TARGET CASE STUDY

To gain valuable insights into Target's operations in Brazil by analyzing various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

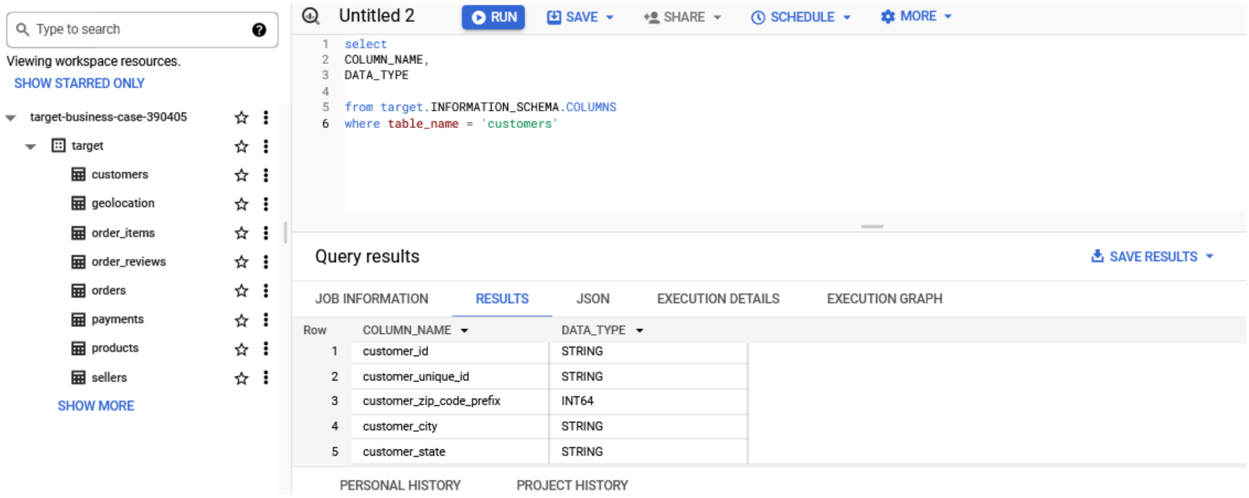
Question 1

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 number of Cities and States in our dataset.

Answer

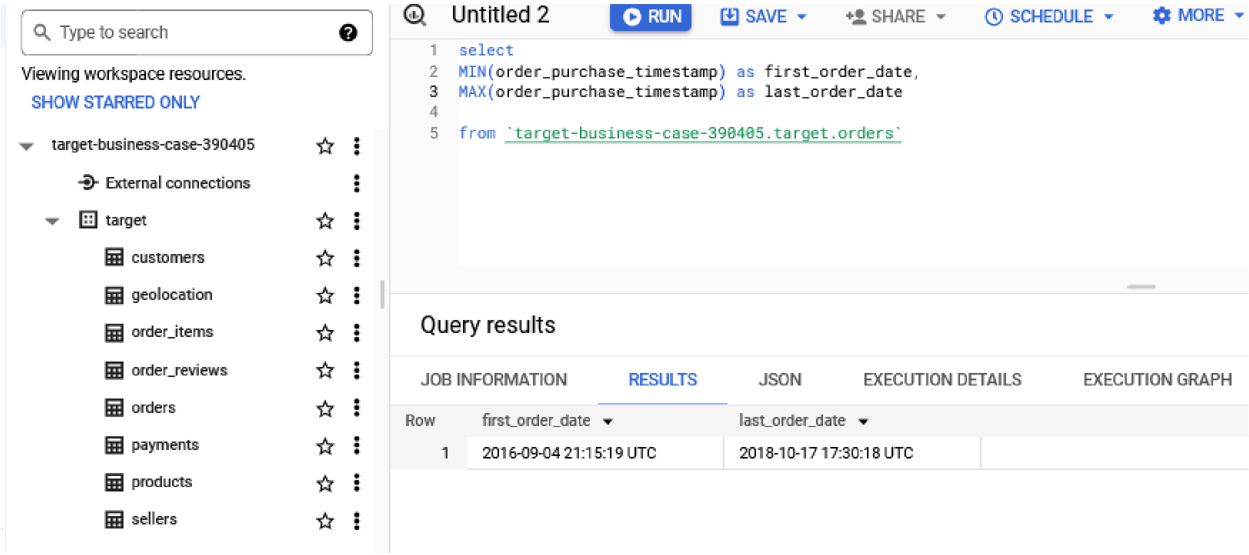
1.



```
select
COLUMN_NAME,
DATA_TYPE

from target.INFORMATION_SCHEMA.COLUMNS
where table_name = 'customers'
```

Row	COLUMN_NAME	DATA_TYPE
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

2.

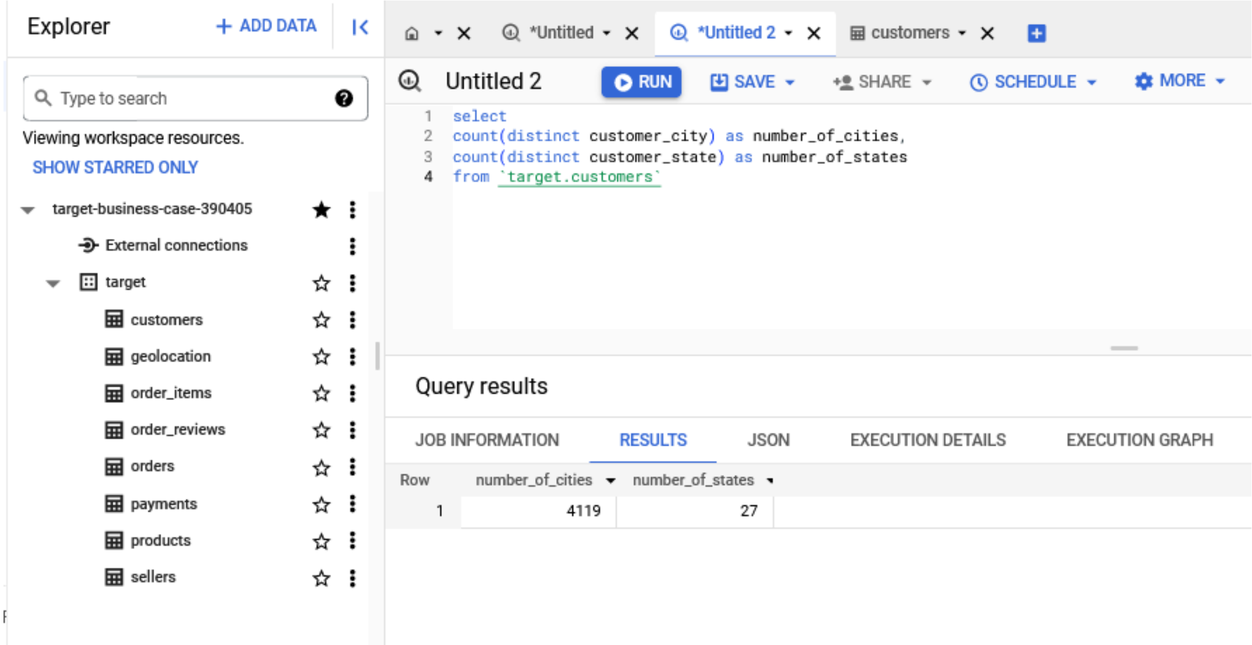


```sql
select
MIN(order_purchase_timestamp) as first_order_date,
MAX(order_purchase_timestamp) as last_order_date

from `target-business-case-390405.target.orders`
```

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

The time range between which orders are placed is between September 2016 and October 2018.

