CASE STUDY OF TARGET

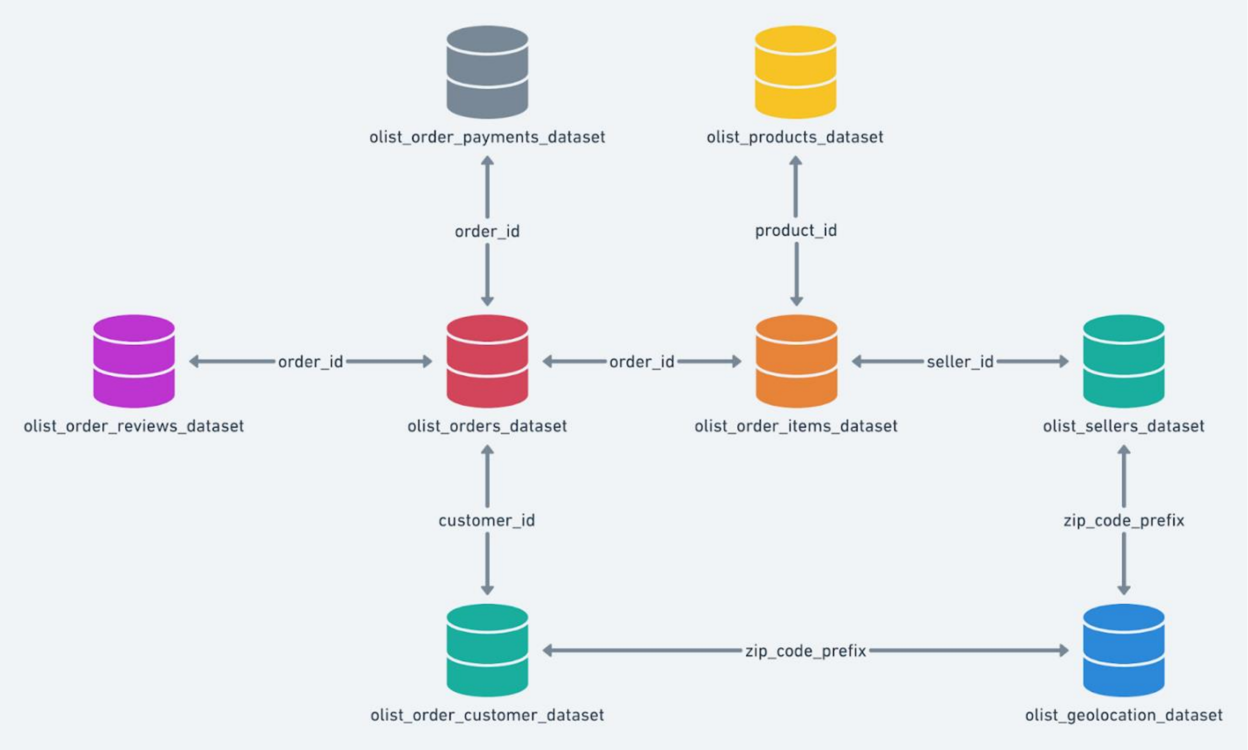
Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

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. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

Problem Statement:

I have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.



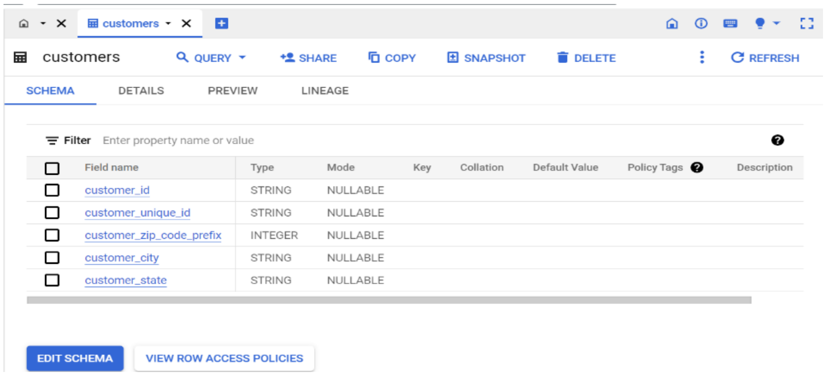
Above is the schema of the dataset of target for the reference.

A) **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**

1. Data type of all columns in the "customers" table.
2. Get the time range between which the orders were placed.
3. Count the Cities & States of customers who ordered during the given period.

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

   Answers : Data type has been mentioned in-front of the columns accordingly in the type column.



customers

SCHEMA DETAILS PREVIEW LINEAGE

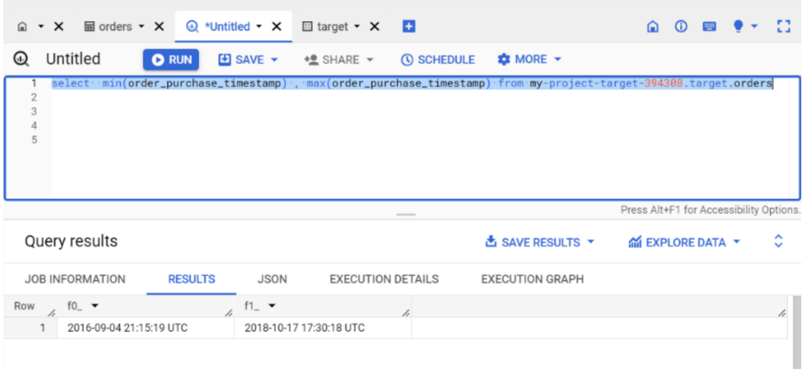
☐	Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
☐	customer_id	STRING	NULLABLE					
☐	customer_unique_id	STRING	NULLABLE					
☐	customer_zip_code_prefix	INTEGER	NULLABLE					
☐	customer_city	STRING	NULLABLE					
☐	customer_state	STRING	NULLABLE					

