THEN total_cost ELSE 0 END) * 100 AS
percentage_increase
FROM
YearlyCost
WHERE
order_year IN (2017, 2018);
```

Query results

SAVE RESULTS EXPLORE DATA

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

Row	year	previous_year	current_year_cost	previous_year_cost	percentage_increase
1	2018	2017	8694733.839999...	3669022.120000...	136.9768716466...

Insights- there has been 136.97%(approx 137) increase in the price from 2017 to 2018.

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

Ans-

```
SELECT c.customer_state,sum(ord.price)as Total,avg(ord.price) as Avg_price
    from `Buisness_case1.customers` c join `Buisness_case1.orders` o
    on c.customer_id=o.customer_id
    join `Buisness_case1.order_items` ord
    on o.order_id=ord.order_id
    group by 1
    order by 2 desc,3 desc
    limit 10;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EX

Row	customer_state	Total	Avg_price
1	SP	5202955.050001...	109.6536291597...
2	RJ	1824092.669999...	125.1178180945...
3	MG	1585308.029999...	120.7485741488...
4	RS	750304.0200000...	120.3374530874...
5	PR	683083.7600000...	119.0041393728...
6	SC	520553.3400000...	124.6535775862...
7	BA	511349.9900000...	134.6012082126...
8	DF	302603.9399999...	125.7705486284...
9	GO	294591.9499999...	126.2717316759...
10	ES	275037.3099999...	121.9137012411...

Insights- Total price for SP State is very much high compared to other State so it more no of products, as well as avg price is also less compare to others so SP may be headquarter of store

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

Ans-
```
SELECT c.customer_state,sum(fr.freight_value)as Total,avg(fr.freight_value) as
Avg_price
    from `Buisness_case1.customers` c join `Buisness_case1.orders` o
    on c.customer_id=o.customer_id
    join `Buisness_case1.order_items` fr
    on o.order_id=fr.order_id
    group by 1
    limit 10;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | E

Row	customer_state	Total	Avg_price
1	MT	29715.43000000...	28.16628436018...
2	MA	31523.77000000...	38.25700242718...
3	AL	15914.58999999...	35.84367117117...
4	SP	718723.0699999...	15.14727539041...
5	MG	270853.4600000...	20.63016680630...
6	PE	59449.65999999...	32.91786267995...
7	RJ	305589.3100000...	20.96092393168...
8	DF	50625.49999999...	21.04135494596...
9	RS	135522.7400000...	21.73580433039...
10	SE	14111.46999999...	36.65316883116...

Insights- the maxm freight charges were charged in RN State and least were charged in MA State.the least avg freight charges were in SP State. The head quarter is in that state.

Q V). Analysis based on sales, freight and delivery time.

-- A. 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.

Ans-
```
select
    order_id,customer_id,DATE_DIFF(order_delivered_customer_date,order_purchase_timesta
    mp, day)as delivery_time ,
DATE_DIFF(order_delivered_customer_date,order_estimated_delivery_date , day)as
    delay_in_days
from `Buisness_case1.orders`
order by 3 desc,4 desc
```

```
limit 10
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRA

Row	order_id	customer_id	delivery_time	delay_in_days
1	ca07593549f1816d26a572e06...	75683a92331068e2d281b11a...	209	181
2	1b3190b2dfa9d789e1f14c05b...	d306426abe5fca15e54b645e4...	208	188
3	440d0d17af552815d15a9e41a...	7815125148cfa1e8c7fee1ff79...	195	165
4	285ab9426d6982034523a855f...	9cf2c3fa2632cee748e1a59ca9...	194	166
5	0f4519c5f1c541ddec9f21b3bd...	1a8a4a30dc296976717f44e78...	194	161
6	2fb597c2f772eca01b1f5c561b...	217906bc11a32c1e470eb7e08...	194	155
7	47b40429ed8cce3aee9199792...	cb2caaaead400c97350c37a3f...	191	175
8	2fe324febf907e3ea3f2aa9650...	65b14237885b3972ebec28c0f...	189	167
9	2d7561026d542c8dbd8f0daea...	8199345f57c6d1cbe9701f924...	188	159
10	c27815f7e3dd0b926b5855262...	f85e9ec0719b16dc4dd0edd43...	187	162

Insights- **the maxm delivery time taken is 209 days , maybe the product was not available at that time and also there has been a delay of 181 days, the product took more time than expected time.**