3.



```sql
select
count(distinct customer_city) as number_of_cities,
count(distinct customer_state) as number_of_states
from `target.customers`
```

Row	number_of_cities	number_of_states
1	4119	27

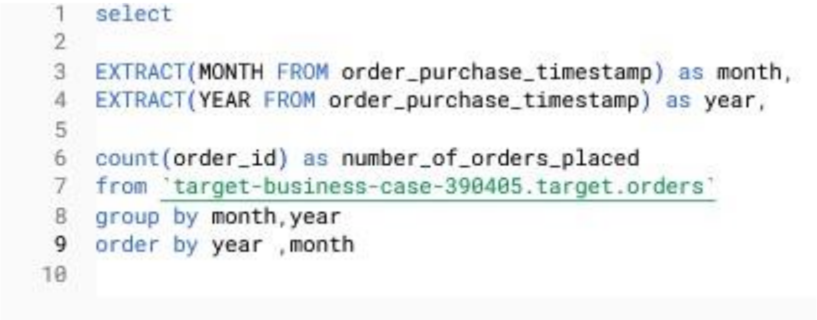
Question 2

In-depth Exploration:

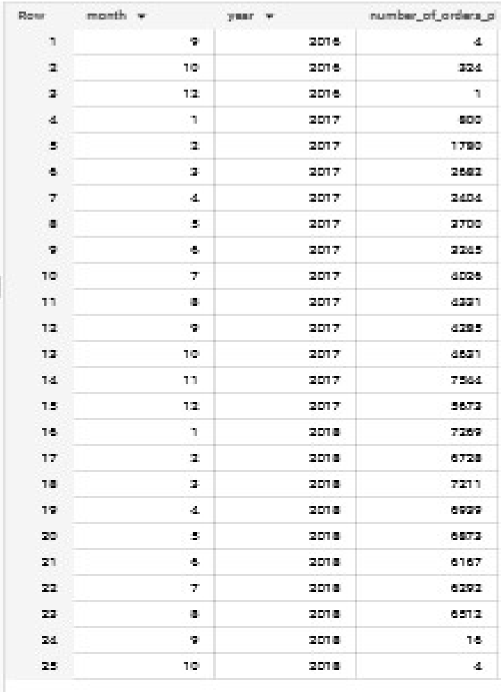
1. Is there a growing trend in the no. of orders placed over the past years?
2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
3. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
    - 0-6 hrs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

Answer 2

1.



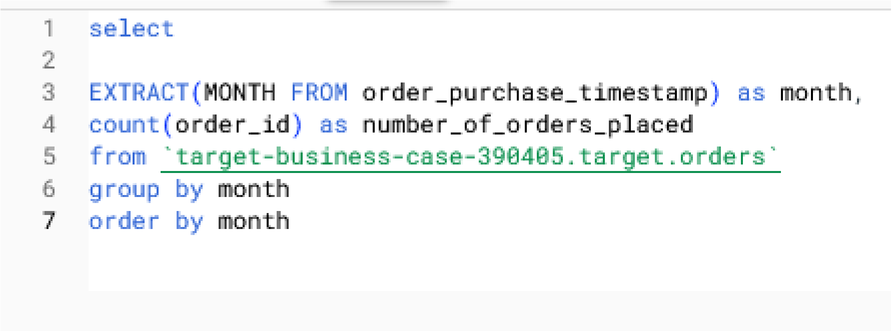
```
select

EXTRACT(MONTH FROM order_purchase_timestamp) as month,
EXTRACT(YEAR FROM order_purchase_timestamp) as year,

count(order_id) as number_of_orders_placed
from `target-business-case-390405.target.orders`
group by month,year
order by year ,month
```



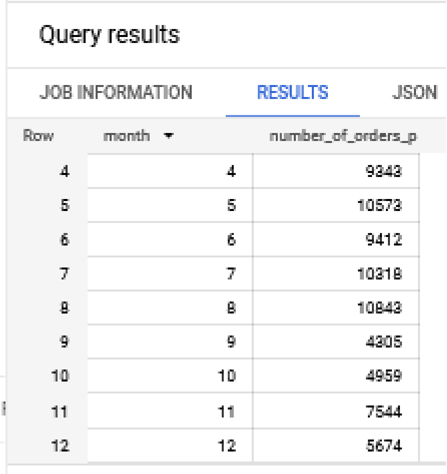
Row	month	year	number_of_orders_p
1	9	2016	4
2	10	2016	324
3	12	2016	1
4	1	2017	800
5	2	2017	1780
6	3	2017	2682
7	4	2017	2404
8	5	2017	3700
9	6	2017	3245
10	7	2017	4026
11	8	2017	4331
12	9	2017	4285
13	10	2017	4631
14	11	2017	7544
15	12	2017	5673
16	1	2018	7269
17	2	2018	6728
18	3	2018	7211
19	4	2018	6939
20	5	2018	6873
21	6	2018	6167
22	7	2018	6292
23	8	2018	6512
24	9	2018	16
25	10	2018	4

The number of orders placed increases from 2016 to the end of 2017 and then remains constant till august 2018

2.



```sql
select

EXTRACT(MONTH FROM order_purchase_timestamp) as month,
count(order_id) as number_of_orders_placed
from `target-business-case-390405.target.orders`
group by month
order by month
```

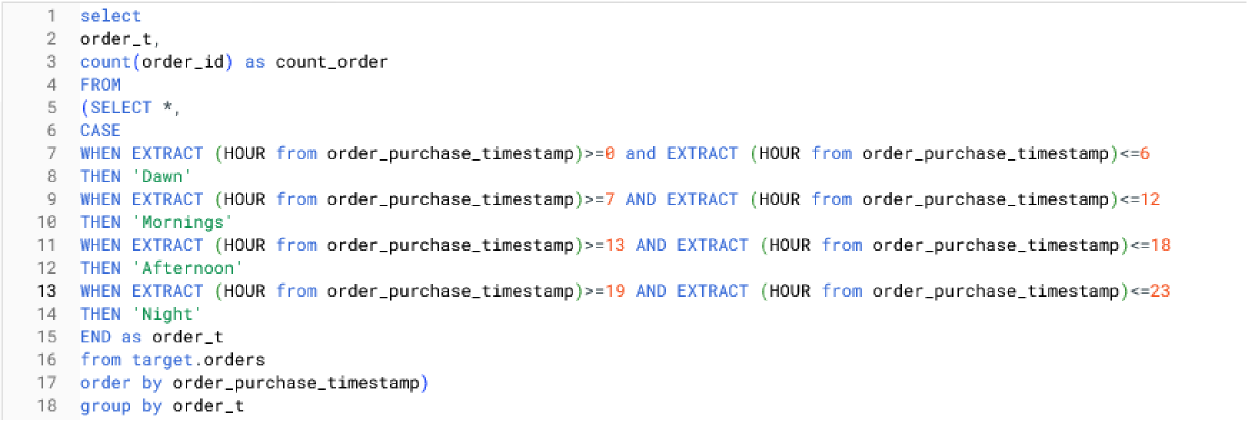


Query results

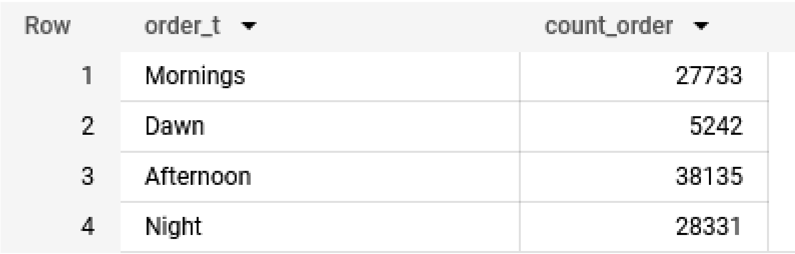
Row	month	number_of_orders_p
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959
11	11	7544
12	12	5674

The no. of orders placed are high during winter season and decreases during summers.