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

   Answer -

   Query : SELECT min(order_purchase_timestamp), max(order_purchase_timestamp)

   FROM `my-project-target-394308.target.orders`



```
select  min(order_purchase_timestamp) , max(order_purchase_timestamp) from my-project-target-394308.target.orders
```

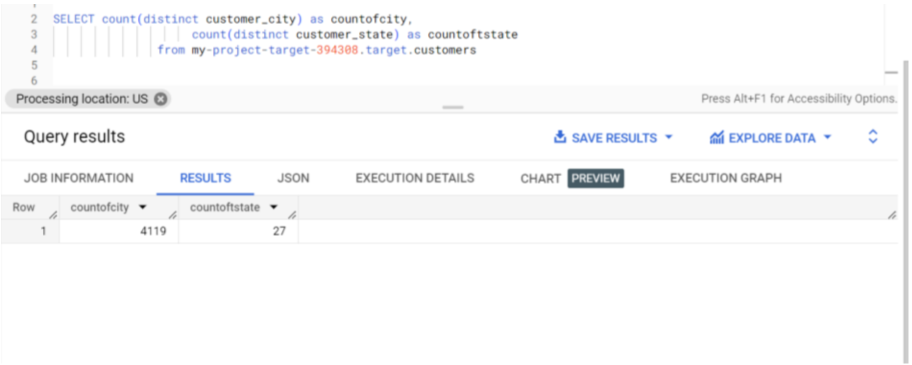
Query results

```
Row  f0_                      f1_
  1  2016-09-04 21:15:19 UTC  2018-10-17 17:30:18 UTC
```

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

Answer :

Query:
```
SELECT count(distinct customer_city) as countofcity,
    count(distinct customer_state) as countoftstate
  from my-project-target-394308.target.customers
```



```
SELECT count(distinct customer_city) as countofcity,
                  count(distinct customer_state) as countoftstate
                from my-project-target-394308.target.customers
```

Row	countofcity	countoftstate
1	4119	27

Analysis: There are 4119 cities and 27 states that we are catering all over.

B) In-depth Exploration:

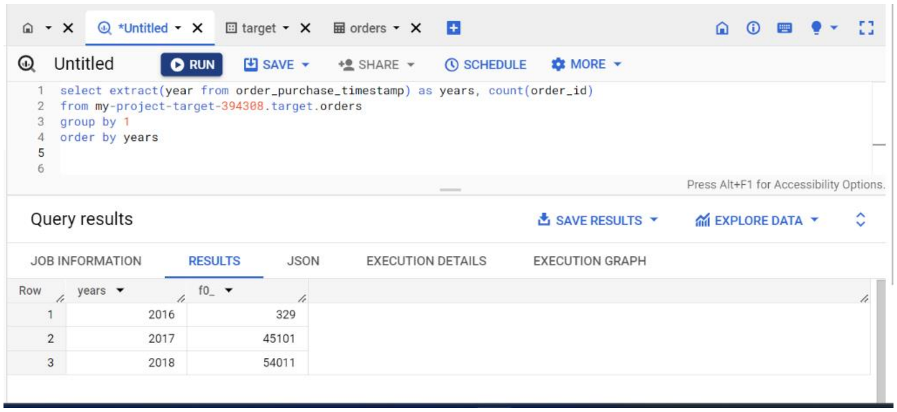
1. Is there a growing trend in the no. of orders placed over the past years?
2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
3. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
    - 0-6 hrs : Dawn
    - 7-12 hrs : Mornings
    - 13-18 hrs : Afternoon
    - 19-23 hrs : Night

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

Answer-

Query :

```
Select extract(year from order_purchase_timestamp) as years, Count(order_id)
from my-project-target-394308.target.orders
 Group by 1
Order by  years
```



```
select extract(year from order_purchase_timestamp) as years, count(order_id)
from my-project-target-394308.target.orders
group by 1
order by years
```

Row	years	f0_
1	2016	329
2	2017	45101
3	2018	54011

Analysis: There is surely a growing trend in the past years , but we cannot compare it with 2016 as it only had data of very few months and not the whole 2016 year.

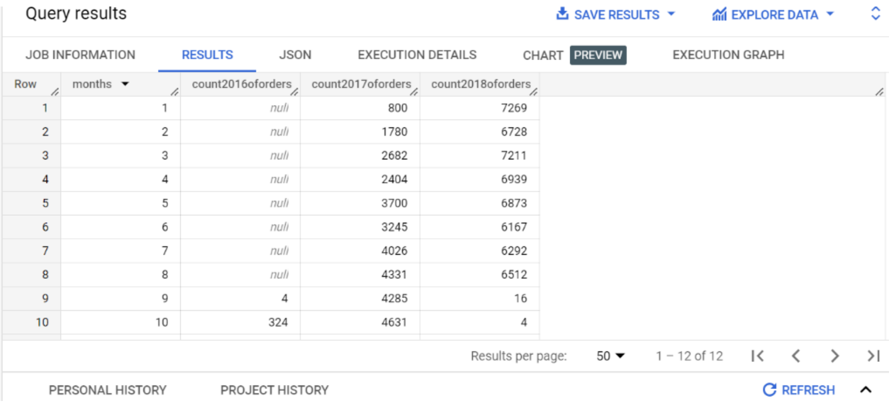
2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