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

Ans-

```
(select c.customer_state,avg(ord.freight_value) as freight_cost from
`Buisness_case1.order_items`ord
join `Buisness_case1.orders`o
on ord.order_id=o.order_id
join `Buisness_case1.customers`c
on o.customer_id=c.customer_id
group by 1
order by 2 desc limit 5)
union all
(select c.customer_state,avg(ord.freight_value) as freight_cost from
`Buisness_case1.order_items`ord
join `Buisness_case1.orders`o
on ord.order_id=o.order_id
join `Buisness_case1.customers`c
on o.customer_id=c.customer_id
group by 1
order by 2 limit 5);
```

Query results

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

Row	customer_state	freight_cost
1	RR	42.98442307692...
2	PB	42.72380398671...
3	RO	41.06971223021...
4	AC	40.07336956521...
5	PI	39.14797047970...
6	SP	15.14727539041...
7	PR	20.53165156794...
8	MG	20.63016680630...
9	RJ	20.96092393168...
10	DF	21.04135494596...

Insights- the top 5 rows show the maximum freight cost (may be farthest from store) and bottom 5 show the minimum freight charges for different states(maybe nearer to store).

Q C). Find out the top 5 states with the highest & lowest average delivery time.

Ans-

```
(select
    c.customer_state,avg(DATE_DIFF(o.order_delivered_customer_date,o.order_purchase_tim
    estamp, day))as delivery_time from `Buisness_case1.orders`o
    join `Buisness_case1.customers`c
    on o.customer_id=c.customer_id
    group by 1
    order by 2 desc limit 5)
    union all
   (select
       c.customer_state,avg(DATE_DIFF(o.order_delivered_customer_date,o.order_purchase_
       timestamp, day))as delivery_time from
     `Buisness_case1.orders`o join `Buisness_case1.customers`c
     on o.customer_id=c.customer_id
     group by 1
     order by 2 limit 5);
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | E

Row	customer_state	delivery_time
1	RR	28.97560975609…
2	AP	26.73134328358…
3	AM	25.98620689655…
4	AL	24.04030226700…
5	PA	23.31606765327…
6	SP	8.298061489072…
7	PR	11.52671135486…
8	MG	11.54381329810…
9	DF	12.50913461538…
10	SC	14.47956019171…

Insights-the top 5 rows show the maximum delivery time for different states(may be farthest from store), and bottom 5 show the minimum delivery time(maybe nearer to store)

Q D). Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

Ans-

```
Select c.customer_state,avg(DATE_DIFF(order_estimated_delivery_date
,order_delivered_customer_date, day))as delivery_time from `Buisness_case1.orders`o
 join `Buisness_case1.customers`c
 on o.customer_id=c.customer_id
 group by 1
 order by 2 desc limit 5;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | E

Row	customer_state	delivery_time
1	AC	19.7625
2	RO	19.13168724279…
3	AP	18.73134328358…
4	AM	18.60689655172…
5	RR	16.41463414634…

Insights- the fastest order delivered in AC State (19 day before estimated delivery). The delivery agent are really good in that state or product availability is also good.

Q VI). Analysis based on the payments:

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

Ans-

```
select extract(month from order_purchase_timestamp)as month,p.payment_type
,COUNT(order_purchase_timestamp) AS no_of_orders
 from `Buisness_case1.orders` o join `Buisness_case1.payments` p
 on o.order_id=p.order_id
 GROUP BY 1,2
 order by 1,3 desc
 limit 10;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | E

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

Insights- the maximum payment was done through credit card and 2nd UPI.and also increase in payment through credit card and UPI, because apart from these two other payments have decreased, maybe they have applied some offer on credit card and UPI payments.

Q B). Find the no. of orders placed on the basis of the payment instalments that have been paid.

Ans-

```
select p.payment_installments
,COUNT(order_purchase_timestamp) AS no_of_orders
from `Buisness_case1.orders` o join `Buisness_case1.payments` p
on o.order_id=p.order_id
where p.payment_installments !=0
GROUP BY 1
order by 1,2 desc
limit 10;
```

Query results

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

Row	payment_installment	no_of_orders
1	1	52546
2	2	12413
3	3	10461
4	4	7098
5	5	5239
6	6	3920
7	7	1626
8	8	4268
9	9	644
10	10	5328

Insights- payment 1 has been paid by maximum customers

NOTE - COMPANY SHOULD WORK ON DELIVERY FACILITY AS IT IS TAKING VERY MUCH TIME TO DELIVER THE PRODUCT, THEY GOOD NO OF CUSTOMER ALSO THE ONLY DRAWBACK I FEEL WAS DELIVERY TIME AND OTHER PRICE ESCALATION BETWEEN 2017 TO 2018.