3.



```
select
order_t,
count(order_id) as count_order
FROM
(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)>=7 AND EXTRACT (HOUR from order_purchase_timestamp)<=12
THEN 'Mornings'
WHEN EXTRACT (HOUR from order_purchase_timestamp)>=13 AND EXTRACT (HOUR from order_purchase_timestamp)<=18
THEN 'Afternoon'
WHEN EXTRACT (HOUR from order_purchase_timestamp)>=19 AND EXTRACT (HOUR from order_purchase_timestamp)<=23
THEN 'Night'
END as order_t
from target.orders
order by order_purchase_timestamp)
group by order_t
```



Row	order_t	count_order
1	Mornings	27733
2	Dawn	5242
3	Afternoon	38135
4	Night	28331

The orders placed are highest during afternoon and lowest during dawn.

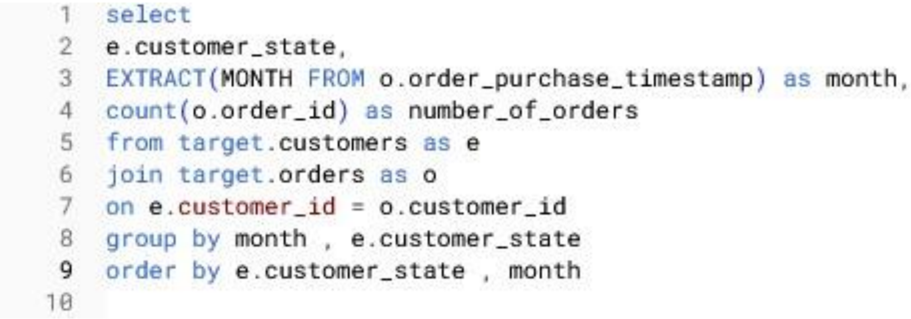
Question 3

Evolution of E-commerce orders in the Brazil region:

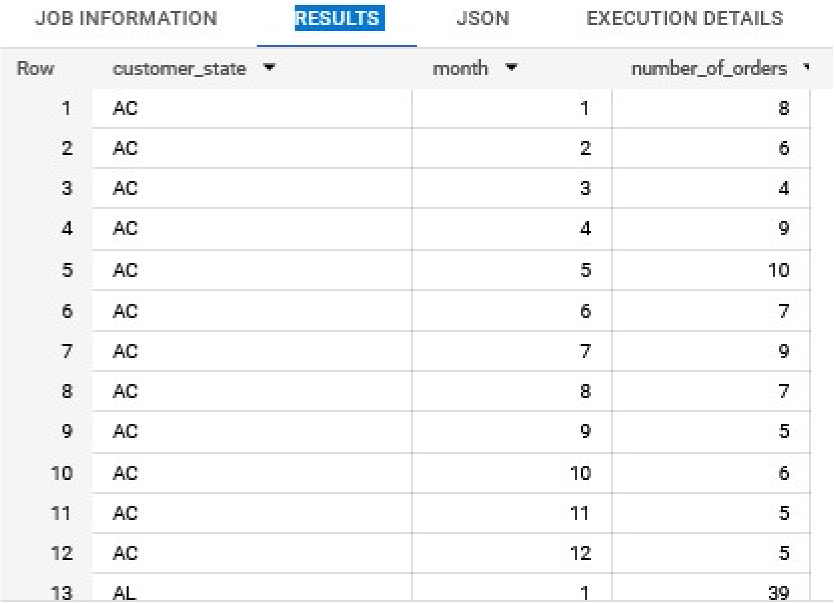
1. Get the month on month no. of orders placed in each state.
2. How are the customers distributed across all the states?

Answer 3

1.

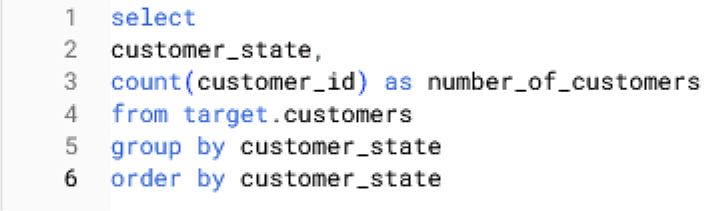


```
select
e.customer_state,
EXTRACT(MONTH FROM o.order_purchase_timestamp) as month,
count(o.order_id) as number_of_orders
from target.customers as e
join target.orders as o
on e.customer_id = o.customer_id
group by month , e.customer_state
order by e.customer_state , month

```

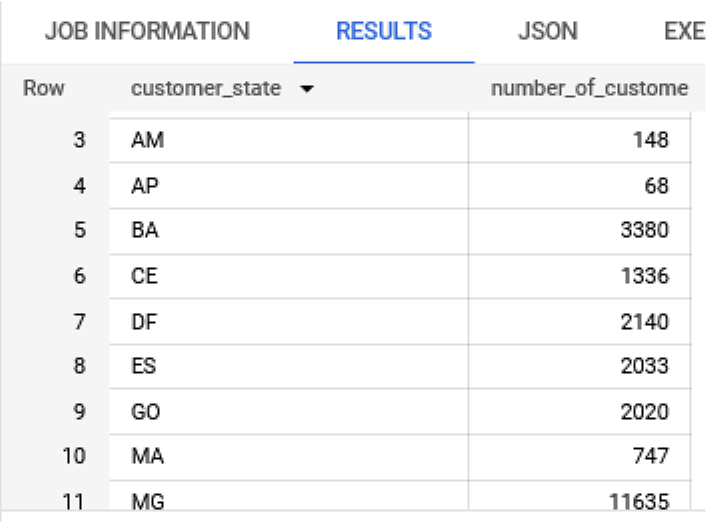


Row	customer_state	month	number_of_orders
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6
11	AC	11	5
12	AC	12	5
13	AL	1	39

2.



```
select
customer_state,
count(customer_id) as number_of_customers
from target.customers
group by customer_state
order by customer_state
```



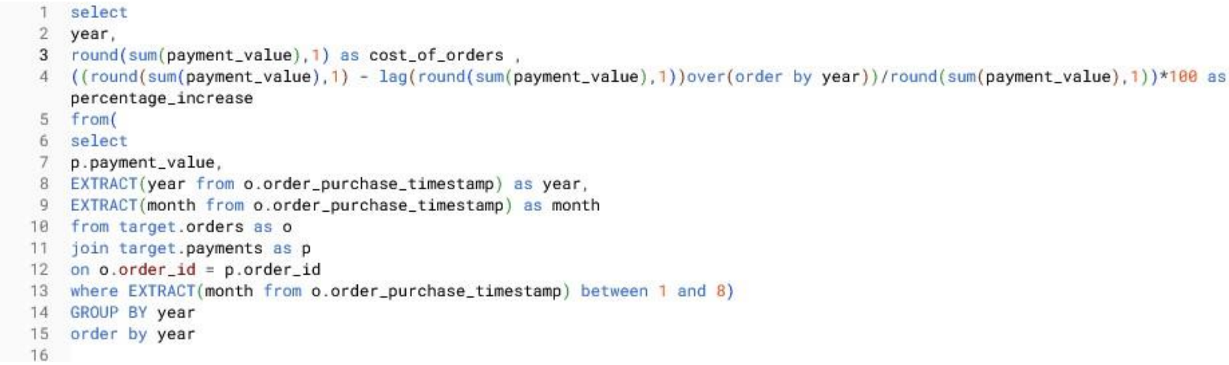
Row	customer_state	number_of_custome
3	AM	148
4	AP	68
5	BA	3380
6	CE	1336
7	DF	2140
8	ES	2033
9	GO	2020
10	MA	747
11	MG	11635

Question 4

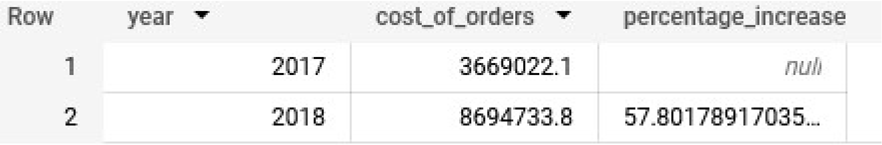
Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

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

1.



