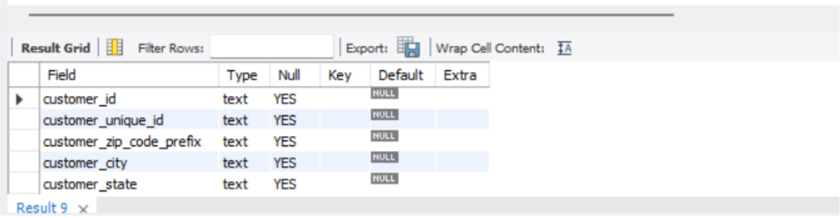
1. **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset**
    1. Data type of columns in a table

```
Show fields from customers;
show fields from geolocation;
show fields from order_items;
show fields from order_reviews;
show fields from orders;
show fields from payments;
show fields from products;
show fields from sellers;
```



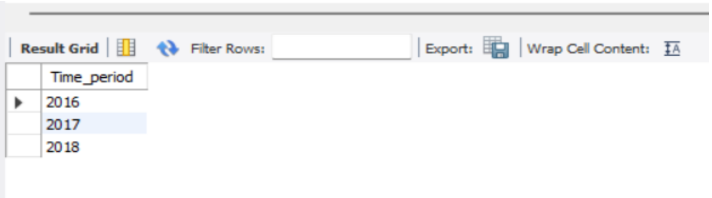
Field	Type	Null	Key	Default	Extra
customer_id	text	YES		NULL	
customer_unique_id	text	YES		NULL	
customer_zip_code_prefix	text	YES		NULL	
customer_city	text	YES		NULL	
customer_state	text	YES		NULL	

2. Time period for which the data is given

```
select Year(order_purchase_timestamp) as Time_period  from orders

group by Time_period

order by Time_period asc;
```



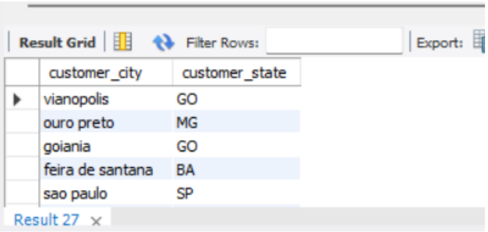
Time_period
2016
2017
2018

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

```
select c.customer_city, c.customer_state

from customers c

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



customer_city	customer_state
vianopolis	GO
ouro preto	MG
goiania	GO
feira de santana	BA
sao paulo	SP

2. 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?

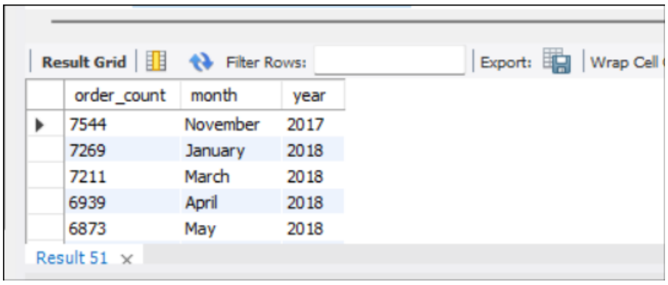
```
select count(order_id) as order_count,

monthname(order_purchase_timestamp) as month,
year(order_purchase_timestamp) as year