   Answer-

Query:

```
With year2016 as
(select extract (year from order_purchase_timestamp) as years,
                  extract (month from order_purchase_timestamp) as months,
                  count(order_id) as count2016oforders
           from my-project-target-394308.target.orders
           where extract (year from order_purchase_timestamp) = 2016
             group by months , years

             order by  years , months),


 year2017 as
(select extract (year from order_purchase_timestamp) as years,
                  extract (month from order_purchase_timestamp) as months,
                  count(order_id) as count2017oforders
           from my-project-target-394308.target.orders
           where extract (year from order_purchase_timestamp) = 2017
           group by months , years
```

```
            order by  years , months),

 year2018 as
(select extract (year from order_purchase_timestamp) as years,
                 extract (month from order_purchase_timestamp) as months,
                 count(order_id) as count2018oforders
           from my-project-target-394308.target.orders
           where extract (year from order_purchase_timestamp)= 2018
            group by months , years
            order by  years , months)



select year2017.months , year2016.count2016oforders, year2017.count2017oforders,
year2018.count2018oforders
from year2016 right join year2017 on year2016.months = year2017.months left join year2018 on
year2017.months = year2018.months
order by year2017.months asc
```



Row	months	count2016oforders	count2017oforders	count2018oforders
1	1	null	800	7269
2	2	null	1780	6728
3	3	null	2682	7211
4	4	null	2404	6939
5	5	null	3700	6873
6	6	null	3245	6167
7	7	null	4026	6292
8	8	null	4331	6512
9	9	4	4285	16
10	10	324	4631	4

Results per page: 50 1 – 12 of 12

Analysis: To analyze this data we need more data of previous years to check the monthly seasonality. To check, it requires to be in a pattern every year and hence the data is of two years there is no such more number of orders being placed in the last two years for the a particular month/ months.

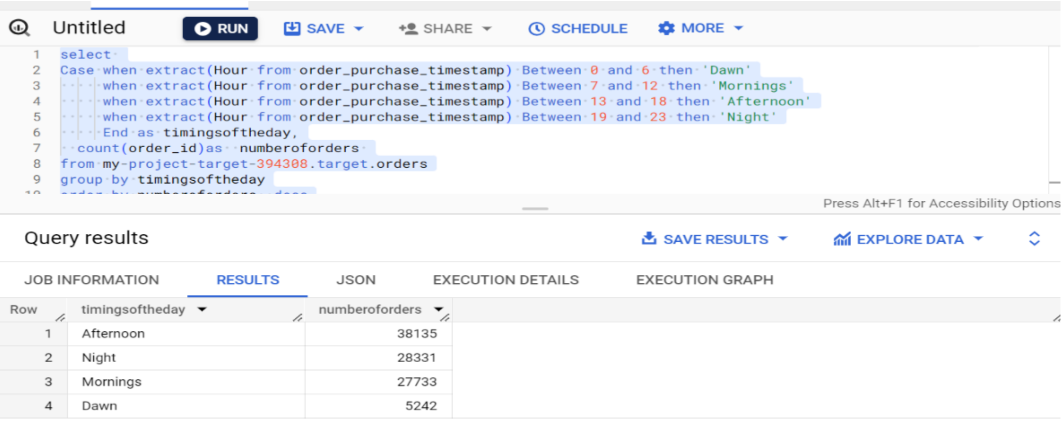
3.During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

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

- 19-23 hrs : Night

ANSWER-
QUERY :

```
select
Select Case when extract (Hour from order_purchase_timestamp) Between 0 and 6 then 'Dawn'
    when extract (Hour from order_purchase_timestamp) Between 7 and 12 then 'Mornings'
   when extract(Hour from order_purchase_timestamp) Between 13 and 18 then 'Afternoon'
   when extract (Hour from order_purchase_timestamp) Between 19 and 23 then 'Night'
       End as timingsoftheday, count (order_id)as numberoforders
from my-project-target-394308.target.orders
group by timingsoftheday
order by numberoforders desc
```



```
select
Case when extract(Hour from order_purchase_timestamp) Between 0 and 6 then 'Dawn'
    when extract(Hour from order_purchase_timestamp) Between 7 and 12 then 'Mornings'
    when extract(Hour from order_purchase_timestamp) Between 13 and 18 then 'Afternoon'
    when extract(Hour from order_purchase_timestamp) Between 19 and 23 then 'Night'
    End as timingsoftheday,
  count(order_id)as  numberoforders
from my-project-target-394308.target.orders
group by timingsoftheday
order by numberoforders desc
```

Row	timingsoftheday	numberoforders
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Analysis: There has been most number of orders placed in afternoon and then night , followed by mornings and lastly dawn.