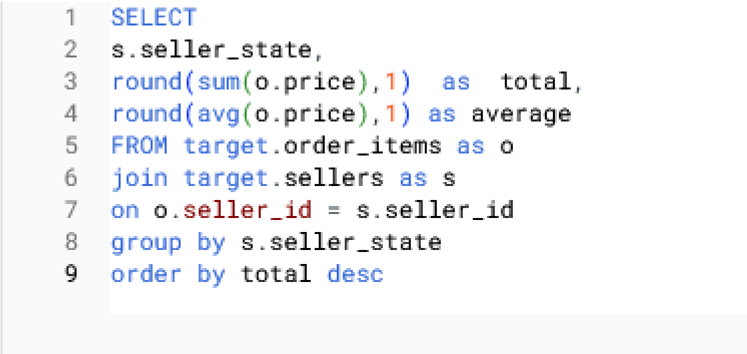
```
select
year,
round(sum(payment_value),1) as cost_of_orders ,
((round(sum(payment_value),1) - lag(round(sum(payment_value),1))over(order by year))/round(sum(payment_value),1))*100 as
percentage_increase
from(
select
p.payment_value,
EXTRACT(year from o.order_purchase_timestamp) as year,
EXTRACT(month from o.order_purchase_timestamp) as month
from target.orders as o
join target.payments as p
on o.order_id = p.order_id
where EXTRACT(month from o.order_purchase_timestamp) between 1 and 8)
GROUP BY year
order by year
```



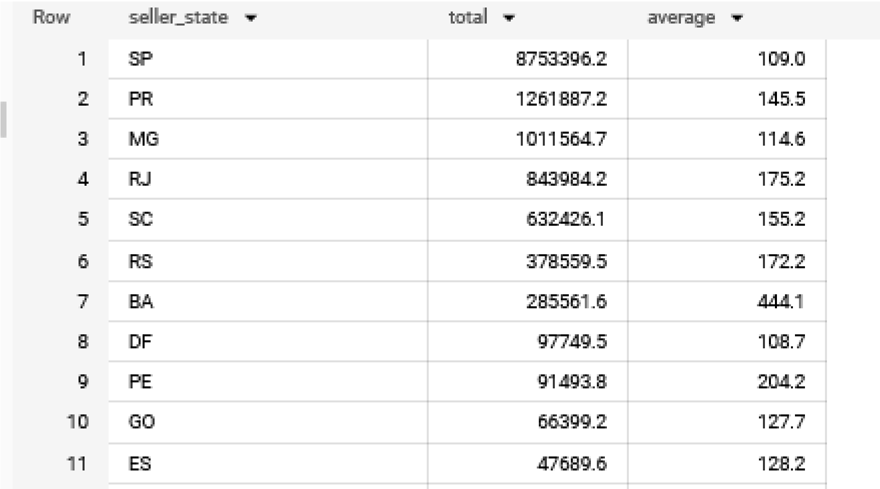
Row	year	cost_of_orders	percentage_increase
1	2017	3669022.1	null
2	2018	8694733.8	57.80178917035...

The cost of orders increase by more than 50 % in 2018 from 2017.

2.



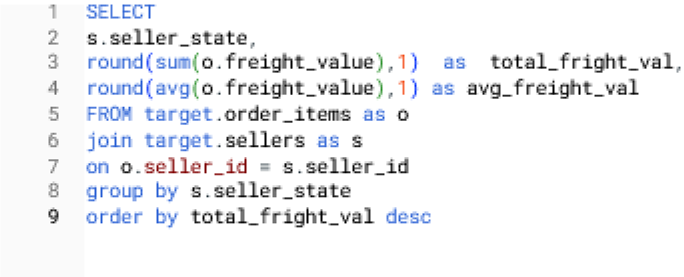
```
SELECT
s.seller_state,
round(sum(o.price),1)  as  total,
round(avg(o.price),1) as average
FROM target.order_items as o
join target.sellers as s
on o.seller_id = s.seller_id
group by s.seller_state
order by total desc
```



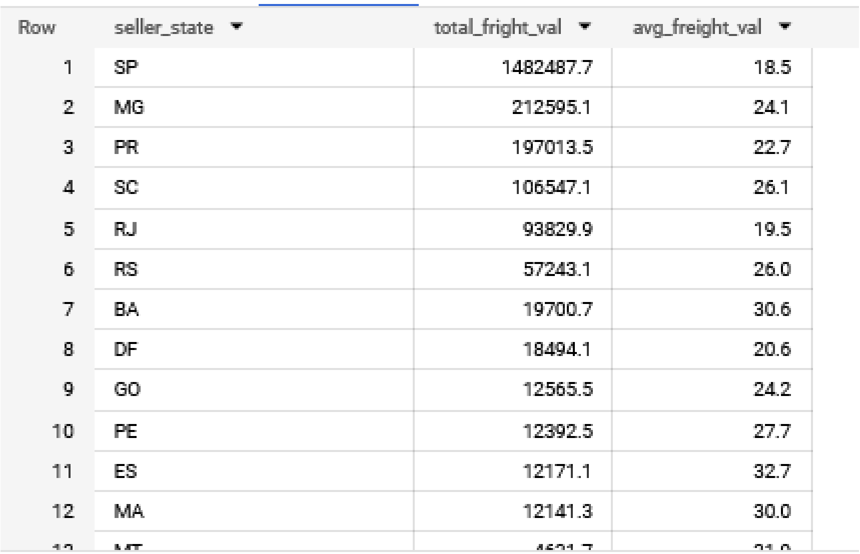
Row	seller_state	total	average
1	SP	8753396.2	109.0
2	PR	1261887.2	145.5
3	MG	1011564.7	114.6
4	RJ	843984.2	175.2
5	SC	632426.1	155.2
6	RS	378559.5	172.2
7	BA	285561.6	444.1
8	DF	97749.5	108.7
9	PE	91493.8	204.2
10	GO	66399.2	127.7
11	ES	47689.6	128.2

The data shows that as cost of orders decreases ,the numbers of orders placed increases

3.



```
SELECT
s.seller_state,
round(sum(o.freight_value),1)  as  total_fright_val,
round(avg(o.freight_value),1) as avg_freight_val
FROM target.order_items as o
join target.sellers as s
on o.seller_id = s.seller_id
group by s.seller_state
order by total_fright_val desc
```



Row	seller_state	total_fright_val	avg_freight_val
1	SP	1482487.7	18.5
2	MG	212595.1	24.1
3	PR	197013.5	22.7
4	SC	106547.1	26.1
5	RJ	93829.9	19.5
6	RS	57243.1	26.0
7	BA	19700.7	30.6
8	DF	18494.1	20.6
9	GO	12565.5	24.2
10	PE	12392.5	27.7
11	ES	12171.1	32.7
12	MA	12141.3	30.0
13	MT	4621.7	31.0

The above data shows high no. of orders placed where avg freight value is low.