from orders

group by month,year

order by order_count desc;
```



order_count	month	year
7544	November	2017
7269	January	2018
7211	March	2018
6939	April	2018
6873	May	2018

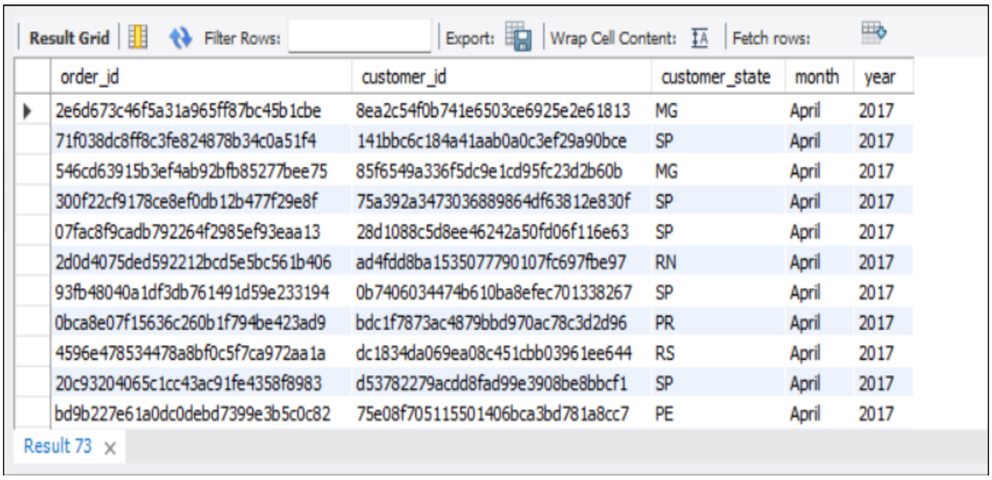
2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

```
Select count(order_id) as order_count, concat(h,":",m,":",s) as time
From
(select customer_id, order_id, hour(order_purchase_timestamp)h,
minute(order_purchase_timestamp)m,
second(order_purchase_timestamp)s
from orders) As
group by time;
```

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

1. Get month on month orders by states

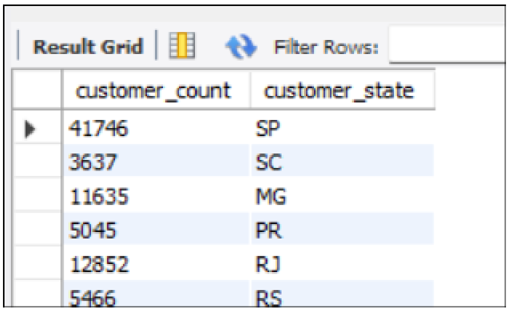
```
select o.order_id, o.customer_id, c.customer_state,
monthname(o.order_purchase_timestamp) as month,
year(o.order_purchase_timestamp) as year
from orders o
join customers c on o.customer_id = c.customer_id
order by month, year;
```



order_id	customer_id	customer_state	month	year
2e6d673c46f5a31a965ff87bc45b1cbe	8ea2c54f0b741e6503ce6925e2e61813	MG	April	2017
71f038dc8ff8c3fe824878b34c0a51f4	141bbc6c184a41aab0a0c3ef29a90bce	SP	April	2017
546cd63915b3ef4ab92bfb85277bee75	85f6549a336f5dc9e1cd95fc23d2b60b	MG	April	2017
300f22cf9178ce8ef0db12b477f29e8f	75a392a3473036889864df63812e830f	SP	April	2017
07fac8f9cadb792264f2985ef93eaa13	28d1088c5d8ee46242a50fd06f116e63	SP	April	2017
2d0d4075ded592212bcd5e5bc561b406	ad4fdd8ba1535077790107fc697fbe97	RN	April	2017
93fb48040a1df3db761491d59e233194	0b7406034474b610ba8efec701338267	SP	April	2017
0bca8e07f15636c260b1f794be423ad9	bdc1f7873ac4879bbd970ac78c3d2d96	PR	April	2017
4596e478534478a8bf0c5f7ca972aa1a	dc1834da069ea08c451cbb03961ee644	RS	April	2017
20c93204065c1cc43ac91fe4358f8983	d53782279acdd8fad99e3908be8bbcf1	SP	April	2017
bd9b227e61a0dc0debd7399e3b5c0c82	75e08f705115501406bca3bd781a8cc7	PE	April	2017

2. Distribution of customers across the states in Brazil