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

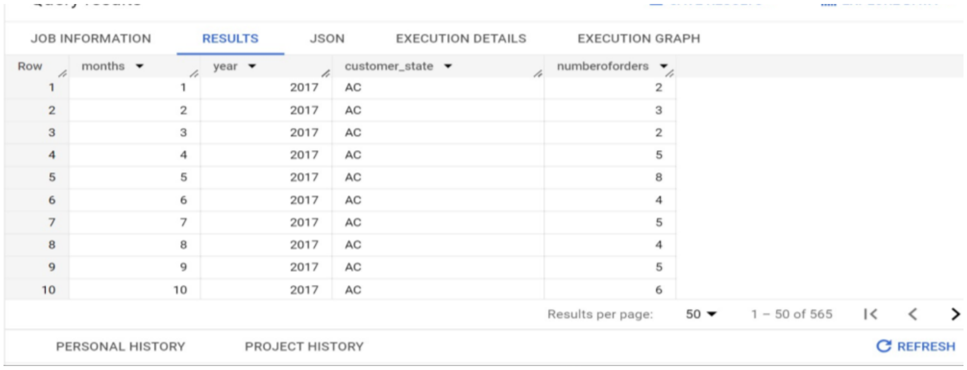
1. Get the month on month no. of orders placed in each state.
2. How are the customers distributed across all the states?

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

Answer –

Query :

```
select extract (month from order_purchase_timestamp) as months,

extract (year from order_purchase_timestamp) as years,
customer_state,
count(order_id) as countoforders
 from my-project-target-394308.target.orders as o join my- project-target-
394308.target.customers as c on o.customer_id = c.customer_id
group by 1,2,3
order by customer_state asc, years asc, months asc
```

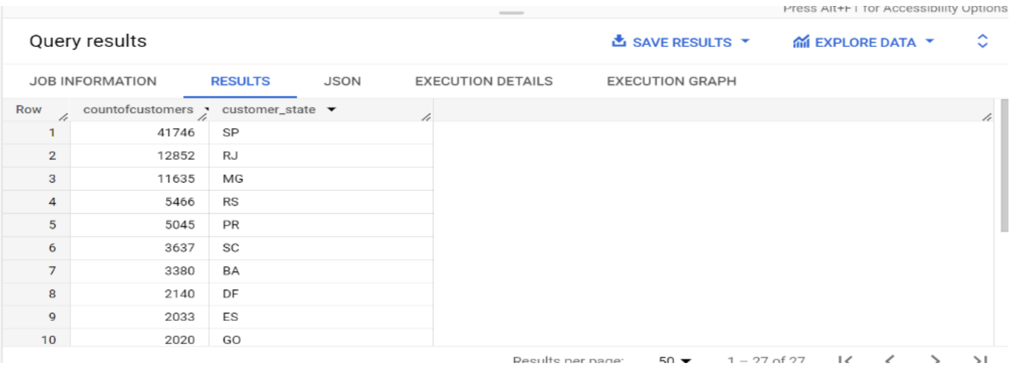


Row	months	year	customer_state	numberoforders
1	1	2017	AC	2
2	2	2017	AC	3
3	3	2017	AC	2
4	4	2017	AC	5
5	5	2017	AC	8
6	6	2017	AC	4
7	7	2017	AC	5
8	8	2017	AC	4
9	9	2017	AC	5
10	10	2017	AC	6

Results per page: 50 1 – 50 of 565

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

```
select count(distinct customer_id) as countofcustomers ,
       customer_state
  from my-project-target-394308.target.customers
  group by customer_state
 order by countofcustomers desc
```



Query results

Row	countofcustomers	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

Results per page: 50 1 – 27 of 27

Analysis: The most number of orders and customers are from the SP state and followed by other states according to the customer locations.

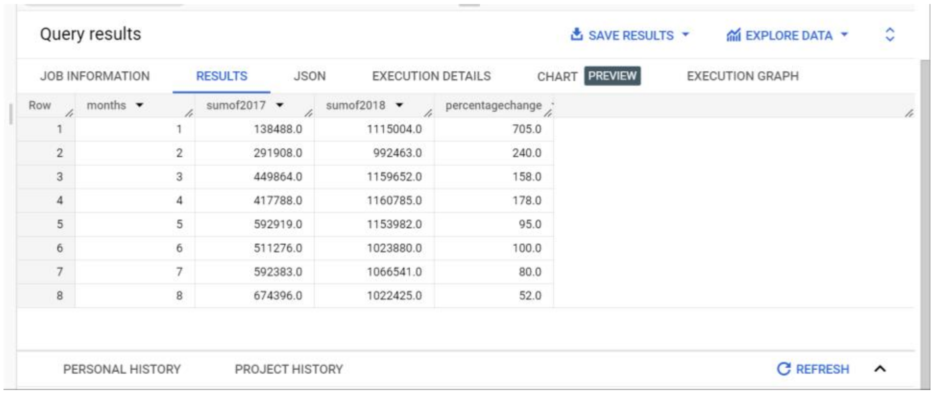
D. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

a. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
You can use the "payment_value" column in the payments table to get the cost of orders.
b. Calculate the Total & Average value of order price for each state.
c. Calculate the Total & Average value of order freight for each state

1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
   You can use the "payment_value" column in the payments table to get the cost of orders.