Question 5

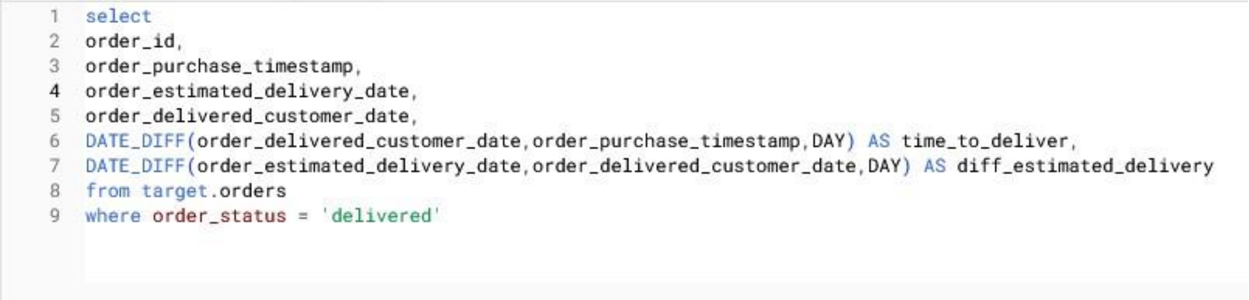
Analysis based on sales, freight and delivery time.

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.

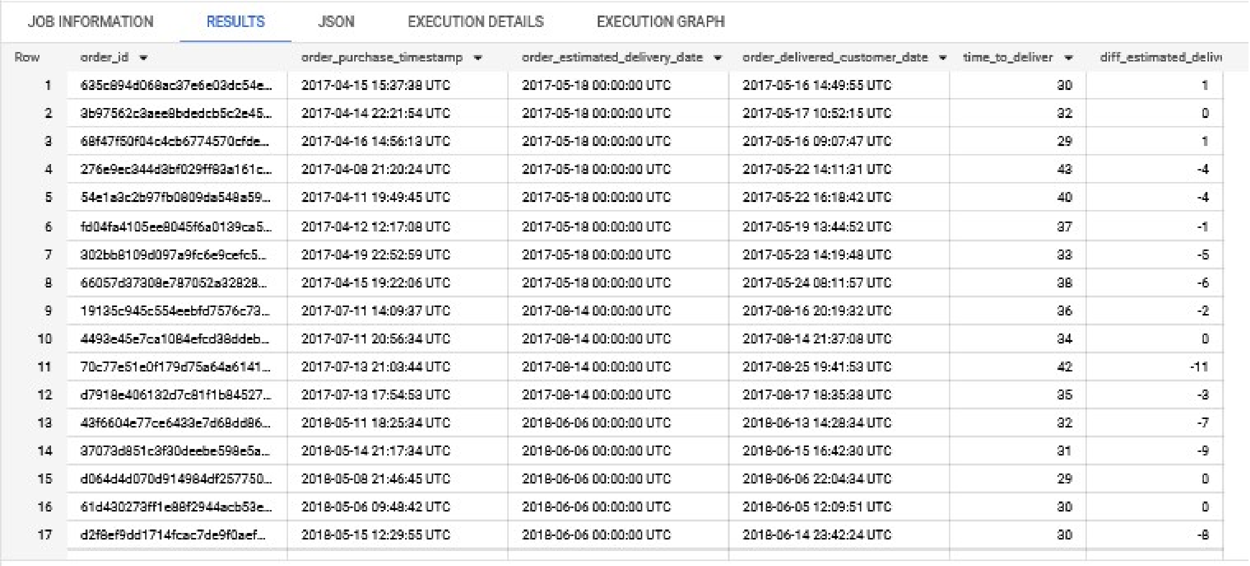
   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
   - **time_to_deliver** = order_delivered_customer_date - order_purchase_timestamp
   - **diff_estimated_delivery** = order_estimated_delivery_date - order_delivered_customer_date
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.

Answer 5

1.



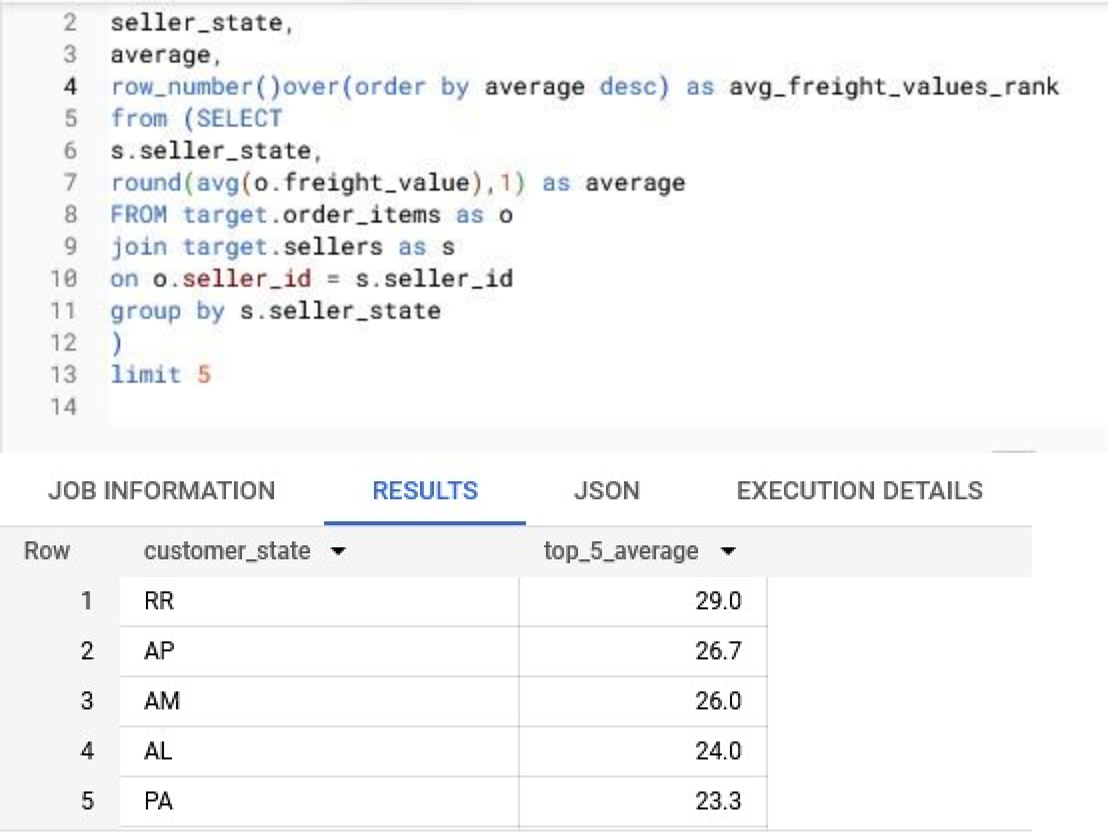
```
select
order_id,
order_purchase_timestamp,
order_estimated_delivery_date,
order_delivered_customer_date,
DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY) AS time_to_deliver,
DATE_DIFF(order_estimated_delivery_date,order_delivered_customer_date,DAY) AS diff_estimated_delivery
from target.orders
where order_status = 'delivered'
```