```
Select count(customer_id) as customer_count,customer_state
from customers
group by customer_state;
```



customer_count	customer_state
41746	SP
3637	SC
11635	MG
5045	PR
12852	RJ
5466	RS

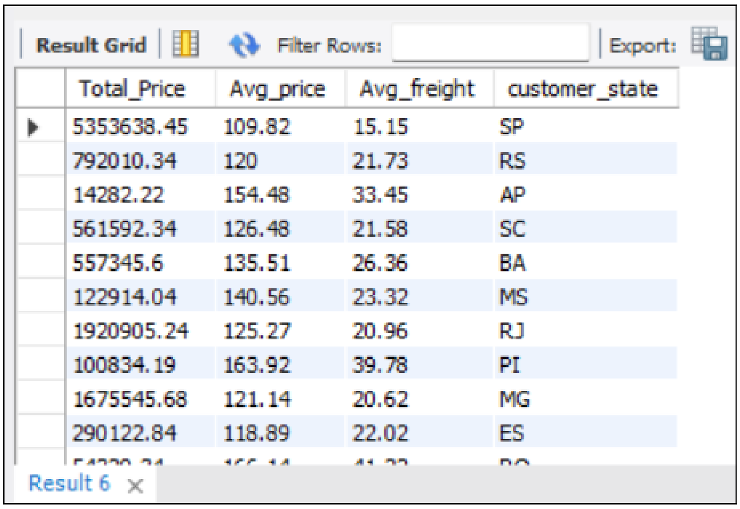
4. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**

   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 round(sum(price+freight_value),2)Total_Price,
round(AVg(price),2) Avg_price,
round(avg(freight_value),2) Avg_freight,
c.customer_state
```

```
from order_items oi
join orders o on oi.order_id = o.order_id
join customers c on o.customer_id = c.customer_id
group by c.customer_state;
```

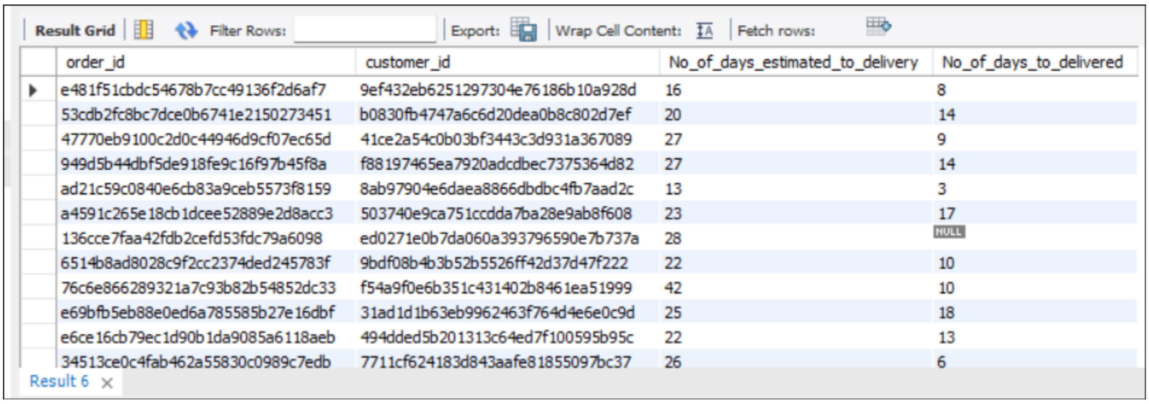


Total_Price	Avg_price	Avg_freight	customer_state
5353638.45	109.82	15.15	SP
792010.34	120	21.73	RS
14282.22	154.48	33.45	AP
561592.34	126.48	21.58	SC
557345.6	135.51	26.36	BA
122914.04	140.56	23.32	MS
1920905.24	125.27	20.96	RJ
100834.19	163.92	39.78	PI
1675545.68	121.14	20.62	MG
290122.84	118.89	22.02	ES