Query:

```
with year2017 as (select

   extract (month from order_purchase_timestamp) as months,
  Round(sum(payment_value)) as sumof2017 from my-project-target-394308.target.orders as o
join
my-project-target-394308.target.payments as p on o.order_id = p.order_id
where extract (year from order_purchase_timestamp) = 2017
and extract (month from order_purchase_timestamp) Between 1 and 8
group by 1
order by months),

 year2018 as (select
   extract (month from order_purchase_timestamp) as months,
  Round(sum(payment_value)) as sumof2018 from my-project-target-394308.target.orders as o
join
my-project-target-394308.target.payments as p on o.order_id = p.order_id
where extract (year from order_purchase_timestamp) = 2018
and extract (month from order_purchase_timestamp) Between 1 and 8
group by 1
order by months)

select year.2017months, sumof2017 , sumof2018 ,round(((sumof2018-
sumof2017)/sumof2017)*100) as percentagechange
 from year2017 join year2018
on year2017.months = year2018.months
order by year2017.months asc
```



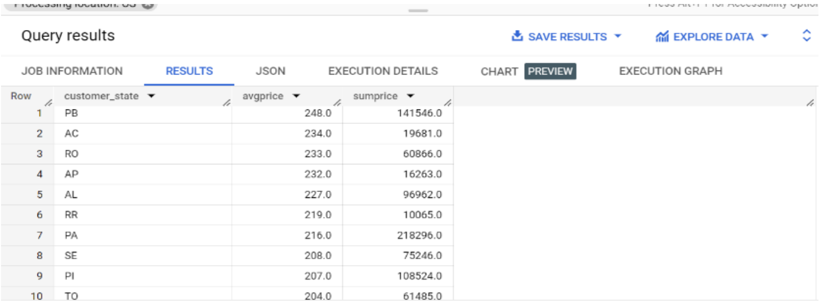
Row	months	sumof2017	sumof2018	percentagechange
1	1	138488.0	1115004.0	705.0
2	2	291908.0	992463.0	240.0
3	3	449864.0	1159652.0	158.0
4	4	417788.0	1160785.0	178.0
5	5	592919.0	1153982.0	95.0
6	6	511276.0	1023880.0	100.0
7	7	592383.0	1066541.0	80.0
8	8	674396.0	1022425.0	52.0

Analysis: There is immense growth from the year 2017 to 2018 , and specifically in the month of jan (growth of 705%) and then followed by others.

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

Query:

```
 select customer_state,

round(avg(payment_value)) as avgprice,
round(sum(payment_value)) as sumprice
from
my-project-target-394308.target.customers as c join
my-project-target-394308.target.orders as o
on c.customer_id = o.customer_id
join my-project-target-394308.target.payments as p
on o.order_id = p.order_id
group by 1
order by avgprice desc
```

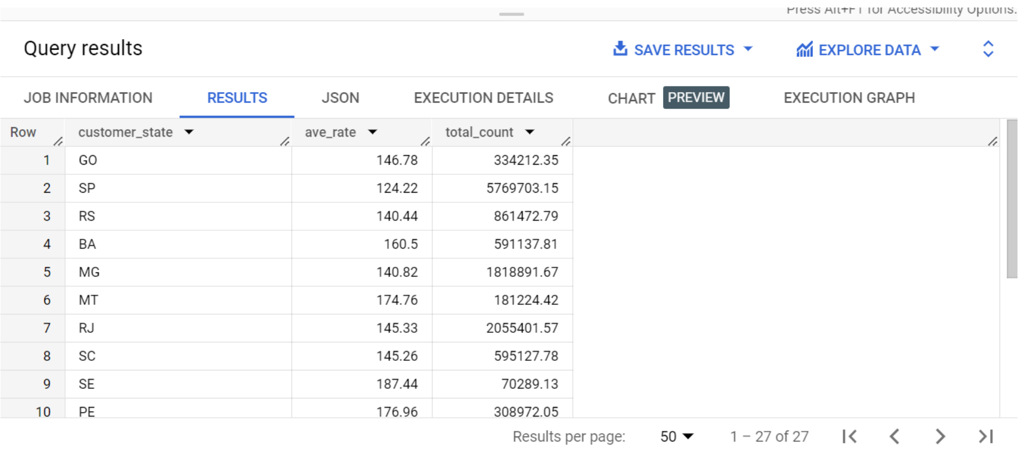


Row	customer_state	avgprice	sumprice
1	PB	248.0	141546.0
2	AC	234.0	19681.0
3	RO	233.0	60866.0
4	AP	232.0	16263.0
5	AL	227.0	96962.0
6	RR	219.0	10065.0
7	PA	216.0	218296.0
8	SE	208.0	75246.0
9	PI	207.0	108524.0
10	TO	204.0	61485.0

Analysis: The average and total order price is highest in the PB state and then AC and then followed by other states.

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

```
SELECT
c.customer_state,
round (avg (oi.price + oi.freight_value), 2) as ave_rate,
 round (sum(price+freight_value), 2)as total_count
from my-project-target-394308.target.customers as c
join
 my-project-target-394308.target.orders as o
on c.customer_id = o.customer_id
join
 my-project-target-394308.target.orders_item as oi
on o.order_id = oi.order_id
where o.order_status = 'delivered'
group by 1
```



Row	customer_state	ave_rate	total_count
1	GO	146.78	334212.35
2	SP	124.22	5769703.15
3	RS	140.44	861472.79
4	BA	160.5	591137.81
5	MG	140.82	1818891.67
6	MT	174.76	181224.42
7	RJ	145.33	2055401.57
8	SC	145.26	595127.78
9	SE	187.44	70289.13
10	PE	176.96	308972.05

E. 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.
   Do this in a single query.

   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

  - **time_to_deliver** = order_delivered_customer_date - order_purchase_timestamp
  - **diff_estimated_delivery** = order_estimated_delivery_date - order_delivered_customer_date

2. Find out the top 5 states with the highest & lowest average freight value.
3. Find out the top 5 states with the highest & lowest average delivery time.
4. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
   You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