Row	order_id	order_purchase_timestamp	order_estimated_delivery_date	order_delivered_customer_date	time_to_deliver	diff_estimated_deliv
1	635c894d068ac37e6e03dc54e...	2017-04-15 15:37:38 UTC	2017-05-18 00:00:00 UTC	2017-05-16 14:49:55 UTC	30	1
2	3b97562c3aee8bdedcb5c2e45...	2017-04-14 22:21:54 UTC	2017-05-18 00:00:00 UTC	2017-05-17 10:52:15 UTC	32	0
3	68f47f50f04c4cb6774570cfde...	2017-04-16 14:56:13 UTC	2017-05-18 00:00:00 UTC	2017-05-16 09:07:47 UTC	29	1
4	276e9ec344d3bf029ff83a161c...	2017-04-08 21:20:24 UTC	2017-05-18 00:00:00 UTC	2017-05-22 14:11:31 UTC	43	-4
5	54e1a3c2b97fb0809da548a59...	2017-04-11 19:49:45 UTC	2017-05-18 00:00:00 UTC	2017-05-22 16:18:42 UTC	40	-4
6	fd04fa4105ee8045f6a0139ca5...	2017-04-12 12:17:08 UTC	2017-05-18 00:00:00 UTC	2017-05-19 13:44:52 UTC	37	-1
7	302bb8109d097a9fc6e9cefc5...	2017-04-19 22:52:59 UTC	2017-05-18 00:00:00 UTC	2017-05-23 14:19:48 UTC	33	-5
8	66057d37308e787052a32828...	2017-04-15 19:22:06 UTC	2017-05-18 00:00:00 UTC	2017-05-24 08:11:57 UTC	38	-6
9	19135c945c554eebfd7576c73...	2017-07-11 14:09:37 UTC	2017-08-14 00:00:00 UTC	2017-08-16 20:19:32 UTC	36	-2
10	4493e45e7ca1084efcd38ddeb...	2017-07-11 20:56:34 UTC	2017-08-14 00:00:00 UTC	2017-08-14 21:37:08 UTC	34	0
11	70c77e51e0f179d75a64a6141...	2017-07-13 21:03:44 UTC	2017-08-14 00:00:00 UTC	2017-08-25 19:41:53 UTC	42	-11
12	d7918e406132d7c81f1b84527...	2017-07-13 17:54:53 UTC	2017-08-14 00:00:00 UTC	2017-08-17 18:35:38 UTC	35	-3
13	43f6604e77ce6433e7d68dd86...	2018-05-11 18:25:34 UTC	2018-06-06 00:00:00 UTC	2018-06-13 14:28:34 UTC	32	-7
14	37073d851c3f30deebe598e5a...	2018-05-14 21:17:34 UTC	2018-06-06 00:00:00 UTC	2018-06-15 16:42:30 UTC	31	-9
15	d064d4d070d914984df257750...	2018-05-08 21:46:45 UTC	2018-06-06 00:00:00 UTC	2018-06-06 22:04:34 UTC	29	0
16	61d430273ff1e88f2944acb53e...	2018-05-06 09:48:42 UTC	2018-06-06 00:00:00 UTC	2018-06-05 12:09:51 UTC	30	0
17	d2f8ef9dd1714fcac7de9f0aef...	2018-05-15 12:29:55 UTC	2018-06-06 00:00:00 UTC	2018-06-14 23:42:24 UTC	30	-8

The negative difference shows the no. of days order delayed from delivery time.

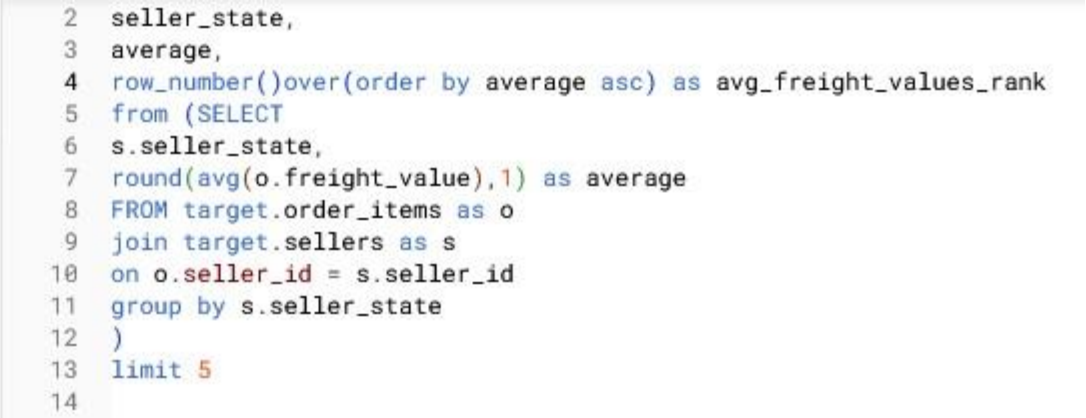
2.



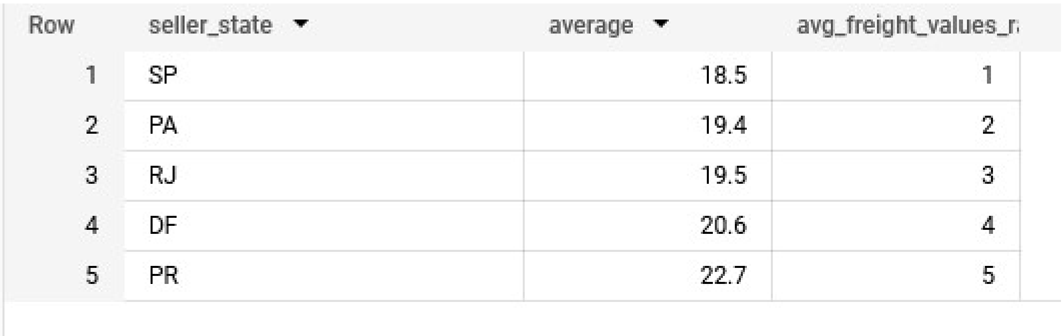
```
seller_state,
average,
row_number()over(order by average desc) as avg_freight_values_rank
from (SELECT
s.seller_state,
round(avg(o.freight_value),1) as average
FROM target.order_items as o
join target.sellers as s
on o.seller_id = s.seller_id
group by s.seller_state
)
limit 5
```

Row	customer_state	top_5_average
1	RR	29.0
2	AP	26.7
3	AM	26.0
4	AL	24.0
5	PA	23.3

Highest avg freight value is in RR i.e. 29



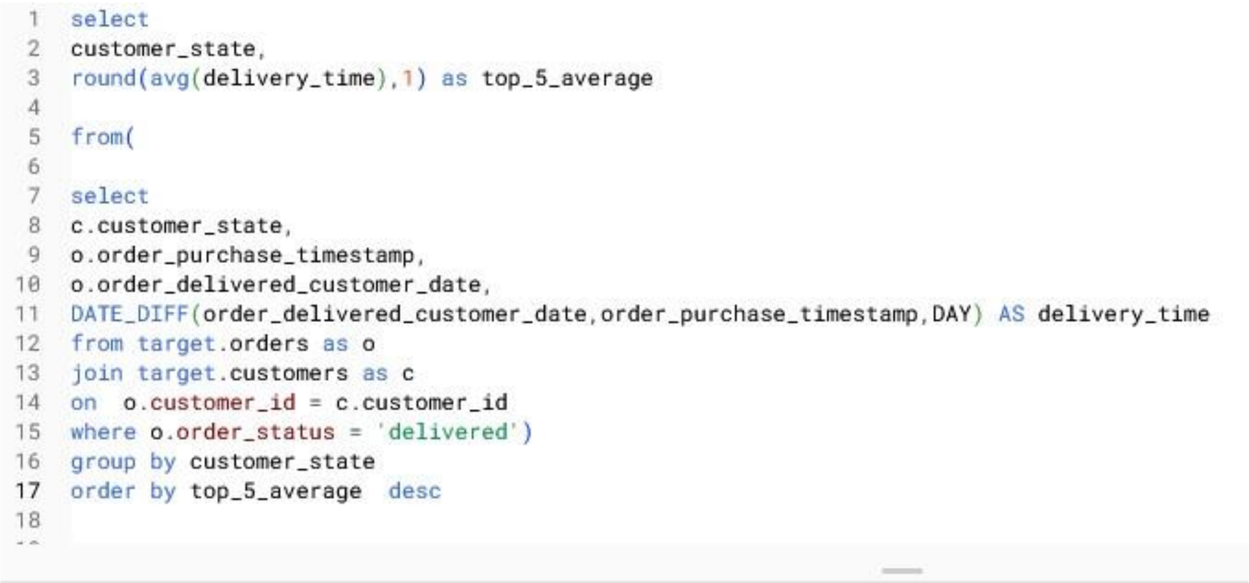
```
seller_state,
average,
row_number()over(order by average asc) as avg_freight_values_rank
from (SELECT
s.seller_state,
round(avg(o.freight_value),1) as average
FROM target.order_items as o
join target.sellers as s
on o.seller_id = s.seller_id
group by s.seller_state
)
limit 5
```