5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery

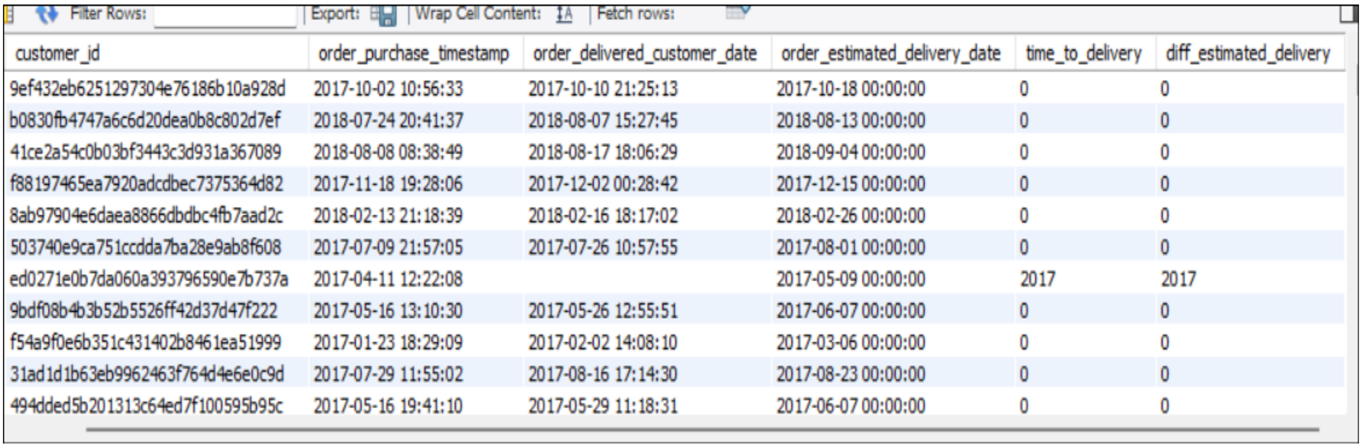
```
select order_id, customer_id,
datediff(order_estimated_delivery_date, order_purchase_timestamp)
No_of_days_estimated_to_delivery,
datediff(order_delivered_customer_date, order_purchase_timestamp)
No_of_days_to_delivered from orders ;
```



order_id	customer_id	No_of_days_estimated_to_delivery	No_of_days_to_delivered
e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	16	8
53cdb2fc8bc7dce0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef	20	14
47770eb9100c2d0c44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089	27	9
949d5b44dbf5de918fe9c16f97b45f8a	f88197465ea7920adcdbec7375364d82	27	14
ad21c59c0840e6cb83a9ceb5573f8159	8ab97904e6daea8866dbdbc4fb7aad2c	13	3
a4591c265e18cb1dcee52889e2d8acc3	503740e9ca751ccdda7ba28e9ab8f608	23	17
136cce7faa42fdb2cefd53fdc79a6098	ed0271e0b7da060a393796590e7b737a	28	NULL
6514b8ad8028c9f2cc2374ded245783f	9bdf08b4b3b52b5526ff42d37d47f222	22	10
76c6e866289321a7c93b82b54852dc33	f54a9f0e6b351c431402b8461ea51999	42	10
e69bfb5eb88e0ed6a785585b27e16dbf	31ad1d1b63eb9962463f764d4e6e0c9d	25	18
e6ce16cb79ec1d90b1da9085a6118aeb	494dded5b201313c64ed7f100595b95c	22	13
34513ce0c4fab462a55830c0989c7edb	7711cf624183d843aafe81855097bc37	26	6

2. Find time_to_delivery & diff_estimated_delivery. Formula for the same given below:
   - time_to_delivery = order_purchase_timestamp-order_delivered_customer_date
   - diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date