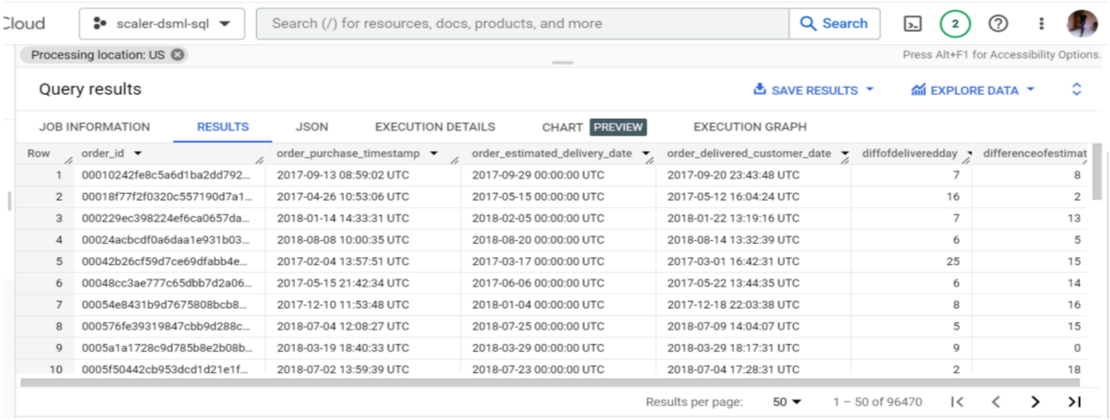
1.Find the no. of days taken to deliver each order from the order's purchase date as delivery time.
Also, calculate the difference (in days) between the estimated & actual delivery date of an order.
Do this in a single query.

You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

- **time_to_deliver** = order_delivered_customer_date - order_purchase_timestamp
- **diff_estimated_delivery** = order_estimated_delivery_date - order_delivered_customer_date

Query:

```
select order_id,

order_purchase_timestamp,order_estimated_delivery_date,

order_delivered_customer_date,

date_diff(order_delivered_customer_date, order_purchase_timestamp,Day) as
diffofdeliveredday,
date_diff(order_estimated_delivery_date,order_delivered_customer_date,Day) as
differenceofestimatedday
from my-project-target-394308.target.customers as c join
my-project-target-394308.target.orders as o
on c.customer_id  = o.customer_id

where  order_status = 'delivered' and order_delivered_customer_date is not null

order by order_id asc
```


Cloud
scaler-dsml-sql
Search (/) for resources, docs, products, and more
Search
Processing location: US
Press Alt+F1 for Accessibility Options.

Query results

Row	order_id	order_purchase_timestamp	order_estimated_delivery_date	order_delivered_customer_date	diffofdeliveredday	differenceofestimat
1	00010242fe8c5a6d1ba2dd792...	2017-09-13 08:59:02 UTC	2017-09-29 00:00:00 UTC	2017-09-20 23:43:48 UTC	7	8
2	00018f77f2f0320c557190d7a1...	2017-04-26 10:53:06 UTC	2017-05-15 00:00:00 UTC	2017-05-12 16:04:24 UTC	16	2
3	000229ec398224ef6ca0657da...	2018-01-14 14:33:31 UTC	2018-02-05 00:00:00 UTC	2018-01-22 13:19:16 UTC	7	13
4	00024acbcdf0a6daa1e931b03...	2018-08-08 10:00:35 UTC	2018-08-20 00:00:00 UTC	2018-08-14 13:32:39 UTC	6	5
5	00042b26cf59d7ce69dfabb4e...	2017-02-04 13:57:51 UTC	2017-03-17 00:00:00 UTC	2017-03-01 16:42:31 UTC	25	15
6	00048cc3ae777c65dbb7d2a06...	2017-05-15 21:42:34 UTC	2017-06-06 00:00:00 UTC	2017-05-22 13:44:35 UTC	6	14
7	00054e8431b9d7675808bcb8...	2017-12-10 11:53:48 UTC	2018-01-04 00:00:00 UTC	2017-12-18 22:03:38 UTC	8	16
8	000576fe39319847cbb9d288c...	2018-07-04 12:08:27 UTC	2018-07-25 00:00:00 UTC	2018-07-09 14:04:07 UTC	5	15
9	0005a1a1728c9d785b8e2b08b...	2018-03-19 18:40:33 UTC	2018-03-29 00:00:00 UTC	2018-03-29 18:17:31 UTC	9	0
10	0005f50442cb953dcd1d21e1f...	2018-07-02 13:59:39 UTC	2018-07-23 00:00:00 UTC	2018-07-04 17:28:31 UTC	2	18

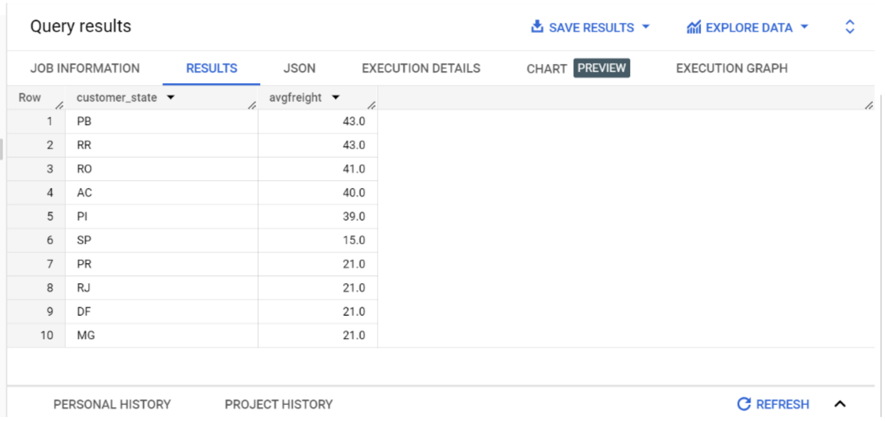
Results per page: 50 1 – 50 of 96470

Analysis: There is a detailed analysis of delivery time from the ordered date and difference between the estimated and delivered time of all the orders across the state.