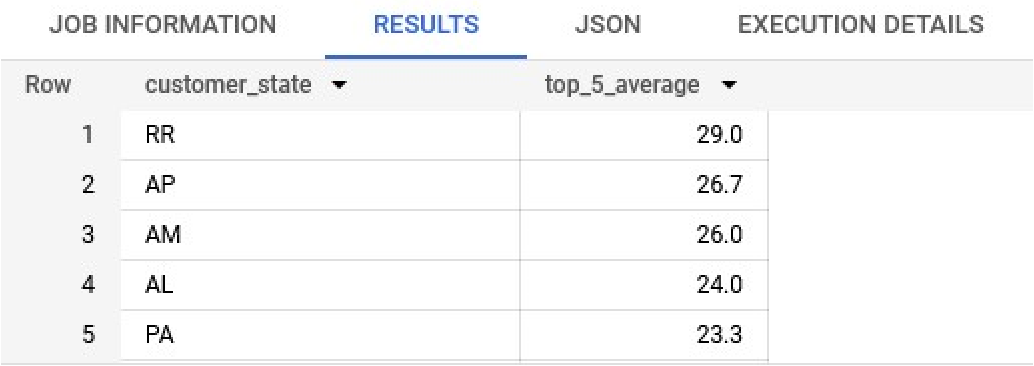
Row	seller_state	average	avg_freight_values_r
1	SP	18.5	1
2	PA	19.4	2
3	RJ	19.5	3
4	DF	20.6	4
5	PR	22.7	5

Lowest avg freight value is in SP i.e. 18.5

3

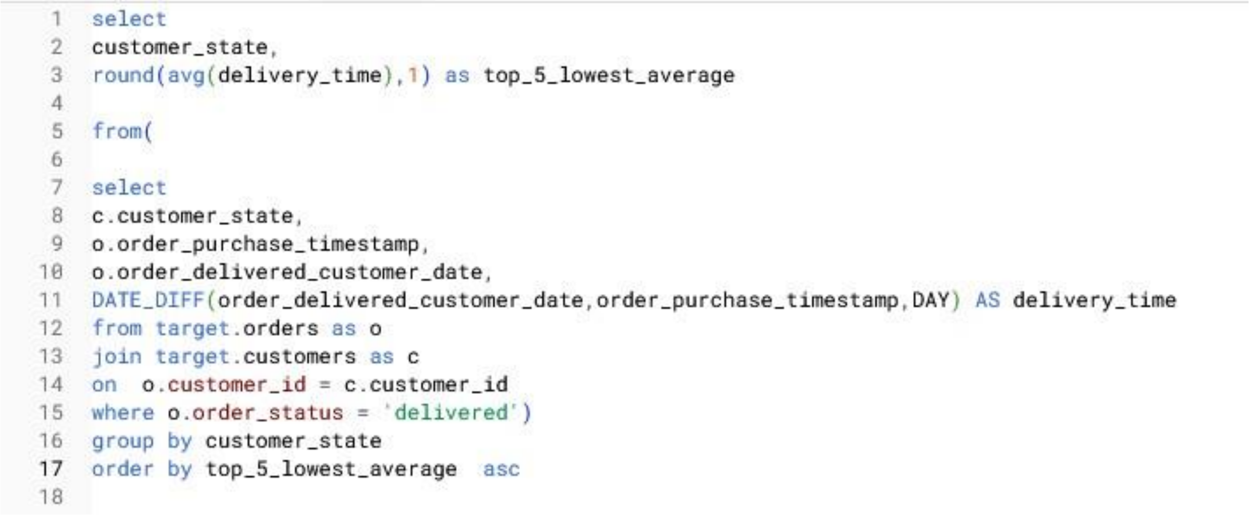


```
select
customer_state,
round(avg(delivery_time),1) as top_5_average

from(

select
c.customer_state,
o.order_purchase_timestamp,
o.order_delivered_customer_date,
DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY) AS delivery_time
from target.orders as o
join target.customers as c
on  o.customer_id = c.customer_id
where o.order_status = 'delivered')
group by customer_state
order by top_5_average  desc

```

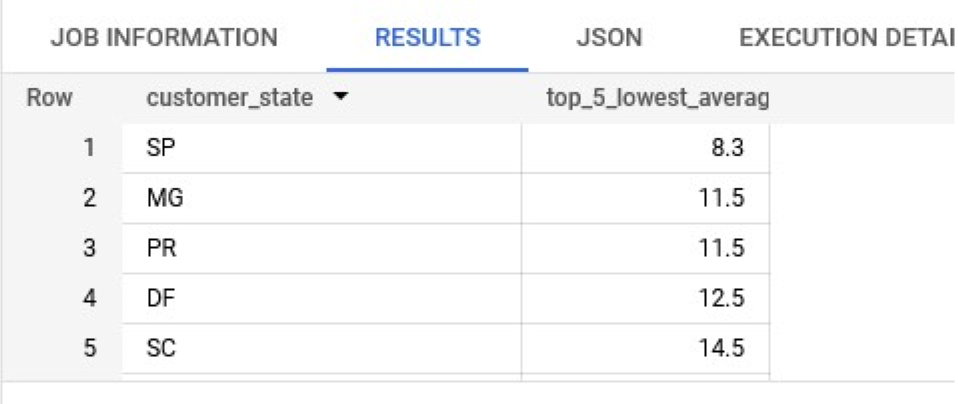


Row	customer_state	top_5_average
1	RR	29.0
2	AP	26.7
3	AM	26.0
4	AL	24.0
5	PA	23.3

Highest avg delivery time is in RR i.e. 29.



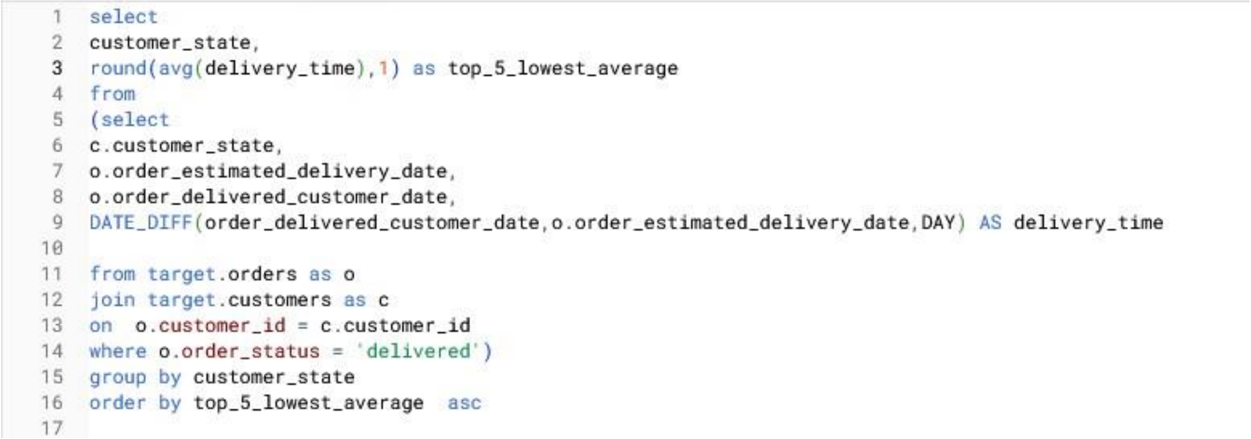
```
select
customer_state,
round(avg(delivery_time),1) as top_5_lowest_average

from(

select
c.customer_state,
o.order_purchase_timestamp,
o.order_delivered_customer_date,
DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY) AS delivery_time
from target.orders as o
join target.customers as c
on  o.customer_id = c.customer_id
where o.order_status = 'delivered')
group by customer_state
order by top_5_lowest_average  asc