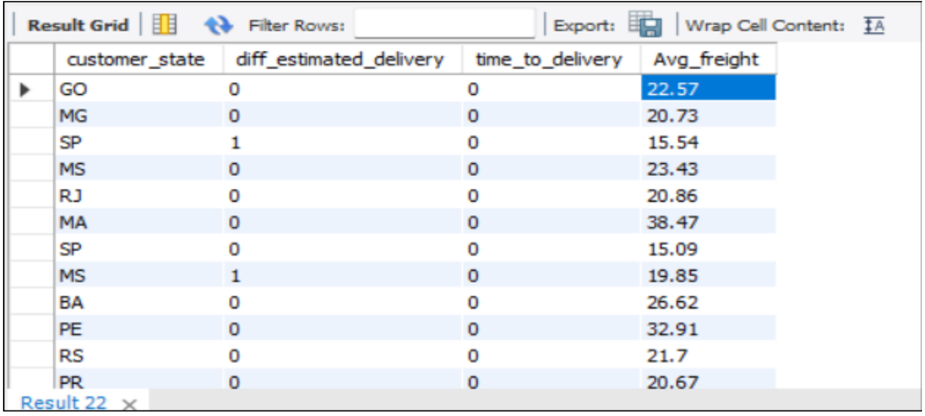
```
select order_id, customer_id,
order_purchase_timestamp, order_delivered_customer_date,
order_estimated_delivery_date,
(order_purchase_timestamp - order_delivered_customer_date)
time_to_delivery,
(order_estimated_delivery_date - order_delivered_customer_date)
diff_estimated_delivery
FROM orders;
```



customer_id	order_purchase_timestamp	order_delivered_customer_date	order_estimated_delivery_date	time_to_delivery	diff_estimated_delivery
9ef432eb6251297304e76186b10a928d	2017-10-02 10:56:33	2017-10-10 21:25:13	2017-10-18 00:00:00	0	0
b0830fb4747a6c6d20dea0b8c802d7ef	2018-07-24 20:41:37	2018-08-07 15:27:45	2018-08-13 00:00:00	0	0
41ce2a54c0b03bf3443c3d931a367089	2018-08-08 08:38:49	2018-08-17 18:06:29	2018-09-04 00:00:00	0	0
f88197465ea7920adcdbec7375364d82	2017-11-18 19:28:06	2017-12-02 00:28:42	2017-12-15 00:00:00	0	0
8ab97904e6daea8866dbdbc4fb7aad2c	2018-02-13 21:18:39	2018-02-16 18:17:02	2018-02-26 00:00:00	0	0
503740e9ca751ccdda7ba28e9ab8f608	2017-07-09 21:57:05	2017-07-26 10:57:55	2017-08-01 00:00:00	0	0
ed0271e0b7da060a393796590e7b737a	2017-04-11 12:22:08		2017-05-09 00:00:00	2017	2017
9bdf08b4b3b52b5526ff42d37d47f222	2017-05-16 13:10:30	2017-05-26 12:55:51	2017-06-07 00:00:00	0	0
f54a9f0e6b351c431402b8461ea51999	2017-01-23 18:29:09	2017-02-02 14:08:10	2017-03-06 00:00:00	0	0
31ad1d1b63eb9962463f764d4e6e0c9d	2017-07-29 11:55:02	2017-08-16 17:14:30	2017-08-23 00:00:00	0	0
494dded5b201313c64ed7f100595b95c	2017-05-16 19:41:10	2017-05-29 11:18:31	2017-06-07 00:00:00	0	0

3. Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery

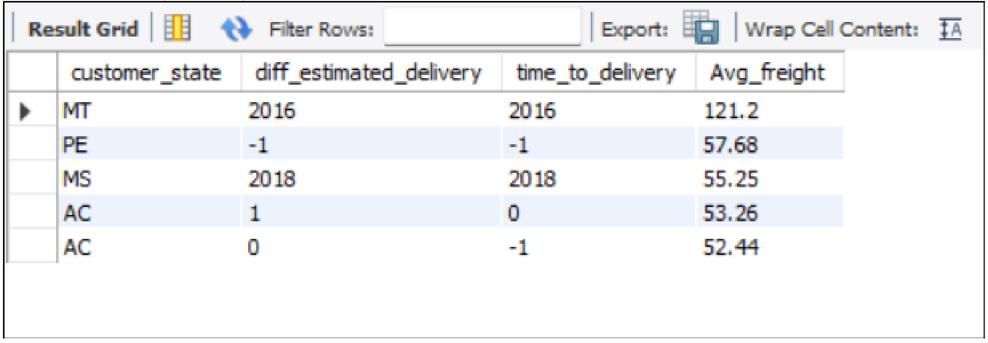
```
select a.*, c.customer_state from
(select order_id, customer_id,
order_purchase_timestamp, order_delivered_customer_date,
order_estimated_delivery_date,
(order_purchase_timestamp - order_delivered_customer_date)
time_to_delivery,
(order_estimated_delivery_date - order_delivered_customer_date)
diff_estimated_delivery
FROM orders)a
join customers c on a.customer_id = c.c
```



customer_state	diff_estimated_delivery	time_to_delivery	Avg_freight
GO	0	0	22.57
MG	0	0	20.73
SP	1	0	15.54
MS	0	0	23.43
RJ	0	0	20.86
MA	0	0	38.47
SP	0	0	15.09
MS	1	0	19.85
BA	0	0	26.62
PE	0	0	32.91
RS	0	0	21.7
PR	0	0	20.67

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

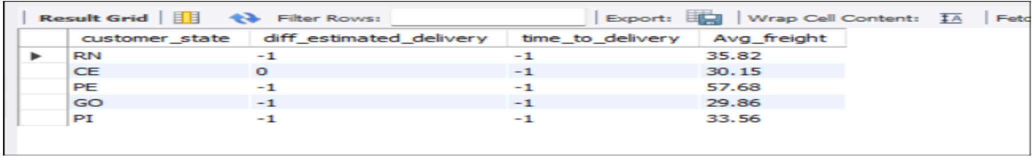
```
select c.customer_state, a.diff_estimated_delivery, a.time_to_delivery,
round(Avg(o.freight_value),2) Avg_freight from
(select order_id, customer_id,
order_purchase_timestamp, order_delivered_customer_date,
order_estimated_delivery_date,
(order_purchase_timestamp - order_delivered_customer_date)
time_to_delivery,
(order_estimated_delivery_date - order_delivered_customer_date)
diff_estimated_delivery
FROM orders)a
join customers c on a.customer_id = c.customer_id
join order_items o on a.order_id = o.order_id
group by c.customer_state, a.diff_estimated_delivery, a.time_to_delivery
order by Avg_freight desc limit 5 ;
```



customer_state	diff_estimated_delivery	time_to_delivery	Avg_freight
MT	2016	2016	121.2
PE	-1	-1	57.68
MS	2018	2018	55.25
AC	1	0	53.26
AC	0	-1	52.44

6. Top 5 states with highest/lowest average time to delivery

```
select c.customer_state, a.diff_estimated_delivery, a.time_to_delivery,
round(Avg(o.freight_value),2) Avg_freight from
(select order_id, customer_id,
order_purchase_timestamp, order_delivered_customer_date,
order_estimated_delivery_date,
(order_purchase_timestamp - order_delivered_customer_date)
time_to_delivery,
(order_estimated_delivery_date - order_delivered_customer_date)
diff_estimated_delivery
FROM orders)a
join customers c on a.customer_id = c.customer_id
join order_items o on a.order_id = o.order_id
group by c.customer_state, a.diff_estimated_delivery, a.time_to_delivery
order by Avg(time_to_delivery) limit 5 ;
```