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

Answer –
Query:

```
With highest as

  (select customer_state, round(Avg(freight_value),0) as avgfreight

from my-project-target-394308.target.customers as c

join my-project-target-394308.target.orders as o
on c.customer_id = o.customer_id
join my-project-target-394308.target.orders_item as oi on o.order_id = oi.order_id
group by customer_state
order by avgfreight desc
limit 5 ),

 lowest as
      (select customer_state, round(Avg(freight_value),0) as avgfreight
from my-project-target-394308.target.customers as c
join my-project-target-394308.target.orders as o
on c.customer_id = o.customer_id
join my-project-target-394308.target.orders_item as oi on o.order_id = oi.order_id
group by customer_state
order by avgfreight asc
limit 5 )

select * from highest
union all
select * from lowest
```



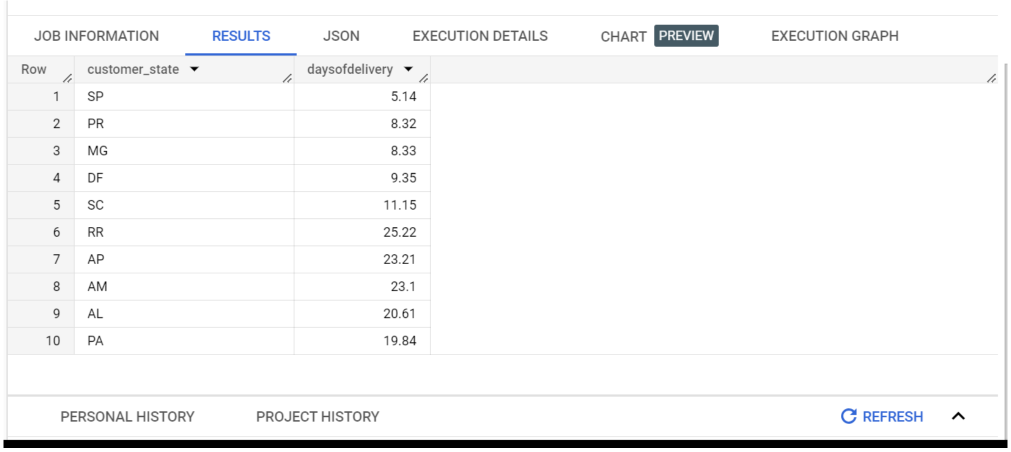
Row	customer_state	avgfreight
1	PB	43.0
2	RR	43.0
3	RO	41.0
4	AC	40.0
5	PI	39.0
6	SP	15.0
7	PR	21.0
8	RJ	21.0
9	DF	21.0
10	MG	21.0

Analysis : There are 5 states like PB, RR, RO, AC , PI which has the highest freight value , and other 5 states like SP,PR,RJ,DF,MG has the lowest freight value.

3.Find out the top 5 states with the highest & lowest average delivery time.

Answer-
Query:

```
With highest as
(select customer_state,
 round(Avg(date_diff(order_delivered_customer_date,order_delivered_carrier_date,Day)),2) as
daysofdelivery
from my-project-target-394308.target.customers as c join
my-project-target-394308.target.orders as o
on c.customer_id  = o.customer_id
where order_delivered_carrier_date is not null and order_delivered_customer_date is not null
group by 1
order by daysofdelivery asc
limit 5),

 lowest as
 (select customer_state,
 round(Avg(date_diff(order_delivered_customer_date,order_delivered_carrier_date,Day)),2) as
daysofdelivery
from my-project-target-394308.target.customers as c join
my-project-target-394308.target.orders as o
on c.customer_id  = o.customer_id
where order_delivered_carrier_date is not null and order_delivered_customer_date is not null
group by 1
order by daysofdelivery desc
limit 5 )

select * from highest
union all
select * from lowest
```



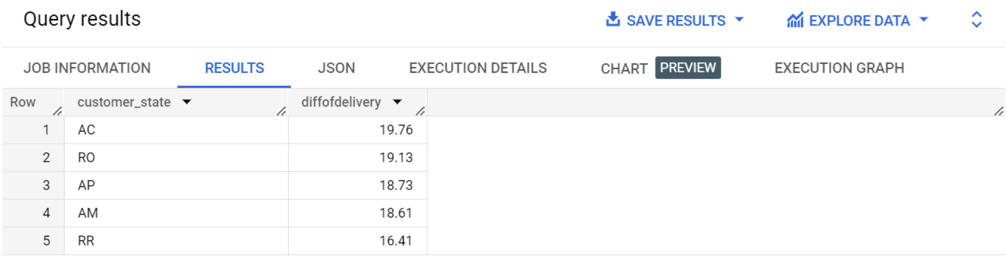
Row	customer_state	daysofdelivery
1	SP	5.14
2	PR	8.32
3	MG	8.33
4	DF	9.35
5	SC	11.15
6	RR	25.22
7	AP	23.21
8	AM	23.1
9	AL	20.61
10	PA	19.84

Analysis: This analysis has been made on the basis of the t ime of delivery it takes from the location of shipment to the delivery location of the customer . So there are top 5 states which requires less time to deliver on an average, there are 5 states which requires too much time to deliver on an average.