```



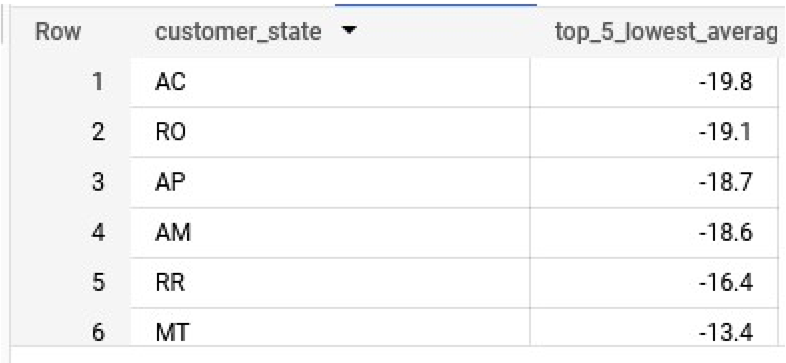
Row	customer_state	top_5_lowest_averag
1	SP	8.3
2	MG	11.5
3	PR	11.5
4	DF	12.5
5	SC	14.5

Lowest avg delivery time is in SP i.e. 8.3

4.



```sql
select
customer_state,
round(avg(delivery_time),1) as top_5_lowest_average
from
(select
c.customer_state,
o.order_estimated_delivery_date,
o.order_delivered_customer_date,
DATE_DIFF(order_delivered_customer_date,o.order_estimated_delivery_date,DAY) AS delivery_time

from target.orders as o
join target.customers as c
on  o.customer_id = c.customer_id
where o.order_status = 'delivered')
group by customer_state
order by top_5_lowest_average  asc
```



Row	customer_state	top_5_lowest_averag
1	AC	-19.8
2	RO	-19.1
3	AP	-18.7
4	AM	-18.6
5	RR	-16.4
6	MT	-13.4

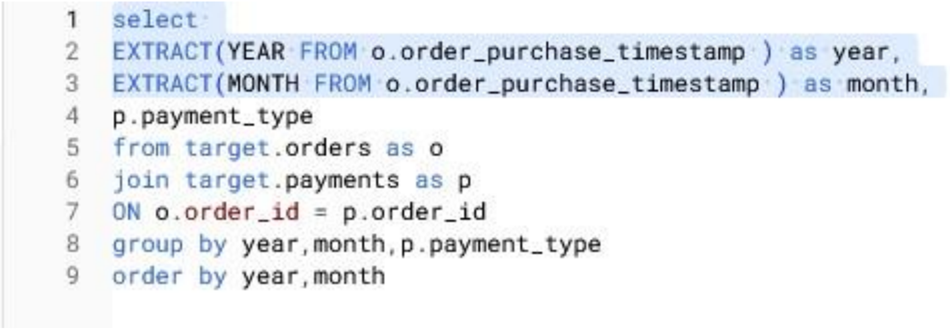
Question 6.

Analysis based on the payments:

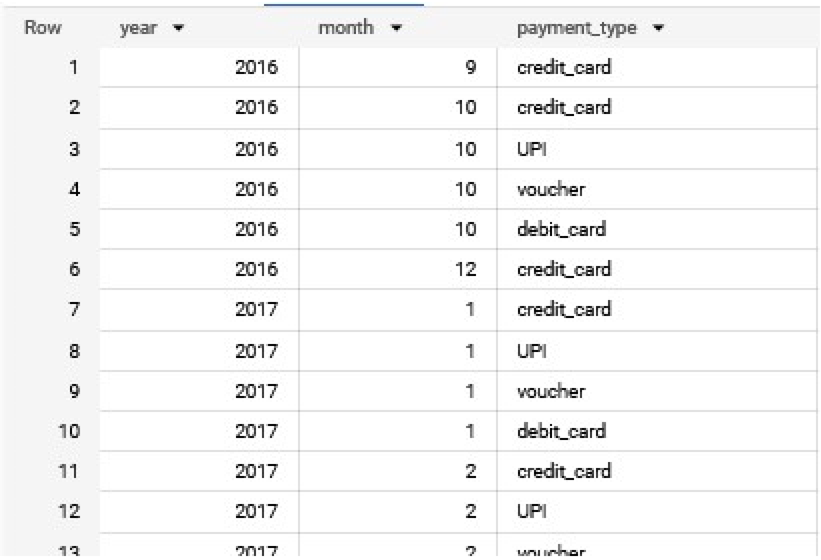
1. Find the month on month no. of orders placed using different payment types.
2. Find the no. of orders placed on the basis of the payment installments that have been paid.

Answer 6.

1.

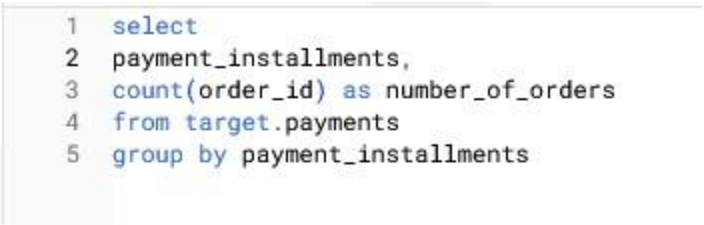


```sql
select 
EXTRACT(YEAR FROM o.order_purchase_timestamp ) as year,
EXTRACT(MONTH FROM o.order_purchase_timestamp ) as month,
p.payment_type
from target.orders as o
join target.payments as p
ON o.order_id = p.order_id
group by year,month,p.payment_type
order by year,month
```

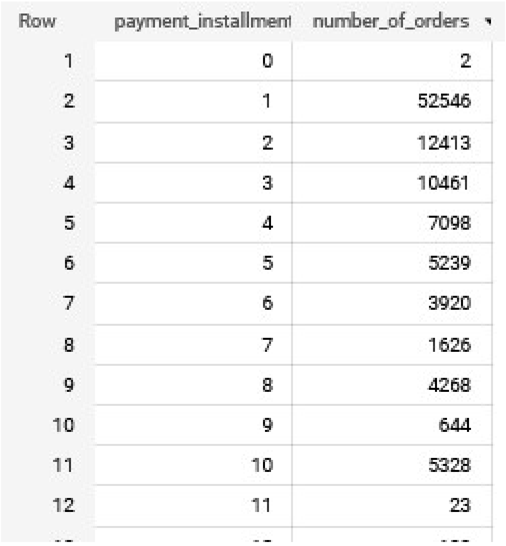


Row	year	month	payment_type
1	2016	9	credit_card
2	2016	10	credit_card
3	2016	10	UPI
4	2016	10	voucher
5	2016	10	debit_card
6	2016	12	credit_card
7	2017	1	credit_card
8	2017	1	UPI
9	2017	1	voucher
10	2017	1	debit_card
11	2017	2	credit_card
12	2017	2	UPI
13	2017	2	voucher

2.



```sql
select
payment_installments,
count(order_id) as number_of_orders
from target.payments
group by payment_installments
```



Row	payment_installment	number_of_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

Recommendation
The cost of order has increased from 2017 to 2018 by more than 50% , but still the number of orders placed is high in areas with low freight value and less orders are placed in areas of high freight values and delivery values , so the company can reduce the price of order in areas with high delivery value and increase the marketing strategies for summer because the no. of orders placed in summer is quite low as compared to winters.