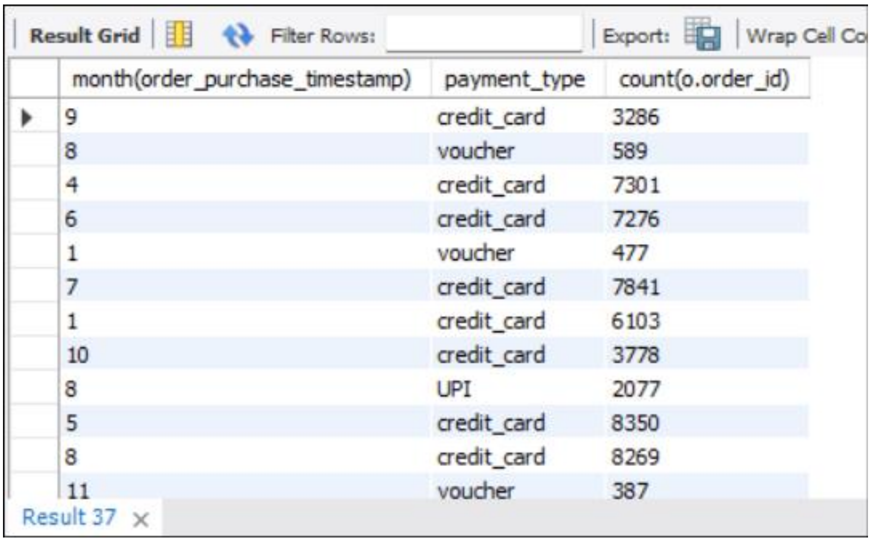
customer_state	diff_estimated_delivery	time_to_delivery	Avg_freight
RN	-1	-1	35.82
CE	0	-1	30.15
PE	-1	-1	57.68
GO	-1	-1	29.86
PI	-1	-1	33.56

7. Top 5 states where delivery is really fast/ not so fast compared to estimated date

6. Payment type analysis:

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

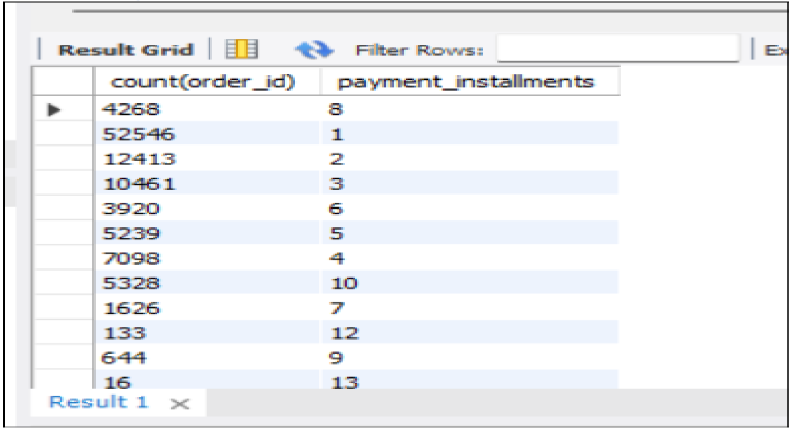
```
select month(order_purchase_timestamp), p.payment_type, count(o.order_id)
from orders o
join payments p on o.order_id = p.order_id
group by month(order_purchase_timestamp), p.payment_type;
```



month(order_purchase_timestamp)	payment_type	count(o.order_id)
9	credit_card	3286
8	voucher	589
4	credit_card	7301
6	credit_card	7276
1	voucher	477
7	credit_card	7841
1	credit_card	6103
10	credit_card	3778
8	UPI	2077
5	credit_card	8350
8	credit_card	8269
11	voucher	387

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

```
select count(order_id), payment_installments
from payments
group by payment_installments;
```



count(order_id)	payment_installments
4268	8
52546	1
12413	2
10461	3
3920	6
5239	5
7098	4
5328	10
1626	7
133	12
644	9
16	13