4.Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

Query :

```
 select  customer_state,


round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,Day)),2) as
diffofdelivery
from my-project-target-394308.target.customers as c join
my-project-target-394308.target.orders as o
on c.customer_id  = o.customer_id
where order_status = 'delivered'
group by 1
order by diffofdelivery desc
limit 5
```



Row	customer_state	diffofdelivery
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

Analysis: These are the 5 states which has the fastest delivery compared to other states .

F. **Analysis based on the payments:**

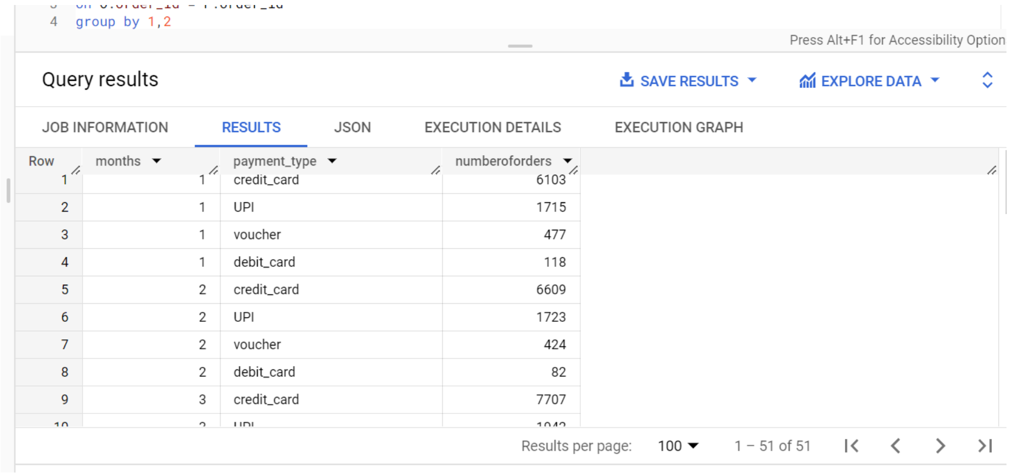
1. Find the month on month no. of orders placed using different payment types.
2. Find the no. of orders placed on the basis of the payment installments that have been paid.

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

Answer-

Query:

```
select extract (month from order_purchase_timestamp) as months,
payment_type, count(O.order_id) numberoforders

from my-project-target-394308.target.orders as O join my-project-target-
394308.target.payments as P

on O.order_id = P.order_id
group by 1,2
order by months, numberoforders desc
```


```
on O.order_id = P.order_id
group by 1,2
```

Row	months	payment_type	numberoforders
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

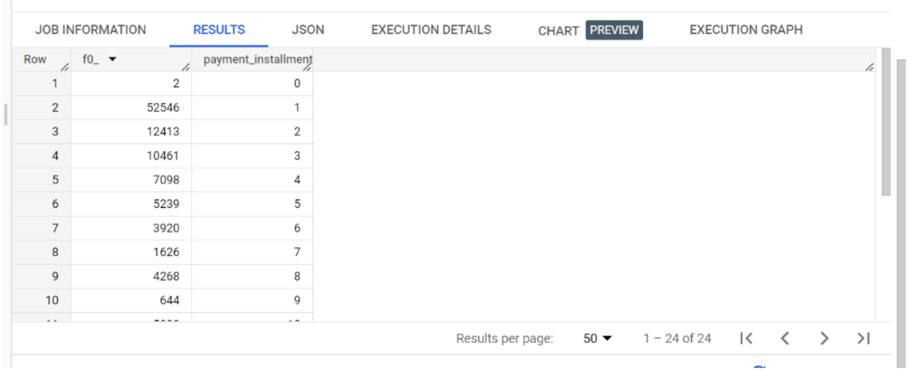
Analysis: My insight on this would be Most of the people have preferred using Credit card and UPI to make make payment every months.

2. Find the no. of orders placed on the basis of the payment installments that have been paid.

Answer-

Query :

```
select count(order_id) , payment_installments
from `my-project-target-394308.target.payments`
group by payment_installments
```



Row	f0_	payment_installment
1	2	0
2	52546	1
3	12413	2
4	10461	3
5	7098	4
6	5239	5
7	3920	6
8	1626	7
9	4268	8
10	644	9

Results per page: 50 ▾ 1 – 24 of 24

Insight/ advices : According to this dataset of target , there has been growth In the number of orders from previous year and there has been increase in reach of delivery to most of the states. Hence the data is small we can't compare it and give a long tern insight or advice on it! We would require to have a big data to have a better comparison.
But as of now ,in the current scenario the growth is good.

Also during the dawn time of the day, the customers engage way too less in buying products , to increase in that hour of the day
We can give some extra discounts or offers, or reduced prices of good so that they engage more in those hours also.

We also need to cut the delayed deliveries and partner it up with some fast deliveries courier services which will improve in delivering the products asap. And specially to those least 5 cities which takes most of the time to deliver the products.

Anshika porwal